A Portrait of
Linear Algebra

Third Edition

Jude Thaddeus Socrates

Pasadena City College

Kendall Hunt
publishing company

To order the print or e-book version of this book, go to:

https://he.kendallhunt.com/product/portrait-linear-algebra

Kendall Hunt
publishing company

www.kendallhunt.com
Send all inquiries to:
4050 Westmark Drive
Dubuque, IA 52004-1840

Copyright © 2016 by Kendall Hunt Publishing Company

ISBN 978-1-4652-9053-3

Printed in the United States of America

Contents

The Answer Key to the Exercises is available as a free download at:
https://he.kendallhunt.com/product/portrait-linear-algebra

Preface to the 3rd Edition

In the three years since the 2nd Edition of ***A Portrait of Linear Algebra*** came out, I have had the privilege of teaching Linear Algebra every semester, and even during most of the summers. All the new ideas, improvements, exercises, and other changes that have been incorporated in the 3rd edition would not have been possible without the lengthy discussions and interactions that I have had with so many wonderful students in these classes, and the colleagues who adopted this book for their own Linear Algebra class.

So let me begin by thanking Daniel Gallup, John Sepikas, Lyman Chaffee, Christopher Strinden, Patricia Michel, Asher Shamam, Richard Abdelkerim, Mark Pavitch, David Matthews, and Guoqiang Song, my colleagues at Pasadena City College who have taught out of my book, for sharing their ideas and experiences with me, their encouragement, and suggestions for improving this text.

I am certain that if I begin to name all the students who have given me constructive criticisms about the book, I will miss more than just a handful. There have been hundreds of students who have gone through this book, and I learned so much from my conversations with many of them. Often, a casual remark or a simple question would prompt me to rewrite an explanation or come up with an interesting new exercise. Many of these students have continued on to finish their undergraduate careers at four-year institutions, and have begun graduate studies in mathematics or engineering. Some of them have kept in touch with me over the years, and the sweetest words they have said to me is how easily they handled upper-division Linear Algebra classes, thanks to the solid education they received from my book. I give them my deepest gratitude, not just for their thoughts, but also for giving me the best career in the world.

It is hard to believe that ten years ago, the idea of this book did not even exist. None of this would have been possible without the help of so many people.

Thank you to Christine Bochniak, Beverly Kraus, and Taylor Knuckey of Kendall Hunt for their valuable assistance in bringing the 3rd edition to fruition.

Many thanks to my long-suffering husband, my best friend and biggest supporter, Juan Sanchez-Diaz, for patiently accepting all the nights and weekends that were consumed by this book. And thank you to Johannes, for your unconditional love and for making me get up from the computer so we can go for a walk or play with the ball. I would have gone bonkers if it weren't for you two.

To the members of the Socrates and Sanchez families all over the planet, *maraming salamat*, *y muchas gracias*, for all your love and support.

Thanks to all my colleagues at PCC, my friends on Facebook, and my *barkada*, for being my unflagging cheering squad and artistic critics.

Thanks to my tennis and gym buddies for keeping me motivated and physically healthy.

Thank you to my late parents, Dr. Jose Socrates and Dr. Nenita Socrates, for teaching me and all their children the love for learning.

And finally, my thanks to our Lord, for showering my life with so many blessings.

Jude Thaddeus Socrates
Professor of Mathematics
Pasadena City College, California
June, 2016

What Makes This Book Different?

A Portrait of Linear Algebra takes a unique approach in developing and introducing the core concepts of this subject. It begins with a thorough introduction of the field properties for real numbers and uses them to guide the student through simple proof exercises. From here, we introduce the Euclidean spaces and see that many of the field properties for the real numbers naturally extend to the properties of vector arithmetic. The core concepts of linear combinations, spans of sets of vectors, linear independence, subspaces, basis and dimension, are introduced in the first chapter and constantly referenced and reinforced throughout the book. This early introduction enables the student to retain these concepts better and to apply them to deeper ideas.

The Four Fundamental Matrix Spaces are encountered at the end of the first Chapter, and transitions naturally into the second Chapter, where we study linear transformations and their standard matrices. The kernel and range of these transformations tells us if our transformations are one-to-one or onto. When they are both, we learn how to find the inverse transformation. We also see that some geometric operations of vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ are examples of linear operators.

Once these core concepts are firmly established, they can be naturally extended to create abstract vector spaces, the most important examples of which are function spaces, polynomial spaces, and matrix spaces. Linear transformations on finite dimensional vector spaces can again be coded using matrices by finding coordinates for our vectors with respect to a basis. Everything we encountered in the first two chapters can now be naturally generalized.

One of the unique features of this book is the use of projections and reflections in $\mathbb{R}^3$, with respect to either a line or a plane, in order to motivate some concepts or constructions. We use them to explore the core concepts of the standard matrix of a linear transformation, the matrix of a transformation with respect to a non-standard basis, and the change of basis matrix. In the case of reflection operators, we see them as motivation for the inverse of a matrix, and as an example of an orthogonal matrix. Projection matrices, on the other hand, are good examples of idempotent matrices.

The second half of the book goes into the study of determinants, eigentheory, inner product spaces, complex vector spaces, the Spectral Theorems, and the material necessary to understand and prove the Fundamental Theorem of Linear Algebra, and its twin, the Singular Value Decomposition. We also see several applications of Linear Algebra in science, engineering, and other areas of mathematics.

Throughout the book, we emphasize clear and precise definitions and proofs of Theorems, constantly encouraging the student to read and understand proofs, and to practice writing their own proofs.

How this Book is Organized

Chapter Zero provides an introduction to sets and set operations, logic, the field axioms for real numbers, and common proof techniques, emphasizing theorems that can be derived from the field axioms. This brief introductory chapter will prepare the student to learn how to read, understand and write basic proofs.

We base our development of the main concepts of Linear Algebra on the following definition:

> ***Linear Algebra*** is the study of ***vector spaces***, their ***structure***, and the ***linear transformations*** that map one vector space to another.

Chapter 1 rigorously examines the archetype vector spaces: Euclidean spaces, their geometry, and the core ideas of spanning, linear independence, subspaces, basis, dimension and orthogonal complements. We will see the Gauss-Jordan Algorithm, the central tool of Linear Algebra, and use it to solve systems of linear equations and investigate the span of a set of vectors. We will also construct the four fundamental matrix spaces: rowspace, columnspace and nullspace for a matrix and its transpose, and find a basis for each space.

Chapter 2 introduces linear transformations on Euclidean spaces as encoded by matrices. We will see how each linear transformation determines special subspaces, namely the kernel and the range of the transformation, and use these spaces to investigate the one-to-one and onto properties. We will define basic matrix operations, including a method to find its inverse when this exists.

Chapter 3 generalizes the concepts from Chapters 1 and 2 in order to construct abstract vector spaces and linear transformations from one vector space to another. We focus most of our examples on function spaces (in particular, polynomial spaces), and linear transformations connecting them, especially those involving derivatives and evaluations. We will see that in the finite-dimensional case, a linear transformation can be encoded by a matrix as well. By focusing on function spaces preserved by the derivative operator, the strong relationship between Linear Algebra and Differential Equations is firmly established.

Chapter 4 investigates the subspace structure of vector spaces, and we will see techniques to fully describe the join and intersection of two subspaces, the image or preimage of a subspace, and the restriction of a linear transformation to a subspace. We will create cosets and quotient spaces, and see one of the fundamental triptychs of modern mathematics: the Isomorphism Theorems of Emy Noether as applied to vector spaces.

Chapter 5 explores the determinant function, its properties, especially its relationship to invertibility, and efficient algorithms to compute it. We will see Cramer’s rule, a technique to solve invertible square systems of equations, albeit not a very practical one.

Chapter 6 introduces the eigentheory of operators both on Euclidean spaces as well as abstract vector spaces. We will see when it is possible to encode operators into the simplest possible form, that is, to diagonalize them. We will study the concept of similarity and its consequences.

Chapter 7 generalizes geometry on a vector space by imposing an inner product on it. This allows us to introduce the concepts of norm and orthogonality in abstract spaces. We will explore orthonormal bases, the Gram-Schmidt Algorithm, orthogonal matrices, the orthogonal diagonalization of symmetric matrices, the method of least squares, and the QR-decomposition.

Chapter 8 applies the constructions thus far to vector spaces over arbitrary fields, especially the field of complex numbers. The main goal of this chapter is to prove the Spectral Theorem of Normal Matrices. One specific case of this Theorem tells us that symmetric matrices can indeed be diagonalized by orthogonal matrices. We also see that commuting diagonalizable matrices can be simultaneously diagonalized by the same invertible matrix, and present an algorithm to do so.

Chapter 9 explores some applications of Linear Algebra in science and engineering. We develop the theories of quadratic forms and positive semi-definite matrices. These enable us to prove The Fundamental Theorem of Linear Algebra, an elegant theorem that ties together the four fundamental matrix spaces and the concepts of eigenspaces and orthogonality. Closely connected to this is the Singular Value Decomposition, which has applications in data processing.

This book is intended to serve as a text for a standard 15-week semester course in introductory Linear Algebra. However, enough material is included in this text for two full semesters. This book is my vision of what today’s student in science and engineering should know about this elegant field.

What is New with the Third Edition?

Over 500 new Exercises have been added since the 2nd edition.

The last two Sections of Chapter 1 in the 2nd Edition were reorganized into three new sections. Section 1.7 introduces the concept of a subspace of $\mathbb{R}^n$ and proves that every non-zero subspace has a basis, leading us to define the concept of dimension. Section 1.8 introduces the four fundamental matrix spaces and the Dimension Theorem for Matrices, the properties and relevance of these spaces, and how to find a basis for each of them. Section 1.9 focuses on finding a basis for the orthogonal complement of a subspace of $\mathbb{R}^n$.

There are three completely new sections in the 3rd edition:

Section 5.5. The Wronskian: a matrix that can determine if a finite set of functions is linearly independent.

Section 6.4. The Exponential of a Matrix: a method to compute e^A, where A is a diagonalizable square matrix. This computation is particularly important in finding the solutions to a System of Linear Differential Equations.

Section 8.7. Simultaneous Diagonalization: an algorithm to find an invertible matrix C that will simultaneously diagonalize two commuting diagonalizable matrices. This is perhaps one of the most elegant ideas presented in this book.

Special Topics and Mini-Projects

Scattered around the Exercises are multi-step problems that guide the student through various topics that probe deeper into Linear Algebra and its connections with Geometry, Calculus, Differential Equations, and other areas of mathematics such as Set Theory, Group Theory and Number Theory.

The Medians of a Triangle: a coordinate-free proof that the three medians of any triangle intersect at a common point which is 2/3 the distance from any vertex to the opposite midpoint (Section 1.1).

The Cross Product: used to create a vector orthogonal to two vectors in $\mathbb{R}^3$, and proves its other properties using the properties of the 3×3 determinant (Sections 1.3, 5.1 and 5.2).

The Uniqueness of the Reduced Row Echelon Form: uses the concepts of the rowspace of a matrix and the Equality of Spans Theorem to prove that the rref of any matrix is unique (Section 1.8).

Drawing Three-Dimensional Objects: applies the concept of a projection in order to show how to draw the edges of a 3-dimensional object as perceived from any given direction (Section 2.2).

The Center of the Ring of Square Matrices: uses basic matrix products to show that the only $n \times n$ matrices that commute with all $n \times n$ matrices are the multiples of the identity matrix (Section 2.4).

The Kernel and Range of a Composition: proves that the kernel of a composition $T_2 \circ T_1$ contains the kernel of T_1, and analogously, the range of $T_2 \circ T_1$ is contained in the range of T_2 (Section 2.5 for Euclidean Spaces and Section 3.7 for arbitrary vector spaces).

The Direct Sum of Matrices: explores the properties of matrices in block-diagonal form (Sections 2.8, 2.9, 5.3, 6.1, 7.6, and 8.7).

The Chinese Remainder Theorem: introduced and applied to construct invertible 2×2 integer matrices whose inverses also have integer entries (Section 2.8).

Cantor's Diagonal Argument: proves that the set of rational numbers is countable by showing how to list its elements in a sequence (Section 3.3).

The Countability of Subintervals of the set of Real Numbers: gives a guided proof that all subintervals of $\mathbb{R}$ that contains at least two points have the same cardinality as $\mathbb{R}$, by explicitly constructing bijections among these subintervals (Section 3.3).

Bisymmetric Matrices: explores the properties and dimensions of this unusual and interesting family of square matrices (Section 3.4).

The Centralizer of a Matrix: proves that the set of matrices that commute with a given square matrix forms a vector space, and finds a basis for it in the 2×2 case (Section 3.4).

Vector Spaces of Infinite Series: proves that the set of absolutely convergent series form a subspace of the space of all infinite series, whereas conditionally convergent and divergent series are not closed under addition (Section 3.4). We also see a natural ***inner product*** which is well-defined on absolutely convergent series but fails for conditionally convergent series (Section 7.1).

Casting Shadows: shows that the shadow on the floor of an image on a window pane is an example of a linear transformation (Section 3.6).

The Vandermonde Determinant: applies row and column operations and cofactor expansions to find a closed formula for the Vandermonde Determinant, and applies it to some Wronskian determinants, proving that certain infinite subsets of function spaces are linearly independent (Sections 5.3 and 5.5).

The Special Linear Group of Integer Matrices: introduces the concept of a ***group***, and proves that the set of all $n \times n$ matrices with integer entries and determinant 1 form a group under matrix multiplication. This project also proves that $\boldsymbol{SL_2}(\mathbb{Z})$ is generated by two special matrices (Section 5.3).

Invertible Triangular Matrices: uses Cramer's rule to prove that the inverse of an invertible upper triangular matrix is again upper triangular, and analogously for lower triangular matrices (Section 5.4).

Eigenspaces of Matrices Related to Rotation Matrices: although a rotation matrix itself does not have real eigenvalues unless the rotation is by 0 or π radians, performing the reflection across the x-axis followed by a rotation matrix always leads to real eigenvalues, and a basis for the eigenspaces that involve the half-angle formula (Section 6.1).

Properties Preserved by Similarity: proves that similar matrices share attributes such as determinants, invertibility, arithmetic and geometric multiplicities, and diagonalizability.

Introduction to Fourier Series: shows that the infinite family of trigonometric functions $\{\sin(nx), \cos(nx) \mid n \in \mathbb{N}\}$ are mutually orthogonal under the inner product defined using the integral of their product over $[0, 2\pi]$ (Section 7.3).

De Morgan's Laws for Subspaces: proves that $(V \cap W)^{\perp} = V^{\perp} \vee W^{\perp}$ and $(V \vee W)^{\perp} = V^{\perp} \cap W^{\perp}$, connecting the ideas of the intersection and join of two subspaces with their orthogonal complements (Section 7.4).

Matrix Decompositions: shows that any square matrix can be decomposed uniquely as the sum of a symmetric and a skew-symmetric matrix, and that the spaces of symmetric and skew-symmetric matrices are orthogonal complements of each other under a naturally defined inner product on all square matrices (Section 7.5).

Right Handed versus Left Handed Orthonormal Bases: uses the cross-product to define and create right-handed orthonormal bases for $\mathbb{R}^3$, and relates the concepts of right-handed versus left-handed orthonormal bases to proper versus improper orthogonal matrices (Section 7.6).

Rotations in Space: explicitly constructs the matrix of the counterclockwise rotation by an angle θ about a fixed unit normal vector $\vec{n}$ in $\mathbb{R}^3$ by elegantly connecting this operator with the concepts of a right-handed coordinate system, orthogonal matrices, and the change of basis formula (Section 7.6).

Finite Fields: introduces finite fields by constructing the addition and multiplication tables for the finite fields $\mathbb{Z}/(5)$ and $\mathbb{Z}/(7)$ (Section 8.1).

The Pauli matrices: an introduction to normal matrices that are important in Quantum Mechanics (Section 8.6).

A Note on Technology

The calculations encountered in modern Linear Algebra would be all but impossible to perform in practice, especially on large matrices, without the advent of the computer. Obviously, it would be tedious to perform calculations on these large matrices by hand. However, we do encourage the student to learn the algorithms and computations first, by practicing on the homework problems by hand (with the help of a scientific calculator, at best), before using technology to perform these computations.

It is easy to find free and downloadable software or apps by typing "Linear Algebra Packages" in a search engine. The following computations and algorithms are relevant for this book:

- Matrix Arithmetic: Addition, Multiplication, Inverse, Transpose, Determinant;
- The Gauss-Jordan Algorithm and the Reduced Row Echelon Form or rref;
- Finding a basis for the Rowspace, Columnspace and Nullspace of a Matrix;
- Characteristic Polynomials, Eigenvalues and Bases for Eigenspaces;
- The QR-decomposition;
- The LU-decomposition;
- The Singular Value Decomposition (SVD).

Some graphing calculators also provide many of these routines. We leave it to the instructor to decide whether or not these will be allowed or required in the classroom, homework, or examinations.

To the Student

You are about to embark on a journey that will introduce you to the inner workings of mathematics. So far, Calculus has prepared you to be a whiz at computations. Please keep an open mind, though, as you struggle with a very different skill — learning abstractions, theorems and proofs. Read the text several times (preferably before the lecture), and familiarize yourself with key definitions and theorems connecting these definitions and concepts. The Section Summaries and Chapter Summaries should be very useful in this regard. They are not substitutes, though, for reading the entire text, especially the examples and the proofs of theorems, which I encourage you to ***imitate***. When you are asked to prove a theorem in the exercises, ***identify*** the key words and the key symbols and ***write down*** their precise definitions or meanings. Identify which conditions are given, and what conditions you are trying to prove or show, and then attempt to tie them together into a well-written proof. Be patient with yourself, and don't give up if you haven't given it an honest ***try***. I hope you enjoy this experience, and in the end, I hope that you discover the beauty of mathematics.

Chapter Zero

The Language of Mathematics:

Sets, Axioms, Theorems & Proofs

Mathematics is a language, and Logic is its grammar.

You are taking a course in Linear Algebra because the major that you have chosen will make use of its techniques, both computational and theoretical, at some points in your career. Whether it is in engineering, computer science, chemistry, physics, economics, or of course, mathematics, you will encounter matrices, vector spaces and linear transformations. For most of you, this will be your first experience in an abstract course that emphasizes theory on an almost equal footing with computation.

The purpose of this introductory Chapter is to familiarize you with the basic components of the mathematical language, in particular, the study of sets (especially sets of numbers), subsets, operations on sets, logic, Axioms, Theorems, and basic guidelines on how to write a coherent and logically correct Proof for a Theorem.

Part I: Set Theory and Basic Logic

The ***set*** is the most basic object that we work with in mathematics:

Definition: A ***set*** is an unordered collection of objects, called the ***elements*** of the set. A set can be described using the ***set-builder notation:***

$$X = \left\{ x \,|\, x \text{ possesses certain determinable qualities} \right\},$$

or the ***roster method:***

$$X = \{a, b, \dots\},$$

where we explicitly ***list*** the elements of X. The bar symbol "|" in set-builder notation represents the phrase "such that."

We will agree that such "objects" are already known to exist. They could consist of people, letters of the alphabet, real numbers, or functions. There is also a special set, called the ***empty set*** or the ***null-set***, that does not contain any elements. We represent the empty set symbolically as:

$$\emptyset \quad \text{or} \quad \{\quad\}.$$

Early in life, we learn how to count using the set of ***natural numbers:***

$$\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}.$$

We learn how to add, subtract, multiply and divide these numbers. Eventually, we learn about ***negative integers***, thus completing the set of all ***integers:***

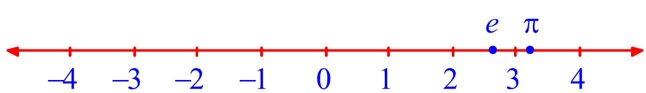
$$\mathbb{Z} = \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

We use the letter $\mathbb{Z}$ from ***Zahlen***, the German word for "number." Later on, we learn that some integers cannot be exactly divided by others, thus producing the concept of a ***fraction*** and the set of ***rational numbers:***

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a \text{ and } b \text{ are integers, with } b \neq 0 \right\}.$$

Notice that we defined $\mathbb{Q}$ using set-builder notation. Still later on, we learn of the number π when we study the circumference and area of a circle. The number π is an ***irrational*** number, although it can be approximated by a fraction like 22/7 or as a decimal like 3.1416. When we learn to take square roots and cube roots, we encounter other examples of irrational numbers, such as $\sqrt{2}$ and $\sqrt[3]{5}$.

By combining the sets of rational and irrational numbers, we get the set of all ***real numbers*** $\mathbb{R}$. We visualize them as corresponding to points on a number line. A point is chosen to be "0," and another point to its right is chosen to be "1." The distance between these two points is the ***unit***, and subsequent integers are marked off using this unit. Real numbers are classified into positive numbers, negative numbers, and zero (which is neither positive nor negative). They are also ordered from left to right by our number line. We show the real number line below along with a couple of famous numbers:



The Real Number Line $\mathbb{R}$

Logical Statements and Axioms

An intelligent development of Set Theory requires us to develop in parallel a ***logical system***. The basic component of such a system is this:

Definition: A ***logical statement*** is a complete sentence that is either ***true*** or ***false***.

Examples: The statement:

The number 2 is an integer.

is a ***true*** logical statement. However:

The number 3/4 is an integer.

is a ***false*** logical statement. The statement:

Gustav Mahler is the greatest composer of all time.

is a sentence but it is ***not*** a logical statement, because the word "greatest" cannot be qualified. Thus, we cannot logically determine if this statement is true or false.□

In everyday life, especially in politics, one person can judge a statement to be true while someone else might decide that it is false. Such judgments depend on one's personal biases, how credible they deem the person who is making the argument, and how they appraise the facts that are carefully chosen (or omitted) to support the case. In mathematics, though, we have a logical system by which to determine the truth or falsehood of a logical statement, so that any two persons using this system will reach the ***same conclusion***. For the sake of sanity, we will need some starting points for our logical process:

> ***Definition:*** An ***Axiom*** is a logical statement that we will ***accept*** as true, that is, as reasonable human beings, we can ***mutually agree*** that such Axioms are true.

You can think of Axioms as analogous to the ***core beliefs*** of a philosophy or religion.

Examples: One of the most important Axioms of mathematics is this:

> The empty set $\emptyset$ exists.

In geometry, we accept as Axioms that ***points*** exist. We symbolize a point with a dot, although it is ***not literally*** a dot. We accept that through two distinct points there must exist a unique ***line***. We accept that any three non-collinear points (that is, three points through which no single line passes) determine a unique ***triangle***. We ***believe*** in the existence of these objects ***axiomatically***. We note, though, that these are Axioms in what we call ***Euclidean Geometry***, but there are other geometric systems that have very different Axioms for points, lines and triangles. $\square$

Quantifiers

Most, if not all of the logical statements that we will encounter in Linear Algebra refer not just to numbers, but also to other objects that we will be constructing, such as vectors and matrices.
We will use what are called ***quantifiers*** in order to specify precisely what kind of object we are referring to:

> ***Definitions — Quantifiers:***
>
> There are two kinds of quantifiers: ***universal*** quantifiers and ***existential*** quantifiers.
>
> Examples of universal quantifiers are the words ***for any, for all*** and ***for every***, symbolized by $\forall$. They are often used in a logical statement to describe ***all*** members of a certain set.
>
> Examples of existential quantifiers are the phrases ***there is*** and ***there exists***, or their plural forms, ***there are*** and ***there exist***, symbolized by $\exists$. Existential quantifiers are often used to claim the existence (or non-existence) of a ***special*** element or elements of a certain set.

Example: In everyday life, we can make the following statement:

Everyone has a mother.

This is certainly a true logical statement. Let us express this statement more precisely using quantifiers:

> ***For every*** human being x, ***there exists*** another human being y who is the mother of x. $\square$

Some of the best examples of logical statements involving quantifiers are found in the Axioms that define the Real Number system. Linear Algebra in a sense is a ***generalization*** of the real numbers, so it is worthwhile to formally study what most of us take for granted.

The Axioms for the Real Numbers

We will assume that the set of real numbers has been ***constructed*** for us, and that this set enjoys certain properties. Furthermore, we will mainly be interested in what are called the Field Axioms:

Axioms — The Field Axioms for the Set of Real Numbers:
There exists a set of Real Numbers, denoted $\mathbb{R}$, together with two binary operations:

$$+ \quad \text{(addition)} \quad \text{and} \quad \cdot \quad \text{(multiplication)}.$$

Furthermore, the members of $\mathbb{R}$ enjoy the following properties:

1. ***The Closure Property of Addition:***

   For all $x, y \in \mathbb{R}$: $x + y \in \mathbb{R}$ as well.

2. ***The Closure Property of Multiplication:***

   For all $x, y \in \mathbb{R}$: $x \cdot y \in \mathbb{R}$ as well.

3. ***The Commutative Property of Addition:***

   For all $x, y \in \mathbb{R}$: $x + y = y + x$.

4. ***The Commutative Property of Multiplication:***

   For all $x, y \in \mathbb{R}$: $x \cdot y = y \cdot x$.

5. ***The Associative Property of Addition:***

   For all $x, y, z \in \mathbb{R}$: $x + (y + z) = (x + y) + z$.

6. ***The Associative Property of Multiplication:***

   For all $x, y, z \in \mathbb{R}$: $x \cdot (y \cdot z) = (x \cdot y) \cdot z$.

7. ***The Distributive Property of Multiplication over Addition:***

   For all $x, y, z \in \mathbb{R}$: $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.

8. ***The Existence of the Additive Identity:***

   There exists $0 \in \mathbb{R}$ such that ***for all*** $x \in \mathbb{R}$: $x + 0 = x = 0 + x$.

9. ***The Existence of the Multiplicative Identity:***

   There exists $1 \in \mathbb{R}$, $1 \neq 0$, such that ***for all*** $x \in \mathbb{R}$: $x \cdot 1 = x = 1 \cdot x$.

10. ***The Existence of Additive Inverses:***

    For all $x \in \mathbb{R}$, ***there exists*** $-x \in \mathbb{R}$, such that: $x + (-x) = 0 = (-x) + x$.

11. ***The Existence of Multiplicative Inverses:***

    For all $x \in \mathbb{R}$, where $x \neq 0$, ***there exists*** $1/x \in \mathbb{R}$, such that:

    $$x \cdot (1/x) = 1 = (1/x) \cdot x.$$

Notice that each of the first seven Axioms begin with the quantifier ***For all***. These Axioms tell us that these properties are valid no matter which two or three real numbers we substitute into the expressions found in that Axiom. On the other hand, Axioms 8 and 9 begin with the quantifier ***There exists***, but in the second phrase, we see the quantifier ***for all***. Axioms 8 and 9 tell us that there are two ***special***, distinct real numbers, 0 and 1, for which two sets of equations are valid ***for all*** real numbers x:

$$x + 0 = x = 0 + x \quad \text{and} \quad x \cdot 1 = x = 1 \cdot x.$$

The numbers 0 and 1 are called ***identities*** because ***every*** $x \in \mathbb{R}$ preserves its identity under the corresponding operation. On the other hand, Axioms 10 and 11 begin with the quantifier ***For all***, but in the second phrase, we see the quantifier ***there exists*** — this is the ***opposite order*** of that found in Axioms 8 and 9. This means that once we ***choose*** x, we can find its additive inverse $-x$, such that:

$$x + (-x) = 0 = (-x) + x.$$

The additive inverse $-x$ ***depends*** on x. Similarly, the reciprocal $1/x$ depends on x, where $x \neq 0$.

We mentioned earlier that we develop our number system by starting with the natural numbers, then constructing negative integers and fractions. After this, though, it is surprisingly difficult to create the full set of real numbers. See Appendix A for a more thorough discussion of how to create numbers, and the complete set of Axioms that the set of real numbers satisfies. These include the ***Order Axioms***, which give us the rules for ***inequalities***, and the ***Completeness Axiom***, which distinguishes the real numbers from the rational numbers. Furthermore, although the 11 Axioms speak only about addition and multiplication, Axioms 10 and 11 allow us to define the related operations of ***subtraction*** and ***division***, and as usual, we will use the notation that is familiar to us:

> ***Definitions — Axioms for Subtraction and Division:***
> For all $x, y \in \mathbb{R}$, define the operation of ***subtraction*** by: $x - y = x + (-y)$.
> Similarly, if $y \neq 0$, define the operation of ***division*** by: $x/y = x \cdot (1/y)$.

Theorems and Implications

Now that we agree that Axioms will be accepted as true, we will be concerned with logical statements which can be ***deduced*** from these Axioms:

> ***Definitions:*** A true logical statement which is not just an Axiom is called a ***Theorem***. Many of the Theorems that we will encounter in Linear Algebra are called ***implications***, and they are of the form: ***if*** p ***then*** q, where p and q are logical statements.
> This implication can also be written symbolically as: $p \Rightarrow q$ (pronounced as: p ***implies*** q).

An implication $p \Rightarrow q$ is true if the statement q is true whenever we know that the statement p is also true. The statements p and q are called ***conditions***. The condition p is called the ***hypothesis*** (or ***antecedent*** or the ***given conditions***), and q is called the ***conclusion*** or the ***consequent***. If such an implication is true, we say that condition p is ***sufficient*** for condition q, and condition q is ***necessary*** for condition p.

Example: In Calculus, we are familiar with the implication:

> ***Theorem:*** ***If*** $f(x)$ is ***differentiable*** at $x = a$, ***then*** $f(x)$ is also ***continuous*** at $x = a$.

Let us use this Theorem to further understand the meaning of the words "necessary" and "sufficient." This Theorem can be interpreted as saying that if we ***want*** $f(x)$ to be continuous at $x = a$, then it is ***sufficient*** that $f(x)$ be differentiable at $x = a$, that is, we have sufficiently paid for the condition of continuity if we have already paid for the ***stricter*** condition of differentiability.

Similarly, if we ***knew*** that $f(x)$ is differentiable at $x = a$, then it is ***necessary*** that $f(x)$ is also continuous at $x = a$: it cannot be discontinuous according to this Theorem.□

Although we will primarily be proving Theorems, it is also important to know when a logical statement is false. An implication $p \Rightarrow q$ can be demonstrated to be false by giving a ***counterexample***, which is a situation where the given condition p is ***true***, but the conclusion q is ***false***.

Example: Let us consider the statement:

If p is a prime number, then $2^p - 1$ is also a prime number.

Recall that an integer $p > 1$ is ***prime*** if the only integers that exactly divide p are 1 and p itself. If we look at the first few prime numbers $p = 2, 3, 5, 7$, we get:

$$\begin{array}{ll}
2^2 - 1 = 4 - 1 = 3 & \text{is prime,} \\
2^3 - 1 = 8 - 1 = 7 & \text{is prime,} \\
2^5 - 1 = 32 - 1 = 31 & \text{is prime, and} \\
2^7 - 1 = 128 - 1 = 127 & \text{is also prime.}
\end{array}$$

This might fool you to believe that the statement is true. However, for $p = 11$, we get:

$$2^{11} - 1 = 2048 - 1 = 2047 = 23 \cdot 89.$$

Thus, we found a ***counterexample*** to the statement above, and so this statement is ***false***.□

In fact, it turns out that the integers of the form $2^p - 1$ where p is a prime number are ***rarely*** prime, and we call such prime numbers ***Mersenne Primes***. As of May 2016, there are only 49 known Mersenne Primes, and the largest of these is $2^{74,207,281} - 1$. This is also the largest known prime number. If this number were expressed in the usual decimal form, it will be 22,338,618 ***digits*** long. Large prime numbers have important applications in ***cryptography***, a field of mathematics which allows us to safely provide personal information such as credit card numbers on the internet.

Negations

> ***Definition:*** The ***negation*** of the logical statement p is written symbolically as: ***not*** p.

The statement ***not*** p is true precisely when p is false, and vice versa. When a negated logical statement is written in plain English, we put the word "not" in a more natural or appropriate place. We can also use related words such as "never" to indicate a negation.

Examples: The statement:

"An integer is ***not*** a rational number."

is a ***false*** logical statement. On the other hand, the statement:

"The function $g(x) = 1/x$ is ***not*** continuous at $x = 0$."

is a ***true*** logical statement.□

Converse, Inverse, Contrapositive and Equivalence

By using negations or reversing the roles of the hypothesis and conclusion, we can construct three implications associated to an implication $p \Rightarrow q$:

> ***Definition:*** For the implication $p \Rightarrow q$, we call:
>
> $q \Rightarrow p$ the ***converse*** of $p \Rightarrow q$,
>
> ***not*** $p \Rightarrow$ ***not*** q the ***inverse*** of $p \Rightarrow q$, and
>
> ***not*** $q \Rightarrow$ ***not*** p the ***contrapositive*** of $p \Rightarrow q$.

Unfortunately, even if we knew that an implication is true, its converse or inverse are not always true.

Example: We saw earlier that the following statement is true:

> "***If*** $f(x)$ is ***differentiable*** at $x = a$, ***then*** $f(x)$ is also ***continuous*** at $x = a$."

The ***converse*** of this statement is:

> "***If*** $f(x)$ is ***continuous*** at $x = a$, ***then*** $f(x)$ is also ***differentiable*** at $x = a$."

This statement is ***false***, as shown by the counterexample $f(x) = |x|$, which is well known to be continuous at $x = 0$, but is ***not differentiable*** at $x = 0$. Similarly, the ***inverse*** of this Theorem is:

> "***If*** $f(x)$ is ***not differentiable*** at $x = a$, ***then*** $f(x)$ is also ***not continuous*** at $x = a$."

The inverse is also ***false***: the same function $f(x) = |x|$ is ***not differentiable*** at $x = 0$, but it ***is*** continuous there. Finally, the ***contrapositive*** of our Theorem is:

> "***If*** $f(x)$ is ***not continuous*** at $x = a$, ***then*** $f(x)$ is also ***not differentiable*** at $x = a$."

The contrapositive is a ***true*** statement: a function which is not continuous cannot be differentiable, because otherwise, it ***has to be*** continuous.□

If we know that $p \Rightarrow q$ and $q \Rightarrow p$ are ***both*** true, then we say that the conditions p and q are ***logically equivalent*** to each other, and we write the ***equivalence*** or ***double-implication:***

$$p \Leftrightarrow q \quad \text{(pronounced as: } p \textbf{\textit{ if and only if }} q\text{).}$$

We saw above that the contrapositive of our Theorem is also true, and in fact, this is no accident. An implication is always logically equivalent to its contrapositive (as proven in Appendix B):

$$(p \Rightarrow q) \Leftrightarrow (\textbf{\textit{not}}\, q \Rightarrow \textbf{\textit{not}}\, p).$$

Later, if we want to prove that the statement $p \Rightarrow q$ is true, we can do so by proving its contrapositive. Similarly, the converse and the inverse of an implication are logically equivalent, and thus they are either both true or both false. We saw this demonstrated above with regards to differentiability versus continuity.

The contrapositive of an equivalence $p \Leftrightarrow q$ is also an equivalence, so we do not have to bother with changing the position of p and q. An equivalence is again equivalent to its contrapositive:

$$(p \Leftrightarrow q) \Leftrightarrow (\textbf{\textit{not}}\, p \Leftrightarrow \textbf{\textit{not}}\, q).$$

Logical Operations

We can combine two logical statements using the common words ***and*** and ***or***:

Definition: If p and q are logical statements, we can form their ***conjunction:***

$$p \textbf{\textit{ and }} q,$$

and their ***disjunction:***

$$p \textbf{\textit{ or }} q.$$

The conjunction p ***and*** q is true precisely if ***both*** conditions p and q are true. Similarly, the disjunction p ***or*** q is true precisely if ***either*** condition p or q is true (or possibly both are true).

Example: The statement:

$\sqrt{2}$ is irrational ***and*** bigger than 1.

is a ***true*** statement. However, the statement:

Every real number is either positive ***or*** negative.

is ***false*** because the real number 0 is neither positive nor negative.□

The negation of a conjunction or a disjunction is sometimes needed in order to understand a Theorem, or more importantly, to prove it. Fortunately, the following Theorem allows us to simplify these compound negations:

Theorem — De Morgan's Laws: For all logical statements p and q:

$$\textbf{\textit{not}}\ (p\ \textbf{\textit{and}}\ q) \Leftrightarrow (\textbf{\textit{not}}\ p)\ \textbf{\textit{or}}\ (\textbf{\textit{not}}\ q), \quad \text{and}$$

$$\textbf{\textit{not}}\ (p\ \textbf{\textit{or}}\ q) \Leftrightarrow (\textbf{\textit{not}}\ p)\ \textbf{\textit{and}}\ (\textbf{\textit{not}}\ q).$$

Note that De Morgan's Laws look very similar to the Distributive Property (with a slight twist), and in fact they are precisely that in the study of ***Boolean Algebras***.

De Morgan's Laws are proven in Appendix B.

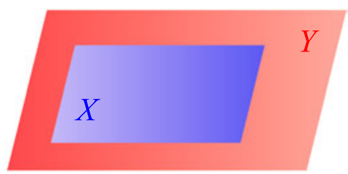
Subsets and Set Operations

We can compare two sets and perform operations on two sets to create new sets.

Definitions: We say that a set X is a ***subset*** of another set Y if every member of X is also a member of Y. We write this symbolically as:

$$X \subseteq Y \Leftrightarrow (x \in X \Rightarrow x \in Y).$$

If X is a subset of Y, we can also say that X is ***contained*** in Y, or Y ***contains*** X. We can visualize sets and subsets using ***Venn Diagrams*** as follows:



We say X ***equals*** Y if X is a subset of Y and Y is a subset of X:

$$(X = Y) \Leftrightarrow \left(X \subseteq Y\ \textbf{\textit{and}}\ Y \subseteq X\right).$$

Equivalently, every member of X is also a member of Y, and every member of Y is also a member of X:

$$(X = Y) \Leftrightarrow \left(x \in X \Rightarrow x \in Y\ \textbf{\textit{and}}\ y \in Y \Rightarrow y \in X\right).$$

We combine two sets into a single set that contains precisely all the members of the two sets using the ***union*** operation:

$$X \cup Y = \left\{z \middle| z \in X\ \textbf{\textit{or}}\ z \in Y\right\}.$$

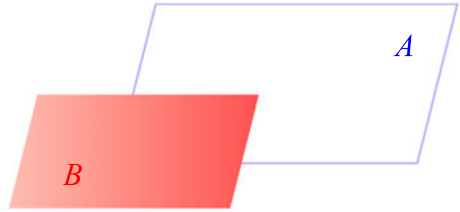
We determine all members common to both sets using the ***intersection*** operation:

$$X \cap Y = \left\{z \middle| z \in X\ \textbf{\textit{and}}\ z \in Y\right\}.$$

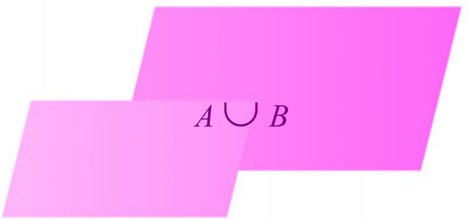
We can also take the ***difference*** or ***complement*** of two sets:

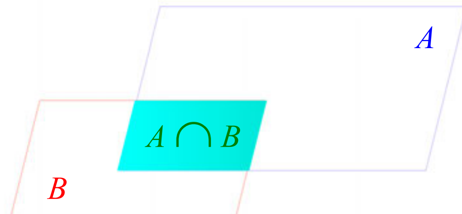
$$X - Y = \left\{z \middle| z \in X\ \textbf{\textit{and}}\ z \notin Y\right\}.$$

Notice the use of ***or*** and ***and*** in the definitions. We can also visualize these set operations using Venn diagrams. We first show two sets A and B below, highlighted separately for clarity:

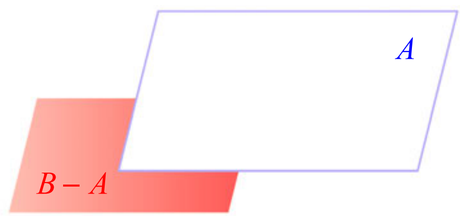


Next, we show their union $A \cup B$, and their intersection $A \cap B$:





Finally, we show the two complements, $A - B$ and $B - A$:



Example: Suppose we have the sets (expressed in roster notation):

$$
\begin{aligned}
A &= \{b, d, e\}, \\
B &= \{a, b, c, d, e, f\}, \\
C &= \{c, e, h, k\}, \quad \text{and} \\
D &= \{d, e, g, k\}.
\end{aligned}
$$

Then $A \subseteq B$ because every member of A is also a member of B, and there are no other subset relationships among the four sets. Now, let us compute the following set operations:

$$
\begin{aligned}
C \cup D &= \{c, e, h, k\} \cup \{d, e, g, k\} = \{c, d, e, g, h, k\}, \\
C \cap D &= \{c, e, h, k\} \cap \{d, e, g, k\} = \{e, k\}, \\
C - D &= \{c, e, h, k\} - \{d, e, g, k\} = \{c, h\}, \quad \text{and} \\
D - C &= \{d, e, g, k\} - \{c, e, h, k\} = \{d, g\}.
\end{aligned}
$$

As a special bonus, notice that:

$$C \cup D = \{c, d, e, g, h, k\} = \{e, k\} \cup \{c, h\} \cup \{d, g\} = (C \cap D) \cup (C - D) \cup (D - C). \square$$

In the course of developing Linear Algebra, we will not just consider sets of real numbers, but also sets of ***vectors***, notably the Euclidean Spaces from Chapter 1, sets of ***polynomials***, and more generally, sets of ***functions*** (such as continuous functions and differentiable functions), and sets of ***matrices***. We will be gradually constructing these objects over time.

Part II: Proofs

Perhaps the most challenging task that you will be asked to do in Linear Algebra is to prove a Theorem. To accomplish this, you need to know what is expected of you:

> ***Definition:*** A ***proof*** for a Theorem is a sequence of true logical statements which ***convincingly*** and ***completely explains*** why a Theorem is true.

In many ways, a proof is very similar to an ***essay*** that you write for a course in Literature or History. It is also similar to a ***laboratory report***, say in Physics or Chemistry, where you have to logically analyze your data and defend your conclusions.

The main difference, though, is that every logical statement in a proof should be true, and must follow as a conclusion from a previously established true statement.

The method of reasoning that we will use is a method of deductive reasoning which is formally called ***modus ponens***. It basically works like this:

> Suppose you already know that an implication $p \Rightarrow q$ is true.
>
> Suppose you also established that condition p is satisfied.
>
> Therefore, it is logical to conclude that condition q is also satisfied.

Example: Let us demonstrate modus ponens on the following logical argument:

> In Calculus, we proved that:
> ***if*** $f(x)$ is a continuous odd function on $[-a, a]$, ***then*** $\int_{-a}^{a} f(x)\,dx = 0$.
>
> The function $f(x) = \sin^5(x)$ is continuous on $\mathbb{R}$, because it is the composition of two continuous functions. It is an odd function on $[-\pi/4, \pi/4]$, since:
>
> $$\sin^5(-x) = (-\sin(x))^5 = -\sin^5(x),$$
>
> where we used the odd property of both the sine function and the fifth power function.
>
> Therefore, $\int_{-\pi/4}^{\pi/4} \sin^5(x)\,dx = 0.$ □

Notice that this reasoning allows us to compute this definite integral without the inconvenience of finding an antiderivative and applying the Fundamental Theorem of Calculus!

A proof often begins by understanding the ***meaning*** of the given conditions and the conclusion that you are supposed to reach. It is therefore important that you can recall and state the ***definitions*** of a variety of ***words*** and ***phrases*** that you will encounter in your study of Linear Algebra. After all, it would be impossible for you to explain how you obtained your conclusion if you do not even know what the conclusion is supposed to mean. We also use special ***symbols*** and ***notation***, so you must be familiar with them. Often, a previously proven Theorem can also be helpful to prove another Theorem. Start by ***identifying*** what is ***given*** (the hypotheses), and what it is that we want to ***show*** (the conclusion).

Rest assured, you will be shown examples which demonstrate proper techniques and reasoning, which you are encouraged to emulate as you learn and develop your own style. In the meantime, we present below some examples of general ***strategies*** and ***techniques*** which will be useful in the coming Chapters. These strategies are certainly not exhaustive: we sometimes combine several strategies to prove a Theorem, and the more difficult Theorems require a ***creative spark***. For our first example, though, let us see how to prove a Theorem using only the Axioms of the Real Number System:

Example: Let us prove the following:

Theorem — The Multiplicative Property of Zero: For all $a \in \mathbb{R}$:

$$0 \cdot a = 0 = a \cdot 0.$$

Proof: Suppose that a is any real number. We want to show that $0 \cdot a = 0$. If we can do this, then we can also conclude by the commutative property of multiplication that $a \cdot 0 = 0$ as well.

We will use a clever idea. We know the ***Identity Property*** of 0, that is, for all $x \in \mathbb{R}$:

$$0 + x = x = x + 0.$$

Since this is true for all real x, it is true in particular for $x = 0$, so we get:

$$0 + 0 = 0.$$

Now, if we multiply both sides of this equation by a, we get the equation:

$$(0 + 0) \cdot a = 0 \cdot a.$$

This equation is again a true equation because of the following Axiom:

Axiom — The Substitution Principle:
If $x, y \in \mathbb{R}$ and $F(x)$ is an arithmetic expression involving x, and $x = y$, then $F(x) = F(y)$.

Simply put, if two quantities are the same, and we do the same arithmetic operations to both quantities, then the resulting quantities are still the same. Continuing now, by the ***Distributive Property***, we get:

$$0 \cdot a + 0 \cdot a = 0 \cdot a.$$

Remember that we want to know exactly what $0 \cdot a$ is. All we know is that $0 \cdot a$ is ***some*** real number, by the ***Closure Property of Multiplication***. Thus it possesses an additive inverse, $-(0 \cdot a)$, by the ***Existence of Additive Inverses***. Let us add this to both sides of the equation:

$$-(0 \cdot a) + (0 \cdot a + 0 \cdot a) = -(0 \cdot a) + 0 \cdot a.$$

By the defining property of the additive inverse, $-(0 \cdot a) + 0 \cdot a = 0$, so we get:

$$-(0 \cdot a) + (0 \cdot a + 0 \cdot a) = 0.$$

But now, by the ***Associative Property of Addition***, the left side is:

$$(-(0 \cdot a) + 0 \cdot a) + 0 \cdot a = 0.$$

Thus, by the additive inverse property, as above, we get:

$$0 + (0 \cdot a) = 0.$$

(we enclosed $0 \cdot a$ in parentheses to emphasize that it is the quantity we are trying to study in our equation). Finally, by the additive property of 0 again, the left side reduces to $0 \cdot a$, so we get:

$$0 \cdot a = 0. \blacksquare$$

Case-by-Case Analysis

We can prove the implication $p \Rightarrow q$ if we can break down p into two or more cases, and every possibility for p is covered by at least one of the cases. If we can prove that q is true in ***each case***, the implication is true. This is also sometimes called ***Proof by Exhaustion***.

Example: Let us prove the following:

> ***Theorem — The Zero-Factors Theorem:*** For all $a,\ b \in \mathbb{R}$:
>
> $a \cdot b = 0$ ***if and only if*** either $a = 0$ ***or*** $b = 0$.

Proof: Since this is an ***if and only if*** Theorem, we must prove two implications. Let us begin with the converse, which is easier:

($\Leftarrow$) Suppose we are given that either $a = 0$ ***or*** $b = 0$. We must show that $a \cdot b = 0$. Since there are two possibilities for the given conditions, we have the following cases:

Case 1. If $a = 0$, then $a \cdot b = 0 \cdot b = 0$ by our previous Theorem.

Case 2. If $b = 0$, then $a \cdot b = a \cdot 0 = 0$, which is the exact same reasoning as Case 1.

Thus, if either $a = 0$ ***or*** $b = 0$, then $a \cdot b = 0$.

($\Rightarrow$) Suppose we are given that $a \cdot b = 0$. We must show that either $a = 0$ ***or*** $b = 0$.

Case 1. Suppose that $a = 0$. Then we are done, since the conclusion "$a = 0$ ***or*** $b = 0$" is satisfied.

Case 2. Suppose that $a \neq 0$. Notice that since this is the exact opposite of Case 1, we have covered ***all*** the possibilities. Now, since a is non-zero, by Axiom 11, it has a ***Multiplicative Inverse*** $1/a$. We are given that:

$$a \cdot b = 0.$$

By ***The Substitution Principle***, we can multiply both sides of the equation by $1/a$ and obtain:

$$(1/a) \cdot (a \cdot b) = (1/a) \cdot 0.$$

Since $1/a$ is again another real number, the right side of this equation is 0, as we already saw above. Now, the left side can be regrouped using Axiom 6, the ***Associative Property of Multiplication***. Thus, we get:

$$(1/a \cdot a) \cdot b = 0.$$

By the ***Multiplicative Inverse Property***, the product of a non-zero number and its reciprocal is 1, so we obtain: $1 \cdot b = 0$.

Finally, by Axiom 9, $1 \cdot b = b$, and thus we get: $b = 1 \cdot b = 0$.

Thus, if $a \neq 0$, then $b = 0$, completing the proof that either $a = 0$ ***or*** $b = 0$.

Notice that the two Cases for the forward implication are different from the two Cases for the converse. This frequently happens. $\blacksquare$

Proof by Contrapositive

We mentioned earlier that an implication $p \Rightarrow q$ is logically equivalent to its contrapositive, which is ***not*** $q \Rightarrow$ ***not*** p. Thus it may be worthwhile to write down the contrapositive of the Theorem we want to prove, and see if we get any ideas on how to prove it. This is the basic idea behind the technique called ***Proof by Contrapositive***, which is also known in Latin as ***modus tollens***.

The example we will discuss below deals with the set of integers, $\mathbb{Z}$. In order to fully appreciate this example, we need to introduce the following Axioms for $\mathbb{Z}$:

> ***Axioms — Closure Axioms for the Set of Integers:***
> If $a, b \in \mathbb{Z}$, then $a + b \in \mathbb{Z}$, $a - b \in \mathbb{Z}$, and $a \cdot b \in \mathbb{Z}$ as well.

> ***Definitions — Even and Odd Integers:***
> An integer $a \in \mathbb{Z}$ is ***even*** if there exists $c \in \mathbb{Z}$ such that $a = 2c$.
> An integer $b \in \mathbb{Z}$ is ***odd*** if there exists $d \in \mathbb{Z}$ such that $a = 2d + 1$.

It is easy to see from these two definitions that every integer is either even or odd, but not both. Now we are ready:

Example: Let us prove the following using the technique of Proof by Contrapositive:

> ***Theorem:*** For all $a, b \in \mathbb{Z}$:
> If the product $a \cdot b$ is ***odd***, then ***both*** a and b are ***odd***.

Proof: Our first step is to write the contrapositive. The conclusion is "both a and b are odd." Since the word ***and*** is in this phrase, we can use De Morgan's Laws to simplify its negation:

not (both a ***and*** b are odd) $\Leftrightarrow$

(a is ***not*** odd) ***or*** (b is ***not*** odd) $\Leftrightarrow$

a is even ***or*** b is even.

Thus, the contrapositive of the Theorem we want to prove is:

> ***Theorem:*** For all $a, b \in \mathbb{Z}$:
> If a is even ***or*** b is even, then $a \cdot b$ is even.

This statement is easier to prove, and all we need is a Case-by-Case analysis:

Case 1. Suppose that a is even. Then a has the form $a = 2 \cdot c$ for some integer c. Thus:

$$a \cdot b = (2 \cdot c) \cdot b = 2 \cdot (c \cdot b),$$

by the Associative Property of Multiplication. Since $c \cdot b \in \mathbb{Z}$ by Closure, $2 \cdot (c \cdot b)$ is even. Thus, $a \cdot b$ is even. A similar argument works for ***Case 2***, where we assume that b is even. ■

Proof by Contradiction

The method of ***Proof by Contradiction*** (or ***reductio ad absurdum***) is often used in order to show that an object does not exist, or in situations when it is difficult to show that an implication is true directly. The idea is to assume that the mythical object does exist, or more generally, the opposite of the conclusion is true. In the course of our reasoning, we should arrive at a condition which ***contradicts*** one of the given conditions, or a condition which has already been concluded to be true (thus producing an ***absurdity*** or contradiction). The only problem with attempting a proof by contradiction is that it is ***not guaranteed*** that you will eventually encounter a contradiction. As in all techniques, give it a ***try***.

Example: One of the best applications of Proof by Contradiction is the classic proof of the following:

> ***Theorem:*** The real number $\sqrt{2}$ is ***irrational***.

Proof: Let us assume the ***opposite*** of the conclusion, that is, $\sqrt{2}$ is ***rational***. Thus, we can write:

$$\sqrt{2} = \frac{a}{b}, \quad \text{where } a \text{ and } b \text{ are positive } \textbf{\textit{integers}}.$$

We must make one important requirement to make the proof work: recall from basic Arithmetic that every fraction can be reduced to ***lowest form***, so we will require that a and b have ***no common factor*** except of course for 1. Now, squaring both sides of this equation, we get: $2 = a^2/b^2$, or $a^2 = 2b^2$.

This last equation tells us that a^2 must be an ***even*** number. But if a^2 is even, then a itself has to be even. To see this convincingly, we can also use Proof by Contradiction: if a^2 were even but a were odd, then $a = 2d + 1$ for some integer d, and we get:

$$a^2 = (2d+1)^2 = 4d^2 + 4d + 1 = 2(2d^2 + 2d) + 1.$$

Since $2d^2 + 2d$ is an integer, a^2 is ***odd***. Thus, we get a contradiction, and so a must be even. Now, we can write $a = 2m$, where m is an integer, and substituting this in the equation $a^2 = 2b^2$, we get:

$$(2m)^2 = 2b^2 \quad \text{or} \quad 4m^2 = 2b^2 \quad \text{or} \quad b^2 = 2m^2.$$

Thus b^2 is also even, and by the same reasoning above, b itself must be even. Therefore, the equation $\sqrt{2} = a/b$ led us to the conclusion that ***both*** a and b are even. This violates the requirement that a and b have no common factor aside from 1. We have reached a contradiction, and so our assumption that $\sqrt{2}$ is rational had to be ***false***, and so its opposite is ***true***: $\sqrt{2}$ must be ***irrational***.■

Proof by Induction

Another technique which is useful in Linear Algebra is the ***Principle of Mathematical Induction***. The Theorems that "induction" (as it is more briefly called) applies to are often about natural numbers or positive integers. Since this statement refers to an integer n, we often write the statement as $\boldsymbol{p(n)}$. As this is seen in Precalculus, let us use an example to review how this technique works.

Example: Use the Principle of Mathematical Induction to prove the following formula:

> ***Theorem:*** For all positive integers n: $1^2 + 2^2 + \cdots + n^2 = \dfrac{n(n+1)(2n+1)}{6}$.

Proof: Induction is accomplished in three major steps:

1. The Basis Step. We will first prove that the statement is true when $n = 1$, that is, $p(1)$ is true. The left side of the equation thus stops at 1^2. The right side is:

$$\frac{1 \cdot (1+1) \cdot (2 \cdot 1 + 1)}{6} = \frac{1 \cdot 2 \cdot 3}{6} = 1,$$

so $p(1)$ is indeed true.

2. The Inductive Hypothesis. In this step, we will simply ***assume*** that the statement is true when n is some positive integer k. In other words, we assume that $p(k)$ is true.

Thus, we rewrite the equation in the statement by replacing n with k:

The Inductive Hypothesis: Assume:

$$1^2 + 2^2 + \cdots + k^2 = \frac{k(k+1)(2k+1)}{6}.$$

Notice that since we have already done Step 1, we have the right to make this assumption, because we have proven it to be true for at least one instance: $k = 1$.

3. The Inductive Step. This is of course where most of the hard work comes in. We must now show that the statement is still true when $n = k+1$, or in other words, that $p(k+1)$ is true.

We begin this step by stating $p(k+1)$, so that we explicitly see what it is we need to prove. Thus, we replace n with $k+1$ (in this case, ***four*** times):

The Inductive Step: Prove:

$$1^2 + 2^2 + \cdots + k^2 + (k+1)^2 = \frac{(k+1)(k+1+1)(2[k+1]+1)}{6}.$$

Notice that the left side of the equation now has ***one more term*** at the end. Now, we can proceed to prove that this equation is true. The Inductive Hypothesis tells us that the first k terms on the left side of this equation can be replaced, as follows:

$$\begin{aligned}
&1^2 + 2^2 + \cdots + k^2 + (k+1)^2 \\
&= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 && \text{(by the Inductive Hypothesis)} \\
&= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} && \text{(combining fractions)} \\
&= \frac{(k+1)[k(2k+1) + 6(k+1)]}{6} && \text{(factoring out } k+1\text{)} \\
&= \frac{(k+1)[2k^2 + k + 6k + 6]}{6} && \text{(distributing the parentheses)} \\
&= \frac{(k+1)(2k^2 + 7k + 6)}{6}.
\end{aligned}$$

However, the right side of our equation in the Inductive Step is:

$$\begin{aligned} &\frac{(k+1)(k+1+1)(2[k+1]+1)}{6} \\ &= \frac{(k+1)(k+2)(2k+3)}{6} && \text{(simplifying)} \\ &= \frac{(k+1)(2k^2+4k+3k+6)}{6} && \text{(distributing two factors)} \\ &= \frac{(k+1)(2k^2+7k+6)}{6}, \end{aligned}$$

thus proving that both sides of $p(k+1)$ are the same. This completes our Proof by Induction.■

Why does this reasoning make sense? We were able to show that the Theorem is true if $n = 1$. If we put Steps 2 and 3 together, then we know that if the statement is true when $n = k$, then it is also true when $n = k+1$. Since we knew that the statement was true when $n = 1$, by modus ponens, it is also true when $n = 2$. But now that we know it is also true when $n = 2$, again, by modus ponens, it is also true when $n = 3$. And so on!

Conjectures and Demonstrations

It might shock you to know that there are ***many*** statements in mathematics which have not been determined to be true or false. They are called ***conjectures***. However, we can try to ***demonstrate*** that it is ***plausible*** for the conjecture to be true by giving examples where the conjecture is satisfied. These demonstrations are ***not*** replacements for a complete proof.

Example: Perhaps the most famous, and certainly one of the oldest and most easily stated conjectures of mathematics is called ***Goldbach's Conjecture***. It was stated in 1742 by the Prussian mathematician Christian Goldbach, in a letter to the great Leonhard Euler. The modern statement is as follows:

> ***Goldbach's Conjecture:*** Every ***even integer*** bigger than 2 can be expressed as the sum of ***two prime numbers***.

We can demonstrate that this conjecture is plausible with the examples:

$$18 = 13 + 5 \quad \text{and} \quad 50 = 3 + 47.$$

Goldbach's Conjecture has been verified for a large range of positive even numbers, but experts feel that we are still a long way from proving it in general.□

Unfortunately, most modern conjectures cannot be understood unless one has spent years studying the background material of their associated fields. Their pursuit falls within the realm of ***mathematical research***. As you learn to understand and prove basic Theorems in Linear Algebra, your skills in learning to read Theorems and prove Theorems on your own will improve over time. It is possible that someday, you will prove a deep and complicated Theorem that nobody has ever proven before.

Chapter Zero Summary:

A ***set*** is an unordered collection of objects called ***elements***. Important sets include the empty set ϕ, the sets of natural numbers $\mathbb{N}$, integers $\mathbb{Z}$, rational numbers $\mathbb{Q}$, and real numbers $\mathbb{R}$.

A ***logical statement*** is a sentence which can be determined to be either ***true*** or ***false***. An ***Axiom*** is a logical statement that we will ***accept*** as true. The ***negation*** of the logical statement p, written as ***not*** p, is true exactly when p is false.

Universal quantifiers are the words ***for any, for all*** and ***for every***. ***Existential quantifiers*** are the phrases ***there is*** and ***there exists*** or their plural forms ***there are*** and ***there exist***.

The ***Field Axioms*** for the set of Real Numbers describe eleven important properties that we agree the set of real numbers possesses.

A true logical statement which is not just an Axiom is called a ***Theorem***. An ***implication*** has the form: ***if*** p ***then*** q, written symbolically as $p \Rightarrow q$. An implication can be demonstrated to be ***false*** by giving a ***counterexample***, a situation where p is true, but q is false.

The ***negation*** of the logical statement p, written as ***not*** p, is true exactly when p is false.

For an implication $p \Rightarrow q$, we call $q \Rightarrow p$ the ***converse*** of $p \Rightarrow q$, ***not*** $p \Rightarrow$ ***not*** q the ***inverse*** of $p \Rightarrow q$, and ***not*** $q \Rightarrow$ ***not*** p the ***contrapositive*** of $p \Rightarrow q$.

If $p \Rightarrow q$ and $q \Rightarrow p$ are ***both*** true, then we say that p and q are ***equivalent*** to each other. We write the ***equivalence*** or ***double-implication*** $p \Leftrightarrow q$, pronounced as p ***if and only if*** q.

The implication $p \Rightarrow q$ is equivalent to its contrapositive ***not*** $q \Rightarrow$ ***not*** p.

The ***conjunction*** p ***and*** q is true precisely if ***both*** conditions p and q are true.

The ***disjunction*** p ***or*** q is true precisely if ***either*** condition p or q is true.

De Morgan's Laws: For all logical statements p and q: ***not*** $(p \textbf{ and } q)$ is logically equivalent to $(\textbf{not } p)$ ***or*** $(\textbf{not } q)$, and similarly, ***not*** $(p \textbf{ or } q)$ is logically equivalent to $(\textbf{not } p)$ ***and*** $(\textbf{not } q)$.

A set X is a ***subset*** of another set Y if every member of X is also a member of Y. We write this symbolically as $X \subseteq Y$. Two sets X and Y are ***equal*** if X is a subset of Y and Y is a subset of X, or equivalently, every member of X is also a member of Y, and vice versa:

$$(X = Y) \Leftrightarrow \left(X \subseteq Y \textbf{ and } Y \subseteq X\right) \Leftrightarrow \left(x \in X \Rightarrow x \in Y \textbf{ and } y \in Y \Rightarrow y \in X\right).$$

Given two sets X and Y, we can find:

- their ***union:*** $X \cup Y = \left\{z \mid z \in X \textbf{ or } z \in Y\right\}$;
- their ***intersection:*** $X \cap Y = \left\{z \mid z \in X \textbf{ and } z \in Y\right\}$; and
- their ***difference*** or ***complement:*** $X - Y = \left\{z \mid z \in X \textbf{ and } z \notin Y\right\}$.

A ***proof*** for a Theorem is a sequence of true logical statements which ***convincingly*** and ***completely explains*** why a Theorem is true.

A good way to begin a proof is by identifying the given conditions and the conclusion that we want to show. It is also a good idea to write down definitions for terms that are found in the Theorem. The main logical technique in writing proofs is ***modus ponens***. We also use techniques such as:

- Case-by-Case Analysis
- Proof by Contrapositive
- Proof by Contradiction
- Proof by Mathematical Induction.

Chapter Zero Exercises

For Exercises (1) to (6): Decide if the following statements are logical statements or not, and if a statement is logical, classify it as True or False.

1. If x is a real number and $|x| < 3$, then $-3 < x < 3$.
2. If x and y are real numbers and $x < y$, then $x^2 < y^2$.
3. If x and y are real numbers and $0 < x < y$, then $1/y < 1/x$.
4. Every real number has a square root which is also a real number.
5. As of March 2016, Roger Federer holds the record for the most number of consecutive weeks as the world's number 1 tennis player.
6. The Golden State Warriors are the best team in the NBA. Why is this different from Exercise 5?

For Exercises (7) to (10): Write the converse, inverse and contrapositive of the following:

7. If you do your homework before dinner, you can watch TV tonight.
8. If it rains tomorrow, we will not go to the beach.
9. If $0 \le x \le \pi/2$, then $\cos(x) \ge 0$. (challenge: write the inverse and contrapositive without using the word "not")
10. If $f(x)$ is continuous on the closed interval $[a,b]$ then $f(x)$ possesses both a maximum and a minimum on $[a,b]$.

For Exercises (11) and (12): For the sets A and B, find $A \cup B$, $A \cap B$, $A - B$ and $B - A$:

11. $A = \{a,c,f,h,i,j,m\}$, $B = \{b,c,g,h,j,p,q\}$.
12. $A = \{a,d,g,h,j,p,r,t\}$, $B = \{b,d,g,h,k,p,q,s,t,v\}$.

For Exercises (13) to (22): Prove the following Theorems concerning Real Numbers using only the 11 Field Axioms (and possibly Theorems that were proven in Chapter Zero). Specify in your proof which Axiom or Theorem you are using at ***each*** step.

13. Prove ***The Cancellation Law for Addition:*** For all $x,\ y,\ c \in \mathbb{R}$:
$$\text{If } x + c = y + c, \text{ then } x = y.$$
14. Prove ***The Cancellation Law for Multiplication:*** For all $x,\ y,\ k \in \mathbb{R},\ k \neq 0$:
$$\text{If } k \cdot x = k \cdot y, \text{ then } x = y.$$
15. Use The Multiplicative Property of Zero to prove that 0 ***cannot*** have a multiplicative inverse. Hint: Use Proof by Contradiction: Suppose 0 ***has*** a multiplicative inverse x. . .
16. Prove ***The Uniqueness of Additive Inverses:*** Suppose $x \in \mathbb{R}$. If $w \in \mathbb{R}$ is any real number with the property that $x + w = 0 = w + x$, then $w = -x$. In other words, $-x$ is the only real number that satisfies the above equations.
17. Use the previous Exercise to show that $-0 = 0$. Hint: which Field Axiom tells us what $0 + 0$ is?
18. Use the Uniqueness of Additive Inverses to prove that for all $x \in \mathbb{R}$: $-x = (-1) \cdot x$.
Hint: simplify $x + (-1) \cdot x$.
19. ***The Double Negation Property:*** Use some of the previous Exercises to show that: For all $x \in \mathbb{R}$: $-(-x) = x$.
20. Prove ***The Uniqueness of Multiplicative Inverses:*** Suppose $x \in \mathbb{R}$ and $x \neq 0$. If $y \in \mathbb{R}$ is any real number with the property that $x \cdot y = 1 = y \cdot x$, then $y = 1/x$. In other $1/x$ is the only real number that satisfies the above equations.

21. Prove ***The Double Reciprocal Property:*** For all $x \in \mathbb{R}$, $x \neq 0$: $1/(1/x) = x$.
22. ***Solving Algebraic Equations:*** Prove that for all $x, a, b \in \mathbb{R}$:
    a. if $x + a = b$, then $x = b - a$.
    b. if $a \neq 0$ and $ax = b$, then $x = b/a$.
23. Prove by Contradiction that there is no ***largest*** positive real number.
24. Prove by Contradiction that there is no ***smallest*** positive real number.
25. Suppose that $n \in \mathbb{Z}$ and n factors as $n = a \cdot b$, where $a, b \in \mathbb{Z}$ and both are positive. Use Proof by Contradiction to show that either $a \leq \sqrt{n}$ or $b \leq \sqrt{n}$.
26. Use the previous Exercise to prove: If n is ***not*** a prime number (that is, n is ***composite***), then n has a prime factor which is at most $\sqrt{n}$.
27. Write the contrapositive of the statement in the previous Exercise. Use this to decide if 11303 is prime or composite.

For Exercises (28) to (31): Use the technique of Proof by Contrapositive to prove the following statements. You may use De Morgan's Law to simplify the contrapositive, when applicable:

28. For all $a, b \in \mathbb{Z}$: if $a \cdot b$ is even, then either a is even ***or*** b is even.
29. For all $a, b \in \mathbb{Z}$: if $a + b$ is even, then either a and b are both odd ***or*** both even.
30. For all $a \in \mathbb{Z}$: a^2 is even ***if and only if*** a is even.
31. For all $x, y \in \mathbb{R}$: if $x \cdot y$ is irrational, then either a is irrational ***or*** b is irrational.

Negating Statements with Quantifiers: A logical statement that begins with a quantifier is negated as follows: ***not*** $(\forall x : p)$ is equivalent to: $\exists x :$ ***not*** (p). This should make sense: if it is ***not true*** that ***all*** x possess property p, then ***at least one*** x does ***not*** possess property p.

Similarly: ***not*** $(\exists x : p)$ is equivalent to: $\forall x :$ ***not*** (p).

Thus, the negation of "All of my friends are Democrats" is "One of my friends is not a Democrat." Notice that "None of my friends are Democrats" is wrong.

Similarly, the negation of "One of my brothers is left-handed" is "All of my brothers are right-handed." It is not "One of my brothers is right-handed."

For Exercises (32) to (35): Write the negation of the following statements, and determine whether the original statement or its negation is true:

32. Every real number x has a multiplicative inverse $1/x$.
33. There exists a real number x such that $x^2 < 0$.
34. There exists a negative number x such that $x^2 = 4$.
35. All prime numbers are odd.
36. Demonstrate Goldbach's Conjecture using: $130 = ? + ?$
37. Rewrite Goldbach's Conjecture using the quantifiers "for every" and "there exist."
38. ***The Twin Prime Conjecture:*** Twin primes are pairs of prime numbers that differ only by 2. For example, $(11, 13)$ are twin primes, as are $(41, 43)$. The Twin Prime Conjecture states that there are an ***infinite*** number of twin primes. What are the next years after 2016 that are twin primes?
39. ***The Fibonacci Prime Conjecture:*** The ***Fibonacci Numbers*** are those in the infinite sequence:

$$0,\ 1,\ 1,\ 2,\ 3,\ 5,\ 8,\ 13,\ 21,\ 34,\ 55,\ 89,\ \ldots$$

where the next number (starting with the third) is the sum of the previous two numbers. Notice that 2, 3, 5, 13 and 89 are primes that appear in this sequence, so they are called ***Fibonacci Primes***. The Fibonacci Prime Conjecture states that there are an infinite number of Fibonacci primes. Find the next Fibonacci prime after 89.

For Exercises (40) to (49): Prove the following by Mathematical Induction:
For all positive integers n:

40. $1^2 + 3^2 + \cdots + (2n-1)^2 = \dfrac{n(2n+1)(2n-1)}{3}$
41. $1^3 + 2^3 + \cdots + n^3 = \left[\dfrac{n(n+1)}{2}\right]^2$
42. $1^3 + 3^3 + \cdots + (2n-1)^3 = n^2(2n^2-1)$
43. $1 \cdot 2 + 2 \cdot 3 + \cdots + n \cdot (n+1) = \dfrac{n(n+1)(n+2)}{3}$
44. $1 \cdot 3 + 2 \cdot 4 + 3 \cdot 5 + \cdots + n(n+2) = \dfrac{n(n+1)(2n+7)}{6}$
45. $\dfrac{1}{1 \cdot 2} + \dfrac{1}{2 \cdot 3} + \cdots + \dfrac{1}{n \cdot (n+1)} = \dfrac{n}{n+1}$
46. $\dfrac{1}{1 \cdot 3} + \dfrac{1}{3 \cdot 5} + \cdots + \dfrac{1}{(2n-1) \cdot (2n+1)} = \dfrac{n}{2n+1}$
47. $1 \cdot 2 + 2 \cdot 2^2 + 3 \cdot 2^3 + \cdots + n \cdot 2^n = 2[(n-1)2^n + 1]$
48. $1 \cdot 3 + 2 \cdot 3^2 + 3 \cdot 3^3 + \cdots + n \cdot 3^n = \frac{3}{4}[(2n-1)3^n + 1]$
49. $n < 2^n$ (this might require a little bit of creativity in Step 3).
50. An ***n-gon*** is a polygon with n vertices (thus a triangle is a 3-*gon* and a quadrilateral is a 4-*gon*). We know from basic geometry that the sum of the angles of any triangle is 180^0. Use Induction to prove that the sum of the interior angles of a convex n-*gon* is $(n-2) \cdot 180^0$ (a polygon is ***convex*** if any line segment connecting two points inside the polygon is entirely within the polygon). Hint: in the inductive step, cut out a triangle using three consecutive vertices. Draw some pictures.
51. Suppose that A and B are subsets of X. Prove that $A \cap B$ is the ***largest*** subset of X which is ***contained*** in both A and B. In other words, prove that if $C \subset A$ and $C \subset B$, then $C \subset A \cap B$.
52. Suppose that A and B are any two subsets of a set X. Prove that $A \cup B$ is the ***smallest*** subset of X which ***contains*** both A and B. In other words, prove that if $A \subset D$ and $B \subset D$, then $A \cup B \subset D$.
53. Suppose that A and B are any two sets. Prove that (a) $(A - B) \cap B = \emptyset$, and (b) $A \cup B = (A \cap B) \cup (B - A) \cup (A - B)$, and each of the three sets in this union have no element in common with the other two. Hint: draw a diagram.
54. ***Properties of Set Union and Intersection:***
    a. If X and Y are two sets, write down the definition of $X \cup Y$.
    b. Similarly, write down the definition of $X \cap Y$.
    c. If A and B are two sets, write down what it means for A to be a subset of B, that is $A \subseteq B$.
    d. Similarly, what does it mean for $A = B$?
    e. Now, use the previous parts to prove that $X \subseteq X \cup Y$ ***and*** $Y \subseteq X \cup Y$
    f. State and prove a similar statement regarding X, Y and $X \cap Y$.
    g. Prove that $X \subseteq Y$ ***if and only if*** $Y = X \cup Y$.
    h. Similarly, prove that $X \subseteq Y$ ***if and only if*** $X = X \cap Y$. Notice that it is now X on the left side of the equation.

55. ***The Method of Descent:*** The Principle of Mathematical Induction goes ***forward***, that is, we start with proving the case when $n = 1$, then we assume that the case when $n = k$ is true, and finally we prove that the case when $n = k+1$, that is, the next ***bigger*** case, is also true. However, sometimes it is useful to go ***backwards*** instead of forward. This is possible because 1 is the ***smallest*** positive integer, and thus if we start with a positive integer n and go lower and lower, we will eventually hit 1 and then we cannot go any lower. We will illustrate this idea, formally called The Method of Infinite Descent or more simply as The Method of Descent, to prove:

Every integer N which is bigger than 1 is either ***prime***

or has a prime factor q which is ***less than*** N.

Recall that an integer p is prime if $p > 1$ and the ***only way*** we can factor p into two positive integers as $p = a \cdot b$ is if ***either*** $a = 1$ and $b = p$ ***or*** $a = p$ and $b = 1$. For example, $7 = 1 \cdot 7 = 7 \cdot 1$, and there are no other ways to factor 7 into two positive integers. It is important to remember that 1 is ***not*** a prime number.

a. Warm-up: List down the first ten prime numbers.

b. Now, suppose N is an integer bigger than 1 and N is not prime. Use the definition above to show that we can factor N as $N = N_1 \cdot N$, where $1 < N_1 < N$ ***and*** $1 < N_2 < N$.

c. Explain why the proof is finished if ***either*** N_1 is prime or N_2 is prime.

d. Suppose now that ***neither*** N_1 nor N_2 is prime. We will ignore N_2 and focus our attention on N_1. Repeat the arguments above and factor N_1 as $N_1 = N_3 \cdot N_4$. What can we now say about N_3 and N_4?

e. We now come to the Method of Descent. Explain why we can keep performing this argument until we produce a list of positive integers: $N > N_1 > N_3 > \ldots$ and explain why this list must ***end*** with some prime number $N_k = q$ which divides N.

f. Explain why we ignored N_2 in part (d). Could we have ignored N_1 instead? How will this affect the list in (e)?

56. Use the previous Exercise to show that every positive integer N can be completely factored into primes: $N = p_1 \cdot p_2 \cdot \cdots \cdot p_k$, for some finite set of primes $\{p_1, p_2, \ldots, p_k\}$.

Note that we take this property for granted when we are first learning Algebra. More precisely, every positive integer N can be factored ***uniquely*** into a product of primes, that is, any two factorizations into primes must have exactly the same primes appearing with the same frequency but possibly in a different order. This is known as the ***Fundamental Theorem of Arithmetic***, and could also be proven by the Method of Descent, but the proof is much more complicated.

57. ***The Infinitude of Primes:*** Our goal in this Exercise is to show that the set of prime numbers is ***infinite***. Thus, if the set of primes is $P = \{2, 3, 5, 7, 11, \ldots\}$, then this list will never terminate.

a. Warm-up: prove that if the integers a and b are both divisible by the integer c, then $a - b$ and $a + b$ are also divisible by c. (We say that an integer x is ***divisible*** by a non-zero integer y if x/y is also an integer).

Now, we will use Proof by Contradiction to prove our main goal. Suppose that P above is a ***finite set***, so the complete set of primes becomes $P = \{2, 3, 5, 7, 11, \ldots, p_L\}$ where p_L is the largest prime number. Let us construct the number $N = (2 \cdot 3 \cdot 5 \cdot \cdots \cdot p_L) + 1$. We will proceed with a Case-by-Case Analysis:

b. Suppose that N is prime. Show that we have a contradiction and thus our proof is finished.

c. Now, suppose that N is not prime (thus we have considered both possibilities about N). The Exercise from The Method of Descent says that N must be divisible by a prime q which is smaller than N. Show that q is ***missing*** from the set P above, and explain why this is a contradiction and our proof is also finished. Hint: (a) could be useful.

58. ***Powersets:*** If $X = \{x_1, x_2, \ldots, x_n\}$ is a finite set, we define $\wp(X)$, the ***powerset*** of X, to be the set of all subsets of X. For example, if $X = \{a, b\}$, then $\wp(X) = \{\emptyset, \{a\}, \{b\}, \{a, b\}\}$, and thus $\wp(X)$ has 4 elements.

   a. If $X = \{a, b, c\}$, list all the members of $\wp(X)$. How many subsets does X have?
   b. Separate the list that you got in part (a) into two columns. Place on the left column those subsets that contain c and place on the right column those that do not contain c.
   c. Now, ***cross out*** c from each subset on the left column. What do you notice?
   d. Prove by induction that if $X = \{x_1, x_2, \ldots, x_n\}$, then $\wp(X)$ has 2^n elements. Hint: in the induction step, we want to show that the number of subsets of $\{x_1, x_2, \ldots, x_{k+1}\}$ is ***double*** the number of subsets of $\{x_1, x_2, \ldots, x_k\}$. Think of how to generalize parts (b) and (c).
   e. Show that the set of subsets of a finite set X has strictly more members than X itself. Hint: Use one of the Exercises above on Induction.

59. The purpose of this Exercise is to prove that for any real number a:

$$\sqrt{a^2} = |a|.$$

Recall that the ***absolute value*** of any real number a is defined by:

$$|a| = \begin{cases} a & \text{if } a \geq 0, \text{ and} \\ -a & \text{if } a < 0. \end{cases}$$

We also know that the function $f(x) = x^2$ is not one-to-one on $(-\infty, \infty)$, but it is one-to-one if the domain is restricted to $[0, \infty)$. In this case, the range of $f(x)$ is also $[0, \infty)$, and so we will define the ***square root*** of a non-negative real number, $b \in [0, \infty)$, as:

$$\sqrt{b} = c, \text{ where } c \in [0, \infty), \text{ and } b = c^2.$$

   a. Warm-up: use the definition above to explain why for any real number a: $|a| \geq 0$.
   b. Again, using the definition, show that $|a|^2 = a^2$.
   c. Our next goal is to show that $\sqrt{b}$ is ***unique***. In other words, prove that if c and d are two real numbers such that $c \geq 0$, and $d \geq 0$, and $b = c^2 = d^2$, then $c = d$. Hint: rewrite this equation into: $c^2 - d^2 = 0$ and use the Zero Factors Theorem.
   d. Rewrite the definition for $\sqrt{b}$ to define $\sqrt{a^2}$.
   e. Put together all the steps above to write a complete proof that $\sqrt{a^2} = |a|$.

Positive Numbers and the Order Axioms: In some of the Exercises above, we assumed that the reader was familiar with the basic properties of positive numbers and inequalities. We can formalize these properties with these additional ***Axioms for Positive Numbers:***

There exists a non-empty subset $\mathbb{R}^+ \subset \mathbb{R}$, consisting of the ***positive real numbers***, such that the following properties are accepted to be true:

a. ***Closure*** under Addition and Multiplication: If $x, y \in \mathbb{R}^+$, then $x + y \in \mathbb{R}^+$, and $x \cdot y \in \mathbb{R}^+$.

b. Zero is ***not*** positive: $0 \notin \mathbb{R}^+$.

c. ***The Dichotomy Property:*** If $x \neq 0$, then either $x \in \mathbb{R}^+$, or $-x \in \mathbb{R}^+$, but ***not*** both.

Using only these three Axioms, prove the following statements (as usual, an earlier Exercise can be used to prove a later Exercise, if applicable).

60. Prove that $1 \in \mathbb{R}^+$. Hint: Use Proof by Contradiction. Suppose instead $-1 \in \mathbb{R}^+$. What do the Closure properties and the Dichotomy Property tell us?
61. Use the previous Exercise to show that the set of positive integers $\{1, 2, 3, \ldots, n, n+1, \ldots\}$ is a subset of $\mathbb{R}^+$. Hint: use the Closure property, and Induction.
62. Prove the ***Reciprocal Property*** for $\mathbb{R}^+$: For all $x \in \mathbb{R}$, $x \neq 0$: $x \in \mathbb{R}^+$ ***if and only if*** $1/x \in \mathbb{R}^+$. See the hints in the two previous Exercises.

The Dichotomy Property creates another set, $\mathbb{R}^-$, consisting of the ***negative real numbers:***

$$\mathbb{R}^- = \{x \in \mathbb{R} \mid -x \in \mathbb{R}^+\}.$$

63. Notice that in the definition for $\mathbb{R}^-$, there is no mention of x being non-zero (unlike in The Dichotomy Property). Use Proof by Contradiction to prove that zero is ***not*** negative either.

This last Exercise tells us that we have three ***disjoint*** and exhaustive subsets of $\mathbb{R}$:

$$\mathbb{R} = \mathbb{R}^- \cup \{0\} \cup \mathbb{R}^+.$$

In other words, every real number is either negative, zero, or positive, and these three sets have no number in common.

64. Prove that $\mathbb{R}^-$ is Closed under Addition.
65. Prove that if $x, y \in \mathbb{R}^-$, then $x \cdot y \in \mathbb{R}^+$. Thus, $\mathbb{R}^-$ is ***not*** closed under Multiplication.
66. Prove that if $x \in \mathbb{R}^-$ and $y \in \mathbb{R}^+$, then $x \cdot y \in \mathbb{R}^-$.
67. Combine the Exercises above to prove: For all $x, y \in \mathbb{R}$:

$$x \cdot y \in \mathbb{R}^+ \ \textbf{\textit{if and only if}}\ x \text{ and } y \in \mathbb{R}^+ \ \textbf{\textit{or}}\ x \text{ and } y \in \mathbb{R}^-.$$

68. Prove the ***Reciprocal Property*** for $\mathbb{R}^-$: For all $x \in \mathbb{R}$, $x \neq 0$: $x \in \mathbb{R}^-$ ***if and only if*** $1/x \in \mathbb{R}^-$.

Next, the set $\mathbb{R}^+$ allow us to establish an ***ordering*** of the real numbers:

We will say that $x > y$ (in words: x is ***greater than*** y) if $x - y \in \mathbb{R}^+$. Similarly, $x < y$ (x is less than y) means $y > x$, $x \leq y$ means $x < y$ ***or*** $x = y$, and $x \geq y$ means $x > y$ ***or*** $x = y$. In the following statements, assume that $x, y, z \in \mathbb{R}$:

69. Prove that $x < y$ ***if and only if*** $x - y \in \mathbb{R}^-$.
70. Prove the ***Trichotomy Property: Exactly one*** of the following three possibilities is true: $x = y$, ***or*** $x < y$, ***or*** $y < x$.
71. Prove the ***Transitive Property:*** If $x > y$ and $y > z$, then $x > z$.
72. Prove that if $x < y$ and $z \in \mathbb{R}^+$, then $x \cdot z < y \cdot z$ and $x \cdot (-z) > y \cdot (-z)$.
73. Prove the ***Order Property for Reciprocals:*** For all $x, y \in \mathbb{R}$:

$$\text{If } x > 0 \ \textbf{\textit{and}}\ y > x, \text{ then } 1/x > 1/y.$$

$$\text{If } y < 0 \ \textbf{\textit{and}}\ y > x, \text{ then } 1/x > 1/y.$$

74. Prove the ***Squeeze Theorem for Inequalities:*** For all $x, y, z \in \mathbb{R}$:

$$\text{If } x \leq y \ \textbf{\textit{and}}\ y \leq x, \text{ then } x = y.$$

75. Let us define the imaginary unit i to be a number (or quantity) with the property that: $i^2 = i \cdot i = -1$. Prove that such a number ***cannot*** be a real number. Hint: if $i \in \mathbb{R}$, then either $i \in \mathbb{R}^+$ or $i \in \mathbb{R}^-$ or $i = 0$. Show that all these possibilities lead to a contradiction.

Chapter 1

The Canvas of Linear Algebra:

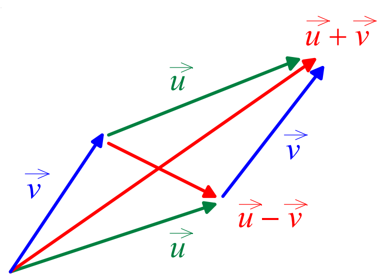
Euclidean Spaces and Subspaces

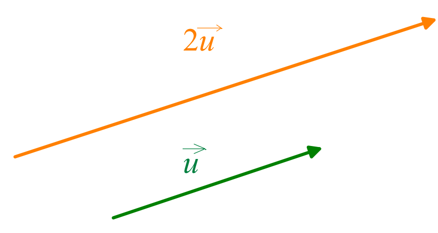
We study Calculus because we are interested in real numbers and functions that operate on them, such as polynomial, rational, radical, trigonometric, exponential and logarithmic functions. We want to study their graphs, derivatives, extreme values, concavity, antiderivatives, Taylor series, and so on..

In the same spirit, we define Linear Algebra as follows:

> ***Linear Algebra*** is the study of sets called ***vector spaces***, which are generalizations of numbers, their ***structure***, and functions with special properties called ***linear transformations*** that map one vector space to another.

In this Chapter, we will look at the basic kind of vector space, called ***Euclidean n-space*** or $\mathbb{R}^n$. Vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$ can be visualized as arrows, and the basic operations of ***vector addition, subtraction*** and ***scalar multiplication*** can be interpreted geometrically:





From these two basic operations, we will construct ***linear combinations*** of vectors, and form the ***Span*** of a set of vectors. We will see that these Spans are the fundamental examples of ***subspaces***, and that we can describe these subspaces as the Span of a finite set of vectors called a ***basis***, which have as few vectors as possible. A basis for a subspace enjoys a special property called ***linear independence***, that allows us to describe subspaces in the most efficient way.

The main computational tool of Linear Algebra is called the ***Gauss-Jordan Algorithm***. We will introduce it in this Chapter, and see that it is useful to solve a general system of linear equations. We will also see the concept of the ***dot product*** and the relationship of ***orthogonality***, and we will see that subspaces of Euclidean n-space come in pairs called ***orthogonal complements***.

1.1 The Main Subject: Euclidean Spaces

In ordinary algebra, we see ordered pairs of numbers such as $(3,-5)$. Our first step will be to generalize these objects:

> ***Definition:*** An ***ordered n-tuple*** or ***vector*** $\vec{v}$ is an ordered list of n real numbers:
>
> $$\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle.$$

Example: $\langle 2,-1,4 \rangle$ is an ordered 3-tuple (more naturally called an ordered ***triple***), and $\langle 5, 7, -3, 0, 6, 2 \rangle$ is an ordered ***6-tuple***.□

> ***Definition:*** The set of all possible n-tuples is called ***Euclidean n-space***, denoted by the symbol $\mathbb{R}^n$:
>
> $$\mathbb{R}^n = \{\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \mid v_1, v_2, \ldots, v_n \in \mathbb{R}\}.$$

Euclidean n-space is the main subject of linear algebra, and it is the fundamental example of a category of objects called ***vector spaces***. Almost all concepts that we will encounter are related to vector spaces. The number $\boldsymbol{n}$ is called the ***dimension*** of the space, and we will refer to $\mathbb{R}^2$ as 2-dimensional space, $\mathbb{R}^3$ as 3-dimensional space, and so on. Euclidean n-spaces are referred to collectively as ***Euclidean spaces***. A vector $\vec{v}$ from $\mathbb{R}^n$ is more specifically called an ***n-dimensional vector***, although we will simply say "vector" when we know which Euclidean space $\vec{v}$ comes from. We use an ***arrow*** on top of a letter to denote that the symbol is a vector. The entries within each vector are called the ***components*** of the vector, and they are numbered with a subscript from 1 to n. We will also agree that $\mathbb{R}^1 = \{\vec{v} = \langle v_1 \rangle \mid v_1 \in \mathbb{R}\} = \mathbb{R}$, the set of real numbers.

Example: Let $\vec{v} = \langle 7,0,-5,1 \rangle \in \mathbb{R}^4$. We say that $v_1 = 7$, $v_2 = 0$, $v_3 = -5$ and $v_4 = 1$.□

To distinguish real numbers from vectors, we will also refer to real numbers as ***scalars***.

> ***Definition:*** Two vectors $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ from $\mathbb{R}^n$ are ***equal*** if all of their components are ***pairwise equal***, that is, $u_i = v_i$ for $i = 1 \ldots n$. Two vectors from ***different*** Euclidean spaces are ***never*** equal.

Example: In $\mathbb{R}^3$, We can say that $\langle 2, 3^2, \cos(\pi) \rangle = \left\langle \sqrt{4}, 9, -1 \right\rangle$, but $\langle -2,5,7 \rangle \neq \langle 5,-2,7 \rangle$.□

Many of the Axioms for Real Numbers that we saw in Chapter Zero have analogs in Euclidean spaces. Let us start by generalizing the scalar zero and the additive inverse of a real number:

> ***Definitions:*** Each $\mathbb{R}^n$ has a special element called the ***zero vector***, also called the ***additive identity***, all of whose components are zero: $\vec{\mathbf{0}}_n = \langle 0, 0, \ldots, 0 \rangle$.
> Every vector $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \in \mathbb{R}^n$ has its own ***additive inverse*** or ***negative:***
>
> $$-\vec{v} = \langle -v_1, -v_2, \ldots, -v_n \rangle.$$

Example: In $\mathbb{R}^5$, the zero vector will be written as $\vec{0}_5 = \langle 0, 0, 0, 0, 0 \rangle$. Notice that we do ***not*** put a subscript on the zeroes. If $\vec{v} = \langle 4, -2, 0, 7, -6 \rangle$, then $-\vec{v} = \langle -4, 2, 0, -7, 6 \rangle$.

Vector Arithmetic

Vectors in $\mathbb{R}^n$ are manipulated in two basic ways:

Definitions: If $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ are vectors in $\mathbb{R}^n$, we define the ***vector sum:***

$$\vec{u} + \vec{v} = \langle u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n \rangle,$$

and if $r \in \mathbb{R}$, we define the ***scalar product:***

$$r \cdot \vec{v} = r\vec{v} = \langle rv_1, rv_2, \ldots, rv_n \rangle.$$

We will call the operation of finding the vector sum as ***vector addition***, and the operation of finding a scalar product as ***scalar multiplication***. We can also define ***vector subtraction*** by:

$$\vec{u} - \vec{v} = \vec{u} + (-\vec{v}) = \langle u_1 - v_1, u_2 - v_2, \ldots, u_n - v_n \rangle.$$

Example: Let $\vec{u} = \langle 3, -5, 6, 7 \rangle$ and $\vec{v} = \langle -4, 2, 3, -2 \rangle$. Then:

$$\begin{aligned}
\vec{u} + \vec{v} &= \langle 3 + (-4), -5 + 2, 6 + 3, 7 + (-2) \rangle \\
&= \langle -1, -3, 9, 5 \rangle, \\
-\vec{v} &= \langle 4, -2, -3, 2 \rangle, \\
5\vec{u} &= \langle 5 \cdot 3, 5(-5), 5 \cdot 6, 5 \cdot 7 \rangle \\
&= \langle 15, -25, 30, 35 \rangle, \text{ and} \\
\vec{u} - \vec{v} &= \langle 3 - (-4), -5 - 2, 6 - 3, 7 - (-2) \rangle \\
&= \langle 7, -7, 3, 9 \rangle. \square
\end{aligned}$$

Something funny happens when we multiply any vector by zero:

Theorem: The Multiplicative Property of the Scalar Zero:
Let $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \in \mathbb{R}^n$. Then:

$$0 \cdot \vec{v} = \vec{\mathbf{0}}_n.$$

Proof: We apply the definition of scalar multiplication:

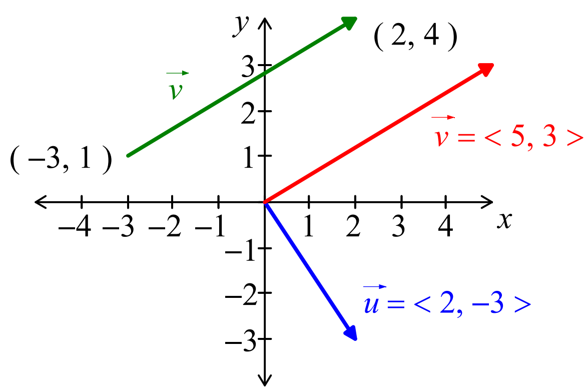
$$\begin{aligned}
0 \cdot \vec{v} &= \left\langle 0 \cdot v_1, 0 \cdot v_2, \ldots, 0 \cdot v_n \right\rangle \\
&= \langle 0, 0, \ldots, 0 \rangle = \vec{\mathbf{0}}_n,
\end{aligned}$$

where we used the Multiplicative Property of Zero in every component, that is, $0 \cdot v_i = 0$ for all $i = 1 \ldots n$. $\blacksquare$

Visualizing Vectors from $\mathbb{R}^2$

The Euclidean space $\mathbb{R}^2$ is easy to visualize. We will use the standard ***xy-plane***, also known as the ***Cartesian plane***. As usual, we will denote by (a, b) the ***point*** on the Cartesian plane with coordinates $x = a$ and $y = b$. A vector in $\mathbb{R}^2$ will be represented using ***arrows*** (also called ***directed line segments***). The arrow representing $\vec{u} = \langle u_1, u_2 \rangle$ is in ***standard position*** if the ***tail*** is at the ***origin*** $(0, 0)$ and the ***head*** is at the point (u_1, u_2). If the tail is not at the origin but at some other point P, we say that the vector has been ***translated to P***. The zero vector $\vec{\mathbf{0}}_2$ is represented by the ***origin*** $(0, 0)$, or any ***point*** on the Cartesian plane for that matter. Notice also that since $-\vec{u} = \langle -u_1, -u_2 \rangle$, we ***reverse*** the arrow for $\vec{u}$ in order to draw the arrow for $-\vec{u}$. A vector whose tail is at P and head is at Q is denoted $\overrightarrow{PQ}$. We also say that $\overrightarrow{PQ}$ is the vector ***from*** P ***to*** Q. We remark that the Cartesian plane is not $\mathbb{R}^2$ but is a framework where we draw the vectors of $\mathbb{R}^2$.

Example: In the picture below, $\vec{u} = \langle 2, -3 \rangle$ is in standard position, and we show $\vec{v} = \langle 5, 3 \rangle$ both in standard position and with its tail translated to $(-3, 1)$.



Plotting Vectors in $\mathbb{R}^2$

Notice that the ***head*** of the second $\vec{v}$ is not at $(5, 3)$, but rather is at $(-3+5, 1+3) = (2, 4)$. □

In general, the signs of u_1 and u_2 tell us the direction that $\vec{u} = \langle u_1, u_2 \rangle$ is pointing, so if both are positive, then $\vec{u}$ is pointing ***right*** and ***up***. Thus, we have the following:

Theorem: Let $\vec{u} = \langle u_1, u_2 \rangle \in \mathbb{R}^2$, and $P(a_1, b_1)$ a point on the Cartesian plane. If $\vec{u}$ is translated to P, then the head of $\vec{u}$ will be located at $Q(a_2, b_2)$, where:

$$a_2 = a_1 + u_1, \text{ and } \quad b_2 = b_1 + u_2.$$

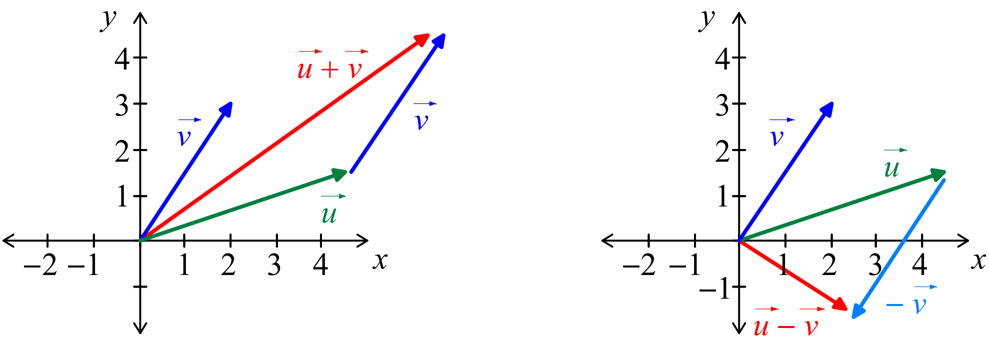
Conversely, if $P(a_1, b_1)$ and $Q(a_2, b_2)$ are two points on the Cartesian plane, then the vector $\vec{u} \in \mathbb{R}^2$ from P to Q is:

$$\vec{u} = \overrightarrow{PQ} = \langle a_2 - a_1, b_2 - b_1 \rangle.$$

We leave the proof as an Exercise.

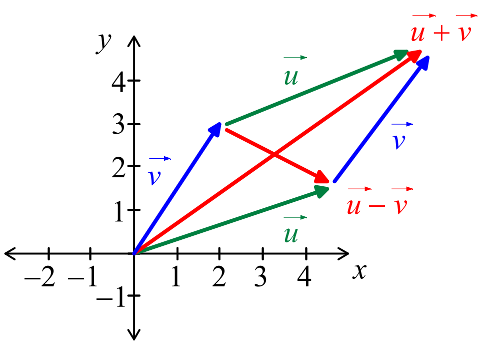
The Geometry of Vector Arithmetic in $\mathbb{R}^2$

Let us think of vector sums and differences from a geometric point of view. To get the vector sum $\vec{u}+\vec{v}$, we put $\vec{u}$ in standard position and translate $\vec{v}$ to the head of $\vec{u}$. We obtain $\vec{u}+\vec{v}$ as the arrow from the origin to the head of $\vec{v}$. Similarly, to get $\vec{u}-\vec{v}$ we put $\vec{u}$ in standard position, reverse $\vec{v}$ (thus getting $-\vec{v}$) and translate it to the head of $\vec{u}$. To obtain $\vec{u}-\vec{v}$, we draw the arrow from the origin to the head of $-\vec{v}$, as we did for $\vec{u}+\vec{v}$.



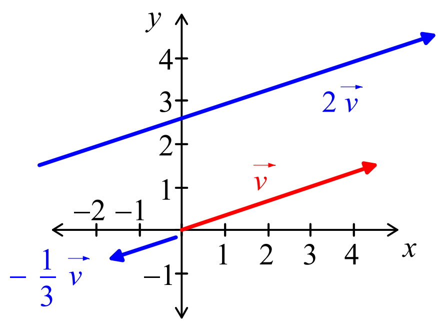
Vector Addition and Subtraction in $\mathbb{R}^2$

These two operations can be seen in a single vector diagram, called ***The Parallelogram Principle*** (where $\vec{u}-\vec{v}$ is translated to the head of $\vec{v}$).



The Parallelogram Principle:

Vector Addition and Subtraction



Scalar Multiplication

Similarly, scalar multiplication has the geometric effect of lengthening or shortening a vector, while preserving or reversing its direction (if the scalar is negative), as shown on the right above.

As we can see, the process of scalar multiplication results in vectors that are pointing in the ***same*** or ***opposite*** directions (we translated $2\vec{v}$ so that it will not overlap with $\vec{v}$). However, we saw that for any vector $\vec{v} \in \mathbb{R}^n$: $0 \cdot \vec{v} = \vec{\mathbf{0}}_n$. It is therefore reasonable to define the following concept:

Definition: Axiom for Parallel Vectors:

We say that two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ are ***parallel to each other*** if there exists either $a \in \mathbb{R}$ ***or*** $b \in \mathbb{R}$ such that:

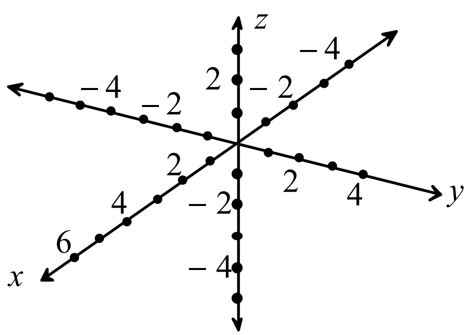
$$\vec{u} = a \cdot \vec{v} \quad \textbf{\textit{or}} \quad \vec{v} = b \cdot \vec{u}.$$

Consequently, this means that $\vec{\mathbf{0}}_n$ is parallel to ***all*** vectors $\vec{v} \in \mathbb{R}^n$, since $\vec{\mathbf{0}}_n = 0 \cdot \vec{v}$.

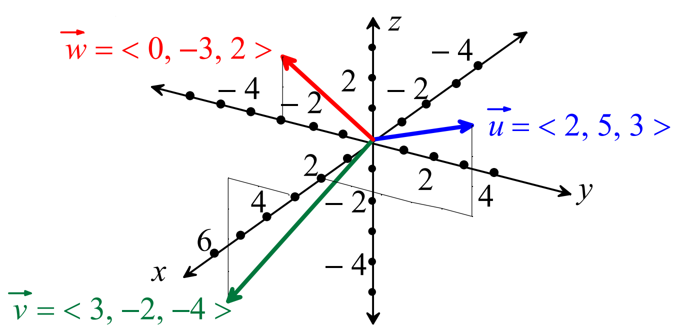
You will show in the Exercises that when $\vec{u}$ and $\vec{v}$ are ***non-zero*** parallel vectors, then a and b both exist and are non-zero scalars, and furthermore, $a = 1/b$.

Visualizing Vectors from $\mathbb{R}^3$

The picture for $\mathbb{R}^3$ requires some imagination. The best way to start is to stand in front of a corner of your room and look down at the corner joining the floor and two walls. The corner will be the origin. The edge on your left is the positive x-axis, the edge on your right is the positive y-axis, and the edge going up is the positive z-axis. To draw this on paper, start by drawing the z-axis as a vertical line. Next, draw the x and y axes as shown below on the left, where the y-axis is slightly rotated clockwise (around 20^0) from the horizontal direction, and the positive x-axis makes an angle of about 120^0 from the positive z-axis. As usual, we mark off a scale on each axis. These three axes determine our ***Cartesian space***. The "floor" determined by the x and y axes is called the xy-plane, the "left side wall" determined by the x and z axes is called the xz-plane, and the "back wall" determined by the y and z axes is called the yz-plane. These three coordinate planes divide Cartesian space into eight ***octants***. The only standard convention is naming the 1st octant as that where the x, y and z coordinates are all positive. As before, we remark that Cartesian space is ***not*** $\mathbb{R}^3$ but it is a framework where we can visualize the vectors of $\mathbb{R}^3$.



Cartesian Space



Plotting Vectors from $\mathbb{R}^3$ in Standard Position

Example: We have plotted in the diagram above on the right three vectors in standard position: $\vec{u} = \langle 2, 5, 3 \rangle$, $\vec{v} = \langle 3, -2, -4 \rangle$ and $\vec{w} = \langle 0, -3, 2 \rangle$.

To plot $\vec{u} = \langle 2, 5, 3 \rangle$, we start at the origin, go forward on the x-axis to 2, go right parallel to the y-axis by 5 units, then go up parallel to the z-axis by 3 units. To plot $\vec{v} = \langle 3, -2, -4 \rangle$, we go forward by 3 units on the x-axis, go left 2 units parallel to the y-axis, and go down 4 units parallel to the z-axis. To plot $\vec{w} = \langle 0, -3, 2 \rangle$, we directly go 3 units left on the y-axis, then go up 2 units up parallel to the z-axis. $\square$

If a vector is translated in $\mathbb{R}^3$, we can make the following statement whose proof is again left as an Exercise:

Theorem: Let $\vec{u} = \langle u_1, u_2, u_3 \rangle \in \mathbb{R}^3$, and $P(a_1, b_1, c_1)$ a point in Cartesian space. If $\vec{u}$ is translated to P, then the head of $\vec{u}$ will be located at $Q(a_2, b_2, c_2)$, where:

$$a_2 = a_1 + u_1, \quad b_2 = b_1 + u_2, \quad \text{and} \quad c_2 = c_1 + u_3.$$

Conversely, let $P(a_1, b_1, c_1)$ and $Q(a_2, b_2, c_2)$ be two points in Cartesian space. Then: the vector $\vec{u} \in \mathbb{R}^3$ from P to Q is:

$$\vec{u} = \overrightarrow{PQ} = \langle a_2 - a_1, b_2 - b_1, c_2 - c_1 \rangle.$$

Properties of Vector Arithmetic

The two basic operations of vector addition and scalar multiplication have properties that naturally follow from the arithmetic properties of real numbers:

Theorem — Properties of Vector Arithmetic:
If $\vec{u}$, $\vec{v}$, and $\vec{w}$ are vectors in $\mathbb{R}^n$ and r and s are scalars, then the following properties are true:

1. ***The Closure Property of Vector Addition*** $\qquad \vec{u} + \vec{v}$ is also in $\mathbb{R}^n$.
2. ***The Closure Property of Scalar Multiplication*** $\qquad r \cdot \vec{u}$ is also in $\mathbb{R}^n$.
3. ***The Commutative Property of Vector Addition*** $\qquad \vec{u} + \vec{v} = \vec{v} + \vec{u}$.
4. ***The Associative Property of Vector Addition*** $\qquad (\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
5. ***The Additive Identity Property*** $\qquad \vec{\mathbf{0}}_n + \vec{v} = \vec{v} = \vec{v} + \vec{\mathbf{0}}_n$.
6. ***The Additive Inverse Property*** $\qquad \vec{v} + (-\vec{v}) = \vec{\mathbf{0}}_n = (-\vec{v}) + \vec{v}$.
7. ***The "Left" Distributive Property*** $\qquad (r + s) \cdot \vec{v} = r \cdot \vec{v} + s \cdot \vec{v}$.
8. ***The "Right" Distributive Property*** $\qquad r \cdot (\vec{u} + \vec{v}) = r \cdot \vec{u} + r \cdot \vec{v}$.
9. ***The Associative Property of Scalar Multiplication*** $\qquad (rs) \cdot \vec{v} = r \cdot (s \cdot \vec{v}) = s \cdot (r \cdot \vec{v})$.
10. ***The Unitary Property of Scalar Multiplication*** $\qquad 1 \cdot \vec{v} = \vec{v}$.

Notice that the additions appearing in the two Distributive Properties are ***different*** additions: in the "Left" Distributive Property, the addition on the left side of the equation is the ordinary addition of real numbers, and all the other additions are vector additions. Similarly, the product rs in the Associative Property is a product of real numbers only, but the five other products are all scalar products. Let us prove one of these properties (the rest will be in the Exercises).

Example: Let us prove that for all $\vec{v} \in \mathbb{R}^n$ and $r, s \in \mathbb{R}$: $(r + s) \cdot \vec{v} = r \cdot \vec{v} + s \cdot \vec{v}$.

Proof: We will express the vectors in component form and follow the Axioms for the Real Numbers. So let us write $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$. Then we have:

$$\begin{aligned}
&(r+s)\cdot\vec{v} \\
&= (r+s)\cdot\langle v_1, v_2, \ldots, v_n\rangle \\
&= \langle (r+s)v_1, (r+s)v_2, \ldots, (r+s)v_n\rangle \quad \text{by the definition of scalar multiplication,} \\
&= \langle rv_1+sv_1, rv_2+sv_2, \ldots, rv_n+sv_n\rangle \quad \text{by the (ordinary) Distributive Property.}
\end{aligned}$$

Let us now work on the right side of the equation we are trying to prove:

$$\begin{aligned}
&r\cdot\vec{v}+s\cdot\vec{v} \\
&= r\cdot\langle v_1, v_2, \ldots, v_n\rangle + s\cdot\langle v_1, v_2, \ldots, v_n\rangle \\
&= \langle rv_1, rv_2, \ldots, rv_n\rangle + \langle sv_1, sv_2, \ldots, sv_n\rangle \quad \text{by the definition of scalar multiplication,} \\
&= \langle rv_1+sv_1, rv_2+sv_2, \ldots, rv_n+sv_n\rangle \quad \text{by the definition of vector addition,}
\end{aligned}$$

and we get the same expanded expression as that of $(r+s)\cdot\vec{v}$. ■

We note that in our proof above, we had to know the ***definitions*** for vector addition and scalar multiplication. We also had to write our vector $\vec{v}$ in proper component form, that is, know the meaning of correct ***notation***. These two key ingredients — definitions and notations — are found in almost ***any*** good proof. Keep these in mind when proving similar properties in the Exercises, especially those dealing with two or more vectors or scalars.

The Length of a Vector:

Since vectors are uniquely determined by directed line segments on the plane or in space, we can speak of the length of a vector:

Definition: Let $\vec{v} = \langle v_1, v_2\rangle \in \mathbb{R}^2$ and $\vec{w} = \langle w_1, w_2, w_3\rangle \in \mathbb{R}^3$.
We define the ***length*** or ***norm*** or ***magnitude*** of these vectors as:

$$\|\vec{v}\| = \sqrt{v_1^2+v_2^2} \quad \text{and} \quad \|\vec{w}\| = \sqrt{w_1^2+w_2^2+w_3^2}\,.$$

We say that $\vec{v}$ is a ***unit vector*** if $\|\vec{v}\| = 1$. Similarly, $\vec{w}$ is a ***unit vector*** if $\|\vec{w}\| = 1$.

Example: The length of $\vec{v} = \langle 3, -2, -4\rangle$, a vector from our previous Example, is:

$$\|\vec{v}\| = \sqrt{9+4+16} = \sqrt{29}\,. \square$$

Notice that the formulas for length come from the ***distance formula*** in 2 and 3 dimensions. From these definitions, we can directly prove:

Theorem: For any scalar $k \in \mathbb{R}$ and vector $\vec{v} \in \mathbb{R}^2$ or $\mathbb{R}^3$: $\|k\cdot\vec{v}\| = |k|\|\vec{v}\|$.
Furthermore, $\|\vec{v}\| \geq 0$, and $\|\vec{v}\| = 0$ ***if and only if*** $\vec{v} = \vec{\mathbf{0}}_2$ or $\vec{\mathbf{0}}_3$.
Consequently, if $\vec{v}$ is a non-zero vector, then:

$$\vec{u}_1 = \frac{1}{\|\vec{v}\|}\cdot\vec{v} \quad \text{and} \quad \vec{u}_2 = \frac{-1}{\|\vec{v}\|}\cdot\vec{v}$$

are ***unit vectors*** parallel to $\vec{v}$.

Proof: We will prove the first equation, and leave the other parts of this Theorem as Exercises. Again, all we need to do is use component notation and in this case, the formula for the norm:

$$\begin{aligned} \| k \cdot \vec{v} \| &= \| \langle kv_1, kv_2 \rangle \| \\ &= \sqrt{(kv_1)^2 + (kv_2)^2} \\ &= \sqrt{k^2 v_1^2 + k^2 v_2^2} \\ &= \sqrt{k^2 (v_1^2 + v_2^2)} \\ &= \sqrt{k^2}\,\sqrt{v_1^2 + v_2^2} \\ &= |k| \| \vec{v} \|. \blacksquare \end{aligned}$$

Note that we used Exercise 59 from Chapter Zero. This Theorem says that scalar multiplication has the effect of shortening or lengthening a vector by the absolute value of k, the scalar factor. If k is positive, then $k\vec{u}$ is the vector in the same direction as $\vec{u}$ whose length is k times the length of $\vec{u}$. If k is negative, we reverse the arrow and multiply its length by $|k|$.

Example: The two unit vectors parallel to $\vec{v} = \langle 3, -2, -4 \rangle$ from our previous Example are:

$$\vec{u}_1 = \frac{1}{\sqrt{29}} \langle 3, -2, -4 \rangle \quad \text{and} \quad \vec{u}_2 = \frac{-1}{\sqrt{29}} \langle 3, -2, -4 \rangle. \square$$

Linear Combinations

More generally, we will combine the operations of scalar multiplication and vector addition involving several vectors and scalars to form a single vector:

> ***Definition:*** If $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in \mathbb{R}^n$, and $x_1, x_2, \ldots, x_k \in \mathbb{R}$, then the vector expression:
> $$x_1 \cdot \vec{v}_1 + x_2 \cdot \vec{v}_2 + \cdots + x_k \cdot \vec{v}_k$$
> is called a ***linear combination*** of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ with ***coefficients*** $x_1, x_2, \ldots, x_k$.

Example: If $\vec{u} = \langle 5, 4, -7 \rangle$, $\vec{v} = \langle -2, 3, 6 \rangle$ and $\vec{w} = \langle 0, 8, -3 \rangle$, then we can compute the linear combination:

$$\begin{aligned} 4\vec{u} - 3\vec{v} + 5\vec{w} &= 4\langle 5, 4, -7 \rangle - 3\langle -2, 3, 6 \rangle + 5\langle 0, 8, -3 \rangle \\ &= \langle 20, 16, -28 \rangle + \langle 6, -9, -18 \rangle + \langle 0, 40, -15 \rangle \\ &= \langle 26, 47, -61 \rangle. \square \end{aligned}$$

Example: If $\vec{u} = \langle 5, -2, 4, 6 \rangle$ and $\vec{v} = \langle 1, -5, 2, -3 \rangle$, is it possible to express $\langle -8, -29, 2, -39 \rangle$ as a linear combination of $\vec{u}$ and $\vec{v}$?

If this is possible, then we can find two coefficients, x and y, such that:

$$\langle -8, -29, 2, -39 \rangle = x\langle 5, -2, 4, 6 \rangle + y\langle 1, -5, 2, -3 \rangle.$$

This would require us to satisfy four equations:

$$
\begin{aligned}
5x + y &= -8, \\
-2x - 5y &= -29, \\
4x + 2y &= 2, \text{ and} \\
6x - 3y &= -39
\end{aligned}
$$

Solving for x and y using only the first two equations, we easily eliminate y and get $23x = -69$, or $x = -3$. Substituting this in the first equation, we get, $y = 7$. But we also have to check that the other two equations are satisfied, and indeed, they are. Thus:

$$\langle -8, -29, 2, -39 \rangle = -3\langle 5, -2, 4, 6 \rangle + 7\langle 1, -5, 2, -3 \rangle,$$

and so $\langle -8, -29, 2, -39 \rangle$ is a linear combination of $\vec{u}$ and $\vec{v}$. $\square$

The Standard Basis Vectors

There is a special set of vectors in each $\mathbb{R}^n$ that we will frequently see:

Definition: The ***standard basis vectors*** in $\mathbb{R}^n$ are the vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ that have 0 in all components except the i^{th} component, which contains 1:

$$\vec{e}_i = \langle 0, 0, \ldots, 0, 1, 0, \ldots, 0 \rangle.$$

In a Multivariable Calculus or Physics course, the standard basis vectors of $\mathbb{R}^2$ and $\mathbb{R}^3$ are known as $\vec{i}$, $\vec{j}$ and $\vec{k}$, respectively, for the unit vectors parallel to the x, y and z axes. When the Euclidean space is specified, this should not lead to any confusion:

The Standard Basis Vectors of $\mathbb{R}^2$ and $\mathbb{R}^3$

The standard basis vectors allow us to represent any vector naturally and uniquely.

Example: The vector $\langle 5, 3, -2, 7 \rangle \in \mathbb{R}^4$ can be written as:

$$
\begin{aligned}
\langle 5, 3, -2, 7 \rangle &= 5\langle 1, 0, 0, 0 \rangle + 3\langle 0, 1, 0, 0 \rangle - 2\langle 0, 0, 1, 0 \rangle + 7\langle 0, 0, 0, 1 \rangle \\
&= 5\vec{e}_1 + 3\vec{e}_2 - 2\vec{e}_3 + 7\vec{e}_4. \square
\end{aligned}
$$

Obviously, we can generalize this example in the following:

Theorem — Uniqueness of Representation: Every vector $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle \in \mathbb{R}^n$ can be expressed ***uniquely*** as a ***linear combination*** of the standard basis vectors:

$$\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n.$$

The Proof Template

One of the most difficult skills to learn and master in Mathematics is how to write a complete and convincing proof. To help introduce you to this skill, begin by following the template below as you go through the proof Exercises in this and in future sections.

> ***Write*** the Theorem you are trying to prove, in its entirety.
>
> ***Paraphrase*** the Theorem by identifying the ***given conditions*** and the ***conclusion:***
>
> > ***We are given that:***
> >
> > ***The conclusion we want to reach is:*** or
> >
> > ***We want to show that:***
>
> ***Write down*** the relevant ***Definitions:***
>
> > ***The given conditions mean that:***
> >
> > ***The conclusions we want to reach mean that:***
> >
> > ***The notations in the Theorem mean:***
>
> Write down any relevant ***Theorems*** that are related to the given conditions or to the conclusion and try to connect everything in your template together in a complete proof.

You have already seen several proofs in this Section as well as in Chapter Zero. You are encouraged to ***imitate*** these and future proofs as you write your own proofs. You will notice that many words and phrases appear a lot in proofs, including but not limited to:

> ***let, consider, assume, if, if and only if, suppose, thus, therefore, this implies, we can conclude that, we know that, according to this Theorem, but, however, we get a contradiction, let us form the contrapositive, conversely . . .***

and so on. A proof is more than just equations and symbols, so learn to incorporate the words and phrases above into your own proofs. As you develop your own writing style, you will most likely come up with your own favorite words and phrases as well. It would also be a good idea to exchange proofs with your study partner and see if you understand each other's proofs.

Like everything in life, you get better at proofs with ***practice***. The proof Exercises that you will encounter in the first few Chapters are really not that hard, so you have no excuse not to ***try***. Get into the habit of attacking proofs with this template and you will be on your way to mastering this skill.

1.1 Section Summary

Euclidean n-space, denoted $\mathbb{R}^n$, is the set of all possible n-tuples or ***vectors***:

$$\mathbb{R}^n = \{\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \mid v_1, v_2, \ldots, v_n \in \mathbb{R}\}.$$

Vectors can be negated, added, subtracted, and multiplied by a scalar:

$$\begin{aligned} -\vec{v} &= \langle -v_1, -v_2, \ldots, -v_n \rangle \\ \vec{u} + \vec{v} &= \langle u_1 + v_1, u_2 + v_2, \ldots, u_n + v_n \rangle \\ \vec{u} - \vec{v} &= \vec{u} + (-\vec{v}) = \langle u_1 - v_1, u_2 - v_2, \ldots, u_n - v_n \rangle \\ r \cdot \vec{v} &= \langle rv_1, rv_2, \ldots, rv_n \rangle. \end{aligned}$$

Each $\mathbb{R}^n$ has a ***zero vector***, $\vec{\mathbf{0}}_n$, and for all $\vec{v} \in \mathbb{R}^n$, the negative of a vector $\vec{v}$ exists and has the property that: $\vec{v} + (-\vec{v}) = \vec{\mathbf{0}}_n = \langle 0, 0, \ldots, 0 \rangle$.

We say that two vectors $\vec{u}$, $\vec{v} \in \mathbb{R}^n$ are ***parallel to each other*** if there exists either $a \in \mathbb{R}$ ***or*** $b \in \mathbb{R}$ such that $\vec{u} = a \cdot \vec{v}$ ***or*** $\vec{v} = b \cdot \vec{u}$. Consequently, this means that $\vec{\mathbf{0}}_n$ is parallel to ***all*** vectors $\vec{v} \in \mathbb{R}^n$, since $\vec{\mathbf{0}}_n = 0 \cdot \vec{v}$.

In $\mathbb{R}^2$ and $\mathbb{R}^3$, we can represent vectors as directed line segments. The ***length*** or ***norm*** or ***magnitude*** of such a vector is the length of this line segment: if $\vec{v} = \langle v_1, v_2 \rangle \in \mathbb{R}^2$ and $\vec{w} = \langle w_1, w_2, w_3 \rangle \in \mathbb{R}^3$, we define: $\|\vec{v}\| = \sqrt{v_1^2 + v_2^2}$ and $\|\vec{w}\| = \sqrt{w_1^2 + w_2^2 + w_3^2}$. We say that $\vec{v}$ is a ***unit vector*** if $\|\vec{v}\| = 1$. For any scalar $k \in \mathbb{R}$ and vector $\vec{v} \in \mathbb{R}^2$ or $\mathbb{R}^3$: $\| k \cdot \vec{v} \| = |k| \| \vec{v} \|$.

Furthermore, $\|\vec{v}\| \geq 0$, and $\|\vec{v}\| = 0$ ***if and only if*** $\vec{v} = \vec{\mathbf{0}}_2$ or $\vec{\mathbf{0}}_3$. Consequently, if $\vec{v}$ is a non-zero vector, then: $\vec{u}_1 = \dfrac{1}{\| \vec{v} \|} \cdot \vec{v}$ and $\vec{u}_2 = \dfrac{-1}{\| \vec{v} \|} \cdot \vec{v}$ are ***unit vectors*** parallel to $\vec{v}$.

If $\vec{u}$, $\vec{v}$, and $\vec{w}$ are vectors in $\mathbb{R}^n$ and r and s are scalars, then the following properties are true:

1. The Closure Property of Vector Addition	$\vec{u} + \vec{v}$ is also in $\mathbb{R}^n$.
2. The Closure Property of Scalar Multiplication	$r \cdot \vec{u}$ is also in $\mathbb{R}^n$.
3. The Commutative Property of Vector Addition	$\vec{u} + \vec{v} = \vec{v} + \vec{u}$.
4. The Associative Property of Vector Addition	$(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w})$.
5. The Additive Identity Property	$\vec{\mathbf{0}}_n + \vec{v} = \vec{v} = \vec{v} + \vec{\mathbf{0}}_n$.
6. The Additive Inverse Property	$\vec{v} + (-\vec{v}) = \vec{\mathbf{0}}_n = (-\vec{v}) + \vec{v}$.
7. The "Left" Distributive Property	$(r + s) \cdot \vec{v} = r \cdot \vec{v} + s \cdot \vec{v}$.
8. The "Right" Distributive Property	$r \cdot (\vec{u} + \vec{v}) = r \cdot \vec{u} + r \cdot \vec{v}$.
9. The Associative Property of Scalar Multiplication	$(rs) \cdot \vec{v} = r \cdot (s \cdot \vec{v}) = s \cdot (r \cdot \vec{v})$.
10. The Unitary Property of Scalar Multiplication	$1 \cdot \vec{v} = \vec{v}$.

If $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ are vectors, and $x_1, x_2, \ldots, x_k$ are scalars, then: $x_1 \cdot \vec{v}_1 + x_2 \cdot \vec{v}_2 + \cdots + x_k \cdot \vec{v}_k$ is called a ***linear combination*** of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ with ***coefficients*** $x_1, x_2, \ldots, x_k$.

The ***standard basis vectors*** in $\mathbb{R}^n$ are the vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$ which have 0 in all components except the *ith* component, which contains 1: $\vec{e}_i = \langle 0, 0, \ldots, 0, 1, 0, \ldots, 0 \rangle$.

Every $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle \in \mathbb{R}^n$ can be written ***uniquely*** as: $\vec{x} = x_1 \cdot \vec{e}_1 + x_2 \cdot \vec{e}_2 + \cdots + x_n \cdot \vec{e}_n$.

1.1 Exercises

1. Write the definitions of the following operations:
   a. vector addition.
   b. scalar multiplication.
   c. the length of a vector $\vec{w}$ in $\mathbb{R}^3$.
   d. parallel vectors.
   e. a linear combination of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$.
2. Suppose $\vec{u} = \langle -4, 7 \rangle$, $\vec{v} = \langle 3, 5 \rangle$ and $\vec{w} = \langle 1, -2 \rangle$.
   a. Draw the three vectors on the Cartesian plane.
   b. Find $\|\vec{u}\|$.
   c. Find the two unit vectors parallel to $\vec{u}$.
   d. Compute the vectors $3\vec{v}$, $5\vec{w}$, $3\vec{v} + 5\vec{w}$ and $3\vec{v} - 5\vec{w}$, and draw them on your diagram.
3. Suppose $\vec{u} = \langle 5, -3, 2 \rangle$, $\vec{v} = \langle 4, 0, -7 \rangle$ and $\vec{w} = \langle -2, 5, 4 \rangle$.
   a. Draw the three vectors in Cartesian space.
   b. Compute the vectors $2\vec{u}$, $3\vec{w}$, $2\vec{u} + 3\vec{w}$ and $2\vec{u} - 3\vec{w}$, and draw them on your diagram.
   c. Find $\|\vec{w}\|$.
   d. Find the two unit vectors parallel to $\vec{w}$.
   e. Compute the linear combinations:
      i. $-\frac{3}{5}\vec{w}$; ii. $2\vec{u} + 5\vec{v}$; iii. $3\vec{w} - 4\vec{u}$; iv. $-4\vec{u} + 7\vec{v} - 2\vec{w}$.
4. Let $\vec{u} = \langle 3, -5, 1, 7 \rangle$, $\vec{v} = \langle -2, 3, 6, -4 \rangle$ and $\vec{w} = \langle -4, 2, 3, -9 \rangle$. Compute the following:
   a. $\vec{u} + \vec{v}$ b. $\vec{u} + \vec{w}$ c. $\vec{v} - \vec{w}$ d. $-2\vec{u}$
   e. $\frac{3}{4}\vec{v}$ f. $-\frac{5}{3}\vec{w}$ g. $5\vec{u} + 3\vec{v}$ h. $-\frac{3}{2}\vec{u} + \frac{5}{4}\vec{v}$
   i. $2\vec{u} - 3\vec{v} + 7\vec{w}$ j. $-5\vec{u} + 2\vec{v} - 4\vec{w}$ k. $-\frac{3}{2}\vec{u} + \frac{3}{4}\vec{v} - \frac{5}{3}\vec{w}$ l. $\frac{3}{2}\vec{u} - \frac{3}{4}\vec{v} + 2\vec{w}$
5. Suppose that $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^3$ such that $3\vec{u} + \vec{v} = \langle -3, 1, 5 \rangle$ and $5\vec{u} + 2\vec{v} = \langle 9, -4, 3 \rangle$. Find $\vec{u}$ and $\vec{v}$. Hint: why can we solve a system of two equations in two unknown vectors, just like in basic algebra?
6. Is it possible to express $\langle -3, 7 \rangle$ as a linear combination of the two vectors $\vec{u} = \langle 5, -2 \rangle$ and $\vec{v} = \langle -7, 3 \rangle$? If so, how? If not, why not?
7. Is it possible to express $\langle -17, -9, 29, -37 \rangle$ as a linear combination of $\langle 3, -5, 1, 7 \rangle$ and $\langle -4, 2, 3, -9 \rangle$? If so, how? If not, why not?
8. Is it possible to express $\langle -30, 47, 50, -60 \rangle$ as a linear combination of $\langle 3, -5, 1, 7 \rangle$ and $\langle -2, 3, 6, -4 \rangle$? If so, how? If not, why not?
9. Suppose that $\vec{u}$ and $\vec{v}$ are vectors from $\mathbb{R}^5$ and you were told that:
$$3\vec{u} - 5\vec{v} = \langle -4, 27, -19, 33, -31 \rangle \quad \text{and}$$
$$2\vec{u} + 7\vec{v} = \langle -13, -13, 39, -9, 0 \rangle$$
   Find $\vec{u}$ and $\vec{v}$.
10. Suppose that $\vec{u} = \langle 5, -7 \rangle$ is translated so that its tail is at $(2, 4)$. Where is its head?

11. Suppose that $\vec{u} = \langle 3, 1, -4 \rangle$ is translated so that its head is at $(-1, 2, 3)$. Where is its tail?

12. Suppose that $P = (8, -3, 6)$ and $Q = (4, 1, -2)$. Find $\vec{u} = \overrightarrow{PQ}$.

For Exercises (13) to (20): You will be asked to prove some general statements about vector operations. Be sure you know the ***definitions*** for the operations involved, and use ***proper notation*** for the vectors. You should also state which of the Field Axioms for Real Numbers from Chapter Zero is applied to some steps:

13. Prove the ***Closure Property for Vector Addition*** and for ***Scalar Multiplication:*** For all vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, and all scalars $r \in \mathbb{R}$: $\vec{u} + \vec{v} \in \mathbb{R}^n$ and $r\vec{u} \in \mathbb{R}^n$.

14. Prove the ***Commutative Property for Vector Addition:*** For all vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$: $\vec{u} + \vec{v} = \vec{v} + \vec{u}$.

15. Prove the ***Associative Property for Vector Addition:*** For all vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$:
$$(\vec{u} + \vec{v}) + \vec{w} = \vec{u} + (\vec{v} + \vec{w}).$$

16. Prove the ***"Right" Distributive Property of Vector Arithmetic:*** For all vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ and scalars $r \in \mathbb{R}$: $r \bullet (\vec{u} + \vec{v}) = r \bullet \vec{u} + r \bullet \vec{v}$.

17. Prove the ***Associative Property of Scalar Multiplication:*** For all vectors $\vec{u} \in \mathbb{R}^n$ and scalars $r, s \in \mathbb{R}$: $r \bullet (s \bullet \vec{u}) = (rs) \bullet \vec{u}$.

18. Prove that for any scalar $k \in \mathbb{R}$ and any vector $\vec{w} \in \mathbb{R}^3$: $\| k \bullet \vec{w} \| = |k| \| \vec{w} \|$.

19. Prove that for any vector $\vec{v} \in \mathbb{R}^2$ or $\mathbb{R}^3$: $\|\vec{v}\| \geq 0$. Furthermore, $\|\vec{v}\| = 0$ ***if and only if*** $\vec{v} = \vec{\mathbf{0}}_2$ or $\vec{\mathbf{0}}_3$. Hint: how would you define a ***non-zero*** vector? Use your definition and Proof By Contrapositive to prove the forward direction.

20. Prove that for any ***non-zero*** vector $\vec{v} \in \mathbb{R}^2$ or $\mathbb{R}^3$: $\vec{u}_1 = \dfrac{1}{\| \vec{v} \|}\vec{v}$ and $\vec{u}_2 = \dfrac{-1}{\| \vec{v} \|}\vec{v}$ are ***unit vectors*** parallel to $\vec{v}$. Hint: let $k = \pm\dfrac{1}{\| \vec{v} \|}$ and apply Exercise 18 and the analogous statement for $\mathbb{R}^2$.

21. The goal of this Exercise is to show ***algebraically*** that if $\vec{u} = \langle u_1, u_2 \rangle$ and $\vec{v} = \langle v_1, v_2 \rangle$ are vectors in $\mathbb{R}^2$, then, $\vec{u}$ and $\vec{v}$ are ***parallel*** to each other ***if and only if*** $u_1 v_2 - u_2 v_1 = 0$.

    a. Begin by stating the ***definition*** of $\vec{u}$ and $\vec{v}$ being ***parallel*** to each other.

    b. ($\Rightarrow$) Prove the forward implication: Show that if $\vec{u} = \langle u_1, u_2 \rangle$ and $\vec{v} = \langle v_1, v_2 \rangle$ are parallel to each other, then $u_1 v_2 - u_2 v_1 = 0$. Hint: just use direct substitution using (a).

    c. ($\Leftarrow$) Prove the backward implication: Show that if $u_1 v_2 - u_2 v_1 = 0$, then $\vec{u} = a \bullet \vec{v}$ for some $a \in \mathbb{R}$. Hint: do a Case-by-Case Analysis with Case 1: $u_1 \neq 0$, and Case 2: $u_1 = 0$. Recall that $\vec{0}_2$ is parallel to all vectors in $\mathbb{R}^2$, and a non-zero number has a ***reciprocal***. Case 2 will have sub-cases: Case 2a: $u_2 \neq 0$, and Case 2b: $u_2 = 0$.

22. Write down the ***contrapositive*** of the Theorem stated in the previous Exercise.

For Exercises (23) to (29): Prove the following Theorems ***without*** using component notation, but instead using only the ten ***Properties of Vector Arithmetic*** on page 31.

23. ***The Uniqueness of The Zero Vector:*** If $\vec{z}$ is any other vector in $\mathbb{R}^n$ with the property that $\vec{z} + \vec{v} = \vec{v}$ for ***any*** $\vec{v} \in \mathbb{R}^n$, then $\vec{z} = \vec{\mathbf{0}}_n$.

24. ***The Uniqueness of the Negative of a Vector:*** Given ***any*** $\vec{v} \in \mathbb{R}^n$, $-\vec{v}$ is the unique vector in $\mathbb{R}^n$ satisfying $\vec{v} + (-\vec{v}) = \vec{\mathbf{0}}_n$, that is, if $\vec{w} \in \mathbb{R}^n$ also satisfies $\vec{v} + \vec{w} = \vec{\mathbf{0}}_n$, then $\vec{w} = -\vec{v}$.

25. ***The Cancellation Law for Vector Addition:*** If $\vec{u}$, $\vec{v}$ and $\vec{w} \in \mathbb{R}^n$ and $\vec{u} + \vec{v} = \vec{u} + \vec{w}$, then $\vec{v} = \vec{w}$.

26. ***The Multiplicative Property of the Scalar 0:*** For all $\vec{v} \in \mathbb{R}^n$: $0 \cdot \vec{v} = \vec{\mathbf{0}}_n$. Hint: $0 + 0 = ?$ Reminder: do ***not*** use component notation.

27. ***The Multiplicative Property of the Zero Vector:*** For all $k \in \mathbb{R}$: $k \cdot \vec{\mathbf{0}}_n = \vec{\mathbf{0}}_n$. Hint: $\vec{\mathbf{0}}_n + \vec{\mathbf{0}}_n = ?$

28. ***The Zero-Factors Theorem for Vectors:*** If $k \in \mathbb{R}$ and $\vec{v} \in \mathbb{R}^n$, then:

$$k \cdot \vec{v} = \vec{\mathbf{0}}_n \textbf{\textit{ if and only if }} \text{either } k = 0 \textbf{\textit{ or }} \vec{v} = \vec{\mathbf{0}}_n.$$

Again, this means that you need to prove both implications:

a. ($\Leftarrow$) This part is easier: Prove that if $k = 0$ or $\vec{v} = \vec{\mathbf{0}}_n$, then $k \cdot \vec{v} = \vec{\mathbf{0}}_n$. You may certainly use some of the Exercises above.

b. ($\Rightarrow$) If $k \cdot \vec{v} = \vec{\mathbf{0}}_n$, prove that either $k = 0$ or $\vec{v} = \vec{\mathbf{0}}_n$.

Hint: review the proof of the Zero-Factors Theorem for Real Numbers from Chapter Zero.

29. Prove that for any $\vec{v} \in \mathbb{R}^n$: $-\vec{v} = (-1) \cdot \vec{v}$. Hint: Use some of the previous Exercises.

30. Let $\vec{u} = \langle u_1, u_2 \rangle \in \mathbb{R}^2$, and $P(a_1, b_1)$ a point on the Cartesian plane. If $\vec{u}$ is translated to P, prove that the head of $\vec{u}$ will be located at $Q(a_2, b_2)$, where $a_2 = a_1 + u_1$, and $b_2 = b_1 + u_2$.

31. Prove that the vector $\vec{u} \in \mathbb{R}^2$ from $P(a_1, b_1)$ to $Q(a_2, b_2)$ is $\vec{u} = \overrightarrow{PQ} = \langle a_2 - a_1, b_2 - b_1 \rangle$.

32. Let $\vec{u} = \langle u_1, u_2, u_3 \rangle \in \mathbb{R}^3$, and $P(a_1, b_1, c_1)$ a point in Cartesian space. If $\vec{u}$ is translated to P, prove that the head of $\vec{u}$ will be located at $Q(a_2, b_2, c_2)$, where $a_2 = a_1 + u_1$, $b_2 = b_1 + u_2$, and $c_2 = c_1 + u_3$.

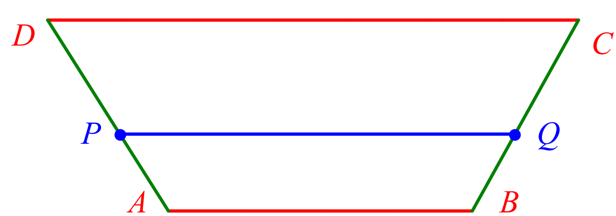
33. Prove that the vector $\vec{u} \in \mathbb{R}^3$ from $P(a_1, b_1, c_1)$ to $Q(a_2, b_2, c_2)$ is:

$$\vec{u} = \overrightarrow{PQ} = \langle a_2 - a_1, b_2 - b_1, c_2 - c_1 \rangle.$$

34. Use the previous Exercise (and scalar multiplication) to prove that the midpoint of the line segment PQ, where $P = (a_1, b_1, c_1)$ and $Q = (a_2, b_2, c_2)$ is:

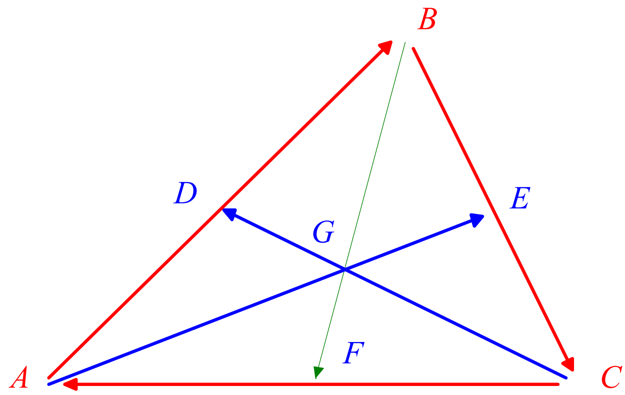
$$\left(\frac{a_1 + a_2}{2}, \frac{b_1 + b_2}{2}, \frac{c_1 + c_2}{2} \right).$$

35. Let $ABCD$ be a trapezoid with parallel sides AB and DC. Suppose that the length of AB is 20 centimeters and the length of DC is 35 centimeters. Let P be a point 2/5 up from A to D and similarly let Q be a point 2/5 up from B to C. Prove using a vector diagram and vector arithmetic that PQ is also parallel to AB and find its length.



36. We will use vectors to prove that ***the three medians of a triangle intersect at one mutual point***. Recall that a median of a triangle is a line segment connecting a vertex to the midpoint of the opposite side. Since we are going to use vectors to prove our Theorem, we will treat all line segments as vectors.

Let us consider $\triangle ABC$ below, with sides $\overrightarrow{AB}$, $\overrightarrow{BC}$ and $\overrightarrow{CA}$. Their midpoints are D, E and F:



The medians are $\overrightarrow{AE}$, $\overrightarrow{CD}$ and $\overrightarrow{BF}$.

Your task is to show that all three intersect at G. Since D is the midpoint of AB, we have $\overrightarrow{AD} = \frac{1}{2}\overrightarrow{AB}$, and similarly for the other three sides. The proof of our Theorem is ***coordinate free***, that is, independent of the Cartesian coordinate system, and in particular, we will not be using the midpoint formula. Fill in the details and complete the proof:

a. Explain why $\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CA} = \vec{\mathbf{0}}_2$.

b. Explain why $\overrightarrow{AE} = \overrightarrow{AB} + \frac{1}{2}\overrightarrow{BC}$.

c. Derive a similar equation for $\overrightarrow{CD}$, also of the form $\overrightarrow{??} + \frac{1}{2}\overrightarrow{??}$.

d. $\overrightarrow{CD}$ and $\overrightarrow{AE}$ clearly intersect at one point, so let us call that G. What is the vector equation (with one side a vector sum) involving $\overrightarrow{AG}$, $\overrightarrow{CG}$ and $\overrightarrow{CA}$?

e. Use (a) to rewrite (b) as $\overrightarrow{AE} = \frac{1}{2}\overrightarrow{AB} - \frac{1}{2}\overrightarrow{CA}$.

f. Your equation in (c) should also involve $\frac{1}{2}\overrightarrow{AB}$. Subtract the two sides of your equations in (c) and (e) to get $\overrightarrow{CD} - \overrightarrow{AE} = \frac{3}{2}\overrightarrow{CA}$.

g. Solve for $\overrightarrow{CA}$ in (f) and substitute it in your equation in (d). Show that you can rewrite the equation you obtain as: $\overrightarrow{AG} - \frac{2}{3}\overrightarrow{AE} = \overrightarrow{CG} - \frac{2}{3}\overrightarrow{CD}$

h. $\overrightarrow{AG}$ and $\overrightarrow{AE}$ are clearly ***parallel*** to each other, and similarly, so are $\overrightarrow{CG}$ and $\overrightarrow{CD}$. However, $\overrightarrow{AG}$ and $\overrightarrow{CG}$ are clearly ***not*** parallel to each other. Now comes the part with the hard thinking: explain why these statements imply that both sides of the equation in (g) must be the ***zero vector***. There are no computations involved in this step.

i. Now that we know from (g) that $\overrightarrow{AG} = \frac{2}{3}\overrightarrow{AE}$, and $\overrightarrow{CG} = \frac{2}{3}\overrightarrow{CD}$, ***solve*** for $\overrightarrow{BG}$ and show that it is also $\frac{2}{3}\overrightarrow{BF}$. This proves that G is also somewhere between B and F, completing the proof that G lies on all three medians. As a bonus, we also get the famous result that G lies 2/3 along the median from each vertex.

37. Using the same diagram as the previous Exercise, show using vector arithmetic, that DE is parallel to AC, and is half of its length. In other words, the line segment connecting the midpoint of two sides of a triangle is parallel to the third side, and is half its length.

1.2 The Span of a Set of Vectors

Now we introduce a central object in Linear Algebra:

Definition: The ***Span*** of a non-empty set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ from $\mathbb{R}^n$ is the set of ***all possible linear combinations*** of the vectors in the set. We write:

$$\begin{aligned} Span(S) &= Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}) \\ &= \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k \,|\, x_1, x_2, \ldots, x_k \in \mathbb{R}\}. \end{aligned}$$

We note that the individual vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ are all members of $Span(S)$, where we let $x_i = 1$ and all the other coefficients 0 in order to produce $\vec{v}_i$. Similarly, the zero vector $\vec{\mathbf{0}}_n$ is also a member of $Span(S)$, where we make all the coefficients x_i zero to produce $\vec{\mathbf{0}}_n$.

The simplest possible set S consists of only the zero vector of $\mathbb{R}^n$. Since $r \cdot \vec{\mathbf{0}}_n = \vec{\mathbf{0}}_n$ for any scalar r, the Span of $\left\{\vec{\mathbf{0}}_n\right\}$ consists only of $\vec{\mathbf{0}}_n$ and no other vector. Thus:

Theorem: In any $\mathbb{R}^n$: $Span\left(\left\{\vec{\mathbf{0}}_n\right\}\right) = \left\{\vec{\mathbf{0}}_n\right\}$.

For the same reason, we can choose to ***exclude*** the zero vector from a set of vectors S, since it is automatically a member of any $Span(S)$, that is:

Theorem: For all $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k \in \mathbb{R}^n$:

$$Span\left(\left\{\vec{\mathbf{0}}_n, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\right\}\right) = Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}).$$

At the other extreme, we saw in the previous Section that every vector $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle \in \mathbb{R}^n$ can be written ***uniquely*** as a linear combination of the standard basis vectors:

$$\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle = x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n.$$

Thus, we can formally say:

Theorem: $\mathbb{R}^n = Span(\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\})$.

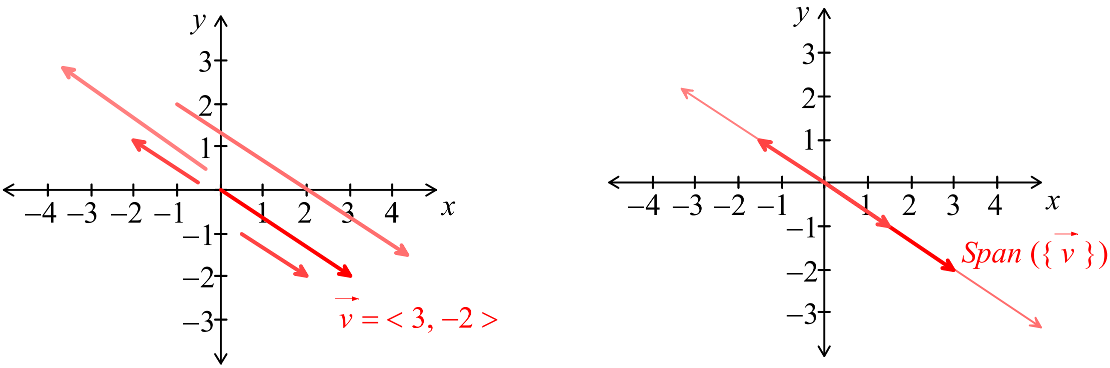
The Span of One or Two Vectors

Let us look at what the Span of a small set of vectors looks like in $\mathbb{R}^2$ and $\mathbb{R}^3$, beginning with a single non-zero vector $\vec{v}$ in $\mathbb{R}^2$.

Example: Suppose that $\vec{v} = \langle 3, -2 \rangle \in \mathbb{R}^2$, and $S = \{\vec{v}\}$. By the definition:

$$Span(S) = \{t\langle 3, -2 \rangle | t \in \mathbb{R}\},$$

that is, all the scalar multiples of $\langle 3, -2 \rangle$. Let us show some of these vectors:



Some Vectors from $Span(\{\langle 3,-2\rangle\})$

We know that every multiple $t\langle 3,-2\rangle$ is ***parallel*** to $\langle 3,-2\rangle$, as we see in the diagram on the left. We deliberately did not put the vectors in standard position so that we can see each individual vector distinctly. However, when we put them all in standard position as we see on the right, then the arrowheads of these multiples lie on what we instinctively call a ***line***, and it passes through the origin since $\vec{\mathbf{0}}_{\mathbf{2}}$ is in $Span(S)$. Indeed, our knowledge of basic algebra will allow us to find the Cartesian equation of this line: any vector $\langle x,y\rangle$ is a member of $Span(\vec{v})$ ***if and only if:***

$$\langle x,y\rangle = t\langle 3,-2\rangle = \langle 3t,-2t\rangle, \quad \text{or in other words:}$$
$$x = 3t \quad \text{and} \quad y = -2t.$$

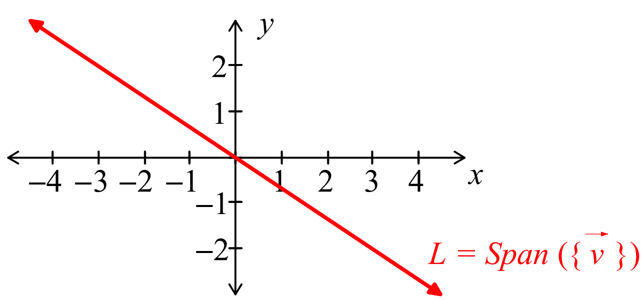
These are examples of what we call ***parametric equations***. The scalar variable t is called the ***parameter***. By eliminating t, we get the usual Cartesian equation of a line:

$$t = \frac{x}{3} = \frac{y}{-2}, \quad \text{and thus } y = -\frac{2}{3}x.$$

Thus, we can conclude that:

$Span(\{\langle 3,-2\rangle\})$ is the line on the Cartesian plane with equation $y = -\frac{2}{3}x$.

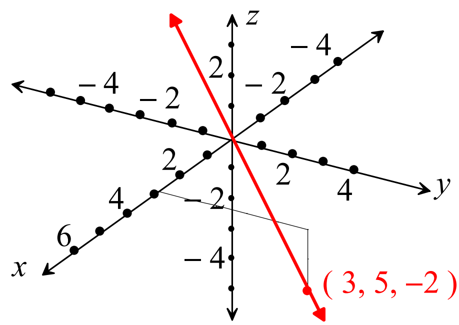
As we do in basic algebra, we will draw this line with an arrowhead on both ends and denote it by an appropriate symbol such as L:



The ***Line*** $L = Span(\{\langle 3,-2\rangle\})$ on the Cartesian Plane. □

Similarly, let us see what the Span of a single non-zero vector in $\mathbb{R}^3$ looks like:

Example: Suppose that $\vec{w} = \langle 3, 5, -2 \rangle \in \mathbb{R}^3$, and $S = \{\vec{w}\}$. As before, $Span(S)$ consists of all the scalar multiples of $\vec{w}$, which are all parallel to $\vec{w}$. Thus, when they are all drawn in standard position, the arrowheads will again form what we intuitively call a line which passes through the origin like before. To draw it, we locate the point $(3, 5, -2)$ and connect this to the origin, as seen below:



The Line $L = Span(\{\langle 3, 5, -2 \rangle\})$ in Cartesian Space

Any vector $\langle x, y, z \rangle$ is a member of L ***if and only if***:

$$\langle x, y, z \rangle = t\langle 3, 5, -2 \rangle = \langle 3t, 5t, -2t \rangle,$$

for some value of the parameter t. If we tried to eliminate t, we get:

$$t = \frac{x}{3} = \frac{y}{5} = \frac{z}{-2}.$$

We call these ***symmetric equations*** for L. Unfortunately, we cannot combine all three expressions into a single equation as we can do in $\mathbb{R}^2$. $\square$

Since our intuition works well in $\mathbb{R}^2$ and $\mathbb{R}^3$, let us make the following generalization:

Definition — Axiom for a Line:
If $\vec{v} \in \mathbb{R}^n$ is a ***non-zero*** vector, then $Span(\{\vec{v}\})$ is geometrically a ***line*** L in $\mathbb{R}^n$ passing through the origin.

We can now proceed with the Span of two non-zero vectors $\vec{u}$ and $\vec{v}$. Suppose first that $\vec{u}$ and $\vec{v}$ were ***parallel*** to each other, that is, $\vec{u} = k\vec{v}$ for some (non-zero) scalar k. Then, a linear combination of $\vec{u}$ and $\vec{v}$ would look like:

$$r\vec{u} + s\vec{v} = r(k\vec{v}) + s\vec{v} = (rk + s)\vec{v} = t\vec{v},$$

where $t = rk + s$ is just another scalar (notice that t can still be any real number). Thus we are back to the Span of a single vector $\vec{v}$. Similarly, by solving for $\vec{v}$ in terms of $\vec{u}$, we can express $r\vec{u} + s\vec{v}$ as a multiple of $\vec{u}$ alone. Notice that the argument above does not require our vectors to be in $\mathbb{R}^2$ or $\mathbb{R}^3$. Thus:

Theorem: If $\vec{u}$ and $\vec{v}$ are ***non-zero*** and ***parallel*** vectors from some $\mathbb{R}^n$, then:

$$Span(\{\vec{u}, \vec{v}\}) = Span(\{\vec{v}\}) = Span(\{\vec{u}\}).$$

Now let us see what happens when $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^2$ which are ***not parallel*** to each other. Note that this automatically includes the restriction that neither $\vec{u}$ nor $\vec{v}$ is the zero vector.

Example: Let us investigate $Span(\{\langle 5, 2\rangle, \langle 3, 4\rangle\})$ in $\mathbb{R}^2$. Clearly these two vectors are not scalar multiples of each other. Let us try to describe all their linear combinations:

$$\vec{w} = \langle x, y\rangle = r\langle 5, 2\rangle + s\langle 3, 4\rangle.$$

In other words, let us try to find any ***restrictions*** on x and y, that will allow us to ***solve*** the equations:

$$\begin{aligned} x &= 5r + 3s, \quad \text{and} \\ y &= 2r + 4s, \end{aligned}$$

for the scalars r and s. This is a linear system of two equations in two variables. As in basic algebra, we can ***eliminate*** one variable, say r, by multiplying each equation by a suitable constant and adding their corresponding sides together. In this case:

$$\begin{aligned} -2x &= -10r - 6s, && \text{and} \\ 5y &= 10r + 20s, && \text{so:} \\ 5y - 2x &= 14s, && \text{or:} \\ s &= \frac{5y - 2x}{14}. \end{aligned}$$

Similarly, by eliminating s, we get:

$$r = \frac{4x - 3y}{14}.$$

Thus, there are ***no restrictions*** on x and y that will prevent us from finding scalars r and s so that:

$$\vec{w} = \langle x, y\rangle = r\langle 5, 2\rangle + s\langle 3, 4\rangle.$$

This means that $Span(\{\langle 5, 2\rangle, \langle 3, 4\rangle\})$ is ***all*** of $\mathbb{R}^2$. □

This is true in general, as we state in the next Theorem. We will guide you through a proof in the Exercises:

Theorem: If $\vec{u}, \vec{v} \in \mathbb{R}^2$ are ***non-parallel*** vectors, then:

$$Span(\{\vec{u}, \vec{v}\}) = \mathbb{R}^2.$$

In other words, ***any*** vector $\vec{w} \in \mathbb{R}^2$ can be expressed as a ***linear combination:***

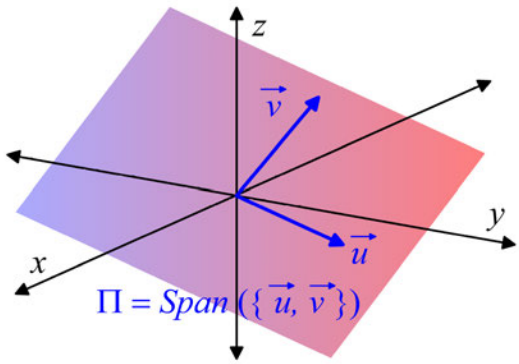
$$\vec{w} = r\vec{u} + s\vec{v},$$

for some scalars r and s.

Geometrically, this means that if we drew all the vectors $r\vec{u} + s\vec{v}$ in $\mathbb{R}^2$, in standard position, their arrowheads will cover ***the entire Cartesian plane.***

Now, suppose that $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^3$ that are not parallel to each other. If we draw them in standard position in $\mathbb{R}^3$, then the origin, together with the tips of $\vec{u}$ and $\vec{v}$ are ***not*** collinear. Thus, they form a ***triangle*** (recall that this is one of the ***Axioms*** of geometry that we mentioned in Chapter Zero). Intuitively, then, we can pretend that $\vec{u}$ and $\vec{v}$ are still vectors in $\mathbb{R}^2$, so once again, their Span is a geometric object that "looks like" all of $\mathbb{R}^2$. In other words, their arrowheads will form an object that we intuitively call a ***plane***. Thus, it is reasonable to accept the following:

Definition — Axiom for a Plane in Cartesian Space:
If $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^3$ that are ***not parallel*** to each other, then $Span(\{\vec{u}, \vec{v}\})$ is geometrically a ***plane*** Π in Cartesian space that passes through the origin.
(Π is the capital form of the lowercase Greek letter π.)



The Span of Two Non-Parallel Vectors is a ***Plane*** in $\mathbb{R}^3$

Example: Let us study $Span(\{\langle 3, 4, -1\rangle, \langle 5, -2, 3\rangle\})$. Notice that we can see immediately that the two vectors are not scalar multiples of each other. A vector $\vec{w} = \langle x, y, z\rangle$ is a member of the Span ***if and only if*** we can write:

$$\begin{aligned} \langle x, y, z\rangle &= r\langle 3, 4, -1\rangle + s\langle 5, -2, 3\rangle \\ &= \langle 3r + 5s, 4r - 2s, -r + 3s\rangle, \end{aligned}$$

for some scalars r and s. In other words:

$$\begin{aligned} x &= 3r + 5s, \\ y &= 4r - 2s, \quad \text{and} \\ z &= -r + 3s. \end{aligned}$$

These are again called ***parametric equations***, but this time, we have ***two*** parameters r and s. Let us see if we can make all this data more compact. Eliminating r from x and y using z, we get:

$$\begin{aligned} x + 3z &= 3r + 5s + 3(-r + 3s) = 14s, \quad \text{and} \\ y + 4z &= 4r - 2s + 4(-r + 3s) = 10s. \end{aligned}$$

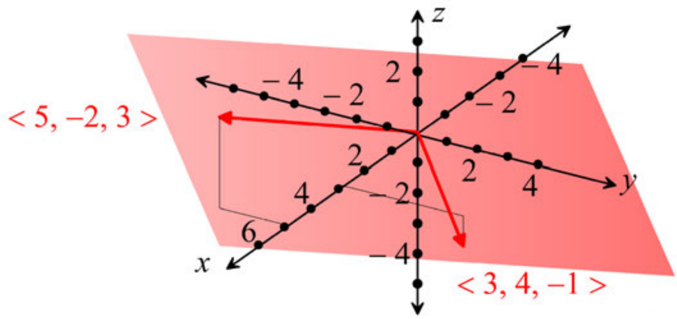
Finally, eliminating s, we get:

$$\begin{aligned} \frac{x + 3z}{14} &= \frac{y + 4z}{10} \quad \Leftrightarrow \\ 5(x + 3z) &= 7(y + 4z) \quad \Leftrightarrow \\ 5x - 7y - 13z &= 0. \end{aligned}$$

We summarize:

$$Span(\{\langle 3, 4, -1\rangle, \langle 5, -2, 3\rangle\}) = \{\langle x, y, z\rangle \in \mathbb{R}^3 \,|\, 5x - 7y - 13z = 0\}.$$

As we see in the diagram above, we will sketch our planes using a ***parallelogram***. In order to give the viewer the illusion that the parallelogram contains the two vectors that determine our plane, we will draw the sides of our parallelogram to be parallel to these two vectors:



The Plane $Span(\{\langle 3, 4, -1\rangle, \langle 5, -2, 3\rangle\})$. □

In general, we can make the following definition, which we will improve later:

Definition: The ***Cartesian equation*** of a plane through the ***origin*** in Cartesian space, given in the form $\Pi = Span(\{\vec{u}, \vec{v}\})$, where $\vec{u}$ and $\vec{v}$ are not parallel, has the form:

$$ax + by + cz = 0,$$

for some constants, a, b and c, where at least one coefficient is non-zero.

General Lines

Although the Span of a set of vectors is an important object in Linear Algebra, it has its limitations. We certainly want to consider lines and planes that do not pass though the origin. Recall that we can translate a vector so that its tail is not at the origin. In general, we will want to ***translate*** an entire Span:

$$Q = \{\vec{q} + \vec{v} | \vec{v} \in Span(S)\},$$

for some ***fixed*** non-zero vector $\vec{q} \in \mathbb{R}^n$, whose head is the location of the tail of $\vec{v}$. In particular, we will make the following:

Definition — Axiom for a General Line:
A ***line*** L in $\mathbb{R}^n$ is the translate of the Span of a single ***non-zero*** vector $\vec{d} \in \mathbb{R}^n$:

$$L = \left\{\vec{x}_p + t\vec{d} | t \in \mathbb{R}\right\},$$

for some vector $\vec{x}_p \in \mathbb{R}^n$. We may think of $\vec{d}$ as a ***direction vector*** of L, and any non-zero multiple of $\vec{d}$ can also be used as a direction vector for L. By setting t to zero, we see that $\vec{x}_p$ is a ***particular vector*** on the line L, whose head is on L.

Example: Consider the line L in Cartesian space passing through the point $(3, -2, 7)$ in the direction of $\langle 4, 1, -3\rangle$. We can think of L as the translate of $\text{Span}(\{\langle 4, 1, -3\rangle\})$ by the vector $\langle 3, -2, 7\rangle$. We can thus form the following parametric equations for L:

$$\langle x, y, z\rangle = \langle 3, -2, 7\rangle + t\langle 4, 1, -3\rangle = \langle 3 + 4t, -2 + t, 7 - 3t\rangle.$$

Notice that we used two different symbols: $(3,-2,7)$ is a ***point*** in Cartesian space, but $\langle 3,-2,7\rangle$ and $\langle 4,1,-3\rangle$ are ***vectors*** in $\mathbb{R}^3$. □

Example: Let us study the line L in Cartesian space passing through the two points $P(2,4,-1)$ and $Q(-1,-3,1)$. This line has direction vector:

$$\vec{d} = \overrightarrow{PQ} = \langle -1-2, -3-4, 1-(-1)\rangle = \langle -3,-7,2\rangle.$$

Now, we can find ***parametric equations*** for L, as before:

$$\langle x,y,z\rangle = \langle -1,-3,1\rangle + t\langle -3,-7,2\rangle.$$

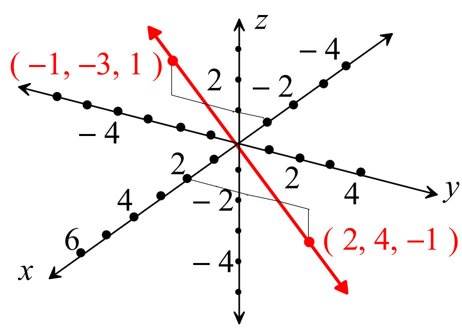
Notice that we could also have used the second point as our particular point, and any scalar multiple of the direction vector that we found, so another correct answer is:

$$\langle x,y,z\rangle = \langle 2,4,-1\rangle + s\langle 6,14,-4\rangle.$$

Again, if we try to solve for t in our first set of parametric equations, then we would get:

$$t = \frac{x+1}{-3} = \frac{y+3}{-7} = \frac{z-1}{2},$$

which we again call ***symmetric equations*** for L. We see its graph below:



The Line Passing through $P(2,4,-1)$ and $Q(-1,-3,1)$

It almost looks as if the ***origin*** is on this line, but if so, then $(0,0,0)$ we must satisfy the symmetric equations for L. This would mean that the value of t at the origin is:

$$t = \frac{1}{-3} = \frac{3}{-7} = \frac{-1}{2},$$

which is obviously false. Thus, it is easy to be fooled by a graph in Cartesian space. □

In general, we can define a general line in Cartesian space as follows:

Definition: A line L in Cartesian space passing through the ***point*** (x_0,y_0,z_0), and with non-zero ***direction vector*** $\vec{d} = \langle a,b,c\rangle$ can be specified using a ***vector equation***, in the form:

$$\langle x,y,z\rangle = \vec{x}_p + t\vec{d} = \langle x_0,y_0,z_0\rangle + t\langle a,b,c\rangle, \text{ where } t\in\mathbb{R}.$$

If ***none*** of the components of $\vec{d}$ are zero, we can obtain ***symmetric equations*** for L, of the form: $\dfrac{x-x_0}{a} = \dfrac{y-y_0}{b} = \dfrac{z-z_0}{c}$.

In the same way that two vectors can be parallel to each other, we will also define the analogous concept for lines. The second statement will be proven in the Exercises.

Definition/Theorem: We will say that two lines are ***parallel*** to each other if they are ***different*** translates of the same line through the origin, that is, they, have ***no point in common***. Consequently, two lines L_1 and L_2 are parallel to each other ***if and only if*** their vector equations can be written as:

$$L_1 = \vec{x}_1 + t\vec{d} \quad \text{and} \quad L_2 = \vec{x}_2 + s\vec{d},$$

for some common non-zero direction vector $\vec{d}$, and where $\vec{x}_1 - \vec{x}_2$ is ***not parallel*** to $\vec{d}$.

General Planes

Let us now turn our attention to general planes in $\mathbb{R}^n$. As with lines, the idea is to allow a plane through the origin to be translated to any other point:

Definition — Axiom for a General Plane:
A ***plane*** Π in $\mathbb{R}^n$ is the translate of a Span of two ***non-parallel*** vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$:

$$\Pi = \{\vec{x} = \vec{x}_p + r\vec{u} + s\vec{v} \mid r, s \in \mathbb{R}\},$$

for some ***fixed*** vector $\vec{x}_p \in \mathbb{R}^n$. However, the choice of $\vec{x}_p$ is ***not*** unique.

We see that by setting r and s to zero that $\vec{x}_p$ is a ***particular vector*** on Π. As with lines, we say that two ***distinct*** planes are ***parallel*** to each other if they are translates of the same plane through the origin.

There are many creative ways to specify a plane in Cartesian space. We can do this, for example, by:

- requiring the plane to contain three non-collinear points.
- requiring the plane to contain two intersecting lines.
- requiring the plane to contain two parallel lines.

We will see more ways in the Exercises, both in this Section and the following one.

Example: Let us find parametric equations and a Cartesian equation for the plane Π passing through $A(1,-3,2)$, $B(-1,-2,1)$ and $C(2,3,-1)$. This time, we will fix any of the three points, say A, and find the vectors connecting A to the other two points:

$$\vec{u} = \overrightarrow{AB} = \langle -1,-2,1\rangle - \langle 1,-3,2\rangle = \langle -2,1,-1\rangle, \text{ and}$$
$$\vec{v} = \overrightarrow{AC} = \langle 2,3,-1\rangle - \langle 1,-3,2\rangle = \langle 1,6,-3\rangle.$$

Notice that these two vectors are ***not parallel*** to each other. Thus the three points do not fall on the same line, and they indeed determine a ***triangle***, and hence a unique plane as before. Now, we can find parametric equations for Π in ***two parameters***, using $(1,-3,2)$ as the base point (converting it to the particular vector $\vec{x}_p$):

$$\begin{aligned}\langle x,y,z\rangle &= \langle 1,-3,2\rangle + r\langle -2,1,-1\rangle + s\langle 1,6,-3\rangle \\ &= \langle 1-2r+s, -3+r+6s, 2-r-3s\rangle.\end{aligned}$$

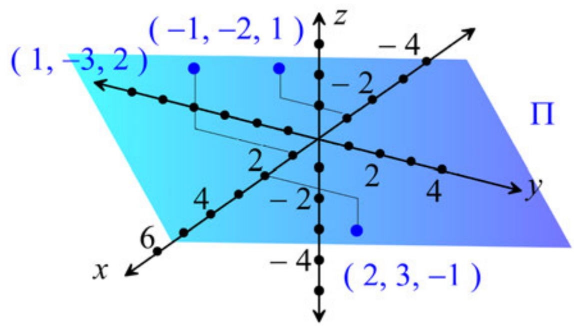
We can find a single equation for Π by eliminating r and s, one at a time, just like we did when the plane passed through the origin. Let us use z to eliminate r in x and y:

$$x - 2z = 1 - 2r + s - 2(2 - r - 3s) = -3 + 7s, \quad \text{and}$$
$$y + z = -3 + r + 6s + (2 - r - 3s) = -1 + 3s.$$

Now we eliminate s:

$$\frac{x - 2z + 3}{7} = s = \frac{y + z + 1}{3} \qquad \Rightarrow$$
$$3(x - 2z + 3) = 7(y + z + 1) \qquad \Rightarrow$$
$$3x - 7y - 13z = -2.$$

This is again called a ***Cartesian equation*** for Π. We see the graph of Π below, where the parallelogram representing Π has sides parallel to the vectors $\vec{u}$ and $\vec{v}$ that we computed above:



A Plane Determined by Three Non-Collinear Points□

It seems reasonable, then, to agree on the following:

Definition: A plane Π in Cartesian space can be specified using a ***Cartesian equation:***

$$ax + by + cz = d,$$

for some constants, a, b, c and d, where either a or b or c is non-zero. It is unique only up to a ***scalar:*** any other Cartesian equation for Π must have the form $kax + kby + kcz = kd$, for some ***non-zero*** constant k.

The plane passes through the origin ***if and only if*** $d = 0$.

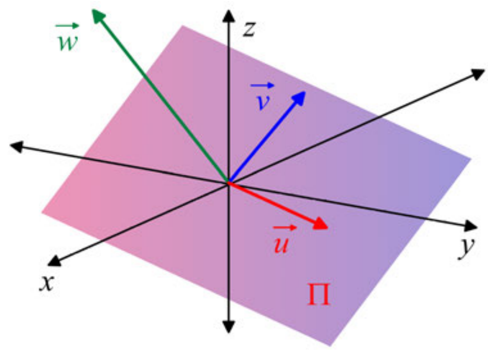
The Span of Three Non-Coplanar Vectors

It should not be a surprise that one of the Theorems that we will prove in the future is a natural analog in $\mathbb{R}^3$ about our Theorem concerning two non-parallel vectors in $\mathbb{R}^2$. You will prove it in the Exercises of Section 1.7:

Theorem: If $\vec{u}$, $\vec{v}$ and $\vec{w}$ are ***non-parallel*** and ***non-coplanar*** vectors in $\mathbb{R}^3$, that is, none of these vectors is on the plane determined by the other two vectors, then:

$$Span(\{\vec{u}, \vec{v}, \vec{w}\}) = \mathbb{R}^3.$$

In other words, any vector $\vec{z} \in \mathbb{R}^3$ can be expressed as a linear combination, $\vec{z} = r\vec{u} + s\vec{v} + t\vec{w}$, for some scalars r, s and t.



If $\vec{u}$, $\vec{v}$ and $\vec{w}$ are ***Non-Coplanar*** Vectors in $\mathbb{R}^3$, then $Span(\{\vec{u}, \vec{v}, \vec{w}\}) = \mathbb{R}^3$

This Theorem will be demonstrated in the Exercises of this Section.

1.2 Section Summary

The ***Span*** of a non-empty set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ from $\mathbb{R}^n$ is the set of ***all possible linear combinations*** of the vectors in the set:

$$Span(S) = Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}) = \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k \,|\, x_1, x_2, \ldots, x_k \in \mathbb{R}\}.$$

If $\vec{v} \in \mathbb{R}^n$ is a ***non-zero*** vector, then $Span(\{\vec{v}\})$ is a ***line*** L in $\mathbb{R}^n$ passing through the origin.

If $\vec{u}, \vec{v} \in \mathbb{R}^2$ are ***non-parallel*** vectors, then $Span(\{\vec{u}, \vec{v}\}) = \mathbb{R}^2$.

If $\vec{u}$, $\vec{v} \in \mathbb{R}^3$ are ***non-parallel*** vectors, then $Span(\{\vec{u}, \vec{v}\})$ is geometrically a ***plane*** Π in Cartesian space that passes through the origin.

More generally, a ***line*** L in $\mathbb{R}^n$ is the translate of the Span of a single ***non-zero*** vector $\vec{d} \in \mathbb{R}^n$: $L = \left\{\vec{x}_p + t\vec{d} \,|\, t \in \mathbb{R}\right\}$, for some vector $\vec{x}_p \in \mathbb{R}^n$.

A line L in Cartesian space passing through the point (x_0, y_0, z_0), and with non-zero direction vector $\vec{d} = \langle a, b, c\rangle$ can be specified using a ***vector equation***, in the form $\langle x, y, z\rangle = \langle x_0, y_0, z_0\rangle + t\langle a, b, c\rangle$, where $t \in \mathbb{R}$. If ***none*** of the components of $\vec{d}$ are zero, we can obtain ***symmetric equations*** for L, of the form:

$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}.$$

A ***plane*** Π in $\mathbb{R}^n$ is the translate of a Span of two ***non-parallel*** vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$: $\Pi = \{\vec{x} = \vec{x}_p + r\vec{u} + s\vec{v} \,|\, r, s \in \mathbb{R}\}$, for some $\vec{x}_p \in \mathbb{R}^n$.

A plane Π in Cartesian space can be specified using a ***Cartesian equation***, in the form:

$$ax + by + cz = d,$$

for some constants, a, b, c and d, where either a, or b, or c, is non-zero. The plane passes through the origin ***if and only if*** $d = 0$.

If $\vec{u}$, $\vec{v}$ and $\vec{w}$ are ***non-coplanar*** vectors in $\mathbb{R}^3$, then $Span(\{\vec{u}, \vec{v}, \vec{w}\}) = \mathbb{R}^3$.

1.2 Exercises

1. Find the Cartesian equation of the line in $\mathbb{R}^2$ determined by $Span(\{\langle 7, 4\rangle\})$. Sketch the line.
2. Find the Cartesian equation of the line in $\mathbb{R}^2$ determined by $Span(\{\langle 3, -5\rangle\})$. Sketch the line.
3. Find parametric and symmetric equations for the line in $\mathbb{R}^3$ determined by $Span(\{\langle 5, -4, 2\rangle\})$. Sketch the line.
4. Find parametric and symmetric equations for the line in $\mathbb{R}^3$ determined by $Span(\{\langle -1, 3, -6\rangle\})$. Sketch the line.
5. Find a Cartesian equation, in the form $ax + by = c$, for the line in $\mathbb{R}^2$ that passes through the point $(-2, 4)$ and has direction vector $\langle -5, 7\rangle$. Sketch the line.
6. Find parametric and symmetric equations for the line in $\mathbb{R}^3$ that passes through the point $(2, -7, 4)$ and has direction vector $\langle -3, 6, 8\rangle$. Sketch the line.
7. Find parametric equations for the line in $\mathbb{R}^3$ that passes through the point $(3, 2, -5)$ and has direction vector $\langle 2, 0, -5\rangle$. Sketch the line. Is it possible to write symmetric equations for this line? Why or why not?
8. Find parametric equations for the line in $\mathbb{R}^3$ that passes through the points $P(-4, 3, -5)$ and $Q(0, 1, -2)$. Sketch the line.
9. Find a Cartesian equation for the plane in $\mathbb{R}^3$ determined by $Span(\{\langle 1, 0, -2\rangle, \langle 3, 1, 5\rangle\})$. Sketch the plane.
10. Find a Cartesian equation for the plane in $\mathbb{R}^3$ determined by $Span(\{\langle 3, 5, -4\rangle, \langle 2, -1, 7\rangle\})$. Sketch the plane.
11. Find a Cartesian equation for the plane in $\mathbb{R}^3$ determined by $Span(\{\langle 3, 0, -2\rangle, \langle 1, 5, 0\rangle\})$. Sketch the plane. Why is this problem different from the last two?
12. Is $Span(\{\langle 4, -10, 6\rangle, \langle -6, 15, -9\rangle\})$ a line or a plane in $\mathbb{R}^3$? Why?
13. Find a Cartesian equation for the plane containing $A(1, -2, 4)$, $B(3, 1, -1)$ and $C(2, 0, 1)$.
14. Find a Cartesian equation for the plane containing $A(-4, 7, 3)$, $B(2, 0, -5)$ and $C(6, -3, -2)$.
15. Do the three points $A(-4, 3, 2)$, $B(11, -17, 12)$ and $C(-10, 11, -2)$ determine a plane or only a line in $\mathbb{R}^3$? Why?
16. Show that the line given by the vector equation $\langle x, y, z\rangle = \langle 3 - 5t, 7 + 2t, 4 - t\rangle$ is exactly the same as the line given by $\langle x, y, z\rangle = \langle 8 + 10s, 5 - 4s, 5 + 2s\rangle$.

    Hint: show that they have at least one point in common, and they have parallel direction vectors.
17. Show that the line L given by $\langle x, y, z\rangle = \langle 3, -4, 7\rangle + t\langle 4, -3, 6\rangle$ is completely contained in the plane with Cartesian equation $3x + 8y + 2z = -9$.
18. Consider the line L with vector equation $\langle x, y, z\rangle = \langle 3, 4, -1\rangle + t\langle 1, -2, 5\rangle$ and the point $P(2, 5, 7)$. Show that P is not on L, and then find a Cartesian equation for the plane that contains both P and L.
19. Find a Cartesian equation for the plane that contains the line $\langle x, y, z\rangle = \langle 3, 5, -4\rangle + t\langle -2, 1, 3\rangle$ and the point $(3, -7, 2)$.
20. At what point does the line with parametric equations $\langle x, y, z\rangle = \langle -2, 0, 3\rangle + t\langle 1, -3, 5\rangle$ intersect the plane $3x - 5y + 8z = 4$?
21. Show that the line with parametric equations $\langle x, y, z\rangle = \langle -2, 1, 7\rangle + t\langle 8, 5, -4\rangle$ does not intersect the plane $2x - 4y - z = 5$.

22. Prove that two lines L_1 and L_2 in $\mathbb{R}^3$ are parallel to each other ***if and only if*** their vector equations can be written as:

$$L_1 = \vec{x}_1 + t\vec{d} \quad \text{and} \quad L_2 = \vec{x}_2 + s\vec{d}$$

for some common non-zero direction vector $\vec{d}$, and where $\vec{x}_1 - \vec{x}_2$ is ***not parallel*** to $\vec{d}$.

23. ***Parallel Lines:*** Let L_1 be the line with symmetric equations:

$$L_1 : \frac{x-7}{3} = \frac{y}{5} = -z+1.$$

a. Show that the point $(5,-2,4)$ is ***not*** on L_1.
b. Find symmetric equations for the line L_2 passing through $(5,-2,4)$ that is parallel to L_1.

24. Find a Cartesian equation for the plane that contains the two parallel lines from Exercise 23.

25. Show that the lines $\langle x,y,z\rangle = \langle 2,1,-5\rangle + s\langle 8,12,-4\rangle$ and $\langle x,y,z\rangle = \langle -4,3,0\rangle + t\langle -6,-9,3\rangle$ are parallel, and find a Cartesian equation for the plane that contains both of them.

26. Show algebraically that if $\vec{u} = \langle u_1, u_2\rangle$ and $\vec{v} = \langle v_1, v_2\rangle$ are vectors in $\mathbb{R}^2$ that are ***not*** parallel to each other, then any vector $\langle x,y\rangle \in \mathbb{R}^2$ can be written as a linear combination of $\vec{u}$ and $\vec{v}$. Hint: Use the contrapositive that you wrote in Exercise 19 from Section 1.1. Show that for any x and y, we can solve the vector equation $\langle x,y\rangle = r\vec{u} + s\vec{v}$ for r and s. You will have to do a Case-by-Case Analysis. We suggest the cases: (1) neither u_1 nor u_2 is zero, and (2) either u_1 or u_2 is zero (explain why they cannot both be zero). Next, explain why we can further conclude that $Span(\{\vec{u},\vec{v}\}) = \mathbb{R}^2$.

27. Now, prove the Theorem in the previous Exercise ***geometrically***. Hint: Draw the two lines determined by $\vec{u}$ and $\vec{v}$. Draw a random vector $\vec{w}$ in standard position. If the tip of $\vec{w}$ is on one of these lines, then it is purely a multiple of either $\vec{u}$ or $\vec{v}$. However, if it is not, draw two lines through the tip of $\vec{w}$, one parallel to $\vec{u}$ and one parallel to $\vec{v}$ (essentially, draw a ***parallelogram***), and explain why this allows us to express $\vec{w}$ as a linear combination of $\vec{u}$ and $\vec{v}$.

28. Let $\vec{u} = \langle 3,-1,0\rangle$, $\vec{v} = \langle 0,1,-2\rangle$, and $\vec{w} = \langle -1,1,1\rangle$.

a. Clearly $\vec{u}$ is not parallel to $\vec{v}$. Find the Cartesian equation of the plane $Span(\{\vec{u},\vec{v}\})$.
b. Show that $\vec{w}$ is ***not*** in $Span(\{\vec{u},\vec{v}\})$.
c. Show that ***any*** vector $\langle x,y,z\rangle \in \mathbb{R}^3$ can be written as a linear combination:

$$\langle x,y,z\rangle = c_1\langle 3,-1,0\rangle + c_2\langle 0,1,-2\rangle + c_3\langle -1,1,1\rangle.$$

Hint: write the right side as a single vector, then write and ***solve*** the 3 equations in 3 unknowns c_1, c_2 and c_3 by comparing the three components.
d. Explain why this shows that $Span(\{\vec{u},\vec{v},\vec{w}\}) = \mathbb{R}^3$.

29. Suppose that $\vec{u} = \langle a,b\rangle$ is a non-zero vector in $\mathbb{R}^2$, where $a \neq 0$. Show that the line $L = Span(\{\vec{u}\})$ has Cartesian equation $y = \frac{b}{a}x$.

30. Suppose that $\vec{u} = \langle 0,b\rangle$ is a non-zero vector in $\mathbb{R}^2$. Show that the line $L = Span(\{\vec{u}\})$ has Cartesian equation $x = 0$. In other words, L is the y-axis.

31. Show that any line L in $\mathbb{R}^2$ with direction vector $\vec{d} = \langle r,s\rangle$ has a general equation of the form $sx - ry = c$, for some real number c. If you are further told that L passes through the point (x_0, y_0), what is c?

32. Suppose that $\vec{u}$, $\vec{v}$ and $\vec{w}$ are non-zero vectors in $\mathbb{R}^3$, where $\vec{u}$ and $\vec{v}$ are not parallel to each other. Thus, $Span(\{\vec{u},\vec{v}\})$ is a plane Π_1 in $\mathbb{R}^3$. Suppose also that $\vec{w}$ is ***not*** on Π_1.

a. Show that $\vec{w}$ ***cannot*** be parallel to either $\vec{u}$ or to $\vec{v}$.
b. Use (a) to explain why $Span(\{\vec{u},\vec{w}\})$ is likewise a plane. Let us call this new plane Π_2.
c. Show that $\vec{v}$ is ***not*** on Π_2.

d. Explain why $Span(\{\vec{v}, \vec{w}\})$ is also a plane, Π_3, and $\vec{u}$ is not on Π_3.

33. Suppose that Π is a plane in $\mathbb{R}^3$, and we assume in this entire problem that Π does ***not*** pass through the origin.

a. Show that Π can be specified by the Cartesian equation $ax + by + cz = 1$.

Now, suppose further in parts (b), (c) and (d) that a, b and c are not zero in our equation above. We define the ***x-intercept*** of Π as the point $(x_0, 0, 0)$ which is on Π, and similarly we can define the y and z intercepts of Π.

b. Show that the intercepts of Π are: $(1/a, 0, 0)$, $(0, 1/b, 0)$, and $(0, 0, 1/c)$.

c. Now, suppose that the intercepts of Π are $(x_0, 0, 0)$, $(0, y_0, 0)$ and $(0, 0, z_0)$. Show that Π can be specified by the Cartesian equation:

$$\frac{x}{x_0} + \frac{y}{y_0} + \frac{z}{z_0} = 1.$$

This is called the ***intercepts form*** of the Cartesian equation for Π.

d. Use (c) to find a Cartesian equation for the plane with intercepts $(10, 0, 0)$, $(0, -12, 0)$ and $(0, 0, 15)$. Simplify your answer so that the equation has the standard form $ax + by + cz = d$, where all of the coefficients are ***integers*** which are as small as possible, and $a > 0$.

e. Now, suppose Π does not have an x-intercept, but has two other intercepts, $(0, y_0, 0)$ and $(0, 0, z_0)$. Show that Π can be specified by the Cartesian equation: $\dfrac{y}{y_0} + \dfrac{z}{z_0} = 1$.

f. Write two statements analogous to the statement in (e) when (1) Π does not have a y-intercept but has two other intercepts, and similarly (2) Π does not have a z-intercept but has two other intercepts.

g. Use (e) or (f) to find a Cartesian equation for the plane with intercepts $(6, 0, 0)$ and $(0, 0, -9)$ but has no y-intercept. Simplify your answer so that the equation has the standard form $ax + by + cz = d$, where all of the coefficients are integers which are as small as possible and the first non-zero coefficient is positive.

h. Prove that a plane Π that does not pass through the origin has ***exactly one*** intercept ***if and only if*** Π has a Cartesian equation of the form $x = x_0$ or $y = y_0$ or $z = z_0$.

i. Find a Cartesian equation for the plane whose only intercept is $(0, 0, -5)$.

34. ***Shortest Distance from a Point to a Line:*** Suppose that L is the line with vector equation:

$$\langle x, y, z\rangle = \langle x_0, y_0, z_0\rangle + t\langle a, b, c\rangle, \text{ where } t \in \mathbb{R}.$$

Let us further assume that the direction vector $\vec{d} = \langle a, b, c\rangle$ is a ***unit vector***, and $Q(x_1, y_1, z_1)$ is a fixed point.

a. Let D be the ***square*** of the distance from Q to an arbitrary point (x, y, z) on L. Express D as a function of t.

b. Find $\dfrac{dD}{dt}$. Simplify your answer so that t appears exactly once.

c. Find the critical values of D, that is, the values of t where $\dfrac{dD}{dt} = 0$.

d. Find $\dfrac{d^2D}{dt^2}$ and show that this second derivative is always positive.

e. Explain why the critical value that you found in (c) must be a ***local*** minimum for D, and then explain further why this critical value must also be the ***absolute*** minimum for D.

35. Use Exercise 34 to find the point on the line $\langle x, y, z\rangle = \langle 5, -3, 2\rangle + t\langle 4, 1, -7\rangle$ which is closest to the origin, and find this distance. Reminder: in Exercise 34, we assumed that $\vec{d}$ was a ***unit vector***.

36. Use Exercise 34 to find the point on the line $\langle x, y, z\rangle = \langle 7, 0, -4\rangle + t\langle -1, 5, 2\rangle$ which is closest to the point $(3, -2, 6)$, and find this distance.

1.3 The Dot Product and Orthogonality

We can draw points, lines and planes on the Cartesian plane and in Cartesian space, and thus we can see the ***geometry*** of $\mathbb{R}^2$ and $\mathbb{R}^3$. However, we often want to study the angles formed by two vectors in higher dimensional spaces which we ***cannot*** visualize. In order to explore this further, we need the following general concept:

Definition: If $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ are vectors from $\mathbb{R}^n$, we define their ***dot product:***

$$\vec{u} \circ \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

Example: If $\vec{u} = \langle 5, -3, 0, 2, -7 \rangle$ and $\vec{v} = \langle 2, 5, 984, -6, -4 \rangle$, then:

$$\vec{u} \circ \vec{v} = 5(2) - 3(5) + 0(984) + 2(-6) - 7(-4) = 10 - 15 + 0 - 12 + 28 = 11. \square$$

The dot product of a vector with ***itself*** has a natural geometric interpretation. The following definition generalizes the concept of length that we introduced in Section 1.1:

Definitions: We define the ***length*** or ***norm*** or ***magnitude*** of a vector $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \in \mathbb{R}^n$ as the non-negative number:

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}\,.$$

It follows directly from the definition of the dot product that:

$$\|\vec{v}\|^2 = \vec{v} \circ \vec{v}, \quad \text{or in other words,} \quad \|\vec{v}\| = \sqrt{\vec{v} \circ \vec{v}}\,.$$

A vector with length 1 is called a ***unit vector***.

Notice that we get exactly the same definition for the length of a vector in $\mathbb{R}^2$ and $\mathbb{R}^3$ as we did in Section 1.1. Similarly, we can easily prove that:

Theorem: For any vector $\vec{v} \in \mathbb{R}^n$ and $k \in \mathbb{R}$: $\|k\vec{v}\| = |k|\|\vec{v}\|$.

In particular, if $\vec{v} \neq \vec{\mathbf{0}}_n$, then $\vec{u}_1 = \dfrac{1}{\|\vec{v}\|}\vec{v}$ is the ***unit vector*** in the ***same*** direction as $\vec{v}$, and $\vec{u}_2 = -\dfrac{1}{\|\vec{v}\|}\vec{v}$ is the ***unit vector*** in the ***opposite*** direction as $\vec{v}$. Furthermore:

$$\|\vec{v}\| = 0 \;\; \textbf{\textit{if and only if}} \;\; \vec{v} = \vec{\mathbf{0}}_n.$$

Example: The vector $\vec{v} = \langle 5, -3, 0, 2, -7 \rangle$ has length:

$$\|\vec{v}\| = \sqrt{25 + 9 + 0 + 4 + 49} = \sqrt{87}\,,$$

and thus the two unit vectors parallel to $\vec{v}$ are:

$$\frac{1}{\sqrt{87}}\langle 5, -3, 0, 2, -7 \rangle \quad \text{and} \quad \frac{-1}{\sqrt{87}}\langle 5, -3, 0, 2, -7 \rangle. \square$$

Notice that in $\mathbb{R}^2$ and $\mathbb{R}^3$, the length of the vector $\vec{u}$ is also the length of the directed line segment or arrow representing $\vec{u}$, as we saw in the Section 1.1. We also see that the ***standard basis vectors*** $\vec{e}_1$, $\vec{e}_2$, $\ldots$, $\vec{e}_n$ are unit vectors in each n-space, since all their components are 0 except for a single "1".

The following properties of the dot product are easily proven, just like we proved the properties of vector arithmetic in the previous section:

Theorem — Properties of the Dot Product:
For any vectors $\vec{u}$, $\vec{v}$, $\vec{w} \in \mathbb{R}^n$ and scalar $k \in \mathbb{R}$, we have:

1. ***The Commutative Property*** $\quad \vec{u} \circ \vec{v} = \vec{v} \circ \vec{u}.$
2. ***The Right Distributive Property*** $\quad \vec{u} \circ (\vec{v} + \vec{w}) = \vec{u} \circ \vec{v} + \vec{u} \circ \vec{w}.$
3. ***The Left Distributive Property*** $\quad (\vec{u} + \vec{v}) \circ \vec{w} = \vec{u} \circ \vec{w} + \vec{v} \circ \vec{w}.$
4. ***The Homogeneity Property*** $\quad (k \cdot \vec{u}) \circ \vec{v} = k(\vec{u} \circ \vec{v}) = \vec{u} \circ (k \cdot \vec{v}).$
5. ***The Zero-Vector Property*** $\quad \vec{u} \circ \vec{\mathbf{0}}_n = 0.$
6. ***The Positivity Property*** $\quad$ If $\vec{u} \neq \vec{\mathbf{0}}_n$, then $\vec{u} \circ \vec{u} > 0.$

The last two properties can be combined into one:

7. ***The Non-Degeneracy Property*** $\quad \vec{u} \circ \vec{u} > 0$ ***if and only if*** $\vec{u} \neq \vec{\mathbf{0}}_n$, and $\vec{\mathbf{0}}_n \circ \vec{\mathbf{0}}_n = 0.$

Example: Suppose we are told that $\vec{u}$ and $\vec{v}$ are two vectors from some $\mathbb{R}^n$ (which $\mathbb{R}^n$ is not really important). Suppose we were provided the information that $\|\vec{u}\| = 5$, $\|\vec{v}\| = 29$, and $\vec{u} \circ \vec{v} = 24$. Let us find $\|5\vec{u} + 3\vec{v}\|$.

This problem will help us appreciate the power of the property that for all vectors $\vec{w} \in \mathbb{R}^n$:

$$\|\vec{w}\|^2 = \vec{w} \circ \vec{w}.$$

To find $\|5\vec{u} + 3\vec{v}\|$, we apply the formula above to $\vec{w} = 5\vec{u} + 3\vec{v}$, along with the distributive and homogeneous properties of the dot product. Thus:

$$\begin{aligned}
\|5\vec{u} + 3\vec{v}\|^2 &= (5\vec{u} + 3\vec{v}) \circ (5\vec{u} + 3\vec{v}) \\
&= (5\vec{u} + 3\vec{v}) \circ 5\vec{u} + (5\vec{u} + 3\vec{v}) \circ 3\vec{v} \\
&= (5\vec{u}) \circ (5\vec{u}) + (3\vec{v}) \circ (5\vec{u}) + (5\vec{u}) \circ (3\vec{v}) + (3\vec{v}) \circ (3\vec{v}) \\
&= 25(\vec{u} \circ \vec{u}) + 15(\vec{u} \circ \vec{v}) + 15(\vec{v} \circ \vec{u}) + 9(\vec{v} \circ \vec{v}) \\
&= 25\|\vec{u}\|^2 + 30(\vec{u} \circ \vec{v}) + 9\|\vec{v}\|^2,
\end{aligned}$$

where in the last step, we again used the property above, to change $\vec{u} \circ \vec{u}$ to $\|\vec{u}\|^2$ and $\vec{v} \circ \vec{v}$ to $\|\vec{v}\|^2$, as well as the ***commutative property*** of the dot product. Also, notice the similarity between the computation above and the FOIL expansion in basic algebra.

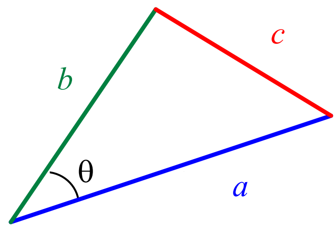
Now, since we know the values of all of these quantities, we can now compute:

$$\|5\vec{u}+3\vec{v}\|^2 = 25\|\vec{u}\|^2+30\vec{u}\circ\vec{v}+9\|\vec{v}\|^2$$
$$= 25(5)^2+30(24)+9(29)^2 = 8914.$$

Thus, $\|5\vec{u}+3\vec{v}\| = \sqrt{8914}$. $\square$

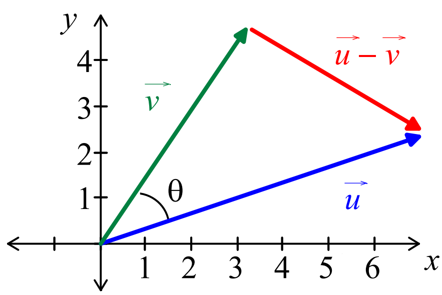
A Geometric Formulation for the Dot Product

The Law of Cosines from Trigonometry says that if a triangle has sides a, b, c and the angle opposite side c is called θ, then:



$$c^2 = a^2+b^2-2ab\cos(\theta)$$

Now, suppose $\vec{u}$ and $\vec{v}$ are two ***non-zero*** vectors of $\mathbb{R}^2$. The vectors $\vec{v}$, $\vec{u}-\vec{v}$ and $\vec{u}$ form a ***triangle*** in $\mathbb{R}^2$:



The Triangle Formed by $\vec{v}$, $\vec{u}-\vec{v}$ and $\vec{u}$

We can easily check that this diagram is correctly labeled, because $\vec{v}+(\vec{u}-\vec{v}) = \vec{u}$. Now, if we let θ be the angle between $\vec{u}$ and $\vec{v}$, as shown in the diagram, then by the Law of Cosines:

$$\|\vec{u}-\vec{v}\|^2 = \|\vec{u}\|^2+\|\vec{v}\|^2-2\|\vec{u}\|\,\|\vec{v}\|\cos(\theta).$$

But recall that:

$$\begin{aligned}\|\vec{u}-\vec{v}\|^2 &= (\vec{u}-\vec{v})\circ(\vec{u}-\vec{v})\\ &= (\vec{u}-\vec{v})\circ\vec{u}-(\vec{u}-\vec{v})\circ\vec{v} && \text{(by the Right Distributive Property)}\\ &= \vec{u}\circ\vec{u}-\vec{v}\circ\vec{u}-\vec{u}\circ\vec{v}+\vec{v}\circ\vec{v} && \text{(by the Left Distributive Property)}\\ &= \|\vec{u}\|^2-2\vec{u}\circ\vec{v}+\|\vec{v}\|^2 && \text{(by the Commutative Property).}\end{aligned}$$

Thus we get:

$$\|\vec{u}\|^2-2\vec{u}\circ\vec{v}+\|\vec{v}\|^2 = \|\vec{u}\|^2+\|\vec{v}\|^2-2\|\vec{u}\|\,\|\vec{v}\|\cos(\theta).$$

Cancelling common terms and dividing both sides by -2, we get:

Definition/Theorem: If $\vec{u}$ and $\vec{v}$ are ***non-zero*** vectors in $\mathbb{R}^2$, then:

$$\vec{u} \circ \vec{v} = \|\vec{u}\| \, \|\vec{v}\| \cos(\theta),$$

where θ is the angle formed by the vectors $\vec{u}$ and $\vec{v}$ in standard position. Thus, we can ***compute*** the angle θ between $\vec{u}$ and $\vec{v}$ by:

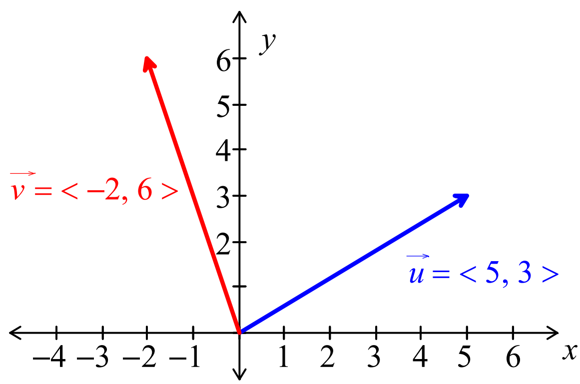
$$\cos(\theta) = \frac{\vec{u} \circ \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|},$$

where $0 \leq \theta \leq \pi$. We will use the exact same formula for two non-zero vectors in $\mathbb{R}^3$.

Example: Let us consider the vectors $\vec{u} = \langle 5, 3 \rangle$ and $\vec{v} = \langle -2, 6 \rangle$. According to our formula, the cosine of the angle θ between them is:

$$\cos(\theta) = \frac{\vec{u} \circ \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|} = \frac{5(-2) + 3(6)}{\sqrt{5^2 + 3^2}\sqrt{(-2)^2 + 6^2}} = \frac{-10 + 18}{\sqrt{34}\sqrt{40}} = \frac{8}{4\sqrt{85}} \approx 0.21693$$

Thus, by using a scientific calculator, we find that $\theta \approx 1.3521$ radians, or about 77.47^0. Let us draw the two vectors together in standard position:



Finding the Angle Between Two Vectors

We can check with a protractor that this answer is reasonable.□

Notice in particular that if $\vec{u} \circ \vec{v} = 0$, then $\cos(\theta) = 0$, hence $\theta = \pi/2$. In other words $\vec{u}$ and $\vec{v}$ are ***perpendicular*** to each other. Since $\vec{\mathbf{0}}_2 \circ \vec{v} = 0$ for all $\vec{v} \in \mathbb{R}^2$, from the definition of the dot product, we will also ***agree*** that $\vec{\mathbf{0}}_2$ is orthogonal to ***all*** vectors in $\mathbb{R}^2$, and similarly for $\vec{\mathbf{0}}_3$ in $\mathbb{R}^3$. We summarize all this with the following:

Definition/Theorem: Two vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^2$ or $\mathbb{R}^3$ are ***perpendicular*** or ***orthogonal*** to each other ***if and only if*** $\vec{u} \circ \vec{v} = 0$.

Example: The vectors $\vec{u} = \langle 5, -3, 1 \rangle$ and $\vec{v} = \langle 4, 9, 7 \rangle$ are orthogonal, since:

$$\vec{u} \circ \vec{v} = 5(4) + (-3)(9) + 1(7) = 20 - 27 + 7 = 0.$$

Sketching these vectors to verify that they are perpendicular would be a futile task, though, because the vectors will not appear to be perpendicular, thanks to the distorted perspective of Cartesian space.□

Revisiting The Cartesian Equation of a Plane

In the previous Section, we saw that:

$$Span(\{\langle 3,4,-1\rangle,\langle 5,-2,3\rangle\}) = \{\langle x,y,z\rangle \in \mathbb{R}^3 \,|\, 5x-7y-13z=0\},$$

a plane Π passing through the origin. If we collect the coefficients of the variables in the equation into a vector, we get $\vec{n} = \langle 5,-7,-13\rangle$, which is called a ***normal vector*** to the plane. It is not unique, but $L = Span(\{\langle 5,-7,-13\rangle\})$, which we call the ***normal line*** to the plane, ***is*** unique. Notice that if we take the two vectors defining our Span and get their dot products with $\vec{n}$, we get:

$$\langle 3,4,-1\rangle \circ \langle 5,-7,-13\rangle = 3(5)+4(-7)-(-13) = 0, \text{ and}$$
$$\langle 5,-2,3\rangle \circ \langle 5,-7,-13\rangle = 5(5)-2(-7)+3(-13) = 0.$$

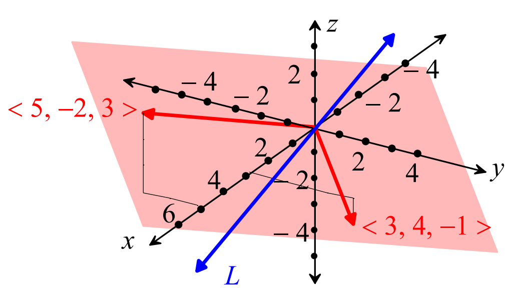
Thus $\vec{n}$ is ***orthogonal*** to ***both*** vectors. However, it also follows from the equation of Π that ***any*** vector $\langle x,y,z\rangle$ which is on the plane is ***orthogonal*** to $\vec{n}$, since:

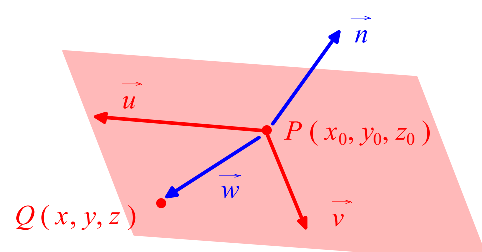
$$0 = 5x-7y-13z = \langle 5,-7,-13\rangle \circ \langle x,y,z\rangle.$$

This argument generalizes to any plane Π passing through the origin:

> ***Definition/Theorem:*** Suppose that $\vec{u},\ \vec{v} \in \mathbb{R}^3$ are non-parallel vectors, and $\Pi = Span(\{\vec{u},\vec{v}\})$ is the plane passing through the origin with Cartesian equation:
>
> $$ax+by+cz = 0.$$
>
> Then: $\vec{n} = \langle a,b,c\rangle$ is a ***normal vector*** to Π, which means that any vector $\langle x,y,z\rangle$ on Π is orthogonal to $\vec{n}$. Although $\vec{n}$ is not unique, the line $L = Span(\{\vec{n}\})$ is unique, and we call L the ***normal line*** to Π.

On the left, below, we show the plane Π and its normal line L from our Example:





$5x-7y-13z = 0$ and $Span(\{\langle 5,-7,-13\rangle\})$ An Arbitrary Plane in Cartesian Space

More generally, suppose that Π is the ***translate*** of $Span(\{\vec{u},\vec{v}\})$ by $\vec{x}_p = \langle x_0,y_0,z_0\rangle$. This is illustrated above, on the right. In this case, Π may no longer pass through the origin. As we saw in the previous Section, Π will have parametric equations:

$$\langle x,y,z\rangle = \langle x_0,y_0,z_0\rangle + t\vec{u} + s\vec{v}.$$

We can rewrite this equation as:

$$\langle x - x_0, y - y_0, z - z_0 \rangle = t\vec{u} + s\vec{v}.$$

In other words, $\langle x - x_0, y - y_0, z - z_0 \rangle$ is a vector in the plane $Span(\{\vec{u}, \vec{v}\})$ which ***does*** pass through the origin. Thus, by our observation above, $\langle x - x_0, y - y_0, z - z_0 \rangle$ must be ***orthogonal*** to any normal vector $\vec{n} = \langle a, b, c \rangle$ for $Span(\{\vec{u}, \vec{v}\})$. Thus:

$$\langle a, b, c \rangle \circ \langle x - x_0, y - y_0, z - z_0 \rangle = 0, \text{ or}$$

$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0, \text{ and expanding:}$$

$$ax + by + cz = ax_0 + by_0 + cz_0.$$

But the right side is a ***constant***, since we ***fixed*** the point (x_0, y_0, z_0) corresponding to $\vec{x}_p$. Thus we obtain, as in the previous Section, the Cartesian equation of any plane in $\mathbb{R}^3$, and we summarize this discussion in the following:

Definition/Theorem: Suppose that $\vec{u}, \vec{v} \in \mathbb{R}^3$ are ***non-parallel*** vectors, and Π is the plane $\vec{x}_p + Span(\{\vec{u}, \vec{v}\})$, where $\vec{x}_p = \langle x_0, y_0, z_0 \rangle$ and (x_0, y_0, z_0) is an arbitrary point in $\mathbb{R}^3$. Then Π has a ***Cartesian equation:***

$$ax + by + cz = d = ax_0 + by_0 + cz_0,$$

where $\vec{n} = \langle a, b, c \rangle$ is a ***normal vector*** to $Span(\{\vec{u}, \vec{v}\})$. Thus, $\vec{n}$ is also normal to ***all*** the translates of $Span(\{\vec{u}, \vec{v}\})$. The plane Π passes through the origin ***if and only if*** $d = 0$.

We will still refer to $L = Span(\{\vec{n}\})$ as the ***normal line*** to Π. Thus, ***all*** the translates of $Span(\{\vec{u}, \vec{v}\})$ have the same normal line L.

Consequently, two ***distinct*** planes Π_1 and Π_2 are ***parallel*** to each other ***if and only if*** they have the same normal line L. In other words, any normal vector to Π_1 is parallel to any normal vector to Π_2.

In the Exercises, we will see a more efficient way to find the Cartesian equation of a plane, by using the ***cross product*** of the vectors $\vec{u}$ and $\vec{v}$. This is a technique seen in any Multivariable Calculus course.

The Cauchy-Schwarz Inequality

Recall that we defined the angle θ between two non-zero vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ using the formula:

$$\cos(\theta) = \frac{\vec{u} \circ \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|}.$$

But since $\cos(\theta)$ is between -1 and 1, we have:

$$-1 \leq \frac{\vec{u} \circ \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|} \leq 1 \qquad \Leftrightarrow$$

$$-\|\vec{u}\| \, \|\vec{v}\| \leq \vec{u} \circ \vec{v} \leq \|\vec{u}\| \, \|\vec{v}\| \qquad \Leftrightarrow$$

$$|\vec{u} \circ \vec{v}| \leq \|\vec{u}\| \, \|\vec{v}\|.$$

It turns out that this final inequality is true for all pairs of vectors in ***any*** $\mathbb{R}^n$, and it is named after two famous mathematicians:

Theorem — The Cauchy-Schwarz Inequality:
For any two vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$: $|\vec{u} \circ \vec{v}| \leq \|\vec{u}\| \|\vec{v}\|$.

Proof: We will separate the proof into two cases:

Case 1: Suppose $\vec{u} = \vec{\mathbf{0}}_n$ or $\vec{v} = \vec{\mathbf{0}}_n$. Then both sides are 0, so the inequality is true.

Case 2: Suppose now that $\vec{u} \neq \vec{\mathbf{0}}_n$ and $\vec{v} \neq \vec{\mathbf{0}}_n$. Thus, by our Positivity Property:

$$\vec{u} \circ \vec{u} = \|\vec{u}\|^2 > 0 \quad \text{and} \quad \vec{v} \circ \vec{v} = \|\vec{v}\|^2 > 0.$$

Let us construct the linear combination:

$$\vec{w} = r\vec{u} + s\vec{v},$$

where r and s are any two scalars, possibly 0. Since $\vec{w}$ ***could*** be $\vec{\mathbf{0}}_n$, the best that we can say is that:

$$\begin{aligned} 0 \leq \|\vec{w}\|^2 &= \vec{w} \circ \vec{w} \\ &= (r\vec{u} + s\vec{v}) \circ (r\vec{u} + s\vec{v}) \\ &= (r\vec{u}) \circ (r\vec{u}) + (r\vec{u}) \circ (s\vec{v}) + (s\vec{v}) \circ (r\vec{u}) + (s\vec{v}) \circ (s\vec{v}) \\ &= r^2(\vec{u} \circ \vec{u}) + 2rs(\vec{u} \circ \vec{v}) + s^2(\vec{v} \circ \vec{v}). \end{aligned}$$

Since this is true for ***any*** s and t, let us first substitute $s = \vec{u} \circ \vec{u}$. Then we get:

$$0 \leq r^2(\vec{u} \circ \vec{u}) + 2r(\vec{u} \circ \vec{u})(\vec{u} \circ \vec{v}) + (\vec{u} \circ \vec{u})^2(\vec{v} \circ \vec{v}),$$

and since $\vec{u} \circ \vec{u} = \|\vec{u}\|^2$ is ***positive***, we can divide it out to obtain the equivalent inequality:

$$0 \leq r^2 + 2r(\vec{u} \circ \vec{v}) + (\vec{u} \circ \vec{u})(\vec{v} \circ \vec{v}).$$

Finally, we substitute $r = -(\vec{u} \circ \vec{v})$, and we get:

$$0 \leq (\vec{u} \circ \vec{v})^2 - 2(\vec{u} \circ \vec{v})(\vec{u} \circ \vec{v}) + (\vec{u} \circ \vec{u})(\vec{v} \circ \vec{v}),$$

which simplifies to:

$$(\vec{u} \circ \vec{v})^2 \leq (\vec{u} \circ \vec{u})(\vec{v} \circ \vec{v}) = \|\vec{u}\|^2 \|\vec{v}\|^2.$$

Since both side are non-negative, by taking square roots, we equivalently obtain:

$$|\vec{u} \circ \vec{v}| \leq \|\vec{u}\| \|\vec{v}\|. \blacksquare$$

Angles and Orthogonality

Thanks to the Cauchy-Schwarz Inequality, we can define the angle between any two vectors in $\mathbb{R}^n$:

Definition: If $\vec{u}$, $\vec{v} \in \mathbb{R}^n$ are ***non-zero*** vectors, we define the ***angle*** θ between $\vec{u}$ and $\vec{v}$ by:

$$\cos(\theta) = \frac{\vec{u} \circ \vec{v}}{\|\vec{u}\| \|\vec{v}\|},$$

where $0 \leq \theta \leq \pi$. Furthermore, we will say that $\vec{u}$ and $\vec{v}$ are ***orthogonal*** to each other ***if and only if*** $\vec{u} \circ \vec{v} = 0$.

Consequently, the zero vector $\vec{\mathbf{0}}_n$ is orthogonal to ***all*** vectors in $\mathbb{R}^n$.

The definition for $\cos(\theta)$ makes sense because the Cauchy-Schwarz Inequality assures us that this quotient is between -1 and 1. Our convention for the zero vector is exactly the same as what we had for $\mathbb{R}^2$ and $\mathbb{R}^3$.

Example: Let us find an approximation for the angle θ between $\vec{u} = \langle 3, -7, 4, 2\rangle$ and $\vec{v} = \langle 5, 2, -3, 9\rangle$, even though it is impossible to visualize vectors in $\mathbb{R}^4$:

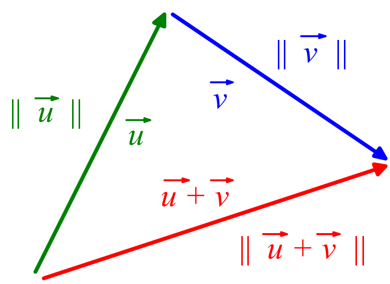
$$\begin{aligned}\cos(\theta) &= \frac{\vec{u} \circ \vec{v}}{\|\vec{u}\|\,\|\vec{v}\|} = \frac{3(5) - 7(2) + 4(-3) + 2(9)}{\sqrt{9 + 49 + 16 + 4}\,\sqrt{25 + 4 + 9 + 81}} \\ &= \frac{7}{\sqrt{78}\,\sqrt{119}} = \frac{7}{\sqrt{9282}} \approx 0.072657,\end{aligned}$$

and thus $\theta \approx \cos^{-1}(0.072657) \approx 1.4981$ radians.□

The following consequence of the Cauchy-Schwarz Inequality will be left as an Exercise:

Theorem — The Triangle Inequality:
For any two vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$: $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$.

Its name comes from the fact that $\vec{u}$, $\vec{v}$ and $\vec{u} + \vec{v}$ form the sides of a triangle, and this Theorem says that the sum of the lengths of two sides is ***at least*** the length of the third.



The Triangle Inequality: $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$

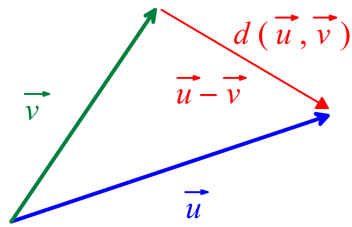
We can say this in everyday language as "the shortest path between two points is along a straight line." Notice that we achieve ***equality*** if $\vec{u}$ and $\vec{v}$ are parallel and in the ***same direction***, in which case the three vectors are ***colinear***. In other words, we have a "degenerate" triangle.

Distance Between Vectors

The distance formula that we see in basic algebra can be generalized using the following:

Definition: If $\vec{u} = \langle u_1, u_2, \ldots, u_n\rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n\rangle$ are vectors from $\mathbb{R}^n$, we define the ***distance between*** $\vec{u}$ and $\vec{v}$ as:

$$d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\| = \sqrt{(u_1 - v_1)^2 + (u_2 - v_2)^2 + \cdots + (u_n - v_n)^2}\,.$$



The Distance Between Two Vectors $\vec{u}$ and $\vec{v}$

This definition makes sense because $\vec{u}-\vec{v}$ connects the head of $\vec{u}$ to the head of $\vec{v}$ when their tails are at the same point, so its length measures the ***separation*** between the ***heads*** of the two vectors.

Example: Let $\vec{u} = \langle 7, 5, -2\rangle$ and $\vec{v} = \langle -2, 3, 4\rangle$. Then:

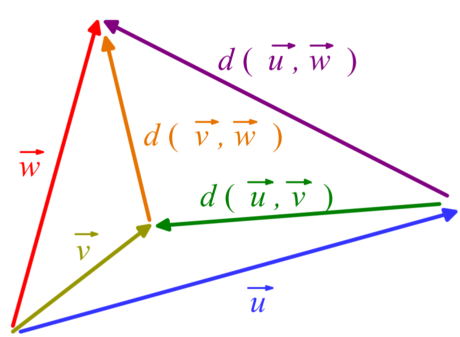
$$\vec{u}-\vec{v} = \langle 7, 5, -2\rangle - \langle -2, 3, 4\rangle = \langle 9, 2, -6\rangle, \text{ and so}$$
$$d(\vec{u},\vec{v}) = \sqrt{81+4+36} = \sqrt{121} = 11. \square$$

The distance function enjoys some properties that are ***inherited*** from the dot product. You will prove them in the Exercises.

Theorem — Properties of Distances:
Let $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ and $k \in \mathbb{R}$. Then, we have the following properties:

1. ***The Symmetric Property for Distances*** $\quad d(\vec{u},\vec{v}) = d(\vec{v},\vec{u})$.
2. ***The Homogeneity Property for Distances*** $\quad d(k\vec{u},k\vec{v}) = |k|\cdot d(\vec{u},\vec{v})$.
3. ***The Triangle Inequality for Distances*** $\quad d(\vec{u},\vec{w}) \le d(\vec{u},\vec{v}) + d(\vec{v},\vec{w})$.

Since the distance between two vectors is the length of the vector connecting one head to the other when the vectors are in standard position (with their tails at the same point), we can visualize the last property with the following diagram:



The Triangle Inequality for Distances: $d(\vec{u},\vec{w}) \le d(\vec{u},\vec{v}) + d(\vec{v},\vec{w})$

1.3 Section Summary

If $\vec{u} = \langle u_1, u_2, \dots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \dots, v_n \rangle$ are from $\mathbb{R}^n$, their ***dot product*** is:

$$\vec{u} \circ \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \|\vec{u}\| \, \|\vec{v}\| \cos(\theta),$$

where the ***length*** of a vector is defined by: $\|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = \sqrt{\vec{u} \circ \vec{u}}$, and where $\theta \in [0, \pi]$ is the ***angle*** between the two vectors.

Properties of the Dot Product: for any vectors $\vec{u}$, $\vec{v}$, $\vec{w} \in \mathbb{R}^n$ and scalar $k \in \mathbb{R}$, we have:

1. ***The Commutative Property*** $\quad \vec{u} \circ \vec{v} = \vec{v} \circ \vec{u}.$
2. ***The Right Distributive Property*** $\quad \vec{u} \circ (\vec{v} + \vec{w}) = \vec{u} \circ \vec{v} + \vec{u} \circ \vec{w}.$
3. ***The Left Distributive Property*** $\quad (\vec{u} + \vec{v}) \circ \vec{w} = \vec{u} \circ \vec{w} + \vec{v} \circ \vec{w}.$
4. ***The Homogeneity Property*** $\quad (k \cdot \vec{u}) \circ \vec{v} = k(\vec{u} \circ \vec{v}) = \vec{u} \circ (k \cdot \vec{v}).$
5. ***The Zero-Vector Property*** $\quad \vec{u} \circ \vec{\mathbf{0}}_n = 0.$
6. ***The Positivity Property*** $\quad$ If $\vec{u} \neq \vec{\mathbf{0}}_n$, then $\vec{u} \circ \vec{u} > 0.$
7. ***The Non-Degeneracy Property*** $\quad \vec{u} \circ \vec{u} > 0$ ***if and only if*** $\vec{u} \neq \vec{\mathbf{0}}_n$, and $\vec{\mathbf{0}}_n \circ \vec{\mathbf{0}}_n = 0.$

For any vector $\vec{v} \in \mathbb{R}^n$ and $k \in \mathbb{R}$: $\|k\vec{v}\| = |k| \|\vec{v}\|$.
In particular, if $\vec{v} \neq \vec{\mathbf{0}}_n$, then $\vec{u}_1 = \dfrac{1}{\|\vec{v}\|} \vec{v}$ is the unit vector in the same direction as $\vec{v}$,

and $\vec{u}_2 = -\dfrac{1}{\|\vec{v}\|} \vec{v}$ is the unit vector in the opposite direction as $\vec{v}$.

Furthermore, $\|\vec{v}\| = 0$ ***if and only if*** $\vec{v} = \vec{\mathbf{0}}_n$.

The Cauchy-Schwarz Inequality says that for any two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$: $|\vec{u} \circ \vec{v}| \leq \|\vec{u}\| \, \|\vec{v}\|$.

If $\vec{u}$ and $\vec{v}$ are ***non-zero*** vectors in $\mathbb{R}^n$, we define the ***angle*** θ between them via:
$\cos(\theta) = \dfrac{\vec{u} \circ \vec{v}}{\|\vec{u}\| \, \|\vec{v}\|}$, where $\theta \in [0, \pi]$.

Two vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$ are ***perpendicular*** or ***orthogonal*** to each other ***if and only if*** $\vec{u} \circ \vec{v} = 0$.
Consequently, $\vec{\mathbf{0}}_n$ is orthogonal to ***all*** vectors in $\mathbb{R}^n$.

A ***Cartesian equation*** for any plane Π in $\mathbb{R}^3$ has the form $ax + by + cz = d$, where $\vec{n} = \langle a, b, c \rangle$ is a non-zero ***normal vector*** to Π.

The plane Π passes through the origin ***if and only if*** $d = 0$.

We call $L = Span(\{\vec{n}\})$ the ***normal line*** to Π.

The Triangle Inequality says that for any two vectors $\vec{u}$ and $\vec{v}$ in $\mathbb{R}^n$: $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$.

If $\vec{u}$ and $\vec{v}$ are vectors in $\mathbb{R}^n$, we define the ***distance*** between them as: $d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|$.

Properties of Distances: Let $\vec{u}$, $\vec{v}$, $\vec{w} \in \mathbb{R}^n$ and $k \in \mathbb{R}$. Then, we have the following properties:

1. ***The Symmetric Property for Distances*** $\quad d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u}).$
2. ***The Homogeneity Property for Distances*** $\quad d(k\vec{u}, k\vec{v}) = |k| \cdot d(\vec{u}, \vec{v}).$
3. ***The Triangle Inequality for Distances*** $\quad d(\vec{u}, \vec{w}) \leq d(\vec{u}, \vec{v}) + d(\vec{v}, \vec{w}).$

1.3 Exercises

1. Find the length of $\vec{u} = \langle 3, -5, 2, 9 \rangle$.
2. Find the exact cosine of the angle between $\langle 5, 2 \rangle$ and $\langle -3, 10 \rangle$ and approximate this angle. Graph both vectors in standard position and verify that your answer is reasonable with a protractor.
3. Let $\vec{u} = \langle -5, 3 \rangle$ and $\vec{v} = \langle 4, -7 \rangle$. Compare $\|2\vec{u} + 5\vec{v}\|$ to $\|2\vec{u}\| + \|5\vec{v}\|$.
4. Find the approximate angle between the two vectors $\langle 7, -5, 3 \rangle$ and $\langle 8, 2, -3 \rangle$.
5. Find the approximate angle between the two vectors $\langle -5, 2, -4, 1 \rangle$ and $\langle 2, -7, 3, 6 \rangle$.
6. Find the angle made by the main diagonal of a cube (i.e. the longest line segment contained in the cube) with any one of the six edges it is adjacent to. Express your answer in degrees, correct to 4 decimal places.
7. A rectangular box has dimensions $2 \times 3 \times 5$ inches. Find the angle made by its main diagonal with each of the three kinds of edges it is adjacent to. Express your answer in degrees, correct to 4 decimal places.

For Exercises (8) to (12): Assume that $\vec{u},\ \vec{v} \in \mathbb{R}^n$, not necessarily $\mathbb{R}^2$ or $\mathbb{R}^3$. The formula $\|\vec{w}\|^2 = \vec{w} \circ \vec{w}$ will be useful in all of these problems. Find the indicated quantities. Do not attempt to find $\vec{u}$ and $\vec{v}$ individually.

8. If $\|\vec{u}\| = 5$ and $\|\vec{v}\| = 7$, find $(3\vec{u} - 8\vec{v}) \circ (3\vec{u} + 8\vec{v})$.
9. If $\|\vec{u}\| = 5$ and $\|\vec{v}\| = 7$ and $\vec{u} \circ \vec{v} = -20$, find $\|4\vec{u} + 11\vec{v}\|$.
10. If $\|\vec{u}\| = 13$, $\|\vec{v}\| = 10$, and $\vec{u} \circ \vec{v} = 32$, find $\|7\vec{u} - 3\vec{v}\|$.
11. If $\|2\vec{u} - 7\vec{v}\| = \sqrt{40{,}637}$, and $\|2\vec{u} + 7\vec{v}\| = \sqrt{41{,}981}$, find $\vec{u} \circ \vec{v}$.
12. If $\|2\vec{u} - 7\vec{v}\| = \sqrt{21{,}305}$, $\|6\vec{u} + 5\vec{v}\| = \sqrt{13{,}801}$, and $\vec{u} \circ \vec{v} = -345$, find $\|\vec{u}\|$, $\|\vec{v}\|$ and $\|4\vec{u} + 9\vec{v}\|$.
13. ***Parallel Planes:*** Consider the plane Π_1 with Cartesian equation $6x - 5y + 2z = -3$.
    a. Show that the point $P(4, 7, -2)$ is not on Π_1.
    b. Find a Cartesian equation for the plane Π_2 passing through P that is parallel to Π_1.
14. Find a Cartesian equation for the plane that contains $(3, -4, -6)$ and is ***parallel*** to the plane $2x + 5y - 9z = 8$.
15. ***The Cross Product:*** Suppose that $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle \in \mathbb{R}^3$.
    Let us define the vector:
    $$\vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2)\vec{i} - (u_1 v_3 - u_3 v_1)\vec{j} + (u_1 v_2 - u_2 v_1)\vec{k},$$
    called the ***cross product*** of $\vec{u}$ and $\vec{v}$.
    Prove that for any $\vec{u},\ \vec{v} \in \mathbb{R}^3$: $\vec{u} \times \vec{v}$ is ***orthogonal*** to both $\vec{u}$ and $\vec{v}$.
16. ***Intersecting Lines:*** Consider the two lines given by:
    $$\langle x, y, z \rangle = \langle 3 - 5t, 7 + 2t, 4 - t \rangle, \text{ and } \langle x, y, z \rangle = \langle 7 + 3s, 5 - s, 10 - 2s \rangle.$$
    a. Show that the two lines intersect at exactly one point, and find the coordinates of that point.
    b. Find a Cartesian equation for the plane that contains the two lines. Hint: use the cross-product on the two direction vectors.
17. ***Orthogonal Lines:*** We will say that two lines L_1 and L_2 are ***orthogonal*** to each other if they intersect ***and*** any direction vector for L_1 is orthogonal to any direction vector for L_2.

Show that the lines $\langle x, y, z\rangle = \langle 3-t, 5t, -7+4t\rangle$, and $\langle x, y, z\rangle = \langle 8+3s, 3-s, 1+2s\rangle$ are orthogonal to each other, and find a Cartesian equation for the plane containing both lines. Where do the lines intersect?

18. Find parametric equations for the line L_1 that is orthogonal to the line $L_2 : \langle x, y, z\rangle = \langle 2+3t, 5-2t, 9+t\rangle$ and passes through the point $P(5, -3, 7)$. Hint: first solve for their point of intersection by connecting P to a general point on L_2 and forcing orthogonality.

19. ***Revisiting Planes Containing Parallel Lines:*** Consider the two lines:

$$L_1 : \langle x, y, z\rangle = \langle 5+2t, 3-t, 3t\rangle, \quad \text{and} \quad L_2 : \langle x, y, z\rangle = \langle 7-4s, 5+2s, 8-6s\rangle.$$

   a. Write down the definition of parallel lines from Section 1.2.
   b. Show that L_1 and L_2 are parallel to each other.
   c. Find a Cartesian equation for the plane that contains both lines. Hint: to find a normal vector for the plane, form the vector from a point P on L_1 to a point Q on L_2 and take the cross product of this vector with either direction vector.

20. ***Skew Lines:*** Two lines in $\mathbb{R}^3$ are called ***skew*** if they do ***not*** intersect and they are ***not*** parallel.
   a. Show that the two lines given by the vector equations below are skew:

$$L_1 : \langle x, y, z\rangle = \langle 3-5t, 7+2t, 4-t\rangle, \quad \text{and} \quad L_2 : \langle x, y, z\rangle = \langle 7-s, 5+3s, 2s\rangle.$$

   b. Use the cross product to find a vector $\vec{n}$ that is orthogonal to the directions of L_1 and L_2.
   c. Use $\vec{n}$ to construct parallel planes Π_1 and Π_2, where Π_1 contains L_1, and Π_2 contains L_2.

21. ***Orthogonal Planes:*** We will define two planes Π_1 and Π_2 in $\mathbb{R}^3$ to be ***orthogonal*** to each other if any normal vector to Π_1 is orthogonal to any normal vector to Π_2.

   Show that the two planes $\Pi_1 : 3x - 5y + 2z = 6$, and $\Pi_2 : 2x + 4y + 7z = -3$ are orthogonal to each other. What is the equation of their line of intersection?

22. Find a Cartesian equation for a plane Π_2 that ***contains*** the line $\langle x, y, z\rangle = \langle 2+t, 5-3t, -7+t\rangle$ and is also ***orthogonal*** to the plane $\Pi_1 : 3x - 5y + 7z = 4$.

23. Suppose that Π_1 is the plane $3x - 5y + 2z = 6$.
   a. Show that the points $(3, 1, 1)$ and $(5, 1, -2)$ are both on Π_1.
   b. Find a Cartesian equation for a plane Π_2 that contains the two points in part (a), such that Π_2 is orthogonal to Π_1. Hint: use the cross product.
   c. Find a vector equation for the line L containing the two points in part (a) and check that every point on L is on both planes.

24. ***Orthogonal Lines and Plane Pairs:*** We will say that a line L in $\mathbb{R}^3$ is ***orthogonal*** to a plane Π if any direction vector for L is ***parallel*** to any normal vector for Π.

   Show that the line $\langle x, y, z\rangle = \langle 3+2t, 5-6t, 8t\rangle$ is orthogonal to the plane $x - 3y + 4z = 7$.

25. Find a Cartesian equation for the plane Π that contains the point $(5, -2, 7)$ and is ***orthogonal*** to the line L with parametric equations: $\langle x, y, z\rangle = \langle -2, 1, 7\rangle + t\langle 8, 5, -4\rangle$.

   At what point does the plane Π intersect the line L?

26. Find parametric equations for the line that contains the point $(5, -2, 1)$ and is ***orthogonal*** to the plane $3x + 7y - 4z = 6$. Where does this line intersect the plane?

27. ***Parallel Lines and Planes:*** We will say that a line L in $\mathbb{R}^3$ is ***parallel*** to a plane Π if the line and the plane do not intersect.

a. Show that the line L with parametric equations $\langle x, y, z\rangle = \langle -2, 1, 7\rangle + t\langle 8, 5, -4\rangle$ is ***parallel*** to the plane $\Pi_1 : 2x - 4y - z = 3$.

b. Find a Cartesian equation for the plane Π_2 that contains L and is ***orthogonal*** to Π_1.

28. ***True or False:*** $\vec{u} \circ \vec{v} = 0$ ***if and only if*** either $\vec{u} = \vec{\mathbf{0}}_n$ or $\vec{v} = \vec{\mathbf{0}}_n$. Explain your answer.

For Exercises (29) to (37): prove the properties. Assume that $\vec{u}$, $\vec{v}$ and $\vec{w} \in \mathbb{R}^n$ and $k \in \mathbb{R}$.

29. ***The Distributive Properties for Dot Products:***

$$\vec{u} \circ (\vec{v} + \vec{w}) = \vec{u} \circ \vec{v} + \vec{u} \circ \vec{w} \text{ and } (\vec{u} + \vec{v}) \circ \vec{w} = \vec{u} \circ \vec{w} + \vec{v} \circ \vec{w}.$$

30. ***The Homogeneity Property for Dot Products:*** $(k \cdot \vec{u}) \circ \vec{v} = k(\vec{u} \circ \vec{v}) = \vec{u} \circ (k \cdot \vec{v})$.

31. ***The Property of the Zero Vector:*** $\vec{u} \circ \vec{\mathbf{0}}_n = 0$.

32. ***The Positivity Property:*** If $\vec{u} \neq \vec{\mathbf{0}}_n$, then $\vec{u} \circ \vec{u} > 0$.

33. ***The Triangle Inequality:*** $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$.
Hint: Use $0 \leq \|\vec{u} + \vec{v}\|^2 = (\vec{u} + \vec{v}) \circ (\vec{u} + \vec{v})$.

34. ***The Symmetric Property for Distances:*** $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$.

35. ***The Homogeneity Property for Distances:*** $d(k \cdot \vec{u}, k \cdot \vec{v}) = |k| \cdot d(\vec{u}, \vec{v})$.

36. ***The Triangle Inequality for Distances:*** $d(\vec{u}, \vec{w}) \leq d(\vec{u}, \vec{v}) + d(\vec{v}, \vec{w})$.
Hint: Use the Triangle Inequality for vectors and rewrite $\vec{u} - \vec{w}$ as $\vec{u} - \vec{v} + \vec{v} - \vec{w}$.

37. Prove that $\vec{\mathbf{0}}_n$ is the ***only*** vector in $\mathbb{R}^n$ that is orthogonal to ***itself***. In other words, if $\vec{v} \in \mathbb{R}^n$ and $\vec{v}$ is orthogonal to $\vec{v}$, then $\vec{v} = \vec{\mathbf{0}}_n$.

38. ***The Parallelogram Law*** states that the sum of the squares of the ***two diagonals*** of a parallelogram is equal to the sum of the squares of the ***four sides***.

 a. On your paper, copy the Parallelogram Principle found on page 29. Be sure you include the labels of all the vectors involved.

 b. Rewrite The Parallelogram Law in terms of the lengths of the vectors in the diagram.

 c. Prove the Law using the identity $\|\vec{w}\|^2 = \vec{w} \circ \vec{w}$.

39. We proved ***The Zero Factors Theorem*** in Section 1.1, which says that $k \cdot \vec{v} = \vec{\mathbf{0}}_n$ ***if and only if*** either $k = 0$ or $\vec{v} = \vec{\mathbf{0}}_n$. Use the dot product to prove ***directly*** that if $\vec{v} \neq \vec{\mathbf{0}}_n$ and $k \cdot \vec{v} = \vec{\mathbf{0}}_n$, then $k = 0$ (that is, ***without*** using Proof by Contradiction or Case-by-Case Analysis).

40. Prove that if $\vec{n}$ is orthogonal to all the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$, then $\vec{n}$ is orthogonal to ***all*** the vectors in $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\})$.

41. Show that in any $\mathbb{R}^n$, the vectors in the standard basis $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ are mutually orthogonal to each other. This means that $\vec{e}_i$ is orthogonal to $\vec{e}_j$ if $i \neq j$.

42. Show that if a line L is orthogonal to a plane Π_1, then any plane Π_2 which ***contains*** L is also orthogonal to Π_1. Note that we want the two ***planes*** to be orthogonal to each other.

43. Show that if a line L is ***parallel*** to a plane Π, then any direction vector for L must be ***orthogonal*** to any normal vector for Π.

44. Suppose that $\vec{u}, \vec{v} \in \mathbb{R}^3$, with $\vec{u}$ not parallel to $\vec{v}$. In Section 1.2, we found the general equation $ax + by + cz = 0$ for the plane $\Pi = Span(\{\vec{u}, \vec{v}\})$. Show that the normal vector $\vec{n} = \langle a, b, c\rangle$ obtained using the method from Section 1.2 is ***parallel*** to $\vec{u} \times \vec{v}$ (though not necessarily equal to it). Hint: include the possibility that one or more of the components of $\vec{u}$ or $\vec{v}$ could be 0.

45. Suppose that $\vec{u}, \vec{v} \in \mathbb{R}^n$ are both ***unit*** vectors, and let θ be the angle between them. Prove that $\|\vec{u} - \vec{v}\| = 2\sin\left(\frac{\theta}{2}\right)$. Review the half angle formulas from Trigonometry.

1.4 Systems of Linear Equations

We saw in Section 1.2 that the Span of a single non-zero vector $\vec{v}$ is a line in $\mathbb{R}^n$ given by parametric equations, and the Span of two non-zero, non-parallel vectors $\vec{u}$ and $\vec{v}$ is a plane in $\mathbb{R}^n$, which can be given by a Cartesian equation when the plane is in $\mathbb{R}^3$. It is not too difficult to check if an arbitrary vector $\vec{b}$ is on a line or a plane given this information.

Example: Let $\vec{b} = \langle -8, -6, 4 \rangle$. Let us determine if $\vec{b}$ is a member of the following Spans:

- L: The line $Span(\{\langle 4, 3, -2 \rangle\})$.
- Π_1: The plane with Cartesian equation $3x - 5y + 2z = 0$.
- Π_2: The plane with Cartesian equation $5x + 4y + 16z = 0$.

Since $\langle -8, -6, 4 \rangle = -2\langle 4, 3, -2 \rangle$, $\vec{b}$ is on L. For the two planes, we can substitute the coordinates of $\vec{b}$ into the Cartesian equations:

$$\begin{aligned} 3(-8) - 5(-6) + 2(4) &= -24 + 30 + 8 = 14, \text{ and} \\ 5(-8) + 4(-6) + 16(4) &= -40 - 24 + 64 = 0, \end{aligned}$$

so $\vec{b}$ is on the plane on Π_2 but not on Π_1. □

The problem becomes much more difficult when we have the Span of three or more vectors. In order to determine if a vector $\vec{b}$ is a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$, we will need to find coefficients $x_1, x_2, \ldots, x_n$ that satisfy the ***vector equation:***

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{b}.$$

In this section, we will study the key computational method of Linear Algebra, the ***Gauss-Jordan Algorithm***, that will help us solve for these coefficients. First, though, let us see an ***equivalent*** form for the vector equation above:

Example: To see if $\vec{b} = \langle 6, 17, -12 \rangle$ is a member of:

$$Span(\{\langle 4, 3, -2 \rangle, \langle -8, -6, 4 \rangle, \langle 3, -4, 3 \rangle, \langle 9, 13, -9 \rangle\}) \subset \mathbb{R}^3,$$

we need to find coefficients x_1 through x_4 such that:

$$\langle 6, 17, -12 \rangle = x_1\langle 4, 3, -2 \rangle + x_2\langle -8, -6, 4 \rangle + x_3\langle 3, -4, 3 \rangle + x_4\langle 9, 13, -9 \rangle.$$

Using the definitions of scalar multiplication and vector addition, we get:

$$\begin{aligned} \langle 6, 17, -12 \rangle &= \langle 4x_1, 3x_1, -2x_1 \rangle + \langle -8x_2, -6x_2, 4x_2 \rangle + \langle 3x_3, -4x_3, 3x_3 \rangle + \langle 9x_4, 13x_4, -9x_4 \rangle \\ &= \langle 4x_1 - 8x_2 + 3x_3 + 9x_4, 3x_1 - 6x_2 - 4x_3 + 13x_4, -2x_1 + 4x_2 + 3x_3 - 9x_4 \rangle. \end{aligned}$$

By the definition of vector equality, the coefficients x_1 through x_4 must satisfy ***all*** of the following equations:

$$\begin{array}{rrrrcr} 4x_1 & -8x_2 & +3x_3 & +9x_4 & = & 6 \\ 3x_1 & -6x_2 & -4x_3 & +13x_4 & = & 17 \\ -2x_1 & +4x_2 & +3x_3 & -9x_4 & = & -12 \end{array}$$

This is called a ***system of linear equations***. If you look at each ***column***, you will see that their coefficients correspond to the four vectors. For example, the column of coefficients for x_3 forms the vector $\langle 3, -4, 3 \rangle$, which is the third vector in our set.

Now, we want to find all vectors $\vec{x} = \langle x_1, x_2, x_3, x_4 \rangle$ that satisfy all three equations. We know from basic algebra that we can solve a system of two equations in two variables using the ***elimination method:*** we can multiply each equation by a suitable constant so that the resulting coefficients of one variable are negatives of each other, then add the equations together to eliminate this variable. For example, by multiplying the third equation above by 2 and adding it to the first equation, we obtain an equation which does not involve x_1.

Notice that if we ***align*** the coefficients of the equations, we will not need to write the variables x_1 through x_4 all the time. We can thus ***encode*** all this information in what is called an ***augmented matrix:***

$$A = \left[\begin{array}{cccc|c} 4 & -8 & 3 & 9 & 6 \\ 3 & -6 & -4 & 13 & 17 \\ -2 & 4 & 3 & -9 & -12 \end{array}\right].$$

We will say that A ***corresponds*** to our system of equations, which we also refer to as ***the system of*** A. Notice that the ***columns*** of A are precisely the vectors in our set, but $\vec{b}$ is on the rightmost column. The augmented matrix is an example of a common object we will use to work with vectors:

Definition: A ***matrix***, A, is a rectangular table of numbers organized into $\boldsymbol{m}$ ***rows*** and $\boldsymbol{n}$ ***columns***. We say that the ***dimension*** of the matrix is $\boldsymbol{m \times n}$, pronounced "***m by n***."
We denote the ***entry*** in row i column j as $(A)_{\boldsymbol{i,j}}$ or $a_{i,j}$. Thus we write:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}.$$

In particular, an $m \times 1$ matrix will be called a ***column matrix***, and a $1 \times n$ matrix will be called a ***row matrix***. An $n \times n$ matrix is called a ***square matrix***. We will treat a 1×1 matrix $[a_{11}]$ as a ***scalar*** a_{11}.
We say that $A = B$ if both matrices are the same dimension and $a_{i,j} = b_{i,j}$ for all i, j.
If $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a set of vectors from $\mathbb{R}^{\boldsymbol{m}}$, and $\vec{b} \in \mathbb{R}^{\boldsymbol{m}}$, we can form the $\boldsymbol{m} \times (\boldsymbol{n}+\mathbf{1})$ ***augmented matrix:***

$$A = \left[\vec{v}_1 \ \vec{v}_2 \ \ldots \ \vec{v}_n \mid \vec{b}\right],$$

where we assemble the vectors in S into ***columns***, and we separate the last column $\vec{b}$ with a dashed line to indicate that it represents the right side of a system of equations.
We use $x_1, x_2, \ldots, x_n$ as the standard variables ***associated*** to each column.

We will now look at the easiest kind of system, where we can describe the solutions with little effort.

The Reduced Row Echelon Form or RREF

A complete description of the solutions to a system of linear equations corresponding to a matrix in the following form can be obtained very easily:

> ***Definition:*** We will say that an $m \times (n+1)$ augmented matrix is in ***row echelon form*** if it satisfies the following conditions:
>
> 1. all the rows consisting entirely of zeroes are at the ***bottom*** of the matrix.
> 2. the first non-zero entry of any row is the number 1. This entry is called a "***leading 1***."
> 3. if the next row is non-zero, its leading 1 is to the ***right*** of the previous leading 1.
>
> Furthermore, we say that the matrix is in ***reduced row echelon form***, or ***rref***, if:
>
> 4. all the entries ***above*** a leading 1 are also zeroes.

If column j contains a leading 1, we call x_j a ***leading variable***, otherwise we call x_j a ***free variable***. We can also call column j a ***leading column***. Conditions 1 and 3 forces all entries ***below*** a leading 1 to be zeroes. Condition 4 forces all entries ***above*** the leading 1 to be zeroes as well. Thus a ***standard basis vector*** $\vec{e}_i$ is found in a leading column. Because a leading 1 is found in every non-zero row, the vectors $\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_k$ will appear in the leading columns, ***in that order***, for some $k \le m$ and $k \le n$.

Example: The following 4×7 augmented matrix A represents a system of equations in the variables x_1 through x_6:

$$A = \left[\begin{array}{cccccc|c} 1 & 5 & -7 & 0 & 3 & 0 & 6 \\ 0 & 0 & 0 & 1 & -2 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right] = \left[\begin{array}{cccccc|c} \boxed{1} & 5 & -7 & 0 & 3 & 0 & 6 \\ 0 & 0 & 0 & \boxed{1} & -2 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 & \boxed{1} & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

Let us check that the four conditions are satisfied. The only row with all zeroes is at the bottom. The first non-zero entry on each row is indeed a "1." They are in $a_{1,1}$, $a_{2,4}$, and $a_{3,6}$, as we boxed above. We write the matrix A again below, but for emphasis, we box each ***leading column:***

$$A = \left[\begin{array}{|c|cc|c|c|c||c} \mathbf{1} & 5 & -7 & 0 & 3 & 0 & 6 \\ 0 & 0 & 0 & \mathbf{1} & -2 & 0 & -9 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{1} & -4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

We see that the leading 1 in the next row is indeed to the right of that in the previous row. All the entries above and below each leading 1 are zeroes, so A is in rref. The ***leading variables*** are thus x_1, x_4 and x_6, hence the ***free variables*** are x_2, x_3 and x_5. We see that the boxed columns 1, 4, and 6 contain $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$, respectively, but $\vec{e}_4$ does ***not*** appear.

Let us write the linear system corresponding to A:

$$\begin{array}{rcrcrcrcrcrcr} x_1 & + & 5x_2 & - & 7x_3 & & & + & 3x_5 & & & = & 6 \\ & & & & & & x_4 & - & 2x_5 & & & = & -9 \\ & & & & & & & & & & x_6 & = & -4 \\ & & & & & & & & & & 0 & = & 0 \end{array}.$$

First, notice that the last equation $0 = 0$ is a ***true*** statement. However, it does not give us further information about the solutions.

The third equation is the simplest one: it says $x_6 = -4$.

The first two are more perplexing. How do we interpret them? This is where the concepts of the leading variables and the free variables come in. We will allow the ***free variables*** to be any value we want them to be. In other words, we will let:

$$x_2 = r,\ x_3 = s, \text{ and } x_5 = t, \text{ where } r,\ s,\ t \in \mathbb{R}.$$

Now that we have established that, we will ***solve*** for the ***leading variables***. Thus, our three non-zero equations are equivalent to:

$$x_1 = 6 - 5x_2 + 7x_3 - 3x_5 = 6 - 5r + 7s - 3t.$$
$$x_4 = -9 + 2x_5 = -9 + 2t, \text{ and}$$
$$x_6 = -4.$$

We can thus summarize the solutions to our system in the following ***vector form:***

$$\langle x_1, x_2, x_3, x_4, x_5, x_6 \rangle = \langle 6 - 5r + 7s - 3t,\ r,\ s, -9 + 2t,\ t, -4 \rangle,$$

where $x_2 = r$, $x_3 = s$ and $x_5 = t$ are free variables. □

With practice, you should be able to write down the final solution from the rref without going through all these steps. In general, though, a system will not be this transparent to solve. Let us now see what we can do to the augmented matrix of a system so that we can find its solutions, if there are any.

Elementary Row Operations

In the same way that we can multiply both sides of an equation by a non-zero constant, exchange two equations, and add a multiple of one equation to another, we are allowed to perform the following types of operations on any matrix:

Definition: An ***elementary row operation*** is any one of the following actions on a matrix:

Type:	Notation:
1. Multiply row i by a nonzero scalar c	$R_i \longrightarrow cR_i$.
2. Exchange row i and row j	$R_i \longleftrightarrow R_j$.
3. Add c times row j to row i	$R_i \longrightarrow R_i + cR_j$.

We pronounce these symbols as "***multiply*** row i by c," "***exchange*** row i and row j," and "***add*** c times row j ***to*** row i .” If c is negative in a Type 3 operation, we can say “***subtract*** c times row j ***from*** row i” instead.

Thus, to add twice the third row to the fifth row, we write:

$$R_5 \longrightarrow R_5 + 2R_3.$$

After we do any of these operations, we expect the new system to have exactly the ***same solution set*** as our original system. We will prove this fact in the following:

Theorem — The Invariance of Solution Sets:

An elementary row operation ***does not change*** the solution set of an augmented matrix. In other words, if A is an augmented matrix and B is obtained from A using an elementary row operation, then the solution set of the system corresponding to A is exactly the same as the solution set of the system corresponding to B.

Before we prove this Theorem, let us recall from Chapter Zero what it means for two ***sets*** to be equal:

Definition — Equality of Sets:

Two sets X and Y are ***equal*** if X is a subset of Y and Y is a subset of X:

$$(X = Y) \Leftrightarrow \left(X \subseteq Y \textbf{\textit{ and }} Y \subseteq X\right).$$

Equivalently, every member of X is also a member of Y, and every member of Y is also a member of X:

$$x \in X \Rightarrow x \in Y \quad \textbf{\textit{and}} \quad y \in Y \Rightarrow y \in X.$$

Proof of the Theorem: Let $\vec{x} = \left\langle x_1, x_2, \ldots, x_n \right\rangle \in \mathbb{R}^n$. We have to show that if $\vec{x}$ is a solution for the system of A then $\vec{x}$ is a solution for the system of B, and vice-versa. We will do a Case-by-Case analysis involving each type of elementary row operation:

A Type 1 elementary row operation only changes row i. Suppose we write the equation corresponding to row i of A as:

$$a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = a_{i,n+1}.$$

Then the corresponding equation for row i for B is:

$$ca_{i,1}x_1 + ca_{i,2}x_2 + \cdots + ca_{i,n}x_n = ca_{i,n+1}.$$

where $c \neq 0$. Thus, we can recover row i of A by dividing both sides of the second equation by c. Hence, $\vec{x}$ satisfies row i of A ***if and only if*** $\vec{x}$ satisfies row i of B. All other rows of A and B are identical, so $\vec{x}$ satisfies the system represented by A ***if and only if*** $\vec{x}$ satisfies the system represented by B.

A Type 2 elementary row operation exchanges two rows (hence, two equations) without changing the coefficients. Both systems have exactly the ***same equations***, only in a different ***order***. Thus $\vec{x}$ satisfies the system represented by A ***if and only if*** $\vec{x}$ it satisfies the system represented by B.

Suppose B is obtained from A using the Type 3 elementary row operation:

$$R_i \rightarrow R_i + c \cdot R_j.$$

Let us make this more explicit:

Row i of B = Row i of A plus c times Row j of A.

All other rows of A are exactly the same as the rows of B: in particular, row j of A is still exactly the same as row j of B. The ***only*** row that changed is row i. Thus the equation:

$$a_{j,1}x_1 + a_{j,2}x_2 + \cdots + a_{j,n}x_n = a_{j,n+1}$$

corresponding to row j is in ***both*** systems. Thus, if $\vec{x}$ satisfies ***either*** the system of A or of B, then $\vec{x}$ also satisfies:

$$ca_{j,1}x_1 + ca_{j,2}x_2 + \cdots + ca_{j,n}x_n = ca_{j,n+1} \qquad (1)$$

for any real number c. Next, the equation corresponding to row i of A is:

$$a_{i,1}x_1 + a_{i,2}x_2 + \cdots + a_{i,n}x_n = a_{i,n+1} \qquad (2)$$

and the equation corresponding to row i of B is:

$$(a_{i,1} + ca_{j,1})x_1 + (a_{i,2} + ca_{j,2})x_2 + \cdots + (a_{i,n} + ca_{j,n})x_n = a_{i,n+1} + ca_{j,n+1} \qquad (3)$$

that we obtain using our Type 3 row operation.

Now, we are ready to complete the proof. Suppose $\vec{x}$ satisfies the system of A, and thus equations (1) and (2). By ***adding*** both sides of equation (1) to the corresponding sides of equation (2) and combining like terms, we see that $\vec{x}$ also satisfies equation (3). Thus it satisfies the system of B. Similarly, if $\vec{x}$ satisfies the system of B, it satisfies equations (1) and (3). By ***subtracting*** the sides of equation (1) from the corresponding sides of equation (3), we see that $\vec{x}$ also satisfies equation (2). Thus it satisfies the system of A. ■

This proof essentially says that each elementary row operation is ***reversible***. Since a single elementary row operation yields a new system with exactly the same solution set as the old system, so will a ***finite sequence*** of elementary row operations. Our goal is to eventually produce an augmented matrix that is in rref so that we can easily read off the solution set of our original system.

Surprisingly, another advantage of the rref is that any two sequences of row operations starting with from A which both yield a matrix in rref must result in the ***same*** rref. The following will be proven in the Exercises of Section 1.8:

Theorem — The Uniqueness of the Reduced Row Echelon Form:

The reduced row echelon form of a matrix is ***unique***.

This means: if we start with a matrix A and arrive at two matrices H and J using two different sequences of row operations, and both H and J are in rref, then $H = J$.

Now that we know there can be only one rref for each matrix, we need an efficient way to find it:

The Gauss-Jordan Algorithm

Let us put it all together: To find out if a vector $\vec{b}$ is a member of the Span of a set of vectors, we need to solve a system of linear equations. To avoid writing the variables $x_1, x_2, \ldots, x_n$ all the time, we form the ***augmented matrix*** corresponding to this system. Elementary row operations do not change the solution set of an augmented matrix, so now we will use a finite sequence of elementary row operations so that we obtain an augmented matrix in ***rref***. We can then read off the solutions to our original system. Thus, we need a ***systematic*** way to perform the required elementary row operations. This is the essence of the following algorithm, which is often given as a programming project in an introductory Computer Science class:

The Gauss-Jordan Algorithm:

1. Ignore all the leftmost columns that contain only zeros, if there are any.
2. Starting from the top row and going downward, find the first non-zero entry, called the ***pivot***.
3. If the pivot is not in the top row, ***exchange*** the top row with the pivot's row (this is a Type 2 row operation).
4. Produce a leading 1 in the top row by ***dividing*** the entire top row by the pivot (this is a Type 1 row operation). We call this step ***normalizing*** the row.
5. Make the entries below the leading 1 ***all zeroes*** by adding suitable multiples of the top row to each row below it (these are Type 3 row operations).
6. Now, cover the top row, the leading column and all columns to its left, and repeat steps 1 through 5 on the smaller submatrix.

If we were to stop at Step 6, the algorithm above would be called ***Gaussian Elimination***. It results in a matrix in ***row echelon form***. Now we continue from ***right to left,*** working ***upwards*** as we go:

7. Starting at the ***rightmost*** leading 1, produce zeroes above the leading 1 by adding suitable multiples of this row to each row above it.
   Again, these are Type 3 row operations.
8. Repeat Step 7 on the next rightmost leading 1, moving ***leftward,*** until the matrix is in ***reduced row echelon form***.

The entire process above is called the ***Gauss-Jordan Algorithm***. If we stop at Step 6, we can still find the solutions to the system using a process called ***back-substitution***, which will be discussed in the Exercises. Many computer packages and graphing calculators can produce the rref of a matrix. Check the internet for a free app.

An Intelligent Modification:

Unfortunately, Step 4 of the algorithm can produce ***fractions***. To avoid this when all the entries are ***integers***, we can produce a leading "1" using a Type 3 operation instead: multiply a row by a suitable constant and add it to another row to get a "1," then swap this row with the top row if necessary. This is possible if two of the leftmost entries are ***relatively prime***. We will call this the ***Modified Gauss-Jordan Algorithm***.

Example: Let us find the solutions to our first example using the Modified Gauss-Jordan Algorithm:

$$\left[\begin{array}{cccc|c} 4 & -8 & 3 & 9 & 6 \\ 3 & -6 & -4 & 13 & 17 \\ -2 & 4 & 3 & -9 & -12 \end{array}\right]$$

$$\xrightarrow{R_2 \to R_2 + R_3} \left[\begin{array}{cccc|c} 4 & -8 & 3 & 9 & 6 \\ \boxed{1} & -2 & -1 & 4 & 5 \\ -2 & 4 & 3 & -9 & -12 \end{array}\right]$$

$$\xrightarrow{R_2 \leftrightarrow R_1} \left[\begin{array}{cccc|c} \boxed{1} & -2 & -1 & 4 & 5 \\ 4 & -8 & 3 & 9 & 6 \\ -2 & 4 & 3 & -9 & -12 \end{array}\right]$$

$$\xrightarrow[R_3 \to R_3 + 2R_1]{R_2 \to R_2 - 4R_1} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 4 & 5 \\ 0 & 0 & 7 & -7 & -14 \\ 0 & 0 & 1 & -1 & -2 \end{array}\right]$$

$$\xrightarrow{R_3 \leftrightarrow R_2} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 4 & 5 \\ 0 & 0 & \boxed{1} & -1 & -2 \\ 0 & 0 & 7 & -7 & -14 \end{array}\right]$$

$$\xrightarrow{R_3 \to R_3 - 7R_2} \left[\begin{array}{cccc|c} 1 & -2 & -1 & 4 & 5 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \text{(row echelon form)}$$

$$\xrightarrow{R_1 \to R_1 + R_2} \left[\begin{array}{cccc|c} \boxed{1} & -2 & 0 & 3 & 3 \\ 0 & 0 & \boxed{1} & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right] \quad \text{(rref)}$$

From the rref, we can see that the leading variables are x_1 and x_3, and the free variables are x_2 and x_4. From this, we can read off the solutions, as before:

$$x_1 = 3 + 2x_2 - 3x_4,$$
$$x_3 = -2 + x_4,$$
$$x_2 = r,\ x_4 = s.$$

and thus:

$$\langle x_1, x_2, x_3, x_4 \rangle = \langle 3 + 2r - 3s,\ r, -2 + s,\ s \rangle, \text{ where } r, s \in \mathbb{R}.$$

Since this system has a non-empty solution set, the vector $\vec{b}$ is a member of $Span(S)$. The easiest solution would be to just set r and s to be 0, and we get:

$$\langle x_1, x_2, x_3, x_4 \rangle = \langle 3, 0, -2, 0 \rangle.$$

We easily check that:

$$3\langle 4, 3, -2 \rangle - 2\langle 3, -4, 3 \rangle = \langle 6, 17, -12 \rangle = \vec{b}.$$

However, since r and s can be any real number, there are infinitely many ways to express $\vec{b}$ as a linear combination of the vectors in S. For instance, if we let $r = 2$ and $s = -1$, then we get:

$$\langle x_1, x_2, x_3, x_4 \rangle = \langle 3 + 2(2) - 3(-1), 2, -2 + (-1), -1 \rangle = \langle 10, 2, -3, -1 \rangle,$$

and in fact, we can verify that:

$$10\langle 4, 3, -2 \rangle + 2\langle -8, -6, 4 \rangle - 3\langle 3, -4, 3 \rangle - \langle 9, 13, -9 \rangle = \langle 6, 17, -12 \rangle = \vec{b}. \square$$

If a system's solution set contains at least one free variable, we can ask if solutions exist that possess some given restrictions. Continuing with our Example above, we can ask: is there a solution where $x_3 = 0$? Since $x_3 = -2 + s$, then necessarily $s = 2$. Thus, the solutions have the form:

$$\langle x_1, x_2, x_3, x_4 \rangle = \langle 3 + 2r - 3(2), r, -2 + 2, 2 \rangle = \langle r - 3, r, 0, 2 \rangle, \text{ where } x_2 = r \in \mathbb{R}.$$

Similarly, if we want a solution where $x_1 = 5$ and $x_3 = -10$, then we have to satisfy the two conditions:

$$3 + 2r - 3s = 5, \text{ and } \; -2 + s = -10.$$

Thus, $s = -8$, and from the first equation, $r = -11$. Thus, we will have only one solution with these restrictions, namely $\langle x_1, x_2, x_3, x_4 \rangle = \langle 5, -11, -10, -8 \rangle$.

On the other hand, if we want a solution where $x_1 = 5$, $x_2 = 3$, and $x_3 = -1$, then we would need $r = 3$ and $-2 + s = -1$, or $s = 1$. But then, $x_1 = 3 + 6 - 3 = 6$, contradicting the first condition. Thus, there is no solution where all three conditions are satisfied.

Success and Failure

In general, $Span(S)$ will not be all of $\mathbb{R}^m$, so we must be able to tell in the course of applying the Gauss-Jordan Algorithm if a particular $\vec{b}$ is ***not*** in $Span(S)$. This happens when we get a ***contradiction***, in the form of a row consisting entirely of zeroes except for a non-zero entry in the rightmost column (where $\vec{b}$ is found originally).

Example: Suppose we are given the following matrix $\left[A \mid \vec{b}\right]$ with corresponding rref R:

$$\left[A \mid \vec{b}\right] = \begin{bmatrix} 5 & 7 & 1 & | & 3 \\ 1 & -1 & 5 & | & -9 \\ 4 & 5 & 2 & | & 2 \end{bmatrix}; \quad R = \begin{bmatrix} 1 & 0 & 3 & | & 0 \\ 0 & 1 & -2 & | & 0 \\ 0 & 0 & 0 & | & 1 \end{bmatrix}.$$

The bottom row is all zeroes except for the 1 on the right side. This tells us that the system has no solutions. Notice also that we have two leading variables and one free variable, so if there were solutions, we would have obtained an infinite number of solutions.

However, let us see what happens when we change just one entry in A. Let us change the 1 in the top row into a 0. We get a new system with corresponding rref:

$$\left[A^{\prime}\,|\,\vec{b}\right] = \left[\begin{array}{ccc|c} 5 & 7 & 0 & 3 \\ 1 & -1 & 5 & -9 \\ 4 & 5 & 2 & 2 \end{array}\right]; \quad R^{\prime} = \left[\begin{array}{ccc|c} 1 & 0 & 0 & -115/9 \\ 0 & 1 & 0 & 86/9 \\ 0 & 0 & 1 & 8/3 \end{array}\right].$$

Because we changed A, we now have three leading variables and no free variables. The rref also tells us that the new system now has ***exactly one solution***, and if the variables are x, y and z, then that solution would be:

$$x = -115/9,\ y = 86/9,\ \text{and } z = 8/3. \ \square$$

The left side of $R^{\prime}$ is an example of a very special matrix, which frequently appears when we solve systems where the number of equations is the same as the number of variables. We introduce them in the following:

Definition — The Identity Matrices:

The $n \times n$ identity matrix, denoted $\boldsymbol{I_n}$, is the matrix which contains $\vec{e}_1$ in column 1, $\vec{e}_2$ in column 2, …, $\vec{e}_n$ in column n:

$$\boldsymbol{I_n} = \left[\,\vec{e}_1\ \vec{e}_2 \cdots \vec{e}_n\,\right] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

In our Example above, $\boldsymbol{I_3}$ appears on the left side of $R^{\prime}$. More generally, if we are working with a linear system with n equations and n variables, then we get exactly one solution to the system if the rref of the corresponding augmented matrix contains $\boldsymbol{I_n}$ in the first n columns. The unique solution is the vector which appears in the final column. But notice also from the same Example that when we had one free variable, we also had one row of zeroes. We generalize this outcome in the following Theorem. We leave its proof as an Exercise:

Theorem — The RREF of an Square Matrix:

If A is an $n \times n$ matrix, then exactly ***one*** of the following two cases happens to the rref R of A:

1. R is the identity matrix $\boldsymbol{I_n}$, or
2. R contains at least one free variable ***and*** at least one row of zeroes.

   Furthermore, if there are r leading 1's, then:

   the number of free variables in $R = n - r =$ the number of rows of zeroes in R.

Systems of linear equations appear in almost all areas of Science, and so the Gauss-Jordan algorithm is extremely useful. In the final Chapter, you can see its application in Balancing Chemical Equations and in Basic Circuit Analysis.

1.4 Section Summary

Our goal in this Section is to determine if $\vec{b} \in Span(S)$, where $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$. This is equivalent to finding $x_1, x_2, \ldots, x_n$ such that $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{b}$.
From this vector equation, we create the ***augmented matrix*** $\left[\vec{v}_1 \ \vec{v}_2 \ \ldots \ \vec{v}_n \mid \vec{b}\right]$ by assembling the vectors in ***columns***. This augmented matrix represents a linear system of m equations in n unknowns.

A matrix is in ***row echelon form*** if it satisfies the following conditions:

1. All the rows consisting entirely of zeroes are at the ***bottom*** of the matrix.
2. The first non-zero entry of any row is the number 1. This entry will be called a ***leading 1***.
3. The leading 1 in the next row is to the ***right*** of the leading 1 in the previous row.
   Furthermore, we say that the matrix is in ***reduced row echelon form***, or ***rref***, if:
4. All the entries ***above*** a leading 1 are zeroes.

The rref of a matrix is ***unique***.

An ***elementary row operation*** is any one of the following actions on a matrix:

1. ***Multiply*** row i by a nonzero scalar c: $R_i \to cR_i$.
2. ***Exchange*** row i and row j: $R_i \leftrightarrow R_j$.
3. ***Add*** c times row j to row i: $R_i \to R_i + cR_j$.

We perform the ***Gauss-Jordan Algorithm*** on the augmented matrix to bring it to rref. In order to avoid producing fractions, we use a ***modified*** version:

1. Ignore all the leftmost columns that contain only zeros, if there are any.
2. Produce a leading 1 on the top row by either exchanging it with a row that already contains a leading one in the first non-zero column, or add a multiple of one row to another row to produce this leading one. If none of these is possible, divide any non-zero row by the first entry (the ***pivot*** entry) to produce a leading 1. This is called ***normalizing*** the row. You may also try adding a multiple of one row to another to produce smaller entries in the column, and consequently, a smaller denominator in the fractions after normalization, if need be.
3. Make the entries below the leading 1 ***all zeroes*** by adding suitable multiples of the top row to each row below it (these are Type 3 row operations).
4. Now, cover the top row, the leading column and all columns to its left, and repeat steps 1 through 5 on the smaller submatrix. When we reach the bottom, the matrix will be in ***row echelon form***.
5. Starting at the rightmost (hence bottom) leading 1, produce zeroes above the leading 1 by adding suitable multiples of this row to each row above it.
6. Repeat Step 5 on the next rightmost leading 1, moving leftward until the matrix is in ***rref***, that is, all entries above and below the leading 1 are zeroes.

If there is a row that consists entirely of zeroes in the rref, except for the entry in the last column, then $\vec{b} \notin Span(S)$. Otherwise, we can find all solutions using the rref, and express $\vec{b}$ as a linear combination of the vectors in S. Variables corresponding to a leading column are called ***leading variables***, otherwise they are called ***free variables***, which may assume as their value any real number. We solve for the leading variables in terms of the free variables. Thus, if there are free variables in the rref when we have at least one solution, then we can find an ***infinite*** number of ways to express $\vec{b}$ as a linear combination of the vectors in S.

The rref R of an $n \times n$ matrix is either: 1) $R = \boldsymbol{I_n} = [\vec{e}_1\,\vec{e}_2 \cdots \vec{e}_n]$, the $n \times n$ ***identity matrix***, or 2) R contains at least one free variable ***and*** at least one row of zeroes. Furthermore, if there are r leading 1's, then the number of free variables in $R = n - r =$ the number of rows of zeroes in R.

1.4 Exercises

For Exercises (1) to (36): Verify that the following augmented matrices are in rref, then find the solutions, if there are any, to the systems of equations corresponding to the rref. Assume that the variables are $x_1, x_2, \ldots, x_n$. As part of your solution, determine the leading and the free variables. Use the letters r, s, t, and u for values of the free variables.

1. $\left[\begin{array}{ccc|c} 1 & 0 & 0 & -3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 6 \end{array}\right]$

2. $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 9 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & 0 \end{array}\right]$

3. $\left[\begin{array}{ccc|c} 1 & 0 & 7 & -3 \\ 0 & 1 & -4 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$

4. $\left[\begin{array}{ccc|c} 1 & -3 & 0 & 6 \\ 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 \end{array}\right]$

5. $\left[\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 0 & 0 \end{array}\right]$

6. $\left[\begin{array}{ccc|c} 1 & -5 & 2 & 8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$

7. $\left[\begin{array}{cccc|c} 1 & 0 & 0 & -5 & 3 \\ 0 & 1 & 0 & 4 & 0 \\ 0 & 0 & 1 & -7 & -2 \end{array}\right]$

8. $\left[\begin{array}{cccc|c} 1 & 0 & 3 & 0 & 5 \\ 0 & 1 & -2 & 0 & 6 \\ 0 & 0 & 0 & 1 & -4 \end{array}\right]$

9. $\left[\begin{array}{cccc|c} 1 & -9 & 0 & 3 & -5 \\ 0 & 0 & 1 & -6 & 2 \\ 0 & 0 & 0 & 0 & 7 \end{array}\right]$

10. $\left[\begin{array}{cccc|c} 1 & -4 & 0 & 0 & 5 \\ 0 & 0 & 1 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

11. $\left[\begin{array}{cccc|c} 1 & -2 & 6 & 0 & 7 \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

12. $\left[\begin{array}{cccc|c} 1 & 0 & 0 & -\frac{2}{3} & \frac{5}{3} \\ 0 & 1 & 0 & \frac{4}{3} & -\frac{7}{3} \\ 0 & 0 & 1 & \frac{1}{3} & \frac{2}{3} \end{array}\right]$

13. $\left[\begin{array}{ccc|c} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \end{array}\right]$

14. $\left[\begin{array}{ccc|c} 1 & 0 & 3 & 2 \\ 0 & 1 & -5 & -4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$

15. $\left[\begin{array}{ccc|c} 1 & -6 & 0 & 7 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$

16. $\left[\begin{array}{cccc|c} 1 & 0 & 0 & 0 & 5 \\ 0 & 1 & 0 & 0 & 6 \\ 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right]$

17. $\left[\begin{array}{cccc|c} 1 & 0 & 0 & -4 & 0 \\ 0 & 1 & 0 & 7 & 3 \\ 0 & 0 & 1 & 3 & -8 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

18. $\left[\begin{array}{cccc|c} 1 & 0 & -6 & 0 & 1 \\ 0 & 1 & 4 & 0 & 5 \\ 0 & 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

19. $\left[\begin{array}{cccc|c} 1 & -5 & 0 & 0 & -2 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

20. $\left[\begin{array}{cccc|c} 1 & 0 & -3 & 2 & -8 \\ 0 & 1 & 4 & -6 & -5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

21. $\left[\begin{array}{cccc|c} 1 & -5 & 0 & -9 & -2 \\ 0 & 0 & 1 & 4 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right]$

22. $\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 7 & 5 & -5 \\ 0 & 1 & 0 & -4 & -3 & 2 \\ 0 & 0 & 1 & 6 & -2 & 4 \end{array}\right]$

23. $\left[\begin{array}{ccccc|c} 1 & 0 & 3 & -4 & -6 & 5 \\ 0 & 1 & -2 & -9 & 8 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

24. $\left[\begin{array}{ccccc|c} 1 & 0 & 0 & 0 & 6 & -5 \\ 0 & 1 & 0 & 0 & -3 & 2 \\ 0 & 0 & 1 & 0 & 2 & 4 \\ 0 & 0 & 0 & 1 & 8 & -1 \end{array}\right]$

25. $\left[\begin{array}{ccccc|c} 1 & 0 & 3 & 0 & 0 & 5 \\ 0 & 1 & -2 & 0 & 0 & 6 \\ 0 & 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & 0 & 1 & 9 \end{array}\right]$

26. $\left[\begin{array}{ccccc|c} 1 & 6 & 0 & 3 & 0 & 2 \\ 0 & 0 & 1 & -8 & 0 & 7 \\ 0 & 0 & 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

27. $\left[\begin{array}{ccccc|c} 1 & -5 & 0 & 0 & 4 & -2 \\ 0 & 0 & 1 & 0 & 7 & 9 \\ 0 & 0 & 0 & 1 & 3 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

28. $\left[\begin{array}{ccccc|c} 1 & 1 & 0 & -3 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

29. $\left[\begin{array}{ccccc|c} 0 & 1 & -3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$ (careful!)

30. $\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 0 & -5 & 3 & 4 \\ 0 & 1 & 0 & 0 & 3 & 0 & 5 \\ 0 & 0 & 1 & 0 & -2 & 4 & -2 \\ 0 & 0 & 0 & 1 & 7 & -6 & 3 \end{array}\right]$

31. $\left[\begin{array}{cccccc|c} 1 & 0 & -9 & 0 & 0 & 4 & 7 \\ 0 & 1 & 3 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 6 & -1 \\ 0 & 0 & 0 & 0 & 1 & 5 & 2 \end{array}\right]$

32. $\left[\begin{array}{cccccc|c} 1 & 6 & 0 & 3 & 5 & 0 & 2 \\ 0 & 0 & 1 & -8 & -2 & 0 & 9 \\ 0 & 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

33. $\left[\begin{array}{ccccc|c} 1 & 0 & 5 & 0 & 0 & 3 \\ 0 & 1 & -2 & 0 & 0 & -7 \\ 0 & 0 & 0 & 1 & 0 & 9 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

34. $\left[\begin{array}{ccccc|c} 1 & 0 & -4 & 0 & 7 & -2 \\ 0 & 1 & 6 & 0 & -3 & 5 \\ 0 & 0 & 0 & 1 & 9 & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

35. $\left[\begin{array}{cccccc|c} 1 & 0 & -8 & 0 & 0 & -1 & -2 \\ 0 & 1 & 5 & 0 & 0 & 7 & 6 \\ 0 & 0 & 0 & 1 & 0 & -4 & 3 \\ 0 & 0 & 0 & 0 & 1 & 9 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$

36. $\left[\begin{array}{cccccc|c} 1 & 0 & 0 & 6 & 0 & 0 & -5 \\ 0 & 1 & 0 & -7 & 0 & 0 & 2 \\ 0 & 0 & 1 & 4 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 0 & -8 \\ 0 & 0 & 0 & 0 & 0 & 1 & 9 \end{array}\right]$

For Exercises (37) to (48): Use the Modified Gauss-Jordan Algorithm to determine if $\vec{b}$ is in Span(S). If so, express $\vec{b}$ as a linear combination of the vectors in S in the simplest possible way (i.e. with all free variables set to 0). Is there only one solution or are there infinitely many?

37. $\vec{b} = \langle -20, -26, 39\rangle$; $S = \{\langle 5, -2, 3\rangle, \langle 7, 4, -6\rangle\}$

38. $\vec{b} = \langle 34, 34, -55\rangle$; $S = \{\langle 5, -2, 3\rangle, \langle 7, 4, -6\rangle\}$

39. $\vec{b} = \langle 3, 14, 15\rangle$, $S = \{\langle 3, 5, 6\rangle, \langle 4, 1, 0\rangle, \langle 2, 1, -3\rangle\}$

40. $\vec{b} = \langle 3, 4, 5\rangle$, $S = \{\langle 3, 5, -2\rangle, \langle 1, 1, 4\rangle, \langle 1, 2, -3\rangle\}$

41. $\vec{b} = \langle 1, 2, 7\rangle$, $S = \{\langle 3, 5, -2\rangle, \langle 1, 1, 4\rangle, \langle 1, 2, -3\rangle\}$

42. $\vec{b} = \langle 9, 7, -8, 2\rangle$; $S = \{\langle 5, -3, 2, 6\rangle, \langle -2, -3, 5, 8\rangle, \langle -5, 4, -2, -3\rangle\}$

43. $\vec{b} = \langle 44, 12, 31, 53\rangle$, $S = \{\langle 2, -3, 2, 3\rangle, \langle 6, -3, 5, 8\rangle, \langle 10, 9, 6, 11\rangle\}$

44. $\vec{b} = \langle -7, 29, -21, -15\rangle$, $S = \{\langle 4, 3, -3, 1\rangle, \langle 2, 0, 1, -1\rangle, \langle -3, 4, -2, -4\rangle, \langle -14, 22, -19, -9\rangle\}$

45. $\vec{b} = \langle 2, -6, 11, -7\rangle$, $S = \{\langle 4, 3, -3, 1\rangle, \langle 2, 0, 1, -1\rangle, \langle 18, 6, -1, -3\rangle, \langle 4, 9, -13, 7\rangle\}$

46. $\vec{b} = \langle 6, 9, -22, 15 \rangle$, $S = \{\langle 4, 3, -3, 1 \rangle, \langle 2, 0, 1, -1 \rangle, \langle 18, 6, -1, -3 \rangle, \langle 4, 1, -3, 2 \rangle\}$

47. $\vec{b} = \langle 13, 0, 1, 18, -2 \rangle$, $S = \{\langle 6, 0, 4, 3, 2 \rangle, \langle 3, 2, 7, 1, -2 \rangle, \langle 5, 1, 2, 8, -3 \rangle\}$

48. $\vec{b} = \langle 18, -8, -8, 11, 18 \rangle$, $S = \{\langle 6, 0, 4, 3, 2 \rangle, \langle 3, 2, 7, 1, -2 \rangle, \langle 12, -4, -2, 7, 10 \rangle\}$

For Exercises (49) to (58): Find the solution/s to the indicated system above, if any, which possess the given restrictions. Use the parameters r, s, t, etc., if need be (i.e. if there is more than one solution with the given restrictions), as before. Use the Answer Key for the correct general solution.

49. The system in Exercise 3, where $x_1 = 0$.
50. The system in Exercise 7, where $x_2 = x_3$.
51. The system in Exercise 10, where $x_1 = x_4$ and $x_2 = x_3$.
52. The system in Exercise 20, where $x_1 = -2x_4$.
53. The system in Exercise 30, where $x_1 = -7$ and $x_4 = 31$.
54. The system in Exercise 32, where $x_3 = -7$. Note: to obtain a standard answer, keep x_5 free and solve for x_4 in terms of x_5.
55. The system in Exercise 35, where $x_1 = x_5$. Note: to obtain a standard answer, keep x_6 free and solve for x_3 in terms of x_6.
56. The system in Exercise 40, where $x_1 = 5$.
57. The system in Exercise 43, where $x_2 = -3$.
58. The system in Exercise 45, where $x_1 = -14$ and $x_2 = -2$.
59. The following matrix is in row echelon form, but is not reduced.

$$\left[\begin{array}{ccccc|c} 1 & 3 & 5 & 6 & 7 & 1 \\ 0 & 0 & 1 & 2 & -3 & -8 \\ 0 & 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right]$$

a. Identify all leading and all free variables, without completing the reduction process.

b. Starting with the bottom row, solve for the leading variables in terms of the free variables ***only***.

c. Write the solutions to the system in standard form.

For Exercises (60) to (62): Apply the Modified Gauss-Jordan Algorithm to describe all the solutions, if there are any, to the following systems of linear equations.

60.
$$\begin{array}{rcrcrcrcr} 3x & - & 15y & - & 5z & + & 2w & = & 27 \\ -2x & + & 10y & + & 3z & - & 4w & = & -28 \\ 5x & - & 25y & - & 2z & - & w & = & 15 \end{array}$$

61.
$$\begin{array}{rcrcrcrcr} 3x & - & 2y & + & 19z & + & 4w & = & 39 \\ 4x & - & y & + & 22z & - & 3w & = & 7 \\ -x & + & 5y & - & 15z & + & 2w & = & -30 \\ x & + & 2y & + & z & + & 4w & = & 5 \end{array}$$

62. $$\begin{aligned} x + y - z &= 4 \\ x - 2y + 5z &= 1 \\ 2x - y + 4z &= 3 \\ 2x + y &= 5 \end{aligned}$$

63. Find parametric equations for the line of intersection of the two planes:

$$\begin{aligned} 3x - 5y + 2z &= 10, \quad \text{and} \\ -2x + 4y + 7z &= 8. \end{aligned}$$

Express your answer in the form: $\langle x, y, z\rangle = \langle x_0, y_0, z_0\rangle + t\langle a, b, c\rangle$, where a, b, and c are all integers. Hint: view this problem as a linear system.

64. Chris takes some shirts, slacks and jackets to the dry cleaners. Four shirts, a pair of slacks and two jackets cost \$25 to clean. On another trip, 6 shirts, 2 pairs of slacks and a jacket cost \$26. On a third trip, two shirts, four pairs of slacks and two jackets cost \$37. How much does it cost to clean each kind of clothing? You may assume that the prices do not change from trip to trip.

65. Farmer Pat wants to feed cows using three kinds of grains. Their nutritional contents, in grams per kilogram, are shown below:

	Fiber	Carbohydrates	Protein
Barley	200	400	50
Oats	100	500	150
Soy	150	250	400

If the cows must be fed so that they receive 800 grams of fiber, 2400 grams of carbohydrates and 1300 grams of protein, how many kilograms of each kind of grain should be used?

66. ***Integer Solutions to Linear Systems:*** Although a linear system may have an infinite number of solutions, it is possible that we are only interested in those solutions where all the variables are ***integers***. This could happen, for example, when the variables represent objects being counted. Consider the following:

A jar of coins contains only dimes (worth 10 cents each), nickels (worth 5 cents each) and pennies (worth 1 cent each). The total value of the coins is \$ 6.49 and the jar has 98 coins.

a. Set up the system of equations representing this system, in the variables d, n, and p.

b. Find the rref of this system.

c. If you did (b) correctly, p should be a free variable. Solve for d and n in terms of p, as usual, then find the ***smallest*** positive integer value for p that will make both d and n positive integers. Use trial and error, if need be. This gives one possible solution to the system, but not necessarily the only one.

d. Find the solution with the ***largest*** number of pennies.

67. Prove the final Theorem in this Section: If A is an $n \times n$ matrix, then exactly ***one*** of the following two cases happens to the rref R of A: (1) R is the identity matrix $\boldsymbol{I_n}$, or (2) R contains at least one free variable ***and*** at least one row of zeroes. Furthermore, if there are r leading 1's, then:

$$\text{the number of free variables in } R = n - r = \text{the number of rows of zeroes in } R.$$

Hint: let Case 1 be that there are n leading ones in R, and let Case 2 be that there are fewer than n leading ones in R.

1.5 Linear Systems and Linear Independence

We saw in the previous section that determining if a vector $\vec{b} \in \mathbb{R}^m$ is a member of a Span of n vectors from $\mathbb{R}^m$ led us to consider systems of m linear equations in n variables. Let us now see some ways by which we can ***categorize*** such systems. We will first define terms that describe the presence or absence of solutions:

> ***Definition:*** A linear system is called ***consistent*** if it has ***at least one*** solution. A system is called ***inconsistent*** if it does ***not*** have any solutions.

We can now recast the problem of determining when a vector $\vec{b}$ is a member of a certain $Span(S)$. The following Theorem follows directly from our construction from the previous Section and the definition above:

> ***Theorem:*** Let $\vec{b} \in \mathbb{R}^m$ and let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors from $\mathbb{R}^m$. Then $\vec{b} \in Span(S)$ ***if and only if*** the system of equations corresponding to the augmented matrix:
>
> $$\left[\begin{array}{cccc|c} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n & \vec{b} \end{array}\right]$$
>
> is ***consistent***.

We also have special words describing the three possible relationships between the number of equations and the number of variables:

> ***Definition:*** A linear system with m equations in n variables is called:
>
> 1. ***underdetermined*** if $m < n$.
> 2. ***square*** if $m = n$.
> 3. ***overdetermined*** if $m > n$.

These three kinds of systems can be visualized as follows:

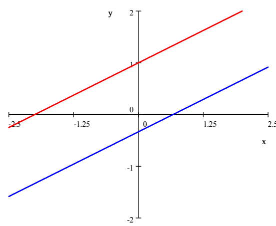
$$\left[\begin{array}{ccccc|c} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{array}\right] \qquad \left[\begin{array}{ccc|c} \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \end{array}\right] \qquad \left[\begin{array}{cc|c} \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots \end{array}\right]$$

Underdetermined System | Square System | Overdetermined System

Geometric Interpretation in $\mathbb{R}^2$ and $\mathbb{R}^3$

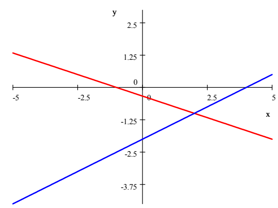
It is easy to understand the meaning of a system of linear equations if there are only two or three variables involved.

A linear system in two variables x and y represents a system of ***lines*** in $\mathbb{R}^2$. If only two lines are involved, the system is ***square***. If these two lines are ***parallel***, then the system is automatically ***inconsistent***, and will remain inconsistent even if more equations are involved. If they are ***coincident***, in other words, the two equations have the same line as their graph, then the system is ***consistent***, with an ***infinite*** number of solutions. If the two lines intersect at exactly one point, then the system is ***consistent***, with a ***unique*** solution.



A Square, Inconsistent System

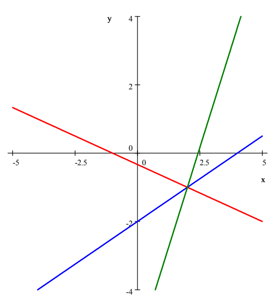
Involving Two Parallel Lines



A Square, Consistent System Involving

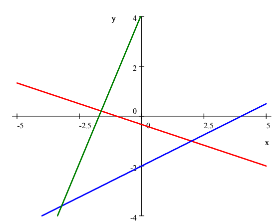
Two Lines With a Unique Intersection Point

If we add a third equation, we will get an ***overdetermined*** system. Suppose we include a third equation in the system above, on the right. Then it will stay ***consistent*** ***if and only if*** the third line ***contains*** the unique point of intersection, otherwise the system will be ***inconsistent:***



An Overdetermined, Consistent System

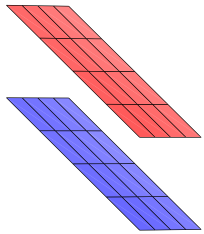
With a Unique Intersection Point

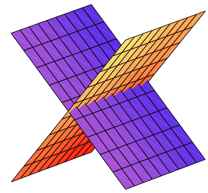


An Overdetermined, Inconsistent System

With No Common Intersection Point

Now, let us consider systems with three variables, say x, y and z. We saw that a single linear equation in these three variables corresponds to a ***plane*** in $\mathbb{R}^3$. Suppose that we had a system of ***two equations*** in these variables. Thus, this system is ***underdetermined***. If these planes are ***parallel***, there will be no intersection, which means that the system will have no solutions, as seen on the left:





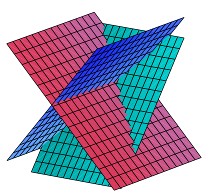
An Inconsistent, Underdetermined
System Involving Two Parallel Planes

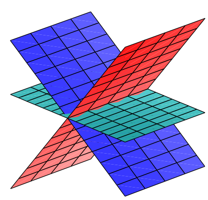
A Consistent, Underdetermined
System Involving Two Intersecting Planes

Thus our system is ***inconsistent***, and will remain inconsistent even if more equations are involved.

If the two planes are ***not parallel***, though, they will intersect in a ***line***, and our system will be ***consistent***, but with an ***infinite*** number of solutions, as we see in the diagram above on the right.

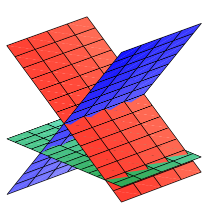
Suppose that we introduce a ***third*** equation to the system on the right above. Thus, this system will now be ***square***, because we have three equations in three variables. There are several possible scenarios with this third plane involved. If the line of intersection of the first two planes intersects the third plane at exactly one point, as seen below on the left, then the system is ***consistent*** and has a ***unique*** solution. In contrast, if the line of intersection is fully contained in the third plane, as seen below on the right, the system is again ***consistent***, but has an ***infinite*** number of solutions.





A Consistent Square System with A Unique Solution, and An Infinite Number of Solutions

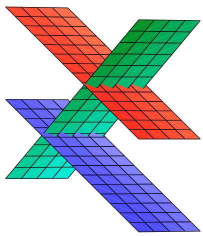
On the other hand, if this line of intersection is ***parallel*** to the third plane, then it will not intersect this plane, so there will be no point in common among all the three planes involved. Thus, our system will be ***inconsistent***, as we see below:



An Inconsistent, Square System
Involving Pairwise Intersecting Planes

Notice that ***each*** pair of planes intersects in a ***line***, but the three planes, ***taken together***, do not have a common point of intersection. It is for this reason that this system is inconsistent. This system is analogous to that of three equations in two variables where pairs of lines have a point of intersection, but all three lines, taken together, do not have a common point of intersection.

Finally, let us see what happens if a third plane is introduced to a system involving two parallel planes, as we see below. We already know that this system is ***inconsistent***, but since it now involves three planes, it is ***square***. The third plane that we chose intersects each of the two parallel planes in a line, but these two lines appear to be parallel to each other. Indeed, they are, as you will be proving in the Exercises.



An Inconsistent, Square System
Involving Two Parallel Planes

Example: Let us investigate the system:

$$\begin{array}{rcrcrcr} 3x & - & y & - & z & = & 3 \\ x & + & 3y & + & z & = & 3 \\ x & - & 2y & - & z & = & 0 \end{array}.$$

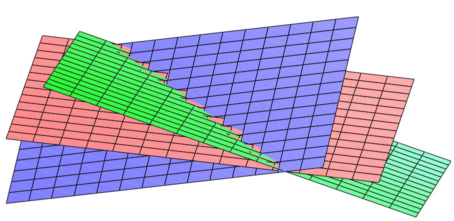
Since we have 3 equations in 3 variables, this system is ***square***. The augmented matrix is:

$$\left[\begin{array}{rrr|r} 3 & -1 & -1 & 3 \\ 1 & 3 & 1 & 3 \\ 1 & -2 & -1 & 0 \end{array}\right] \text{ with rref } \left[\begin{array}{rrr|r} 1 & 0 & -\frac{1}{5} & \frac{6}{5} \\ 0 & 1 & \frac{2}{5} & \frac{3}{5} \\ 0 & 0 & 0 & 0 \end{array}\right].$$

From the rref, we see that x and y are the leading variables, and z is a free variable. This system has solutions, therefore it is ***consistent***. The solutions are:

$$\langle x, y, z \rangle = \left\langle \frac{6}{5}, \frac{3}{5}, 0 \right\rangle + \frac{t}{5}\langle 1, -2, 5 \rangle, \text{ where } z = t \in \mathbb{R}.$$

These solutions form a ***line***, with direction vector $\langle 1, -2, 5 \rangle$. We graph the three planes below:



Three Planes Intersecting at a Common Line.□

Homogeneous Systems

Let us now study a type of system that plays a central role in Linear Algebra:

Definition: A ***homogeneous*** system of m equations in n unknowns is a system of linear equations where the right side of the equations consists entirely of ***zeroes***. In other words, the augmented matrix has the form:

$$\left[A \,|\, \vec{\mathbf{0}}_m \right],$$

where A is an $m \times n$ matrix. If the right side $\vec{b}$ is not the zero vector, we call the system ***non-homogeneous***.

Clearly, $\vec{x} = \vec{\mathbf{0}}_n = \langle 0, 0, \ldots, 0 \rangle$ is a solution to the homogeneous system.

We call this the ***trivial solution*** to a homogeneous system, and any other solution is called a ***non-trivial solution***.

Using the Gauss-Jordan algorithm, we can easily prove the following:

Theorem: A ***consistent*** linear system represented by the augmented matrix $\left[A \,|\, \vec{b} \right]$ has an ***infinite*** number of solutions ***if and only if*** the rref of A has at least one ***free variable***.

However, we can also see that we are guaranteed to produce a free variable if we have more variables than equations:

Theorem: An ***underdetermined homogeneous system*** always has an ***infinite*** number of solutions. In other words, a homogeneous system with ***more variables than equations*** has an infinite number of solutions.

Proof: If we have a linear system with m equations, then we can have at most m leading variables (possibly fewer). So if we have n unknowns where $m < n$, then we are guaranteed at least $n - m \geq 1$ free variable/s, and thus an infinite number of solutions.■

Example: In the previous section, we saw the augmented matrix:

$$\left[\begin{array}{rrrr|r} 4 & -8 & 3 & 9 & 6 \\ 3 & -6 & -4 & 13 & 17 \\ -2 & 4 & 3 & -9 & -12 \end{array} \right], \text{ with rref } \left[\begin{array}{rrrr|r} 1 & -2 & 0 & 3 & 3 \\ 0 & 0 & 1 & -1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right].$$

To review: this system is ***consistent***, since it has solutions. It has 4 variables but only 3 equations, and therefore it is ***underdetermined***.

Now, to solve the corresponding ***homogeneous*** system, we change the rightmost columns entirely to ***zeroes***. This is valid because all of the row operations will preserve the zeroes on the rightmost column. This also means that in future computations we need not bother writing down the zero vector on the right side. Now, this homogeneous system has free variables x_2 and x_4, and the solutions are:

$$\langle x_1, x_2, x_3, x_4 \rangle = \langle 2r - 3s, r, s, s \rangle, \text{ where } r, s \in \mathbb{R}. \square$$

We want to point out something very interesting: Any member of the solution set can be rewritten as:

$$\langle x_1, x_2, x_3, x_4\rangle = \langle 2r, r, 0, 0\rangle + \langle -3s, 0, s, s\rangle = r\langle 2, 1, 0, 0\rangle + s\langle -3, 0, 1, 1\rangle,$$

and we recognize this as an arbitrary member of the ***Span*** of two vectors. This is not a coincidence, and we will say more about this in a future section. But before we go further, let us introduce a fundamental operation in Linear Algebra:

Matrix Products

We have now reached a point where we will combine a vector and a matrix in an arithmetic operation. To begin, suppose that $\vec{x} = \langle x_1, x_2, \ldots, x_n\rangle \in \mathbb{R}^n$. We can arrange the components of $\vec{x}$ ***vertically*** instead of horizontally, and thus think of $\vec{x}$ as an $n \times 1$ or a ***column matrix***. We can call this column matrix $[\vec{x}\,]$, to signify that it comes from the vector $\vec{x}$, but for the sake of brevity, we will also refer to this column matrix as $\vec{x}$ so as to avoid the effort of writing the brackets. When the context is clear, this should not result in confusion. Thus:

$$\vec{x} = \langle x_1, x_2, \ldots, x_n\rangle = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

Similarly, we can view an $m \times n$ matrix A as being ***partitioned*** into n columns:

$$A = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix},$$

where each column $\vec{c}_i \in \mathbb{R}^m$ can also be viewed as an $m \times 1$ matrix. Using this idea, let us define the following operation:

Definition — Matrix Product:

If $A = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix}$ is an $m \times n$ matrix and $\vec{x} \in \mathbb{R}^n$, we define the ***matrix product*** $A\vec{x}$ to be the ***linear combination:***

$$A\vec{x} = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{c}_1 + x_2\vec{c}_2 + \cdots + x_n\vec{c}_n.$$

Notice that since each column is an $m \times 1$ matrix, the matrix product is again an $m \times 1$ matrix. Thus, $A\vec{x}$ is a linear combination of the ***columns*** of A with ***coefficients*** from $\vec{x}$, and so $A\vec{x} \in \mathbb{R}^m$.

In Section 1.7, we will see another way to compute this product, and in Chapter 2, we will also see that we can multiply two matrices, in general, if they satisfy a certain compatibility requirement.

Example: Let us compute the matrix product $A\vec{x}$, where:

$$A = \begin{bmatrix} 4 & 1 & 2 & -5 \\ -2 & 5 & 3 & 4 \\ 3 & 7 & -2 & 1 \end{bmatrix}, \text{ and } \vec{x} = \begin{bmatrix} 3 \\ -2 \\ 7 \\ 0 \end{bmatrix}.$$

By definition, we get:

$$A\vec{x} = \mathbf{3}\begin{bmatrix} 4 \\ -2 \\ 3 \end{bmatrix} \mathbf{-2}\begin{bmatrix} 1 \\ 5 \\ 7 \end{bmatrix} \mathbf{+7}\begin{bmatrix} 2 \\ 3 \\ -2 \end{bmatrix} \mathbf{+0}\begin{bmatrix} -5 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 24 \\ 5 \\ -19 \end{bmatrix}.$$

Notice that although $\vec{x} \in \mathbb{R}^4$, the product $A\vec{x} \in \mathbb{R}^3$, since A is 3×4. □

Matrix multiplication enjoys many properties that are similar to those of vector arithmetic, since it is defined using linear combinations. However, the two most important properties are shown below. You will prove them in the Exercises:

Theorem — Properties of Matrix Multiplication:

For all $m \times n$ matrices A, for all $\vec{x}, \vec{y} \in \mathbb{R}^n$, and for all $k \in \mathbb{R}$, matrix multiplication enjoys the following properties:

The Additivity Property $\quad A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$.

The Homogeneity Property $\quad A(k\vec{x}) = k(A\vec{x})$.

The Matrix Product Form of Linear Systems

Now, let us go back to the problem of determining if a vector $\vec{b}$ is a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$, that is, if we can find coefficients $x_1, x_2, \ldots, x_n$ such that:

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{b}.$$

We formed the augmented matrix $\left[\, \vec{v}_1 \; \vec{v}_2 \; \cdots \; \vec{v}_n \mid \vec{b} \,\right]$, and its rref told us whether or not $\vec{b}$ is a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$. But now that we have the concept of a matrix product, we can turn the left side of the equation above into a matrix product, yielding:

$$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \vec{b}.$$

We can write this symbolically as the ***matrix equation:***

$$A\vec{x} = \vec{b},$$

where the ***coefficient matrix*** A has as its columns $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$. As before, we say that this system is ***consistent*** if it has at least one solution $\vec{x}$, otherwise we say it is ***inconsistent***. Our discussion above can be summarized in the following Theorem:

Theorem: Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors from $\mathbb{R}^{\boldsymbol{m}}$, and $\vec{b} \in \mathbb{R}^{\boldsymbol{m}}$. Let us form the $m \times n$ matrix: $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_n \end{bmatrix}$.
Then: $\vec{b} \in \mathrm{Span}(S)$ ***if and only if*** the matrix equation $A\vec{x} = \vec{b}$ is ***consistent***.

Note the similarities and differences between this Theorem and the first one in this Section.

Linear Dependence and Independence

Homogeneous systems appear in the following important concept:

Definition: A set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ from $\mathbb{R}^{\boldsymbol{m}}$ is ***linearly dependent*** if we can find a ***non-trivial solution*** $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle \in \mathbb{R}^{\boldsymbol{n}}$, where at least one component is ***not zero***, to the vector equation:

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{\mathbf{0}}_{\boldsymbol{m}}.$$

We will call this equation the ***dependence test equation*** for S. An equation of this form where at least one coefficient is ***not zero*** will be referred to as a ***dependence equation***. Thus, for S to be linearly dependent, we must find a ***non-trivial*** solution $\vec{x}$ to the homogeneous system:

$$A\vec{x} = \vec{\mathbf{0}}_{\boldsymbol{m}},$$

where $A = \begin{bmatrix} \vec{v}_1 \mid \vec{v}_2 \mid \ldots \mid \vec{v}_n \end{bmatrix}$ is the matrix with the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ as its ***columns***. This is equivalent to the presence of a ***free variable*** in the rref of the matrix A.

However, if only the trivial solution $\vec{x} = \vec{\mathbf{0}}_{\boldsymbol{n}}$ exists for the dependence test equation, we say that S is ***linearly independent***. We often drop the adjective "linearly" and simply say that a set S is ***dependent*** or ***independent***.

Example: The ***standard basis*** $S = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_m\}$ from $\mathbb{R}^{\boldsymbol{m}}$ is a ***linearly independent*** set, because:

$$x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_m\vec{e}_m = \langle x_1, x_2, \ldots, x_m \rangle,$$

so the linear combination is $\vec{\mathbf{0}}_{\boldsymbol{m}}$ ***if and only if*** $x_1 = 0,\ x_2 = 0,\ \ldots, x_m = 0.\ \square$

Example: Suppose that $\vec{v}_1 = \langle 3, 7, -2 \rangle$, $\vec{v}_2 = \langle -1, 5, -4 \rangle$ and $\vec{v}_3 = \langle 15, 13, 4 \rangle$. To determine if $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent, we need to check whether or not the ***dependence test equation:***

$$x_1\langle 3, 7, -2 \rangle + x_2\langle -1, 5, -4 \rangle + x_3\langle 15, 13, 4 \rangle = \langle 0, 0, 0 \rangle,$$

has a ***nontrivial*** solution. We saw from the introductory subsection that the rightmost column of zeroes will just stay as zeroes throughout the Gauss-Jordan Algorithm. Thus, it is pointless to include this column of zeroes. Therefore, all we have to do is assemble the three vectors into columns, to get the corresponding matrix:

$$A = \begin{bmatrix} 3 & -1 & 15 \\ 7 & 5 & 13 \\ -2 & -4 & 4 \end{bmatrix} \text{ with rref } R = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, x_1 and x_2 are leading variables, x_3 is a free variable, and we get ***nontrivial solutions:***

$$x_1 = -4t,\ x_2 = 3t,\ x_3 = t, \text{ where } t \in \mathbb{R}.$$

Thus, if we let $t = 1$, we get the ***dependence equation:***

$$-4\langle 3, 7, -2\rangle + 3\langle -1, 5, -4\rangle + \langle 15, 13, 4\rangle = \langle 0, 0, 0\rangle.$$

We can now conclude that the set of three vectors $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is ***dependent***. Notice that we can rewrite the corresponding dependence equation as:

$$\langle 15, 13, 4\rangle = 4\langle 3, 7, -2\rangle - 3\langle -1, 5, -4\rangle, \text{ i.e.}$$
$$\vec{v}_3 = 4\vec{v}_1 - 3\vec{v}_2. \square$$

Classifying Small Sets of Vectors

We saw in Section 1.2 how to geometrically describe the Span of a set of one or two vectors as a ***point***, a ***line***, or a ***plane*** in $\mathbb{R}^2$ or $\mathbb{R}^3$. Let us now see what the concepts of linear independence and dependence mean for these small sets of vectors. First, let us make some general observations:

Theorem: Any set $S = \left\{\vec{\mathbf{0}}_m, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\right\} \subset \mathbb{R}^m$ containing $\vec{\mathbf{0}}_m$ is a ***dependent*** set.

Proof: $1 \cdot \vec{\mathbf{0}}_m + 0 \cdot \vec{v}_1 + 0 \cdot \vec{v}_2 + \cdots + 0 \cdot \vec{v}_n = \vec{\mathbf{0}}_m$ is a dependence equation with a ***non-zero coefficient*** 1 as the coefficient of $\vec{\mathbf{0}}_m$.■

With that out of the way, let us investigate a set with ***one vector:***

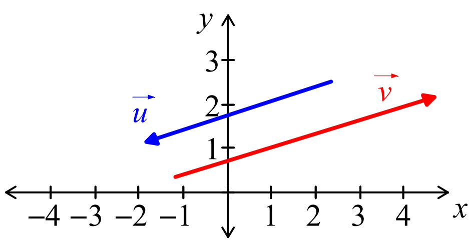
Theorem: A set $S = \{\vec{v}\} \subset \mathbb{R}^m$ is ***independent if and only if*** $\vec{v}$ is a ***non-zero*** vector .

Proof: ($\Rightarrow$) Let us use the contrapositive: if $\vec{v} = \vec{\mathbf{0}}_m$, then $S = \left\{\vec{\mathbf{0}}_m\right\}$ is dependent, which we already know to be true from the previous Theorem. ($\Leftarrow$) Consider the dependence test equation $k \cdot \vec{v} = \vec{\mathbf{0}}_n$. We know from ***The Zero Factors Theorem*** in Exercise 28 of Section 1.1 that $k \cdot \vec{v} = \vec{\mathbf{0}}_n$ ***if and only if*** either $k = 0$ or $\vec{v} = \vec{\mathbf{0}}_n$. Since $\vec{v} \neq \vec{\mathbf{0}}_n$, we must have $k = 0$, hence we only have the trivial solution. Thus $S = \{\vec{v}\}$ is independent.■

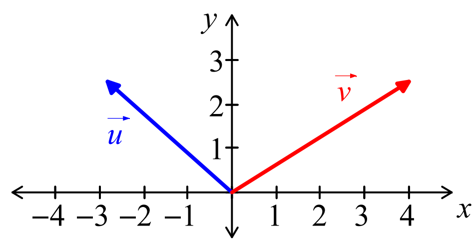
Now, let us think of sets with ***two vectors***, say, $S = \{\vec{u}, \vec{v}\}$. We already know that if one of them is the zero vector, then the set is dependent. So suppose both vectors are ***non-zero***, and we find a dependence equation for them. For example, suppose: $3\vec{u} + 5\vec{v} = \vec{\mathbf{0}}_n$.

But then we can solve for one of them, say $\vec{u}$, and conclude that: $\vec{u} = -\frac{5}{3}\vec{v}$. This says that $\vec{u}$ and $\vec{v}$ are ***parallel*** to each other! Let us prove in general:

Theorem: A set $S = \{\vec{u}, \vec{v}\} \subset \mathbb{R}^m$ is ***dependent*** ***if and only if*** $\vec{u}$ and $\vec{v}$ are ***parallel*** to each other.



Two Dependent (Parallel) Vectors



Two Independent (non-parallel) Vectors

Proof: ($\Rightarrow$) Suppose $\vec{u}$ and $\vec{v}$ are vectors, and $x_1 \cdot \vec{u} + x_2 \cdot \vec{v} = \vec{\mathbf{0}}_n$.
If we had a non-trivial solution, then either x_1 or x_2 is not zero. Suppose x_1 is not zero. Then we can solve for $\vec{u}$, obtaining $\vec{u} = -\frac{x_2}{x_1} \cdot \vec{v}$. Thus, $\vec{u}$ and $\vec{v}$ are parallel to each other.
Similarly, if x_2 is not zero, $\vec{v}$ will be a multiple of $\vec{u}$.

($\Leftarrow$) If $\vec{u}$ and $\vec{v}$ are parallel to each other, then $\vec{u} = a \cdot \vec{v}$, or $\vec{v} = b \cdot \vec{u}$, for some scalars a or b. We can rewrite the first equation as:

$$1 \cdot \vec{u} - a \cdot \vec{v} = \vec{\mathbf{0}}_n.$$

This is a dependence equation, since the coefficient of $\vec{u}$ is not zero, and therefore $\vec{u}$ and $\vec{v}$ are linearly dependent. A similar argument holds for $\vec{v} = b \cdot \vec{u}$. ■

Example: The set $\{\langle 9, 15, -12\rangle, \langle -12, -20, 32\rangle\}$ is linearly ***dependent***, since:

$$\langle -12, -20, 32\rangle = -\frac{4}{3}\langle 9, 15, -12\rangle. \square$$

Now, suppose we had a set of ***three vectors***, say, $S = \{\vec{u}, \vec{v}, \vec{w}\}$ from some $\mathbb{R}^m$. We know that if any two of the vectors were ***parallel***, then S will also be ***dependent***, so let us assume that no two vectors in S are parallel to each other (this also excludes the possibility that one of the vectors is $\vec{\mathbf{0}}_m$). By definition, S is dependent ***if and only if*** we have a ***non-trivial*** solution to the ***dependence test equation:***

$$c_1 \cdot \vec{u} + c_2 \cdot \vec{v} + c_3 \cdot \vec{w} = \vec{\mathbf{0}}_3.$$

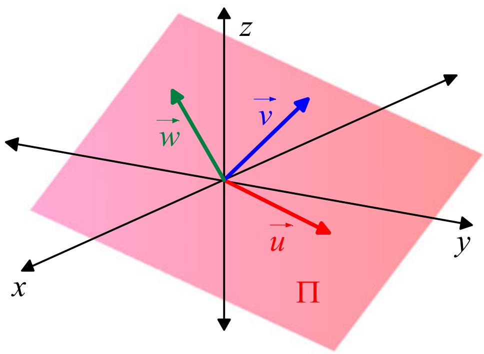
Without loss of generality, suppose $c_3 \neq 0$. Then, we get:

$$\vec{w} = -\frac{c_1}{c_3}\vec{u} - \frac{c_2}{c_3}\vec{v}.$$

But this means that $\vec{w} \in Span(\{\vec{u}, \vec{v}\}) = \Pi$, which is a ***plane*** through the origin. Thus, the three vectors are ***coplanar***. Let us take this a little further: since no two vectors are parallel to each other, neither coefficient $-c_1/c_3$ nor $-c_2/c_3$ can be zero. This means that $\vec{u} \in Span(\{\vec{v}, \vec{w}\})$ and similarly, $\vec{v} \in Span(\{\vec{u}, \vec{w}\})$. Thus, any two of the three vectors will Span the same plane Π. Conversely, we see that if $\vec{w}$ is ***not*** a member of $Span(\{\vec{u}, \vec{v}\})$, then the set $\{\vec{u}, \vec{v}, \vec{w}\}$ is ***independent***.

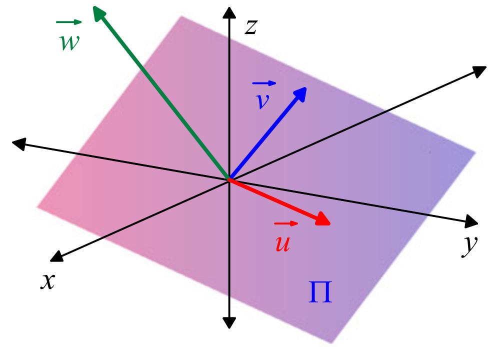
We note that if two of the vectors are ***parallel*** to each other, then $Span(S)$ would be a ***line*** instead of a plane, but this still means that the three vectors are ***coplanar***. We can thus summarize our observations concisely in the following:

Theorem: A set $S = \{\vec{u}, \vec{v}, \vec{w}\} \subset \mathbb{R}^m$ is ***dependent*** ***if and only if*** $\vec{u}$, $\vec{v}$ and $\vec{w}$ are ***coplanar***, that is, all three vectors are on at least one plane Π.



Three Dependent Vectors

where $\vec{w} \in Span(\{\vec{u}, \vec{v}\}) = \Pi$



Three Independent Vectors

where $\vec{w} \notin Span(\{\vec{u}, \vec{v}\}) = \Pi$

Example: Consider the set $S = \{\langle 1, -3, 2\rangle, \langle 2, 3, -1\rangle, \langle 1, 2, 4\rangle\}$.

It is clear that no two vectors are parallel to each other. Let us see if the third vector is a member of the Span of the first two. Thus, let us try to solve the equation:

$$\langle 1, 2, 4\rangle = c_1\langle 1, -3, 2\rangle + c_2\langle 2, 3, -1\rangle.$$

Comparing the first two components, we must have:

$$c_1 + 2c_2 = 1 \quad \text{and} \quad -3c_1 + 3c_2 = 2.$$

Using the Addition Method from basic algebra, we multiply the 1st equation by 3 and add to the 2nd equation to get $9c_2 = 5$. Thus $c_2 = 5/9$ and $c_1 = 1 - 2c_2 = -1/9$.

However, checking the third component, we get:

$$2c_1 - c_2 = -2/9 - 5/9 = -7/9 \neq 4.$$

Thus the three vectors are ***not coplanar***, and the set S is ***independent***.$\square$

The last few scenarios suggest the following generalization. We leave its proof as an Exercise.

Theorem: Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a set of non-zero vectors from some $\mathbb{R}^m$, and S contains at least two vectors. Then: S is linearly ***dependent*** ***if and only if*** at least one vector $\vec{v}_i$ from S can be expressed as a ***linear combination*** of the other vectors in S.

Clearly, if we have a set with four or more vectors, the possibilities get too complicated to list, so we will stop with sets consisting of three vectors. However, the following Theorem says that if a set is "too big," it is automatically dependent:

Theorem: A set $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ of n vectors from $\mathbb{R}^m$ is automatically linearly ***dependent*** if $n > m$.

Proof: Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subseteq \mathbb{R}^m$, with $n > m$. Assembling the vectors into the columns of an $m \times n$ matrix, we obtain a homogeneous system with more variables than equations, making it an ***underdetermined*** homogeneous system. Therefore, the system has non-trivial solutions. Thus, S is dependent.■

Example: The set:

$$S = \{\langle 5, -3, 2\rangle, \langle 2, 3, -8\rangle, \langle 1, 2, 4\rangle, \langle -5, 1, 6\rangle\}$$

consists of 4 vectors from $\mathbb{R}^3$, and therefore S is automatically ***dependent***.□

1.5 Section Summary

A linear system is called ***consistent*** if it has at least one solution. A system is called ***inconsistent*** if it does not have any solutions.

Let $\vec{b} \in \mathbb{R}^m$ and let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$. Then $\vec{b} \in Span(S)$ ***if and only if*** the system of equations corresponding to the augmented matrix $A = \left[\begin{array}{cccc|c} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n & \vec{b} \end{array}\right]$ is ***consistent***.

A linear system with m equations in n variables is called (1) ***underdetermined*** if $m < n$, (2) ***square*** if $m = n$, and (3) ***overdetermined*** if $m > n$.

A ***homogeneous*** system of m equations in n unknowns is a system of linear equations where the right side of the equations consists entirely of ***zeros***. In other words, the augmented matrix has the form $\left[A \mid \vec{\mathbf{0}}_m\right]$, where A is an $m \times n$ matrix.

A homogeneous system has an ***infinite*** number of solutions ***if and only if*** the reduced row echelon form of the coefficient matrix A has ***free variables***.

An ***underdetermined homogeneous*** system always has an ***infinite*** number of solutions, i.e., a homogeneous system with more unknowns than equations has an infinite number of solutions.

We define a ***matrix product*** of an $m \times n$ matrix A with an $n \times 1$ matrix $\vec{x}$ by the linear combination:

$$A\vec{x} = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1\vec{c}_1 + x_2\vec{c}_2 + \cdots + x_n\vec{c}_n.$$

Matrix multiplication possesses the following properties: for all $m \times n$ matrices A, for all $\vec{x}, \vec{y} \in \mathbb{R}^n$, and for all $k \in \mathbb{R}$:

The Additivity Property: $A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$, and ***The Homogeneity Property:*** $A(k\vec{x}) = k(A\vec{x})$.

The set $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$ is ***linearly dependent*** if we can find a ***non-trivial*** solution to the equation $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{\mathbf{0}}_m$. This is called the ***dependence test equation*** for S.

If only the ***trivial*** solution $x_1 = 0,\ x_2 = 0, \ldots,\ x_n = 0$ exists, we say that S is ***linearly independent***.

Any set S containing $\vec{\mathbf{0}}_m$ is a ***dependent*** set. A set $S = \{\vec{v}\}$ consisting of a single ***non-zero*** vector is ***independent***. A set $S = \{\vec{u}, \vec{v}\}$ consisting of ***two*** vectors from $\mathbb{R}^m$ is ***dependent*** ***if and only if*** $\vec{u}$ and $\vec{v}$ are ***parallel*** to each other. A set $S = \{\vec{u}, \vec{v}, \vec{w}\}$ consisting of ***three*** vectors from $\mathbb{R}^m$ is ***dependent*** ***if and only if*** $\vec{u}$, $\vec{v}$ and $\vec{w}$ are ***coplanar***. A set of n vectors from $\mathbb{R}^m$ is linearly ***dependent*** if $n > m$.

1.5 Exercises

For Exercises (1) to (10): Find the rref of each system, then classify the system according to being (a) consistent or inconsistent; and (b) underdetermined, overdetermined or square.

1. $$\begin{aligned} x + 2y &= 4 \\ 3x - 7y &= -3 \end{aligned}$$

2. $$\begin{aligned} x + 2y &= -7 \\ 2x - 5y &= 31 \\ -3x + 4y &= -29 \end{aligned}$$

3. $$\begin{aligned} x + 2y &= -7 \\ 2x - 5y &= 31 \\ x - y &= 12 \end{aligned}$$

4. $$\begin{aligned} x + 2y - 5z &= 4 \\ 3x - 7y + 2z &= -3 \end{aligned}$$

5. $$\begin{aligned} x + 2y - 5z &= 4 \\ 3x + 6y - 15z &= -12 \end{aligned}$$

6. $$\begin{aligned} x + 2y - 5z &= 4 \\ 3x + y - z &= -2 \\ -2x + y - 4z &= 6 \end{aligned}$$

7. $$\begin{aligned} x + 2y - 5z &= 4 \\ 3x + y - z &= -2 \\ -2x + 5y + 3z &= 7 \end{aligned}$$

8. The system in Section 1.4, Exercise 60.
9. The system in Section 1.4, Exercise 61.
10. The system in Section 1.4, Exercise 62.

For Exercises (11) to (17): In the Exercises of Section 1.4, you determined whether or not a vector $\vec{b}$ was a member of $Span(S)$. Based on the computations you already performed in these Exercises (and without any further work), determine if the set S that appears in the indicated Exercise from Section 1.4 is dependent or independent. Use the Answer Key for Section 1.4.

11. Exercise 39.
12. Exercise 42.
13. Exercise 43.
14. Exercise 45.

15. Exercise 46.

16. Exercise 47.

17. Exercise 48.

For Exercises (18) to (24): Determine whether the following sets of vectors are dependent or independent by assembling the vectors as the ***columns*** of a matrix A and finding the rref of A. If the set is dependent, give an example of a dependence equation with ***integer*** coefficients relating the vectors in the set.

18. $S = \{\langle -4, 7, -3\rangle, \langle -5, 16, -11\rangle, \langle 3, 2, -5\rangle\}$

19. $S = \{\langle 0, -6, 9\rangle, \langle 3, 6, 2\rangle, \langle 1, 4, 7\rangle\}$

20. $S = \{\langle 1, 9, 2, -7\rangle, \langle 3, -5, 2, 1\rangle, \langle 5, 7, 4, -2\rangle\}$

21. $S = \{\langle 5, 3, -6, -2\rangle, \langle -5, -4, 13, 16\rangle, \langle -3, -2, 5, 4\rangle\}$

22. $S = \{\langle 5, 3, -6, -2\rangle, \langle -3, -2, 5, 2\rangle, \langle 1, -2, 11, 6\rangle, \langle -5, -4, 1, 6\rangle\}$

23. $S = \{\langle 2, 3, -1, 4, 5\rangle, \langle 9, 1, 23, 18, -20\rangle, \langle 3, 2, 4, 6, -1\rangle\}$

24. $S = \{\langle 2, 3, -1, 4, 5\rangle, \langle 3, 2, 0, -2, -1\rangle, \langle 1, 4, -2, 1, 3\rangle, \langle 9, -4, 6, -2, -5\rangle\}$

For Exercises (25) to (29): The following sets $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$ are all ***dependent***. As before, assemble the vectors as the columns of a matrix A and find the rref of A. Unlike the sets in Exercises 18 to 24, each associated rref has more than one free variable. Find a dependence equation with the required conditions in each sub-item, again with only ***integer*** coefficients.

25. $S = \{\langle 5, 3, -6, -2\rangle, \langle -3, -2, 5, 2\rangle, \langle 1, -2, 11, 6\rangle, \langle -5, -4, 1, 6\rangle\}$.
    a. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$.
    b. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_4$.
    c. Find a dependence equation involving only $\vec{v}_2$, $\vec{v}_3$ and $\vec{v}_4$. Hint: use (a) to eliminate $\vec{v}_1$ from (b).

26. $S = \{\langle 2, 3, -1, 4, 5\rangle, \langle 3, 2, 4, -2, -1\rangle, \langle 1, 4, -6, 10, 11\rangle, \langle 1, 9, -17, 26, 28\rangle\}$.
    a. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$.
    b. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_4$.
    c. Find a dependence equation involving only $\vec{v}_2$, $\vec{v}_3$ and $\vec{v}_4$. Hint: use (a) to eliminate $\vec{v}_1$ from (b).

27. $S = \{\langle 5, 3, -4, -2\rangle, \langle -3, -2, 4, 2\rangle, \langle 1, -1, 2, 2\rangle, \langle -6, 5, -4, -8\rangle, \langle -4, 5, -6, -8\rangle\}$.
    a. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$ and $\vec{v}_4$.
    b. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$ and $\vec{v}_5$.
    c. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_3$, $\vec{v}_4$ and $\vec{v}_5$.

28. $S = \{\langle 5, 3, -4, -2\rangle, \langle -3, -2, 4, 2\rangle, \langle 5, 2, 4, 2\rangle, \langle 1, -1, 2, 2\rangle, \langle 4, 3, -3, -2\rangle\}$.
    a. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$.
    b. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_4$ and $\vec{v}_5$.
    c. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_3$, $\vec{v}_4$ and $\vec{v}_5$.

29. $S = \{\langle 0,-2, 6, 4, 2\rangle, \langle 0, 5,-15,-10,-5\rangle, \langle 3,-2, 4, 5, 0\rangle, \langle -12, 3,-1,-10, 5\rangle, \langle 9,-1,-3, 5,-5\rangle\}$.

   a. Find a dependence equation involving only $\vec{v}_1$ and $\vec{v}_2$.

   b. Find a dependence equation involving only $\vec{v}_1$, $\vec{v}_3$ and $\vec{v}_5$.

   c. Find a dependence equation involving only $\vec{v}_3$, $\vec{v}_4$ and $\vec{v}_5$.

30. Without doing ***any*** computations whatsoever, decide whether or not the set:

$$S = \{\langle 1,-6, 4, 2\rangle, \langle -3, 1, 6, 2\rangle, \langle 3, 8, 5,-9\rangle, \langle 1, 0, 4, 7\rangle, \langle 1, 6,-7, 3\rangle\}$$

is linearly independent, and explain how you got your conclusion.

31. Suppose that $\vec{u}$, $\vec{v}$ and $\vec{w}$ are ***any*** three vectors in some $\mathbb{R}^{n}$. Show that the set of three vectors:

$$S = \{2\vec{u} + \vec{v}, 4\vec{u} + 5\vec{v} - 4\vec{w}, \vec{u} - \vec{v} + 2\vec{w}\}$$

will always be a ***dependent*** set. Hint: find a dependence equation involving these three new vectors.

32. For what value/s of r and s, if any, will the following system have: (1) no solution, (2) exactly one solution, (3) an infinite number of solutions?

$$\begin{array}{rcrcrcr} x & + & y & - & z & = & 8 \\ x & - & y & + & 3z & = & -5 \\ -2x & + & y & + & rz & = & s \end{array}$$

33. For what value/s of r, s and t, if any, will the following system have: (1) no solution, (2) exactly one solution, (3) an infinite number of solutions involving exactly one free variable, and (4) an infinite number of solutions involving exactly two free variables?

$$\begin{array}{rcrcrcr} x & - & 2y & + & 4z & = & -2 \\ 3x & + & ry & + & 2z & = & 7 \\ -2x & + & 4y & + & sz & = & t \end{array}$$

34. What value of c will make the set $S = \{\langle 3, 5,-2\rangle, \langle 3, c,-25\rangle, \langle 1,-4, 7\rangle\}$ dependent?

35. Suppose that $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\} \subseteq \mathbb{R}^{m}$ is linearly independent.

   a. Prove that $S' = \{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ is also linearly independent.

   b. Prove that $S'' = \{\vec{v}_2, \vec{v}_4\}$ is also linearly independent.

36. We will generalize the previous Exercise: Prove that if $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subseteq \mathbb{R}^{m}$ is linearly independent, then ***any*** subset of S is still linearly independent.
   What is the contrapositive of this statement?

37. Give an example of an overdetermined homogeneous system which only has the trivial solution.

38. Give an example of an overdetermined homogeneous system with an infinite number of solutions.

39. Prove that if A is an $m \times n$ matrix, $\vec{x}$ and $\vec{y}$ are $n \times 1$ matrices, and $k \in \mathbb{R}$, then: (a) $A(\vec{x} + \vec{y}) = A\vec{x} + A\vec{y}$, and (b) $A(k\vec{x}) = k(A\vec{x})$.

40. Put together two of the Theorems in this section to prove: A set $S = \{\vec{v}\}$ from $\mathbb{R}^{m}$ consisting of exactly one vector is dependent ***if and only if*** $\vec{v} = \vec{\mathbf{0}}_{m}$.

41. Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a set of non-zero vectors from some $\mathbb{R}^m$, and S contains at least two vectors. Prove that S is linearly dependent ***if and only if*** at least one vector $\vec{v}_i$ from S can be expressed as a linear combination of the other vectors in S.

42. Prove that the ***ordered*** list of non-zero vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ is ***linearly independent*** ***if and only if*** no vector $\vec{v}_i$ can be written as a ***linear combination*** of $\vec{v}_1$ up to $\vec{v}_{i-1}$, that is, using only the vectors ***preceding*** $\vec{v}_i$. How is this Theorem different from that in the previous Exercise?

43. ***The Extension Theorem:*** Suppose that $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subseteq \mathbb{R}^n$ is a ***linearly independent*** set, and $\vec{v} \in \mathbb{R}^n$ is any vector which is ***not*** in Span(S). Prove that the bigger set $S' = S \cup \{\vec{v}\}$ is still ***linearly independent***. Hint: Form the dependence test equation for S' and divide your analysis into two Cases, depending on whether the coefficient of $\vec{v}$ is 0 or ***not*** 0.

44. In this Exercise, we will guide you to prove that if A is an $m \times n$ matrix, $\vec{b}$ is an $m \times 1$ matrix, and we can find ***at least two distinct solutions*** to the system $A\vec{x} = \vec{b}$, then we can find an ***infinite*** number of solutions to this system.

    a. First, show that if $\vec{x}$ and $\vec{y}$ are two such ***distinct*** solutions to $A\vec{x} = \vec{b}$, then $\vec{x} - \vec{y}$ is a solution of the ***homogenous*** system $A\vec{x} = \vec{\mathbf{0}}_m$.

    b. Next, we will use scalar multiples of $\vec{x} - \vec{y}$ to construct an infinite number of solutions to the homogeneous system in (a). Explain why $t_1(\vec{x} - \vec{y}) \neq t_2(\vec{x} - \vec{y})$ if $t_1 \neq t_2$. You may want to use the Zero Factors Theorem. Explain why this means that the set $\{t(\vec{x} - \vec{y}) \,|\, t \in \mathbb{R}\}$ is an infinite set.

    c. Use $\vec{x}$ and part (b) to construct an infinite number of solutions to the original system $A\vec{x} = \vec{b}$, and show that the vectors you constructed are indeed solutions to this system.

    d. As a bonus, prove that any linear system $A\vec{x} = \vec{b}$ either has: (1) no solutions, (2) exactly one solution, or (3) an infinite number of solutions. Hint: Use a Case-by-Case Analysis, but be careful how you begin the 3rd Case.

45. Suppose that Π_1 and Π_2 are ***parallel planes*** in $\mathbb{R}^3$, given by:

$$\Pi_1 : a_1x + b_1y + c_1z = d_1, \quad \text{and} \quad \Pi_2 : a_1x + b_1y + c_1z = d_2,$$

where $d_1 \neq d_2$. Notice that we can use the same normal vector $\vec{n}_1 = \langle a_1, b_1, c_1 \rangle = \vec{n}_2$ for both planes. Now, suppose that Π_3 is another plane given by:

$$\Pi_3 : a_3x + b_3y + c_3z = d_3,$$

and $\vec{n}_3 = \langle a_3, b_3, c_3 \rangle$ is ***not parallel*** to $\vec{n}_1$. Show that the line of intersection between Π_1 and Π_3 is ***parallel*** to the line of intersection of Π_2 and Π_3. Consult Section 1.2 for the definition of parallel lines in $\mathbb{R}^3$.

46. ***True or False:*** Determine whether each statement is true or false, and briefly explain your answer by either applying a Theorem or providing a counterexample or a convincing argument.

    a. A consistent square system has exactly one solution.
    b. An underdetermined linear system has an infinite number of solutions.
    c. An underdetermined homogeneous linear system has an infinite number of solutions.
    d. A consistent system of 4 equations and 7 variables will have exactly 3 free variables.
    e. A consistent system of 4 equations and 7 variables will have at least 3 free variables.
    f. A consistent system of 4 equations and 7 variables could have exactly 2 free variables.
    g. A homogeneous system of 7 equations in 10 unknowns has non-trivial solutions.
    h. A homogeneous system of 10 equations in 7 unknowns only has the trivial solution.
    i. A set of 8 non-zero vectors from $\mathbb{R}^5$ is always linearly dependent.
    j. A set of 5 non-zero vectors from $\mathbb{R}^8$ is always linearly independent.

1.6 Independent Sets versus Spanning Sets

We have spent much time studying and describing the Span of a set of vectors S, and likewise, determining if S is linearly ***dependent*** or ***independent***. We will now see that there are ***strong relationships*** between Spanning sets and linearly dependent or independent sets. Every Theorem that we will see in this Section (except two) will involve ***both*** Spanning and dependence conditions. Let us start with the lone exceptions:

Equality of Spans

If S is a set containing a non-zero vector, the Span of S contains an infinite number of vectors. So it is very possible to choose a different set of vectors, say S', so that $Span(S')$ produces exactly the same set of vectors as $Span(S)$. Here is a very easy way to do this:

Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subseteq \mathbb{R}^{\boldsymbol{m}}$, and $k_1,\ k_2, \ldots,\ k_n \in \mathbb{R}$ a list of n ***non-zero scalars.*** Let us form a new set: $S' = \{k_1\vec{v}_1, k_2\vec{v}_2, \ldots, k_n\vec{v}_n\}$. Then: $Span(S) = Span(S')$.

Proof: We must show that every linear combination of vectors from S also looks like a linear combination of vectors from S', and vice versa. Starting with a linear combination of vectors from S, we get:

$$\begin{aligned} c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n &= c_1\frac{k_1}{k_1}\vec{v}_1 + c_2\frac{k_2}{k_2}\vec{v}_2 + \cdots + c_n\frac{k_n}{k_n}\vec{v}_n \\ &= \frac{c_1}{k_1}(k_1\vec{v}_1) + \frac{c_2}{k_2}(k_2\vec{v}_2) + \cdots + \frac{c_n}{k_n}(k_n\vec{v}_n), \end{aligned}$$

which is a linear combination of vectors from S'. This is possible, since every non-zero k_i has a ***reciprocal***. Thus, every member of $Span(S)$ is also a member of $Span(S')$. Similarly, a linear combination of $k_1\vec{v}_1,\ k_2\vec{v}_2,\ \ldots,\ k_n\vec{v}_n$ has the form:

$$c_1(k_1\vec{v}_1) + c_2(k_2\vec{v}_2) + \cdots + c_n(k_n\vec{v}_n) = (c_1k_1)\vec{v}_1 + (c_2k_2)\vec{v}_2 + \cdots + (c_nk_n)\vec{v}_n,$$

so every member of $Span(S')$ is also a member of $Span(S)$. Thus, the two Spans are the same. ■

Example: Let $S = \left\{\langle 15,-35, 10,-30, 25\rangle, \left\langle -\frac{16}{3}, 12, 4, -\frac{8}{3}, \frac{4}{3}\right\rangle, \left\langle -\frac{35}{4}, -14, -\frac{63}{4}, 7, -\frac{21}{2}\right\rangle\right\}$.

Notice that we can factor out 5 from the 1st vector, 4/3 from the 2nd, and –7/4 from the 3rd. If we do so, we obtain a new set:

$$S' = \{\langle 3,-7, 2,-6, 5\rangle, \langle -4, 9, 3,-2, 1\rangle, \langle 5, 8, 9,-4, 6\rangle\},$$

and by our Theorem above, $Span(S) = Span(S')$. In general, if all the vectors in S have ***rational*** components, we can find a set of vectors S' with ***integer*** components with the same Span as S. □

The previous Theorem is a very easy case because the pairs of vectors are ***parallel*** to each other. We can ask ***in general***, though, when are the Spans of two sets of vectors S and S' exactly the same, even when the vectors do not look obviously related to each other? Here is the complete answer:

Theorem — The Equality of Spans Theorem:
Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ and $S' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be two sets of vectors from some Euclidean space $\mathbb{R}^k$, where n, m, and k are ***any*** positive integers. Then: $Span(S) = Span(S')$ ***if and only if*** every $\vec{v}_i$ can be written as a ***linear combination*** of the $\vec{w}_1$ through $\vec{w}_m$, ***and*** every $\vec{w}_j$ can also be written as a ***linear combination*** of the $\vec{v}_1$ through $\vec{v}_n$.

Proof: ($\Rightarrow$) Suppose that $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}) = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\})$. We must show that every $\vec{v}_i$ can be written as a linear combination of the $\vec{w}_1$ through $\vec{w}_m$, ***and*** every $\vec{w}_j$ can also be written as a linear combination of the $\vec{v}_1$ through $\vec{v}_n$. But since $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$ includes $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ themselves, and the two Spans are equal, this means that each $\vec{v}_i$ is indeed a linear combination of the $\vec{w}_1$ through $\vec{w}_m$. Exactly the same reasoning also applies to the vectors $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m$.

($\Leftarrow$) Now, suppose that every $\vec{v}_i$ can be written as a linear combination of the $\vec{w}_1$ through $\vec{w}_m$, ***and*** every $\vec{w}_j$ can also be written as a linear combination of the $\vec{v}_1$ through $\vec{v}_n$. We must show that $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}) = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\})$. Since these are two sets, we must show that the first Span is a subset of the second Span, and vice versa. Any member of the first Span looks like:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n.$$

However, we are told that each $\vec{v}_i$ can be written in terms of the $\vec{w}_j s$. We will use what is called ***double-index notation*** to express these linear combinations:

$$\begin{aligned}
\vec{v}_1 &= a_{1,1}\vec{w}_1 + a_{1,2}\vec{w}_2 + \cdots + a_{1,m}\vec{w}_m,\\
\vec{v}_2 &= a_{2,1}\vec{w}_1 + a_{2,2}\vec{w}_2 + \cdots + a_{2,m}\vec{w}_m, \ldots\ldots\\
\vec{v}_n &= a_{n,1}\vec{w}_1 + a_{n,2}\vec{w}_2 + \cdots + a_{n,m}\vec{w}_m,
\end{aligned}$$

for some real numbers $a_{i,j}$. We will ***substitute*** these expressions into our original linear combination above to get:

$$\begin{aligned}
c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n &= c_1(a_{1,1}\vec{w}_1 + a_{1,2}\vec{w}_2 + \cdots + a_{1.m}\vec{w}_m) +\\
&\quad c_2(a_{2,1}\vec{w}_1 + a_{2,2}\vec{w}_2 + \cdots + a_{2.m}\vec{w}_m) + \cdots +\\
&\quad c_n(a_{n,1}\vec{w}_1 + a_{n,2}\vec{w}_2 + \cdots + a_{n.m}\vec{w}_m).
\end{aligned}$$

Upon distributing and collecting like terms (which we conveniently see in ***columns*** above), we see that this we will obtain a linear combination of the $\vec{w}_1$ through $\vec{w}_m$:

$$\begin{aligned}
c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n &= c_1a_{1,1}\vec{w}_1 + c_1a_{1,2}\vec{w}_2 + \cdots + c_1a_{1.m}\vec{w}_m +\\
&\quad c_2a_{2,1}\vec{w}_1 + c_2a_{2,2}\vec{w}_2 + \cdots + c_2a_{2.m}\vec{w}_m + \cdots +\\
&\quad c_na_{n,1}\vec{w}_1 + c_na_{n,2}\vec{w}_2 + \cdots + c_na_{n.m}\vec{w}_m\\
&= (c_1a_{1,1} + c_2a_{2,1} + \cdots + c_na_{n,1})\vec{w}_1 +\\
&\quad (c_1a_{1,2} + c_2a_{2,2} + \cdots + c_na_{n,2})\vec{w}_2 + \cdots +\\
&\quad (c_1a_{1,m} + c_2a_{2,m} + \cdots + c_na_{n,m})\vec{w}_m.
\end{aligned}$$

Thus, a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$ is also a member of $Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\})$. A similar argument shows that a member of $Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\})$ is also a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$, and will be left as an Exercise.■

Unfortunately, this Theorem says that to check if two Spans are equal, we need to solve for $n+m$ dependence test equations. This is a huge task, especially if we have large sets of vectors with no obvious dependence relationships among them. Let us demonstrate this Theorem with just a small number of vectors:

Example: Let us show that:

$$Span(\{\langle 3,-2,5\rangle, \langle 1,7,-4\rangle\}) = Span(\langle 10,1,11\rangle, \langle -6,4,-10\rangle, \langle 17,4,17\rangle).$$

We need to show that every vector in the second set is a linear combination of the vectors in the first set, and vice versa. Let us show it for the first element:

$$\langle 10,1,11\rangle = x\langle 3,-2,5\rangle + y\langle 1,7,-4\rangle,$$

so we must have, using only the first two coordinates, that:

$$3x+y = 10 \quad \text{and} \quad -2x+7y = 1.$$

Solving this system of equations, we get $x = 3$ and $y = 1$. We check that indeed:

$$\langle 10,1,11\rangle = 3\langle 3,-2,5\rangle + \langle 1,7,-4\rangle.$$

Repeating this process for the other two vectors, we also get:

$$\langle -6,4,-10\rangle = -2\langle 3,-2,5\rangle, \text{ and } \langle 17,4,17\rangle = 5\langle 3,-2,5\rangle + 2\langle 1,7,-4\rangle.$$

Now for the other set:

$$\langle 3,-2,5\rangle = -\frac{1}{2}\langle -6,4,-10\rangle, \text{ and } \langle 1,7,-4\rangle = \frac{1}{2}\langle 17,4,17\rangle + \frac{5}{4}\langle -6,4,-10\rangle.$$

Thus, the two Spans are the same.□

Now, as promised, let us look at some of the deeper connections between the Span of a set of vectors and linearly independent sets of vectors.

The Elimination Theorem

We saw in the previous Section that the set $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \{\langle 3,7,-2\rangle, \langle -1,5,-4\rangle, \langle 15,13,4\rangle\}$ is linearly dependent. In fact, we saw that:

$$\vec{v}_3 = 4\vec{v}_1 - 3\vec{v}_2.$$

This means that we can simplify a linear combination of the three vectors as follows:

$$\begin{aligned} c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 &= c_1\vec{v}_1 + c_2\vec{v}_2 + c_3(4\vec{v}_1 - 3\vec{v}_2) \\ &= c_1\vec{v}_1 + c_2\vec{v}_2 + 4c_3\vec{v}_1 - 3c_3\vec{v}_2 \\ &= (c_1+4c_3)\vec{v}_1 + (c_2-3c_3)\vec{v}_2, \end{aligned}$$

or in other words, we can express any member of Span(S) exclusively as a linear combination of $\vec{v}_1$ and $\vec{v}_2$ ***only***. Similarly, we can express $\vec{v}_2$ in terms of $\vec{v}_1$ and $\vec{v}_3$, and $\vec{v}_1$ in terms of $\vec{v}_2$ and $\vec{v}_3$. Consequently:

$$Span(S) = Span(\{\vec{v}_1, \vec{v}_2\}) = Span(\{\vec{v}_2, \vec{v}_3\}) = Span(\{\vec{v}_1, \vec{v}_3\}).$$

In general, we have the following:

Theorem — The Elimination Theorem:

Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a ***linearly dependent*** set of vectors from $\mathbb{R}^{\boldsymbol{m}}$, and $\vec{v}_n = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_{n-1}\vec{v}_{n-1}$. Then:

$$Span(S) = Span(S - \{\vec{v}_n\}).$$

In other words, we can ***eliminate*** $\vec{v}_n$ from S and still maintain the ***same Span***.

More generally, if $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{\mathbf{0}}_{\boldsymbol{m}}$, where ***none*** of the coefficients in this dependence equation is 0, then:

$$Span(S) = Span(S - \{\vec{v}_i\}),$$

for all $i = 1..n$. Thus, we can ***eliminate*** any vector from S and maintain the ***same*** Span.

We leave the proof as an Exercise. Recall from Chapter Zero that $S - \{\vec{v}_i\}$ means the set S with $\vec{v}_i$ ***removed***. This Theorem says that if we start with a set of dependent vectors, we can remove one vector from this set, and the Span of the remaining vectors is identical to the original Span. Consequently, if this smaller set is ***still*** dependent, we can remove another vector, and so on, removing one vector at a time from the set until we can no longer find a dependence equation for the remaining vectors. At this point, the set is ***independent***, and has the ***same Span*** as the original set.

Example: Let $S = \{\langle 3, 3, 5, 2, 4\rangle, \langle 3, 2, 0, -1, 6\rangle, \langle 12, 12, 20, 8, 16\rangle, \langle 0, 1, 5, 3, -2\rangle\}$, and let us call these vectors $\vec{v}_1$, $\vec{v}_2$, $\vec{v}_3$, and $\vec{v}_4$, in that order.

If you stare at these four vectors long enough, you will probably see the dependency relations. It is obvious, though, that $\vec{v}_1$ is not parallel to $\vec{v}_2$. However, $\vec{v}_3 = 4\vec{v}_1$, so we can Eliminate $\vec{v}_3$. Next, $\vec{v}_4 = \vec{v}_1 - \vec{v}_2$, so we can also Eliminate $\vec{v}_4$. Thus:

$$Span(\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}) = Span(\{\vec{v}_1, \vec{v}_2\}). \square$$

Obviously, this process is inefficient if there are more than a handful of vectors, or if the components are large, with no obvious relationships. The next Theorem gives us a more efficient way:

The Minimizing Theorem

The Gauss-Jordan Algorithm will now enable us to express the Span of a set of vectors with as few vectors as possible:

Theorem — The Minimizing Theorem:

Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors from $\mathbb{R}^{\boldsymbol{m}}$, and let $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix}$ be the $m \times n$ matrix with $\vec{v}_1$, $\vec{v}_2$, $\ldots$, $\vec{v}_n$ as its ***columns***.

Suppose that R is the rref of A, and i_1, i_2, $\ldots$, i_k are the columns of R that contain the ***leading variables***. Then the set $S' = \{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}$, that is, the subset of vectors of S consisting of the corresponding columns of A, is a ***linearly independent*** set, and:

$$Span(S) = Span(S').$$

Furthermore, ***every*** $\vec{v}_i \in S - S'$, that is, the vectors of S corresponding to the ***free variables*** of R, can be expressed as a ***linear combination*** of the vectors of S', using the ***coefficients*** found in the corresponding column of R.

We call this The Minimizing Theorem because we immediately go from a Spanning set S to a linearly independent subset S' that has the ***same Span*** as S. We have mentioned that linearly independent subsets are as small as possible, so in a sense, we have reduced the size of S until we can reduce it no further — that is, S' is as ***efficient*** as possible to describe Span(S).

Proof of the Theorem: The Gauss-Jordan Algorithm does not change the solution set of a homogeneous system, so $\vec{x}$ is a solution to $A\vec{x} = \vec{\mathbf{0}}_m$ ***if and only if*** $\vec{x}$ is a solution to $R\vec{x} = \vec{\mathbf{0}}_m$, where R is the rref of A. Thus, the components of $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle$ give us a dependence equation for the columns of A ***if and only if*** they give us a dependence equation for the columns of R:

$$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{\mathbf{0}}_m \textbf{\textit{ if and only if}}$$
$$x_1\vec{c}_1 + x_2\vec{c}_2 + \cdots + x_n\vec{c}_n = \vec{\mathbf{0}}_m \text{ also,}$$

where $\vec{c}_1$ through $\vec{c}_n$ are the columns of R. The same reasoning works for just a ***subset*** of the columns of A and the corresponding columns of R, by setting some of the components of $\vec{x}$ to 0. Thus, a ***subset*** of the columns of A is linearly independent ***if and only if*** the corresponding columns of R also form a linearly independent set.

Notice that the columns of R that contain the leading 1's are precisely $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_k\}$ for some $k \leq m$. For example, the rref of A could be:

$$R = \begin{bmatrix} 1 & a & b & 0 & d & 0 & f \\ 0 & 0 & 0 & 1 & g & 0 & h \\ 0 & 0 & 0 & 0 & 0 & 1 & k \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \left[\boxed{\begin{matrix} 1 \\ 0 \\ 0 \\ 0 \end{matrix}} \; \begin{matrix} a \\ 0 \\ 0 \\ 0 \end{matrix} \; \begin{matrix} b \\ 0 \\ 0 \\ 0 \end{matrix} \; \boxed{\begin{matrix} 0 \\ 1 \\ 0 \\ 0 \end{matrix}} \; \begin{matrix} d \\ g \\ 0 \\ 0 \end{matrix} \; \boxed{\begin{matrix} 0 \\ 0 \\ 1 \\ 0 \end{matrix}} \; \begin{matrix} f \\ h \\ k \\ 0 \end{matrix} \right].$$

We see that $\vec{e}_1$, $\vec{e}_2$ and $\vec{e}_3$ are in columns 1, 4 and 6, which we boxed on the right. Thus, these three columns are linearly independent. By the reasoning above, the corresponding columns in A must also be linearly independent.

Now, suppose that column i of R does not contain a leading 1. By definition of the rref, every non-zero entry of column i has a leading 1 to its ***left***. Thus, we can express column i as a linear combination of the columns of R that contain only the leading 1's to the ***left*** of column i. Thus:

$$\vec{c}_2 = a \bullet \vec{e}_1 = a \bullet \vec{c}_1,$$

$$\vec{c}_3 = b \bullet \vec{e}_1 = b \bullet \vec{c}_1,$$

$$\vec{c}_5 = d \bullet \vec{e}_1 + g \bullet \vec{e}_2 = d \bullet \vec{c}_1 + g \bullet \vec{c}_4, \quad \text{and}$$

$$\vec{c}_7 = f \bullet \vec{e}_1 + h \bullet \vec{e}_2 + k \bullet \vec{e}_3 = f \bullet \vec{c}_1 + h \bullet \vec{c}_4 + k \bullet \vec{c}_6.$$

According to our reasoning above, columns 2, 3, 5 and 7 of the ***original*** matrix A will be linear combinations of columns 1, 4 and 6 of A, using the ***same coefficients:*** $\vec{v}_2 = a \bullet \vec{v}_1$, $\vec{v}_3 = b \bullet \vec{v}_1$, $\vec{v}_5 = d \bullet \vec{v}_1 + g \bullet \vec{v}_4$, and $\vec{v}_7 = f \bullet \vec{v}_1 + h \bullet \vec{v}_4 + k \bullet \vec{v}_6$.

By our reasoning in the first part of the proof, the same coefficients for these linear combinations will also enable us to express each column of A that does not correspond to a leading 1 as a linear combination of the columns of A that correspond to the leading 1's. ■

Example: Suppose that:

$$S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} = \{\langle 1,-2,-3,4\rangle, \langle 2,-1,-1,2\rangle, \langle 1,4,7,-8\rangle, \langle 3,-3,-2,5\rangle, \langle 5,-4,7,2\rangle\}.$$

This is a set of 5 vectors from $\mathbb{R}^4$, so it is certainly ***dependent***, but it is not obvious just how many vectors we will need to eliminate. We form the matrix A using the vectors as ***columns:***

$$A = \begin{bmatrix} 1 & 2 & 1 & 3 & 5 \\ -2 & -1 & 4 & -3 & -4 \\ -3 & -1 & 7 & -2 & 7 \\ 4 & 2 & -8 & 5 & 2 \end{bmatrix} \text{ with rref } R = \begin{bmatrix} 1 & 0 & -3 & 0 & -5 \\ 0 & 1 & 2 & 0 & -4 \\ 0 & 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The leading ones are in c_1, c_2 and c_4, so $Span(S) = Span(\{\vec{v}_1, \vec{v}_2, \vec{v}_4\})$**.** Notice that:

$$\vec{c}_3 = \begin{bmatrix} \mathbf{-3} \\ \mathbf{2} \\ 0 \\ 0 \end{bmatrix} = \mathbf{-3}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} + \mathbf{2}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{-3}\vec{c}_1 + \mathbf{2}\vec{c}_2 \quad \text{and}$$

$$\vec{c}_5 = \begin{bmatrix} \mathbf{-5} \\ \mathbf{-4} \\ \mathbf{6} \\ 0 \end{bmatrix} = \mathbf{-5}\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} - \mathbf{4}\begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} + \mathbf{6}\begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \mathbf{-5}\vec{c}_1 - \mathbf{4}\vec{c}_2 + \mathbf{6}\vec{c}_4.$$

Thus, we can express $\vec{v}_3$ and $\vec{v}_5$ as linear combinations of $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_4$ using exactly the same coefficients:

$$\vec{v}_3 = \begin{bmatrix} 1 \\ 4 \\ 7 \\ -8 \end{bmatrix} = \mathbf{-3}\begin{bmatrix} 1 \\ -2 \\ -3 \\ 4 \end{bmatrix} + \mathbf{2}\begin{bmatrix} 2 \\ -1 \\ -1 \\ 2 \end{bmatrix} = \mathbf{-3}\vec{v}_1 + \mathbf{2}\vec{v}_1 \quad \text{and}$$

$$\vec{v}_5 = \begin{bmatrix} 5 \\ -4 \\ 7 \\ 2 \end{bmatrix} = \mathbf{-5}\begin{bmatrix} 1 \\ -2 \\ -3 \\ 4 \end{bmatrix} - \mathbf{4}\begin{bmatrix} 2 \\ -1 \\ -1 \\ 2 \end{bmatrix} + \mathbf{6}\begin{bmatrix} 3 \\ -3 \\ -2 \\ 5 \end{bmatrix} = \mathbf{-5}\vec{v}_1 - \mathbf{4}\vec{v}_2 + \mathbf{6}\vec{v}_4. \square$$

The Size of Dependent Sets from Spanning Sets

We saw in the previous Section that if we have a set S of m vectors from $\mathbb{R}^n$, where $m > n$, then S is definitely ***dependent***. This is a particular instance of the following Theorem that allows us to compare the relative ***size*** of certain linearly dependent sets generated by a Spanning set S:

Theorem — The Dependent Sets from Spanning Sets Theorem:

Suppose we have a set of n vectors:

$$S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\},$$

from some Euclidean space $\mathbb{R}^k$, and we form $Span(S)$. Now, suppose we randomly choose a set of m vectors from $Span(S)$ to form a new set:

$$L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}.$$

We can now conclude that ***if*** $m > n$, ***then*** L is automatically linearly ***dependent***.

In other words, if we chose ***more*** vectors from $Span(S)$ than the number of vectors we used to ***generate*** the Span, then this new set will certainly be ***dependent***.

Proof: This will be the ***deepest*** Theorem of this Chapter, but the strategy behind the proof is simple: create an ***underdetermined homogeneous*** system of n equations in m variables.

Our goal is to show that L, as constructed above, is linearly ***dependent***. The members of $L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$ are vectors from $Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\})$, so we can write $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m$ as linear combinations of $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n$.

However, since we will be writing m different linear combinations, one for each $\vec{u}_i$, we will again use the ***double-index notation*** that we saw in the Equality of Spans Theorem. Let us write:

$$\begin{aligned}
\vec{u}_1 &= a_{1,1}\vec{w}_1 + a_{1,2}\vec{w}_2 + \cdots + a_{1,n}\vec{w}_n, \\
\vec{u}_2 &= a_{2,1}\vec{w}_1 + a_{2,2}\vec{w}_2 + \cdots + a_{2,n}\vec{w}_n, \ \ldots\ldots \\
\vec{u}_m &= a_{m,1}\vec{w}_1 + a_{m,2}\vec{w}_2 + \cdots + a_{m,n}\vec{w}_n.
\end{aligned}$$

Let us form the ***dependence test equation:*** $c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_m\vec{u}_m = \vec{\mathbf{0}}_k$.

We will show that there must be a ***non-trivial*** solution to this dependence test equation. As in the Equality of Spans Theorem, we will substitute the expressions that we wrote down above for $\vec{u}_1$ to $\vec{u}_m$ into the dependence test equation. We get:

$$\begin{aligned}
\vec{\mathbf{0}}_k &= c_1(a_{1,1}\vec{w}_1 + a_{1,2}\vec{w}_2 + \cdots + a_{1,n}\vec{w}_n) + \\
&\quad c_2(a_{2,1}\vec{w}_1 + a_{2,2}\vec{w}_2 + \cdots + a_{2,n}\vec{w}_n) + \cdots + \\
&\quad c_m(a_{m,1}\vec{w}_1 + a_{m,2}\vec{w}_2 + \cdots + a_{m,n}\vec{w}_n) \\
&= c_1a_{1,1}\vec{w}_1 + c_1a_{1,2}\vec{w}_2 + \cdots + c_1a_{1,n}\vec{w}_n + \\
&\quad c_2a_{2,1}\vec{w}_1 + c_2a_{2,2}\vec{w}_2 + \cdots + c_2a_{2,n}\vec{w}_n + \cdots + \\
&\quad c_ma_{m,1}\vec{w}_1 + c_ma_{m,2}\vec{w}_2 + \cdots + c_ma_{m,n}\vec{w}_n \\
&= (c_1a_{1,1} + c_2a_{2,1} + \cdots + c_ma_{m,1})\vec{w}_1 + \\
&\quad (c_1a_{1,2} + c_2a_{2,2} + \cdots + c_ma_{m,2})\vec{w}_2 + \cdots + \\
&\quad (c_1a_{1,n} + c_2a_{2,n} + \cdots + c_ma_{m,n})\vec{w}_n.
\end{aligned}$$

Now, we can ***force*** a solution to this equation if we set ***all*** of the coefficients of the vectors $\vec{w}_1$ through $\vec{w}_n$ to be zero, that is:

$$\begin{aligned}
c_1a_{1,1} + c_2a_{2,1} + \cdots + c_ma_{m,1} &= 0, \\
c_1a_{1,2} + c_2a_{2,2} + \cdots + c_ma_{m,2} &= 0, \ \ldots\ldots \ \text{and} \\
c_1a_{1,n} + c_2a_{2,n} + \cdots + c_ma_{m,n} &= 0.
\end{aligned}$$

The only thing we need to notice about this homogeneous system of equations is that there are ***n* equations** and ***m* unknowns** (the coefficients c_1 through c_m). But since $m > n$, we have more unknowns than equations, that is, we have an ***underdetermined homogeneous system***. Such a system would have a ***free variable*** and therefore have an ***infinite*** number of solutions. Thus, we can obtain ***non-trivial*** values for $c_1, c_2, \ldots, c_m$ that will satisfy our original dependence equation for $\vec{u}_1$ through $\vec{u}_m$. Thus $L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$ is a ***dependent*** set.■

The ***contrapositive*** of our Theorem is also worth stating:

Theorem — The Independent Sets from Spanning Sets Theorem:
Suppose we have a set of n vectors $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ from some Euclidean space $\mathbb{R}^{\boldsymbol{k}}$, and we form $Span(S)$.
Suppose now we randomly choose a set of m vectors from $Span(S)$ to form a new set:

$$L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}.$$

We can now conclude that ***if*** L is ***independent***, ***then*** $m \leq n$.

Example: A set of 10 vectors from $Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_8\})$ must be linearly dependent, since $10 > 8$. However, a set with 8 or fewer vectors from this same Span is not necessarily linearly independent (i.e., they could be all be parallel to each other).□

The Extension Theorem

One of the Exercises in the previous Section is of such importance that we mention it again here. It tells us how to ***extend*** a linearly independent set by one vector so that the new set is still linearly independent:

Theorem — The Extension Theorem:
Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***linearly independent*** set of vectors from $\mathbb{R}^{\boldsymbol{m}}$, and suppose $\vec{v}_{n+1}$ is ***not*** a member of $Span(S)$. Then, the ***extended*** set:

$$S' = S \cup \{\vec{v}_{n+1}\} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{v}_{n+1}\}$$

is ***still linearly independent***.

Proof: Let us construct the ***dependence test equation*** for the extended set:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n + c_{n+1}\vec{v}_{n+1} = \vec{\mathbf{0}}_{\boldsymbol{m}}.$$

We must show that we can only get the trivial solution: $c_1 = 0,\ c_2 = 0,\ \ldots\ c_{n+1} = 0$. At this point, let us break up the analysis into two cases:

Case 1. Suppose we can find a solution where $c_{n+1} = 0$. Then we get a (shorter) dependence equation:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{\mathbf{0}}_{\boldsymbol{m}}.$$

However, we know that S is linearly independent, so all the coefficients c_1 through c_n of this equation must be 0. Thus $c_1 = c_2 = \cdots = c_{n+1} = 0$, so S' is linearly independent.

Case 2. Suppose we can find a solution where $c_{n+1} \neq 0$. Then the ***dependence equation*** can be used to solve for $\vec{v}_{n+1}$:

$$\begin{aligned} c_{n+1}\vec{v}_{n+1} &= -c_1\vec{v}_1 - c_2\vec{v}_2 - \cdots - c_n\vec{v}_n, \text{ or} \\ \vec{v}_{n+1} &= -\frac{c_1}{c_{n+1}}\vec{v}_1 - \frac{c_2}{c_{n+1}}\vec{v}_2 - \cdots - \frac{c_n}{c_{n+1}}\vec{v}_n. \end{aligned}$$

But this equation implies that $\vec{v}_{n+1}$ is a member of $Span(S)$, and thus Case 2 leads to a ***contradiction***. Thus, only Case 1 is possible, so S' is linearly ***independent***. ■

Example: Let $S = \{\langle 2, 0, -7, 4\rangle, \langle 0, -3, 2, 5\rangle\} \subseteq \mathbb{R}^4$. The two vectors of S are clearly not parallel, so S is linearly independent. Let us extend S by one more vector, say $\vec{v}_3 = \langle 0, 0, 3, -5\rangle$. We need only check that $\vec{v}_3$ is not a linear combination of $\vec{v}_1$ and $\vec{v}_2$. Suppose:

$$\langle 0, 0, 3, -5\rangle = c_1\langle 2, 0, -7, 4\rangle + c_2\langle 0, -3, 2, 5\rangle.$$

The first two coordinates tell us that $0 = 2 \cdot c_1 + 0 \cdot c_2$, and $0 = 0 \cdot c_1 - 3 \cdot c_2$.

Thus $c_1 = 0 = c_2$. But these values will imply in the equation above that:

$$\langle 0, 0, 3, -5\rangle = 0 \cdot \langle 2, 0, -7, 4\rangle + 0 \cdot \langle 0, -3, 2, 5\rangle = \langle 0, 0, 0, 0\rangle,$$

which is clearly a contradiction. Thus $\langle 0, 0, 3, -5\rangle$ is not in $Span(S)$, so the extended set $S' = \{\langle 2, 0, -7, 4\rangle, \langle 0, -3, 2, 5\rangle, \langle 0, 0, 3, -5\rangle\}$ is still a linearly independent subset of $\mathbb{R}^4$. □

1.6 Section Summary

Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subseteq \mathbb{R}^m$, and $k_1, k_2, \ldots, k_n \in \mathbb{R}$ a list of n ***non-zero scalars.*** Let us form a new set: $S' = \{k_1\vec{v}_1, k_2\vec{v}_2, \ldots, k_n\vec{v}_n\}$. Then: $Span(S) = Span(S')$.

The Equality of Spans Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^k$ and $S' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\} \subset \mathbb{R}^k$, where n, m, and k are ***any*** positive integers. Then: $Span(S) = Span(S')$ ***if and only if*** every $\vec{v}_i$ can be written as a ***linear combination*** of the $\vec{w}_1$ through $\vec{w}_m$, ***and*** every $\vec{w}_j$ can also be written as a ***linear combination*** of the $\vec{v}_1$ through $\vec{v}_n$.

The Elimination Theorem: Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a ***linearly dependent*** set of vectors from $\mathbb{R}^m$, and $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{\mathbf{0}}_m$, where ***none*** of the coefficients in the dependence equation is 0. Then: $Span(S) = Span(S - \{\vec{v}_i\})$, for all $i = 1..n$. Thus, we can ***eliminate*** any of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ and still maintain the ***same*** Span.

The Minimizing Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$, and let $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix}$. Suppose that R is the rref of A, and $i_1, i_2, \ldots, i_k$ are the ***leading columns*** of R. Then the set $S' = \{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}$, that is, the subset of vectors of S consisting of the corresponding columns of A, is a ***linearly independent*** set, and $Span(S) = Span(S')$. Furthermore, ***every*** $\vec{v}_i \in S - S'$, that is, the vectors of S corresponding to the ***free variables*** of R, can be expressed as a ***linear combination*** of the vectors of S', using the ***coefficients*** found in the corresponding column of R.

The Dependent/Independent Sets from Spanning Sets Theorem:
Suppose $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\} \subset \mathbb{R}^k$, and we form $Span(S)$. Suppose now we form a new set $L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$ consisting of m vectors randomly chosen from $Span(S)$. We can then conclude that: ***if*** $m > n$, ***then*** L is automatically linearly ***dependent***. Consequently, ***if*** L is linearly ***independent***, ***then*** $m \leq n$.

The Extension Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***linearly independent*** set of vectors from $\mathbb{R}^m$, and suppose $\vec{v}_{n+1}$ is ***not*** a member of $Span(S)$. Then, the ***extended*** set $S' = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{v}_{n+1}\}$ is ***still linearly independent***.

1.6 Exercises

1. Show that the sets:

$$S = \{\langle 3, 6, -2, 7\rangle, \langle -1, 2, -4, 5\rangle, \langle -15, 0, -25, 20\rangle\}, \text{ and}$$
$$S' = \{\langle -6, -12, 4, -14\rangle, \langle -3, 6, -12, 15\rangle, \langle 3, 0, 5, -4\rangle\}$$

have exactly the same Span, by comparing corresponding pairs of vectors. Very few computations should be necessary.

2. Apply the Equality of Spans Theorem to show that the sets:

$$S = \{\langle 3, 6, -2, 7\rangle, \langle -1, 2, -4, 5\rangle\} \text{ and } S' = \{\langle 4, 10, -5, 13\rangle, \langle 3, 0, 5, -4\rangle, \langle 7, 10, 0, 9\rangle\}$$

have exactly the same Span. In other words, show that every vector in S is a linear combination of the vectors in S', and vice versa. Why can't we apply the same reasoning as Exercise 1?

3. Show that the sets:

$$S = \{\langle 3, 6, 5, -2, 7\rangle, \langle 13, 34, 39, -18, 45\rangle\} \text{ and } S' = \{\langle 11, 14, 1, 2, 11\rangle, \langle -1, 2, 7, -4, 5\rangle\}$$

have exactly the same Span. Which of the first two Theorems should we apply?

4. Show that the sets:

$$S = \{\langle 3, 6, 5, -2, 7\rangle, \langle 4, -8, -28, 16, -20\rangle\} \text{ and } S' = \{\langle 15, 30, 25, -10, 35\rangle, \langle -1, 2, 7, -4, 5\rangle\}$$

have exactly the same Span. Which of the first two Theorems should we apply?

5. Consider the set $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6\}$,

$$S = \{\langle 3, 6, -12\rangle, \langle -12, -24, 48\rangle, \langle 7, -3, 4\rangle, \langle 5, 10, -20\rangle, \langle 19, 4, -12\rangle, \langle 0, 1, 0\rangle\}.$$

   a. Explain why S is certainly linearly dependent.
   b. Scan the vectors for all parallel vectors, if any.
   c. Eliminate some vectors from S using your findings in (b) to obtain a smaller set S' with the same Span as S.
   d. Show that $\vec{v}_5$ can be expressed as a linear combination of $\vec{v}_1$ and $\vec{v}_3$.
   e. Use The Elimination Theorem to construct a subset S'' of S with as few vectors as possible so that $Span(S) = Span(S'')$. How many vectors are left?
   f. Check that the final set S'' is now linearly independent.

6. Consider the set:

$$S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\}$$
$$= \{\langle 1, 0, -1, 3\rangle, \langle -2, 3, 1, -1\rangle, \langle 7, -6, -5, 11\rangle, \langle 10, -15, -5, 5\rangle, \langle 0, 3, -1, 5\rangle\}.$$

   a. Explain why S is certainly linearly dependent.
   b. Scan the vectors for all parallel vectors, if any.
   c. Eliminate some vectors from S using your findings in (b) to obtain a smaller set S' with the same Span as S.
   d. Show that $\vec{v}_3$ and $\vec{v}_5$ can be expressed as linear combination of $\vec{v}_1$ and $\vec{v}_2$.
   e. Use The Elimination Theorem to construct a subset S'' of S with as few vectors as possible so that $Span(S) = Span(S'')$. How many vectors are left?
   f. Show that the final set S'' is now linearly independent.

7. Suppose that $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} \subset \mathbb{R}^7$, and you are told that:

$$3\vec{v}_1 - 8\vec{v}_2 + 2\vec{v}_3 = \vec{\mathbf{0}}_7, \quad \text{and} \quad 5\vec{v}_4 - 6\vec{v}_5 = \vec{\mathbf{0}}_7.$$

Use the Elimination Theorem to decide which of the following sets of vectors ***definitely*** have the same Span as S :

a. $S_1 = \{\vec{v}_1, \vec{v}_2, \vec{v}_4, \vec{v}_5\}$ b. $S_2 = \{\vec{v}_1, \vec{v}_3, \vec{v}_4\}$ c. $S_3 = \{\vec{v}_1, \vec{v}_4, \vec{v}_5\}$
d. $S_4 = \{\vec{v}_1, \vec{v}_2, \vec{v}_5\}$ e. $S_5 = \{\vec{v}_3, \vec{v}_5\}$

8. Suppose that $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6, \vec{v}_7, \vec{v}_8, \vec{v}_9, \vec{v}_{10}\} \subset \mathbb{R}^{10}$, and you are told that:

$$2\vec{v}_1 + 5\vec{v}_2 - \vec{v}_3 = \vec{\mathbf{0}}_{10}, \quad 8\vec{v}_4 - 3\vec{v}_5 + 7\vec{v}_6 = \vec{\mathbf{0}}_{10}, \quad \text{and} \quad 4\vec{v}_7 - 2\vec{v}_8 + 5\vec{v}_9 - 6\vec{v}_{10} = \vec{\mathbf{0}}_{10}.$$

Use the Elimination Theorem to decide which of the following sets of vectors ***definitely*** have the same Span as S :

a. $S_1 = \{\vec{v}_1, \vec{v}_2, \vec{v}_4, \vec{v}_6, \vec{v}_7, \vec{v}_8, \vec{v}_9\}$ b. $S_2 = \{\vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_7, \vec{v}_{10}\}$
c. $S_3 = \{\vec{v}_1, \vec{v}_3, \vec{v}_5, \vec{v}_8, \vec{v}_9, \vec{v}_{10}\}$ d. $S_4 = \{\vec{v}_1, \vec{v}_2, \vec{v}_5, \vec{v}_6, \vec{v}_7, \vec{v}_9, \vec{v}_{10}\}$
e. $S_5 = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_6, \vec{v}_7, \vec{v}_8, \vec{v}_9\}$

9. Suppose that $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} \subset \mathbb{R}^6$, and you are told that:

$$7\vec{v}_1 + 3\vec{v}_2 - 8\vec{v}_3 = \vec{\mathbf{0}}_6, \quad 4\vec{v}_1 - 5\vec{v}_2 + 9\vec{v}_4 = \vec{\mathbf{0}}_6, \quad \text{and} \quad 2\vec{v}_1 - 4\vec{v}_3 + 2\vec{v}_5 = \vec{\mathbf{0}}_6.$$

Use the Elimination Theorem to decide which of the following sets of vectors ***definitely*** have the same Span as S:

a. $S_1 = \{\vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\}$ b. $S_2 = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ c. $S_3 = \{\vec{v}_2, \vec{v}_3, \vec{v}_4\}$
d. $S_4 = \{\vec{v}_2, \vec{v}_3, \vec{v}_5\}$ e. $S_5 = \{\vec{v}_2, \vec{v}_5\}$ f. $S_6 = \{\vec{v}_1, \vec{v}_2\}$

Assisted Computation: For Exercises (10) to (34): Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$, and $A = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \ldots & \vec{v}_n \end{bmatrix}$ is the $m \times n$ matrix obtained by assembling the vectors into columns. Suppose that R is the rref of A, as shown in each item. Use the Minimizing Theorem to find a linearly independent subset $S' = \{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}$ of S such that $Span(S) = Span(S')$, and express each vector $\vec{v}_i \in S - S'$ as a linear combination of the vectors in S'. You may express your answers in terms of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$.

10. $A = \begin{bmatrix} 2 & -3 & 3 & -12 \\ -3 & 0 & -1 & -5 \\ 4 & -5 & -2 & 10 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & -4 \end{bmatrix}$

11. $A = \begin{bmatrix} 3 & 2 & 5 & 12 \\ -1 & 1 & -5 & 1 \\ 1 & -2 & 7 & -4 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

12. $A = \begin{bmatrix} -2 & 10 & 3 & 9 \\ 4 & -20 & -2 & 2 \\ -3 & 15 & 4 & 11 \end{bmatrix}$; $R = \begin{bmatrix} 1 & -5 & 0 & 3 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

13. $A = \begin{bmatrix} 5 & -2 & -1 & 9 & 5 \\ -2 & 3 & -3 & 12 & 8 \\ 3 & -4 & 2 & -11 & -8 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 0 & 3 & 2 \\ 0 & 1 & 0 & 4 & 3 \\ 0 & 0 & 1 & -2 & -1 \end{bmatrix}$

14. $A = \begin{bmatrix} 5 & -2 & 6 & -1 & 9 \\ -2 & 3 & 13 & -3 & 12 \\ 3 & -4 & -16 & 2 & -11 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & 0 & 3 \\ 0 & 1 & 7 & 0 & 4 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix}$

15. $A = \begin{bmatrix} -1 & 4 & -2 & 5 & 5 \\ -3 & 12 & 3 & 8 & -2 \\ 2 & -8 & -4 & -8 & 3 \end{bmatrix}$; $R = \begin{bmatrix} 1 & -4 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 0 & -\frac{3}{2} \\ 0 & 0 & 0 & 1 & \frac{1}{2} \end{bmatrix}$

16. $A = \begin{bmatrix} -1 & -3 & -2 & -8 & 5 \\ -3 & -9 & 3 & -6 & -2 \\ 2 & 6 & -4 & 0 & 3 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 3 & 0 & 4 & 0 \\ 0 & 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

17. $A = \begin{bmatrix} -2 & 5 & 7 & 12 & -1 \\ 3 & -2 & 6 & -7 & -4 \\ -5 & 3 & -11 & 11 & 7 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & -1 & -2 \\ 0 & 1 & 3 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

18. $A = \begin{bmatrix} 3 & 2 & 0 \\ -2 & 1 & -7 \\ 1 & -2 & 8 \\ -4 & -2 & -2 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

19. $A = \begin{bmatrix} 3 & -2 & 5 \\ 7 & 4 & -6 \\ 1 & 0 & 8 \\ -9 & -5 & 2 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

20. $A = \begin{bmatrix} 2 & 3 & 8 & 3 \\ 1 & -2 & -3 & -16 \\ -2 & 1 & 0 & 17 \\ -2 & -4 & -10 & -8 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 1 & -6 \\ 0 & 1 & 2 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

21. $A = \begin{bmatrix} 3 & -2 & 1 & 5 \\ 7 & -4 & 7 & -6 \\ 1 & 0 & 5 & 8 \\ -9 & 6 & -3 & 2 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 5 & 0 \\ 0 & 1 & 7 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

22. $A = \begin{bmatrix} -3 & 9 & 1 & -11 \\ 7 & -21 & -4 & 19 \\ 5 & -15 & 2 & 33 \\ 4 & -12 & -3 & 8 \end{bmatrix}$; $R = \begin{bmatrix} 1 & -3 & 0 & 5 \\ 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

23. $A = \begin{bmatrix} 0 & 2 & 9 & -4 & -6 \\ -7 & 1 & -13 & 3 & 15 \\ 8 & -2 & 11 & -1 & -11 \\ -2 & -2 & -14 & 6 & 14 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & \frac{5}{2} & 0 & 1 \\ 0 & 1 & \frac{9}{2} & 0 & 7 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

24. $A = \begin{bmatrix} 3 & -2 & 5 & -3 & 3 \\ 7 & -4 & 6 & 7 & -2 \\ 1 & 0 & 3 & -1 & 0 \\ -9 & 6 & -9 & -3 & 4 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 0 & 5 & 0 \\ 0 & 1 & 0 & 4 & 0 \\ 0 & 0 & 1 & -2 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

25. $A = \begin{bmatrix} 3 & -2 & 5 & -3 & -9 \\ 7 & -4 & 6 & 7 & 5 \\ 1 & 0 & 3 & -1 & -5 \\ -9 & 6 & -9 & -3 & 3 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 0 & 5 & 7 \\ 0 & 1 & 0 & 4 & 5 \\ 0 & 0 & 1 & -2 & -4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

26. $A = \begin{bmatrix} 15 & 3 & 0 \\ -3 & -2 & 1 \\ 13 & 4 & -1 \\ -9 & 1 & -2 \\ -11 & 2 & -3 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & \frac{1}{7} \\ 0 & 1 & -\frac{5}{7} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

27. $A = \begin{bmatrix} 3 & -2 & 5 \\ 7 & -4 & 6 \\ 1 & 0 & 3 \\ -9 & 6 & -9 \\ 4 & 3 & 7 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

28. $A = \begin{bmatrix} 15 & 3 & 21 & 0 \\ -3 & -2 & -10 & 1 \\ 13 & 4 & 24 & -1 \\ -9 & 1 & -1 & -2 \\ -11 & 2 & 2 & -3 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & \frac{4}{7} & \frac{1}{7} \\ 0 & 1 & \frac{29}{7} & -\frac{5}{7} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

29. $A = \begin{bmatrix} 3 & -2 & -1 & 2 \\ 7 & -4 & 3 & 6 \\ 1 & 0 & 5 & 1 \\ -9 & 6 & 3 & -8 \\ 4 & -3 & -4 & 9 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 5 & 0 \\ 0 & 1 & 8 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

30. $A = \begin{bmatrix} 5 & 3 & 2 & -7 \\ -3 & -2 & -1 & 4 \\ 3 & 4 & 3 & -18 \\ -9 & 1 & -1 & -17 \\ -1 & 2 & 2 & -16 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & -4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

31. $A = \begin{bmatrix} 12 & 2 & 3 & -13 & 5 \\ -6 & -1 & -2 & 11 & -3 \\ 18 & 3 & 4 & -15 & 0 \\ -6 & -1 & 1 & -16 & 7 \\ 12 & 2 & 2 & -4 & -1 \end{bmatrix}$; $R = \begin{bmatrix} 1 & \frac{1}{6} & 0 & \frac{7}{6} & 0 \\ 0 & 0 & 1 & -9 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

32. $A = \begin{bmatrix} -2 & -4 & -8 & -5 & 2 \\ 1 & 2 & 4 & 3 & -1 \\ 3 & 5 & 7 & -1 & 0 \\ -1 & 0 & 6 & 0 & -5 \\ 4 & 6 & 6 & 1 & 2 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & -6 & 0 & 5 \\ 0 & 1 & 5 & 0 & -3 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

33. $A = \begin{bmatrix} 3 & -2 & -1 & 2 & 5 \\ 7 & -4 & 3 & 6 & 6 \\ 1 & 0 & 5 & 1 & 3 \\ -9 & 6 & 3 & -8 & -9 \\ 4 & -3 & -4 & 9 & 7 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 5 & 0 & 0 \\ 0 & 1 & 8 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

34. $A = \begin{bmatrix} 2 & 3 & -13 & 1 & 12 & -13 \\ -1 & -2 & 11 & -3 & -19 & 17 \\ 3 & 4 & -15 & 2 & 20 & -18 \\ -1 & 1 & -16 & -2 & -11 & -4 \\ 2 & 2 & -4 & -1 & 1 & -1 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 7 & 0 & 2 & 4 \\ 0 & 1 & -9 & 0 & 1 & -6 \\ 0 & 0 & 0 & 1 & 5 & -3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

For Exercises (35) to (45): Use ***The Minimizing Theorem*** to find a subset S' of the set S so that S' is as small as possible and $Span(S) = Span(S')$. Next, express the vectors in $S - S'$ as linear combinations of the vectors in S'. Use technology to compute the rrefs, if permitted by your instructor.

35. The set S in Exercise 5.
36. The set S in Exercise 6.
37. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} = \{\langle 1, 14, 9\rangle, \langle 2, 1, -3\rangle, \langle 4, -7, -13\rangle, \langle 1, 5, 2\rangle, \langle 1, -4, -5\rangle\}$
38. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} = \{\langle 1, 4, -3\rangle, \langle 4, -2, 10\rangle, \langle -6, 3, -15\rangle, \langle 1, -5, 2\rangle, \langle 3, 2, 7\rangle\}$
39. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\}$
$= \{\langle 3, -2, 5, 1\rangle, \langle 2, 6, 3, -4\rangle, \langle -2, -28, -2, 18\rangle, \langle 2, -37, 5, 24\rangle, \langle 0, -22, 1, 14\rangle\}$
40. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6\}$
$= \{\langle 2, -1, 3, -1\rangle, \langle 5, -3, 1, 0\rangle, \langle -6, 5, 17, -7\rangle, \langle 1, 0, -1, 2\rangle, \langle -1, 1, 2, -2\rangle, \langle -15, 10, -2, 10\rangle\}$
41. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6\}$
$= \{\langle 1, 0, -1, 2\rangle, \langle -1, 1, 2, -2\rangle, \langle -3, 8, 11, -6\rangle, \langle 5, -3, 1, 0\rangle, \langle 18, -5, 13, -4\rangle, \langle 11, -3, 4, 2\rangle\}$
42. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} = \{\langle 3, 2, 5, -1, 7\rangle, \langle 2, -3, -4, 1, 6\rangle, \langle -6, -4, -10, 2, -14\rangle,$
$\langle -1, -5, -9, 2, -1\rangle, \langle 4, -19, -30, 7, 16\rangle\}$
43. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} = \{\langle 3, -2, 5, -1, 3\rangle, \langle -2, 3, -4, 1, -2\rangle, \langle 6, 1, 8, -1, 6\rangle,$
$\langle -1, -5, 2, 2, -1\rangle, \langle -7, -4, -4, 9, -7\rangle\}$
44. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6\} = \{\langle 3, -2, 1, 2, 0\rangle, \langle -2, 1, 0, -3, 2\rangle, \langle 5, -4, 3, 0, 2\rangle,$
$\langle -9, 6, -3, -6, 4\rangle, \langle -3, 1, -1, 1, -1\rangle, \langle 8, 1, 0, -3, 3\rangle\}$
45. $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6\} = \{\langle 3, -2, 1, 2, 0\rangle, \langle -2, 1, 0, -3, 2\rangle, \langle 5, -4, 3, 0, 2\rangle,$
$\langle -1, 1, -1, 1, 2\rangle, \langle -3, 1, -1, 1, -1\rangle, \langle -1, 1, 0, 3, -2\rangle\}$

46. Consider the set: $S = \{\vec{v}_1, \vec{v}_2\} = \{\langle 4, 5, -2\rangle, \langle -6, 7, 3\rangle\} \subset \mathbb{R}^3$, and let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the standard basis vectors for $\mathbb{R}^3$, as usual.
    a. Explain why S is linearly independent.
    b. Find the Cartesian equation of the plane $\Pi = Span(\{\vec{v}_1, \vec{v}_2\})$
    c. Which of the standard basis vectors, if any, are in $Span(S)$?
    The rest of this Exercise concerns the application of The Extension Theorem:
    d. Is $S' = \{\vec{v}_1, \vec{v}_2, \vec{e}_1\}$ still linearly independent? Why or why not?
    e. Is $S'' = \{\vec{v}_1, \vec{v}_2, \vec{e}_2\}$ still linearly independent? Why or why not?
    f. Is $S''' = \{\vec{v}_1, \vec{v}_2, \vec{e}_3\}$ still linearly independent? Why or why not?

47. Consider the set:
$$S^{(1)} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = \{\langle 3,-2,5,4,-6\rangle, \langle 3,4,5,4,-6\rangle, \langle 3,-2,2,4,-6\rangle\},$$
and let $\{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4, \vec{e}_5\}$ be the standard basis vectors for $\mathbb{R}^5$, as usual.
    a. Show that $S^{(1)}$ is linearly independent. (take this opportunity to look at the similarities among the vectors in S)
    b. Is $S^{(2)} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{e}_1\}$ still linearly independent?

c. Is $S^{(3)} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{e}_2\}$ still linearly independent?

d. Is $S^{(4)} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{e}_3\}$ still linearly independent?

e. Is $S^{(5)} = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{e}_1, \vec{e}_5\}$ still linearly independent? Think carefully!

48. Let $\vec{v}_1 = \langle 5,-3,4,2,-6,-1\rangle$, $\vec{v}_2 = \langle 3,-4,2,5,-7,1\rangle$ and $\vec{v}_3 = \langle 5,10,6,-17,11,-9\rangle$.

a. Form the matrix $A = [\vec{v}_1\, \vec{v}_2\, \vec{v}_3\, \vec{e}_1\, \vec{e}_2\, \vec{e}_3\, \vec{e}_4\, \vec{e}_5\, \vec{e}_6]$

b. Find the rref R of A. Use technology if allowed by your instructor.

Now, decide which of the following sets are linearly independent, and explain why:

c. $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{e}_1\}$

d. $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{e}_1, \vec{e}_2\}$

e. $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{e}_1, \vec{e}_3, \vec{e}_4\}$

f. $\{\vec{v}_1, \vec{v}_2, \vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$

g. $\{\vec{v}_2, \vec{v}_3, \vec{e}_1, \vec{e}_2\}$

49. Use the Dependent Sets from Spanning Sets Theorem to give a different proof that a set S of n vectors from $\mathbb{R}^m$ is linearly dependent if $n > m$. Hint: express $\mathbb{R}^m$ as a Span of a set of vectors. Which set could you use?

50. Prove that if $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a linearly independent set from $\mathbb{R}^m$, then $\vec{v}_n$ is ***not*** a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-1}\})$. Hint: Use Proof by Contradiction. This is basically the converse of The Extension Theorem.

51. Complete the last part of ***The Equality of Spans Theorem:***

Suppose that every $\vec{w}_j$ can also be written as a linear combination of the $\vec{v}_1$ through $\vec{v}_n$. Show that every member of $Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\})$ is likewise a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$. Imitate the proof of the first part found in the text, but be careful not to confuse m and n.

52. Prove ***The Elimination Theorem:*** Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a linearly ***dependent*** set of vectors from $\mathbb{R}^m$, and $\vec{v}_n = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_{n-1}\vec{v}_{n-1}$. Then:

$$Span(S) = Span(S - \{\vec{v}_n\}).$$

More generally, if $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_{n-1}\vec{v}_{n-1} + c_n\vec{v}_n = \vec{\mathbf{0}}_{\boldsymbol{m}}$, where ***none*** of the coefficients in the dependence equation is 0, then $Span(S) = Span(S - \{\vec{v}_i\})$, for all $i = 1..n$. Hint: Use the ***Equality of Spans Theorem***.

53. ***True or False:*** Determine whether each statement is true or false, and briefly explain your answer by either applying a Theorem or providing a counterexample or a convincing argument.

a. If S has 5 non-zero vectors, then any set of 4 vectors from $Span(S)$ will always be linearly independent.

b. If S has 5 non-zero vectors, then any set of 7 vectors from $Span(S)$ will always be linearly dependent.

c. If S has 9 non-zero vectors, then any set of 9 vectors from $Span(S)$ will always be linearly independent.

d. If a set of 7 vectors from $Span(S)$ is linearly independent, then S has at least 7 vectors.

e. If a set of 7 vectors from $Span(S)$ is linearly independent, then S has exactly 7 vectors.

f. If a set of 7 vectors from $Span(S)$ is linearly independent, then S has at most 7 vectors.

g. If a set of 5 vectors from $Span(S)$ is linearly dependent, then S has at least 5 vectors.

h. If a set of 5 vectors from $Span(S)$ is linearly dependent, then S has exactly 5 vectors.

i. If a set of 5 vectors from $Span(S)$ is linearly dependent, then S has at most 5 vectors.

1.7 Subspaces of Euclidean Spaces; Basis and Dimension

We know that the sum of two integers is again an integer, and similarly, the sum of two rational numbers is again a rational number. We say that the set of integers is ***closed*** under addition, and so is the set of rational numbers. Similarly, we want to define special subsets of Euclidean space that have analogous closure properties:

Definition: A ***subspace*** W of $\mathbb{R}^n$ is a non-empty subset of vectors of $\mathbb{R}^n$ such that if $\vec{u}$, $\vec{v} \in W$, and $r \in \mathbb{R}$, then we also have:

$$\vec{u} + \vec{v} \in W \quad \textbf{\textit{and}} \quad r \bullet \vec{v} \in W.$$

We say W is ***closed*** under vector addition and scalar multiplication, and write:

$$W \trianglelefteq \mathbb{R}^n$$

to indicate that W is a subspace of $\mathbb{R}^n$. We call $\mathbb{R}^n$ the ***ambient space*** of W.

The symbol $\trianglelefteq$ is a stylized version of the subset symbol $\subseteq$. We also say informally that "W ***lives*** in $\mathbb{R}^n$." Next, let us see that any subspace contains a familiar friend:

Theorem: The zero vector $\vec{\mathbf{0}}_n$ is always a member of any subspace $W \trianglelefteq \mathbb{R}^n$.

Proof: Since W is not empty, W has at least one member, say $\vec{v}$. But any scalar multiple of $\vec{v}$ must be a member of W, so $0 \bullet \vec{v} = \vec{\mathbf{0}}_n$ is a member of W also. ■

Next, any $\mathbb{R}^n$ has at least two subspaces (we leave the proof as an Exercise):

Definition/Theorem: For any $\mathbb{R}^n$, there are two ***trivial subspaces:***
(1) the subspace $\left\{\vec{\mathbf{0}}_n\right\}$ consisting only of the zero vector, and (2) all of $\mathbb{R}^n$ itself.

The opposite of trivial is ***non-trivial*** or ***proper***. Now let us look at one particular non-trivial Example that we have already seen.

Span(S) as a Subspace

The primary example of a subspace of $\mathbb{R}^n$ is the Span of a set of vectors:

Theorem: If $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is a non-empty set of vectors from $\mathbb{R}^n$, then $W = Span(S)$ is a ***subspace*** of $\mathbb{R}^n$.

Proof: First, W is non-empty since it contains $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k$. Next, we have to show that W is closed under vector addition and scalar multiplication. This means that the sum of two linear combinations of this set of vectors is again a linear combination of the same set, and a scalar multiple of a linear combination of this set is again a linear combination of the same set.

Once again, we just need to write this symbolically, using proper notation. Let us write two arbitrary vectors in $Span(S)$ as: $\vec{v} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$, and $\vec{w} = d_1\vec{u}_1 + d_2\vec{u}_2 + \cdots + d_k\vec{u}_k$. Thus:

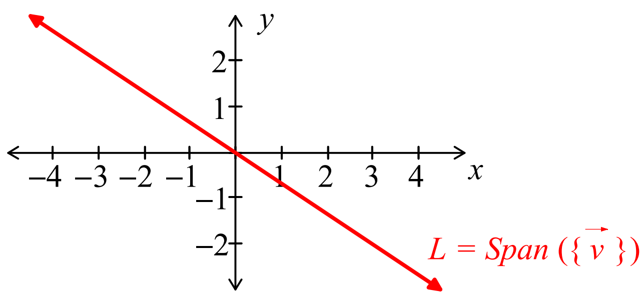
$$\begin{aligned}\vec{v} + \vec{w} &= (c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k) + (d_1\vec{u}_1 + d_2\vec{u}_2 + \cdots + d_k\vec{u}_k) \\ &= c_1\vec{u} + d_1\vec{u}_1 + c_2\vec{u}_2 + d_2\vec{u}_2 + \cdots + c_k\vec{u}_k + d_k\vec{u}_k \\ &= (c_1 + d_1)\vec{u}_1 + (c_2 + d_2)\vec{u}_2 + \cdots + (c_k + d_k)\vec{u}_k,\end{aligned}$$

which is again a linear combination from S. Notice that we used the Associative, Commutative and Distributive Properties of vector arithmetic. Similarly:

$$\begin{aligned}r\vec{v} &= r(c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k) = r(c_1\vec{u}_1) + r(c_2\vec{u}_2) + \cdots + r(c_k\vec{u}_k) \\ &= (rc_1)\vec{u}_1 + (rc_2)\vec{u}_2 + \cdots + (rc_k)\vec{u}_k,\end{aligned}$$

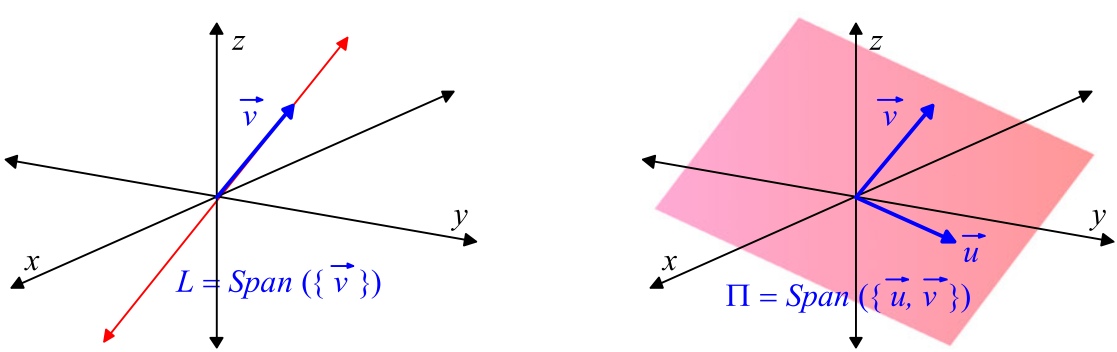
is a linear combination from S. This time, we used the Distributive Property and the Associative Property of Scalar Multiplication. Thus, $Span(S)$ is closed under addition and scalar multiplication. ■

Examples: In $\mathbb{R}^2$, we already know the two trivial subspaces: $\left\{\vec{\mathbf{0}}_2\right\}$ and $\mathbb{R}^2$. Now, if $\vec{v}$ is any non-zero vector, then $Span(\{\vec{v}\})$ is a line L passing through the origin:



A Line Through the Origin is a Subspace of $\mathbb{R}^2$

Thus L is a subspace of $\mathbb{R}^2$. Similarly, in $\mathbb{R}^3$, we have the trivial subspaces $\left\{\vec{\mathbf{0}}_3\right\}$ and $\mathbb{R}^3$. If $\vec{v}$ is a non-zero vector, then $Span(\{\vec{v}\})$ is again a line L passing through the origin, and if $\vec{v}$ and $\vec{w}$ are non-zero, non-parallel vectors, then $Span(\{\vec{v}, \vec{w}\})$ is a plane Π passing through the origin:



Lines or Planes Through the Origin are Subspaces of $\mathbb{R}^3$. □

Basis for a Subspace

We encountered the ***standard basis*** $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ for $\mathbb{R}^n$ in Section 1.1, and saw that any vector in $\mathbb{R}^n$ can be expressed as a linear combination of these vectors in exactly one way. We will now set the stage to generalize this concept:

> ***Definition:*** A ***basis*** for a non-zero ***subspace*** $W \trianglelefteq \mathbb{R}^n$ is a non-empty set of vectors $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset W$ which ***Spans*** W and is also ***linearly independent***.

Since Span(B) includes the vectors $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$, it is clear that these vectors must automatically belong to W to begin with. We will also agree that the trivial subspace $\left\{\vec{\mathbf{0}}_n\right\}$ does ***not*** have a basis, because any set containing $\vec{\mathbf{0}}_n$ (which is the only possible candidate) is automatically dependent.

Example: Since we know that the "standard basis" $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ is linearly independent and Spans $\mathbb{R}^n$, this set is a basis for $\mathbb{R}^n$, according to our definition above. □

Example: If $L = Span(\{\vec{v}\})$ is any line through the origin of $\mathbb{R}^n$, where $\vec{v} \neq \vec{\mathbf{0}}_n$, then any vector in L has the form $k \cdot \vec{v}$, for some scalar k. Thus, $\{\vec{v}\}$ is a basis for L.

Now, suppose Π is the plane Π through the origin in $\mathbb{R}^3$ given by:

$$\Pi : 3x - 5y + 2z = 0.$$

To find a basis for Π, we need to find a set of vectors that Spans Π and is also linearly independent. But notice that we can think of this single equation as a homogeneous system of one equation in three variables, and therefore it is consistent with ***two free variables***, which in this case are y and z.

Solving for the leading variable x, we get: $x = \frac{5}{3}y - \frac{2}{3}z$. Thus we have:

$$\langle x, y, z\rangle = \left\langle \frac{5}{3}y - \frac{2}{3}z, y, z \right\rangle = \left\langle \frac{5}{3}y, y, 0 \right\rangle + \left\langle -\frac{2}{3}z, 0, z \right\rangle = \frac{y}{3}\langle 5, 3, 0\rangle + \frac{z}{3}\langle -2, 0, 3\rangle.$$

This shows that ***every*** vector on Π is a linear combination of the vectors $\langle 5, 3, 0\rangle$ and $\langle -2, 0, 3\rangle$, and thus these two vectors ***Span*** Π. But we can see immediately that these two vectors are ***not parallel***, and so they are ***linearly independent***. Thus, a possible basis for Π is:

$$B = \{\langle 5, 3, 0\rangle, \langle -2, 0, 3\rangle\}.$$

We note that this is not the only basis for Π. In fact, there are an ***infinite*** number of choices for B. □

A Basis for Span(S)

So far, we have been working with very simple examples, that is, lines or planes through the origin. In our constructive Theorem above, though, we now know that ***every*** subspace W of $\mathbb{R}^n$ can be written as $Span(S)$, so we will next focus our attention on how to efficiently construct a basis for $Span(S)$, whenever possible. Let us start with the easiest case:

> ***Theorem:*** The set $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a ***basis*** for $W = Span(B)$ ***if and only if*** B is ***linearly independent***.

Proof: $(\Rightarrow)$ If B is a basis for W, then B Spans W ***and*** is linearly independent.

$(\Leftarrow)$ Conversely, if B is linearly independent, and $W = Span(B)$, then B is a basis for W. ■

Example: Suppose $S = \{\langle 1, 7, 3, -8, 2\rangle, \langle 4, -2, 5, 3, -4\rangle\}$. The two vectors in S are from $\mathbb{R}^5$ and are obviously not parallel to each other. Therefore, S is linearly independent. Thus, S is a basis for $W = Span(S)$, a subspace of $\mathbb{R}^5$. □

With more than two vectors, though, there is a lot more work to do in order to check whether or not S is already linearly independent. However, ***The Minimizing Theorem*** from Section 1.6 tells us that if we assemble the vectors of S into the columns of a matrix A, and S' are the columns of A corresponding to the columns of the rref R of A that contain the leading 1's, then:

$$Span(S) = Span(S').$$

Moreover, this Theorem also says that S' is ***linearly independent***, and thus S' is a ***basis*** for $W = Span(S)$. We can thus rephrase this Theorem as follows:

Theorem — The Minimizing Theorem (Basis for a Subspace Version):
Suppose $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset \mathbb{R}^n$, and $W = Span(S)$. If $A = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_k \end{bmatrix}$, and R is the rref of A, then the columns of A corresponding to the ***leading columns*** of R form a ***basis*** for W.

Example: Suppose $W = Span(S)$, where:

$$S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\} = \{\langle 11, -13, -8, 17\rangle, \langle -4, 7, 3, -6\rangle, \langle 10, -5, -7, 16\rangle\} \subset \mathbb{R}^4.$$

We assemble these vectors as the ***columns*** of a 4×3 matrix:

$$A = \begin{bmatrix} 11 & -4 & 10 \\ -13 & 7 & -5 \\ -8 & 3 & -7 \\ 17 & -6 & 16 \end{bmatrix} \quad \text{with rref} \quad R = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The leading 1's of R are in the 1st and 2nd columns, and thus a basis for $Span(S)$ is:

$$B = \{\langle 11, -13, -8, 17\rangle, \langle -4, 7, 3, -6\rangle\},$$

the corresponding columns in A. Notice that the coefficients in column 3 tell us how the 3rd vector is a ***linear combination*** of the first two:

$$\langle 10, -5, -7, 16\rangle = 2\langle 11, -13, -8, 17\rangle + 3\langle -4, 7, 3, -6\rangle. \ \square$$

Constructing a Basis for Any Subspace

We already know that $\mathbb{R}^n$ itself has a basis, namely the standard basis $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$, and we saw above that it was easy to find a basis for a plane through the origin, or for $W = Span(S)$. However, suppose that W is not given as the solution to an equation or as the Span of a set of vectors.

The following Theorem tells us that for ***any*** non-zero subspace W of $\mathbb{R}^n$, we can always construct a basis for W, one vector at a time:

Theorem — Existence of a Basis Theorem:

If W is any ***non-zero*** subspace of $\mathbb{R}^n$, then ***there exists*** a ***basis*** $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset W$ for W. In other words, we can write:

$$W = Span(B) = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}),$$

where B is a ***linearly independent*** set that ***Spans*** W. Furthermore, we must have $k \leq n$.

Proof: We will prove this Theorem by explicitly constructing a basis for W using repeated applications of ***The Extension Theorem***. Since W is not the zero subspace, we can pick any non-zero vector $\vec{w}_1 \in W$, and form the set $S_1 = \{\vec{w}_1\}$. Since $\vec{w}_1$ is a non-zero vector, S_1 is ***independent***. Now, if $W = Span(S_1)$, then we are finished. Otherwise, consider any $\vec{w}_2 \in W$ which is not a member of $Span(S_1)$. The new set $S_2 = \{\vec{w}_1, \vec{w}_2\}$ must be linearly ***independent*** by The Extension Theorem. We ask again if $W = Span(S_2)$. If so, we are done, otherwise we extend S_2 using another vector $\vec{w}_3 \in W$ which is not in $Span(S_2)$, thus producing another ***independent*** set $S_3 = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$.

In general, let us proceed by Induction: suppose we have constructed a linearly ***independent*** subset $S_i = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_i\}$ of W. If $W = Span(S_i)$, then we are done, otherwise we find another vector $\vec{w}_{i+1} \in W$ which is not in $Span(S_i)$, and create the extended set $S_{i+1} = S_i \cup \{\vec{w}_{i+1}\}$, which is still independent by the Extension Theorem. Since any subset of $\mathbb{R}^n$ with $n+1$ vectors is linearly ***dependent***, this process must terminate with some set $S_k = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$, for some $k \leq n$, which is linearly ***independent*** and ***Spans*** W. Thus S_k is a basis for W. ■

This Theorem tells us that in fact ***every*** subspace of $\mathbb{R}^n$ can be written as the Span of a set of vectors. Since we know exactly what the Spans of vectors look like in $\mathbb{R}^2$ and $\mathbb{R}^3$, we can now formally list ***all*** their subspaces:

Theorem — The Subspaces of Euclidean 2-Space and 3-Space:

The only subspaces of $\mathbb{R}^2$ are:

(a) the zero subspace $\left\{\vec{\mathbf{0}}_2\right\}$,

(b) the ***lines*** through the origin, and

(c) all of $\mathbb{R}^2$.

Similarly, the only subspaces of $\mathbb{R}^3$ are:

(a) the zero subspace $\left\{\vec{\mathbf{0}}_3\right\}$,

(b) the ***lines*** through the origin,

(c) the ***planes*** through the origin, and

(d) all of $\mathbb{R}^3$.

Let us take a second look at our proof concerning the existence of a basis for a subspace $W \trianglelefteq \mathbb{R}^n$. Since we randomly choose the vectors $\vec{w}_1$, then $\vec{w}_2$, and so on, to form our basis, it is certainly possible to create two completely different ***bases*** (the plural of basis, but pronounced bay-***sees***) for W. In fact, there are an ***infinite*** number of them. However, they all have something in common.

The Dimension of a Subspace

We said in Section 1.1 that we call n the ***dimension*** of $\mathbb{R}^n$. We intuitively think of a line as a ***one-dimensional*** object. We think of the Cartesian plane, or any plane for that matter, as a ***two-dimensional*** object, and the universe that we live in as ***three-dimensional*** space. The notion of ***dimension*** is therefore of fundamental importance in Linear Algebra (and in fact, most of Mathematics), and we are now ready to rigorously define this concept:

Theorem/Definition — The Dimension of a Subspace:

If B and B' are any two ***bases*** for the same non-zero ***subspace*** $W \trianglelefteq \mathbb{R}^n$, then B and B' contain ***exactly the same number of vectors***. We call this number the ***dimension*** of W, and we write $\boldsymbol{dim(W) = k}$. We also say that W is ***k-dimensional***.

We agreed that the trivial subspace $\left\{\vec{\mathbf{0}}_n\right\}$ does ***not*** have a basis. By convention, $dim\left(\left\{\vec{\mathbf{0}}_n\right\}\right) = 0$. Conversely, $dim(W)$ is a ***positive integer*** for a ***non-zero*** subspace W.

Proof: By the ***Independent Sets from Spanning Sets Theorem*** from Section 1.6, a linearly independent subset of $W = Span(S)$ has ***at most*** as many members as a Spanning subset S for W. However, both B and B' are bases, and so ***both sets*** are linearly independent ***and*** Span W. Thus, the number of members of B is at most the number of members of B'. Similarly, the number of elements of B' is at most the number of elements of B. Hence, they have the same number of elements.■

Example: Since the set $S = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ is a basis for $\mathbb{R}^n$, we can now formally say that $dim(\mathbb{R}^n) = n$.□

Example: Suppose that Π is the plane $3x - 5y + 2z = 0$ from our second Example. We saw that the set $\{\langle 5, 3, 0\rangle, \langle -2, 0, 3\rangle\}$ is a basis for Π. Thus, $dim(\Pi) = 2$, that is, Π is a 2-dimensional subspace of $\mathbb{R}^3$. This matches our intuition of a 2-dimensional object is: one that is flat and lacking depth.

Similarly, if L is the normal line to Π, then $L = Span(\{\langle 3, -5, 2\rangle\})$, and so L is a 1-dimensional subspace of $\mathbb{R}^3$. Again, this matches our intuition of a line as a 1-dimensional object.□

Example: We saw that if $W = Span(S)$, where:

$$S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\} = \{\langle 11, -13, -8, 17\rangle, \langle -4, 7, 3, -6\rangle, \langle 10, -5, -7, 16\rangle\},$$

then $B = \{\langle 11, -13, -8, 17\rangle, \langle -4, 7, 3, -6\rangle\}$ is a basis for W. Note that S has ***three*** vectors in it, but the basis B only has ***two***. Thus, W is only ***2-dimensional***.□

More generally, the ***Minimizing Theorem*** tells us that if $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset \mathbb{R}^n$ and $W = Span(S)$, then W is ***at most*** k-dimensional. The exact value of $dim(W)$ is the number of ***leading ones*** in the rref R of the $n \times k$ matrix $A = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_k \end{bmatrix}$.

Example: Suppose that $W = Span(S)$, where:

$$S = \{\langle 3, -2, 5, 4, 1, -6\rangle, \langle 5, -3, 6, 3, 2, -8\rangle, \langle -11, 5, -2, 11, -6, 8\rangle,$$
$$\langle -2, 1, -4, -2, 0, 6\rangle, \langle -4, 1, -1, 7, -1, 6\rangle\} \subset \mathbb{R}^6.$$

Since S has 5 vectors, we can only say at this point that W is at most 5-dimensional. To find its true dimension, we assemble these five vectors from $\mathbb{R}^6$ into the columns of the 6×5 matrix:

$$A = \begin{bmatrix} 3 & 5 & -11 & -2 & -4 \\ -2 & -3 & 5 & 1 & 1 \\ 5 & 6 & -2 & -4 & -1 \\ 4 & 3 & 11 & -2 & 7 \\ 1 & 2 & -6 & 0 & -1 \\ -6 & -8 & 8 & 6 & 6 \end{bmatrix}, \quad \text{with rref:} \quad R = \begin{bmatrix} 1 & 0 & 8 & 0 & 5 \\ 0 & 1 & -7 & 0 & -3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

We see that the leading columns are columns 1, 2 and 4, and thus:

$$B = \{\langle 3,-2,5,4,1,-6\rangle, \langle 5,-3,6,3,2,-8\rangle, \langle -2,1,-4,-2,0,6\rangle\}$$

is a basis for W. We can also conclude that W is 3-dimensional.$\square$

1.7 Section Summary

A ***subspace*** W of $\mathbb{R}^n$ is a non-empty subset of vectors of $\mathbb{R}^n$ such that if $\vec{u},\ \vec{v} \in W$, and if $r \in \mathbb{R}$, then $\vec{u} + \vec{v} \in W$ also, and $r\vec{v} \in W$ also.

We say that W is ***closed*** under vector addition and scalar multiplication, and write $W \trianglelefteq \mathbb{R}^n$ to indicate that W is a subspace of $\mathbb{R}^n$. We call $\mathbb{R}^n$ the ***ambient space*** of W.

The zero vector $\vec{\mathbf{0}}_n$ is always a member of any subspace $W \trianglelefteq \mathbb{R}^n$.

If $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\} \subset \mathbb{R}^n$ is a non-empty set, then $W = Span(S)$ is a ***subspace*** of $\mathbb{R}^n$.

A ***basis*** for a subspace $W \trianglelefteq \mathbb{R}^n$ is a set of vectors $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ which ***Spans*** W and is also ***linearly independent***.

The set $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a ***basis*** for $W = Span(B)$ ***if and only if*** B is ***linearly independent***.

The Minimizing Theorem — Basis for a Subspace Version: Suppose $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset \mathbb{R}^n$, and $W = Span(S)$. If $A = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_k \end{bmatrix}$ is the matrix with the vectors assembled in ***columns***, and R is the rref of A, then the columns of A corresponding to the ***leading columns*** of R form a ***basis*** for W. Consequently, $dim(W)$ is the number of ***leading ones*** of R, and $dim(W) \leq k$.

If W is any non-zero subspace of $\mathbb{R}^n$, then ***there exists*** a ***basis*** $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset W$ for W. In other words, we can write $W = Span(B)$, where B is a ***linearly independent*** set that ***Spans*** W. Furthermore, we must have $k \leq n$.

The only subspaces of $\mathbb{R}^2$ are: (a) $\left\{\vec{\mathbf{0}}_2\right\}$, (b) a ***line*** through the origin, and (c) all of $\mathbb{R}^2$.

The only subspaces of $\mathbb{R}^3$ are: (a) $\left\{\vec{\mathbf{0}}_3\right\}$, (b) a ***line*** through the origin, (c) a ***plane*** through the origin, and (d) all of $\mathbb{R}^3$.

If B and B' are any two ***bases*** for the same non-zero subspace $W \trianglelefteq \mathbb{R}^n$, then B and B' contain exactly the ***same number of vectors***. We call this number the ***dimension*** of W, and we write ***dim(W) = k***. We also say that W is ***k-dimensional***.

By convention, $W = \left\{\vec{\mathbf{0}}_n\right\}$ has no basis, and $dim\left(\left\{\vec{\mathbf{0}}_n\right\}\right) = 0$. Conversely, $dim(W)$ is a ***positive integer*** for a ***non-zero*** subspace W.

1.7 Exercises

1. Find a basis for the line through the origin in $\mathbb{R}^2$ with equation $y = \frac{5}{7}x$, consisting of a single vector with integer coordinates, whose entries are as small as possible.
2. Explain why the line with equation $y = \frac{3}{4}x + 2$ is ***not*** a subspace of $\mathbb{R}^2$.
3. Find a basis for the plane $\Pi : 3x - 7y + 4z = 0$ in $\mathbb{R}^3$, consisting of two vectors with integer coordinates, where one component in each vector is 0. There is more than one correct answer.
4. Find a basis for the plane $\Pi : 2x - 5z = 0$ in $\mathbb{R}^3$, consisting of two vectors with integer coordinates, where at least one component in each vector is 0.
5. Explain why the plane with equation $5x + 2y - 8z = 7$ is ***not*** a subspace of $\mathbb{R}^3$.

For Exercises (6) to (13): State a basis for $W = Span(S)$, where S is the set of vectors in the indicated Exercise from Section 1.6, and state $dim(W)$. You may use your computations or the Answer Key for Section 1.6:

6. Exercise 29.
7. Exercise 30.
8. Exercise 31.
9. Exercise 32.
10. Exercise 33.
11. Exercise 34.
12. Exercise 35.
13. Exercise 36.

For Exercises (14) to (25): Use the ***Minimizing Theorem*** (Basis for a Subspace Version) to find a basis for the subspace $W = Span(S)$, for each of the sets S below. State $dim(W)$. Use technology if permitted by your instructor.

14. $S = \{\langle 5,-3,6,7\rangle, \langle 3,-1,4,5\rangle, \langle 5,1,8,11\rangle, \langle 1,5,4,7\rangle\}$
15. $S = \{\langle 5,-3,6,7\rangle, \langle 3,-1,4,5\rangle, \langle 7,-9,6,5\rangle, \langle 5,1,8,-3\rangle\}$
16. $S = \{\langle 5,-3,6,7\rangle, \langle 3,-1,4,5\rangle, \langle 7,-5,8,9\rangle, \langle 1,3,-1,1\rangle, \langle 1,3,-9,-7\rangle\}$
17. $S = \{\langle 7,5,-4,3,9\rangle, \langle 4,3,-2,1,5\rangle, \langle 3,1,-4,7,5\rangle, \langle -8,-7,2,3,-9\rangle\}$
18. $S = \{\langle 7,5,-4,3,9\rangle, \langle 4,3,-2,1,5\rangle, \langle 1,2,2,-6,0\rangle, \langle 4,3,-5,9,5\rangle, \langle 5,2,0,-6,8\rangle\}$
19. $S = \{\langle 7,5,-4,3,9\rangle, \langle 4,3,-2,1,5\rangle, \langle 4,3,-5,4,5\rangle, \langle 4,5,-16,9,3\rangle, \langle 2,3,-7,2,1\rangle\}$
20. $S = \{\langle 5,-3,7,-4,6,3\rangle, \langle 9,-7,8,-9,4,7\rangle, \langle 4,-5,-3,-6,-7,5\rangle,$
    $\langle 3,2,14,2,19,-2\rangle, \langle 7,-6,-1,-7,-3,6\rangle\}$
21. $S = \{\langle 7,-3,4,2,-5,2\rangle, \langle 5,-2,3,3,-4,1\rangle, \langle 5,-3,2,-8,-1,4\rangle,$
    $\langle 7,-4,3,-9,-2,5\rangle, \langle -4,1,-3,-8,5,1\rangle\}$

22. $S = \{\langle 7,-3,4,2,-5,2\rangle, \langle 5,-2,3,3,-4,1\rangle, \langle 5,-3,2,-8,-1,4\rangle,$
$\langle 6,-4,3,-9,-2,5\rangle, \langle -4,1,-3,-8,5,1\rangle\}$

Note: the only change from the previous Exercise is in the 1st component of the 4th vector, which was changed from 7 to 6.

23. $S = \{\langle 7,-3,4,2,-5,2\rangle, \langle 5,-2,3,3,-4,1\rangle, \langle 5,-3,2,-8,-1,4\rangle,$
$\langle 7,-4,3,-9,-2,5\rangle, \langle -4,1,-3,-2,4,-1\rangle, \langle 1,-5,-4,-1,6,0\rangle\}$

Note: the first three vectors are identical to those in the last two Exercises.

24. $S = \{\langle 7,-3,4,2,-5,2\rangle, \langle 5,-2,3,3,-4,1\rangle, \langle 5,-3,2,-8,-1,4\rangle,$
$\langle 8,-4,3,-9,-2,5\rangle, \langle -4,1,-3,-2,4,-1\rangle, \langle 1,-5,-4,-1,6,0\rangle\}$

Note: The only change from the previous Exercise is in the 1st component of the 4th vector, which was changed from 7 to 8.

25. $S = \{\langle 7,-3,4,2,-5,2\rangle, \langle 5,-2,3,3,-4,1\rangle, \langle 5,-3,2,-8,-1,3\rangle,$
$\langle 8,-4,3,-9,-2,5\rangle, \langle -4,1,-3,-2,4,-1\rangle, \langle 1,-5,-4,-1,6,0\rangle\}$

Note: The only change from the previous Exercise is in the last component of the 3rd vector, which was changed from 4 to 3.

26. Show that $W = \{\langle x,y,z\rangle \in \mathbb{R}^3 \,|\, y = 0\}$ is a subspace of $\mathbb{R}^3$. Describe in words what this subspace is. Find a basis for W consisting of vectors with integer coordinates, and state its dimension.

27. Show that $W = \left\{\langle x,y,z\rangle \in \mathbb{R}^3 \,|\, y = 0 \textbf{\textit{ and }} z = 0\right\}$ is a subspace of $\mathbb{R}^3$. Describe in words what this subspace is, and find a basis for W consisting of vectors with integer coordinates, and state its dimension.

28. Suppose that $W = \left\{\langle x,y,z\rangle \in \mathbb{R}^3 \,|\, y = 0 \textbf{\textit{ or }} z = 0\right\}$. Decide whether or not W is a subspace of $\mathbb{R}^3$. Note the difference between this Exercise and the previous one.

29. Show that $W = \left\{\langle x_1,x_2,x_3,x_4\rangle \in \mathbb{R}^4 \,|\, x_1 = 5x_3 \text{ and } x_2 = -x_4\right\}$ is a subspace of $\mathbb{R}^4$. Find a basis for W consisting of vectors with integer coordinates, and state its dimension.

30. Show that $W = \left\{\langle x_1,x_2,x_3,x_4,x_5\rangle \in \mathbb{R}^5 \,|\, x_2 - 5x_3 = 6x_4 \text{ and } x_1 = -7x_5\right\}$ is a subspace of $\mathbb{R}^5$. Find a basis for W consisting of vectors with integer coordinates, and state its dimension.

31. Explain why $W = \left\{\langle x_1,x_2,x_3,x_4\rangle \in \mathbb{R}^4 \,|\, x_1 = 2x_4 - 3 \text{ and } x_3 = 0\right\}$ is ***not*** a subspace of $\mathbb{R}^4$.

32. Suppose that $W = \left\{\langle x_1,x_2\rangle \in \mathbb{R}^2 \,|\, x_1 \text{ and } x_2 \text{ are } \textbf{\textit{integers}}\right\}$. Decide if W is a subspace of $\mathbb{R}^2$.

33. Show that in any Euclidean space $\mathbb{R}^n$: (a) The set $\left\{\vec{\mathbf{0}}_n\right\}$ consisting only of the zero vector is a subspace of $\mathbb{R}^n$. (b) $\mathbb{R}^n$ is itself a subspace of $\mathbb{R}^n$. Hint: show that both subsets are non-empty and closed under vector addition and scalar multiplication.

34. Prove that for any subspace W of $\mathbb{R}^n$: $W = \mathbb{R}^n$ ***if and only if*** $\mathit{dim}(W) = n$.
Hint: the forward direction is obvious. For the other direction, use Proof by Contradiction and The Extension Theorem.

35. Prove that a non-empty ***subset*** U of $\mathbb{R}^n$ is a ***subspace*** of $\mathbb{R}^n$ ***if and only if*** $U = \mathit{Span}(S)$ for some non-empty subset S of $\mathbb{R}^n$. Warning: read this statement several times before attempting to

prove it. There is one special Case that you also have to treat separately.

36. ***Tying Up A Loose End:*** Prove the last Theorem from Section 1.2: If $\vec{u}$, $\vec{v}$ and $\vec{w} \in \mathbb{R}^3$ are three ***non-parallel*** vectors which are ***not coplanar***, that is, none of these vectors is on the plane determined by the two others, then: $Span(\{\vec{u}, \vec{v}, \vec{w}\}) = \mathbb{R}^3$.

    In other words, any vector $\vec{z} \in \mathbb{R}^3$ can be expressed as a ***linear combination:*** $\vec{z} = r\vec{u} + s\vec{v} + t\vec{w}$, for some scalars r, s and t. Hint: interpret this Theorem in terms of basis and dimension. The Extension Theorem can be useful.

37. Suppose that $W \trianglelefteq \mathbb{R}^n$ with $dim(W) = m$, and $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is a subset of W. Prove that S is a basis for W ***if and only if*** we can find ***at least one*** vector $\vec{b} \in W$ such that the equation:

$$\vec{b} = c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_m\vec{w}_m$$

    has ***exactly one*** solution in $c_1, c_2, \ldots, c_m$. Hint: what does this say about the augmented matrix that represents this equation? Which direction (forwards or backwards) is obvious?

38. ***Equivalent Conditions for a Basis of a Subspace:*** Suppose that $W \trianglelefteq \mathbb{R}^n$ is a non-zero subspace, and $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is any subset of W (we are ***not*** assuming that S is a basis for W, and neither are we assuming that $dim(W) = m$). Prove the following statements:

    a. S is a basis for W ***if and only if*** S is a ***maximal*** linearly independent subset of W. This means that if $S^/$ is another subset of W, with more vectors than S, then $S^/$ must be linearly ***dependent***. Hint: use Proof by Contradiction and the Extension Theorem to show that S must also Span W in the converse direction.

    b. S is a basis for W ***if and only if*** S is a ***minimal*** Spanning set of W. This means that if $S^{//}$ is another subset of W, with fewer vectors than S, then $S^{//}$ ***cannot Span*** W. Hint: use Proof by Contradiction and the Elimination Theorem to show that S must also be linearly independent in the converse direction.

    c. S is a basis for W ***if and only if*** for ***every*** $\vec{w} \in W$, the equation:

$$\vec{w} = c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_m\vec{w}_m$$

    has ***exactly one*** solution in $c_1, c_2, \ldots, c_m$.

    This is called the ***Uniqueness of Representation Property*** of a basis.

    Note/Hint: Why is this different from Exercise 37?

    d. Use (a) to show that $dim(W) = k$ ***if and only if*** there exists a maximal linearly independent subset of W consisting of k vectors.

    e. Use (b) to show that $dim(W) = k$ ***if and only if*** there exists a minimal Spanning subset of W consisting of k vectors.

39. ***Nested Subspaces:*** Suppose that V and W are both subspaces of $\mathbb{R}^n$. We say that V is a ***subspace*** of W, and write $V \trianglelefteq W$, if as ***subsets*** of $\mathbb{R}^n$, $V \subseteq W$. We say that V is ***nested*** inside W, or that $V \trianglelefteq W \trianglelefteq \mathbb{R}^n$ is a ***nesting of subspaces***.

    a. Let L be the line $Span(\{\langle 4,-2,-11\rangle\})$, and Π the plane $3x - 5y + 2z = 0$, both subspaces of $\mathbb{R}^3$. Show that $L \trianglelefteq \Pi \trianglelefteq \mathbb{R}^3$.

    b. Prove in general that if $V \trianglelefteq W \trianglelefteq \mathbb{R}^n$, then $dim(V) \leq dim(W)$. Furthermore, $dim(V) = dim(W)$ ***if and only if*** $V = W$. Hint: imitate the proof of The Existence of a Basis Theorem (starting with V), and use the Extension Theorem to construct a basis for W.

1.8 The Fundamental Matrix Spaces

We will now define four important subspaces associated with any matrix:

Definitions/Theorem — The Four Fundamental Matrix Spaces:
Let A be an $m \times n$ matrix. The ***rowspace*** of A is the Span of the rows of A.
The ***columnspace*** of A is the Span of the columns of A.
The ***nullspace*** of A is the set of all solutions to $A\vec{x} = \vec{\mathbf{0}}_m$:

$$\begin{aligned} rowspace(A) &= Span(\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m\}), \\ colspace(A) &= Span(\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}), \text{ and} \\ nullspace(A) &= \left\{ \vec{x} \in \mathbb{R}^n \mid A\vec{x} = \vec{\mathbf{0}}_m \right\}, \end{aligned}$$

where $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m$ are the rows of A (considered as vectors from $\mathbb{R}^n$),
and $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ are the columns of A (considered as vectors from $\mathbb{R}^m$).
Let us define the ***transpose*** matrix operation, where $A^\top$ ("***A transpose***") is the $n \times m$ matrix obtained from A by writing row 1 of A as column 1 of $A^\top$, writing row 2 of A as column 2 of $A^\top$, and so on. Consequently, column 1 of A becomes row 1 of $A^\top$ as well, and so on.
The fourth fundamental matrix space is:

$$nullspace(A^\top) = \left\{ \vec{y} \in \mathbb{R}^m \mid A^\top\vec{y} = \vec{\mathbf{0}}_n \right\}.$$

Under these definitions, the subspaces and the corresponding ambient spaces are:

$$rowspace(A) = colspace(A^\top) \trianglelefteq \mathbb{R}^n, \quad colspace(A) = rowspace(A^\top) \trianglelefteq \mathbb{R}^m,$$
$$nullspace(A) \trianglelefteq \mathbb{R}^n, \text{ and } nullspace(A^\top) \trianglelefteq \mathbb{R}^m.$$

Proof: The rowspace and columnspace of A are expressed as Spans, and are therefore subspaces of their respective ambient spaces. Since each row has n entries, and each column has m entries, we obtain the ambient spaces: $rowspace(A) \trianglelefteq \mathbb{R}^n$ and $colspace(A) \trianglelefteq \mathbb{R}^m$.
Thus, we only need to show that the nullspace of A is a subspace of $\mathbb{R}^n$ (the proof is similar for $A^\top$). The nullspace is non-empty, since it always contains $\vec{\mathbf{0}}_n$. Next, we must show that if $\vec{x}_1$ and $\vec{x}_2$ are both members of the nullspace of A, and k is any scalar, then $\vec{x}_1 + \vec{x}_2$ and $k\vec{x}_1$ are also members of the nullspace. So suppose $A\vec{x}_1 = \vec{\mathbf{0}}_m$ and $A\vec{x}_2 = \vec{\mathbf{0}}_m$. Then:

$$A(\vec{x}_1 + \vec{x}_2) = A\vec{x}_1 + A\vec{x}_2 = \vec{\mathbf{0}}_m + \vec{\mathbf{0}}_m = \vec{\mathbf{0}}_m,$$

and similarly, $A(k\vec{x}_1) = k(A\vec{x}_1) = k\vec{\mathbf{0}}_m = \vec{\mathbf{0}}_m$. Thus $\vec{x}_1 + \vec{x}_2$ and $k\vec{x}_1$ are both in $nullspace(A)$.■

We will now find a basis for each of these matrix spaces. As expected, the rref of A (and $A^\top$) will be useful, because of the following:

Theorem — Basis for the Rowspace:
Elementary row operations do not change the rowspace of a matrix. Thus, if B is obtained from A using an elementary row operation, then $rowspace(A) = rowspace(B)$.
Consequently, if R is the rref of A, then the ***non-zero*** rows of R form a ***basis*** for $rowspace(A)$.

The proof will be left as an Exercise. Next, we can rephrase The Minimizing Theorem from the previous Section to help us describe the columnspace:

Theorem — The Minimizing Theorem (Basis for the Columnspace Version):
If an $m \times n$ matrix A has reduced row echelon form R, then the columns of A that correspond to the ***leading columns*** of R form a ***basis*** for $colspace(A)$.

Finally, we know that the rref R also lets us describe $nullspace(A)$:

$$A\vec{x} = \vec{\mathbf{0}}_m \quad \textbf{\textit{if and only if}} \quad R\vec{x} = \vec{\mathbf{0}}_m.$$

As in Section 1.5, we write the equations obtained from R and solve for the leading variables in terms of the free variables. Recall that we were able to describe the solutions as a linear combination of several vectors, where each vector in the linear combination corresponds to a free variable x_i.
The vector we obtain from x_i contains 1 in the *ith* component. However, there will be a ***zero*** in the *ith* component of another vector corresponding to a ***different*** free variable x_j of R. Thus, a linear combination of these vectors can result in $\vec{\mathbf{0}}_m$ ***if and only if*** all the coefficients are 0. Thus, these vectors will ***Span*** the nullspace and will also be ***linearly independent***. This proves the following:

Theorem — Basis for the Nullspace:
Let A be an $m \times n$ matrix with reduced row echelon form R. Then:

$$nullspace(A) = nullspace(R).$$

Furthermore, if R has ***k free variables***, then $nullspace(A)$ will be ***k-dimensional***, and we obtain a basis for $nullspace(A)$ by solving for the leading variables in terms of the free variables, as usual. A similar equation applies to $A^\top$ and its rref.

Warning: We can ***directly*** use the entries of the rref of A to find a basis only for the ***rowspace*** and ***nullspace*** of A. However, we have to go back to the ***original*** columns of A to find a basis for the ***columnspace*** of A, using the leading ones to guide us.

Example: Suppose we have the matrix:

$$A = \begin{bmatrix} 5 & 7 & 1 & 5 & 6 & 0 \\ 3 & 4 & 1 & 4 & 3 & -1 \\ 1 & 2 & -1 & -2 & 3 & 3 \\ 2 & 4 & -2 & -4 & 5 & 2 \end{bmatrix} \text{ with rref } R = \begin{bmatrix} 1 & 0 & 3 & 8 & 0 & 5 \\ 0 & 1 & -2 & -5 & 0 & -7 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The three non-zero rows of R form a basis for $rowspace(A)$, and we write:

$$rowspace(A) = Span(\{\langle 1, 0, 3, 8, 0, 5 \rangle, \langle 0, 1, -2, -5, 0, -7 \rangle, \langle 0, 0, 0, 0, 1, 4 \rangle\}) \trianglelefteq \mathbb{R}^6.$$

These rows are linearly independent because we only find ***zeroes*** above and below each leading 1. Thus, to produce a linear combination that adds up to the zero vector, all coefficients must be zero. Let us see this concretely: if we create the dependence test equation for these three vectors, we get:

$$\begin{array}{cccccccccc} & c_1 & < \mathbf{1}, & 0, & 3, & 8, & 0, & 5 & > & \\ + & c_2 & < 0, & \mathbf{1}, & -2, & -5, & 0, & -7 & > & \\ + & c_3 & < 0, & 0, & 0, & 0, & \mathbf{1}, & 4 & > & = \vec{\mathbf{0}}_6. \end{array}$$

If c_1 were non-zero, we would have a non-zero entry in the 1st component, due to the leading 1 in row 1. Thus, c_1 must be zero. By the same reasoning, c_2 must also be zero because of the leading 1 in the 2nd component of the 2nd row. Finally, c_3 must also be zero by the Zero-Factors Theorem.

We can also verify that each of the ***original*** rows of A can be expressed as a linear combination of these three vectors. Again, thanks to the placement of the leading 1's, we can easily eyeball the correct coefficients. For example, row 1 of A can be written as:

$$\begin{bmatrix} \boxed{5} & \boxed{7} & 1 & 5 & \boxed{6} & 0 \end{bmatrix} = \mathbf{5}\begin{bmatrix} \boxed{1} & 0 & 3 & 8 & 0 & 5 \end{bmatrix} + \mathbf{7}\begin{bmatrix} 0 & \boxed{1} & -2 & -5 & 0 & -7 \end{bmatrix} + \mathbf{6}\begin{bmatrix} 0 & 0 & 0 & 0 & \boxed{1} & 4 \end{bmatrix}.$$

Thus, these three rows Span $rowspace(A)$. Since we already know that they are linearly independent, we have verified that they form a basis for $rowspace(A)$.

Now let us investigate $colspace(A)$. We will denote the ***original*** columns of A as $\vec{c}_1$ through $\vec{c}_7$. The rref R tells us that the leading variables correspond to columns 1, 2 and 5, and the free variables correspond to columns 3, 4, and 6. Thus, $\vec{c}_1$, $\vec{c}_2$ and $\vec{c}_5$ form a basis for $colspace(A)$, and we write:

$$colspace(A) = Span(\{\vec{c}_1, \vec{c}_2, \vec{c}_5\}) = Span(\{\langle 5, 3, 1, 2\rangle, \langle 7, 4, 2, 4\rangle, \langle 6, 3, 3, 5\rangle\}) \trianglelefteq \mathbb{R}^4.$$

Notice that we wrote the columns horizontally as ***vectors*** in $\mathbb{R}^4$. The Minimizing Theorem tells us that not only are these three columns linearly independent, but we can also write $\vec{c}_2$, $\vec{c}_4$, $\vec{c}_6$ and $\vec{c}_7$ in terms of these three columns, using the coefficients in these columns in R:

$$\begin{aligned} \vec{c}_3 &= 3\vec{c}_1 - 2\vec{c}_2 \\ \vec{c}_4 &= 8\vec{c}_1 - 5\vec{c}_2, \quad \text{and} \\ \vec{c}_6 &= 5\vec{c}_1 - 7\vec{c}_2 + 4\vec{c}_5. \end{aligned}$$

This verifies that $\vec{c}_1$, $\vec{c}_2$ and $\vec{c}_5$ Span $colspace(A)$. Since we already know that they are linearly independent, we have verified that they form a basis for $colspace(A)$.

Next, to find a basis for $nullspace(A)$, we set up the three homogeneous equations represented by R:

$$\begin{array}{rcrcrcrcrcl} x_1 & & & + & 3x_3 & + & 8x_4 & & & + & 5x_6 = 0 \\ & & x_2 & - & 2x_3 & - & 5x_4 & & & - & 7x_6 = 0 \\ & & & & & & & & x_5 & + & 4x_6 = 0 \end{array}$$

The leading variables are x_1, x_2, and x_5, and the free variables are x_3, x_4, and x_6. We solve for the leading variables in terms of the free variables:

$$x_1 = -3x_3 - 8x_4 - 5x_6,$$
$$x_2 = 2x_3 + 5x_4 + 7x_6, \quad \text{and}$$
$$x_5 = -4x_6.$$

Instead of using a parameter, such as $x_3 = r$, let us just use the names of the free variables. Thus, we will write our solutions as:

$$\begin{aligned}
&\langle x_1, x_2, x_3, x_4, x_5, x_6 \rangle \\
&= \langle -3x_3 - 8x_4 - 5x_6, 2x_3 + 5x_4 + 7x_6, x_3, x_4, -4x_6, x_6 \rangle \\
&= \langle -3x_3, 2x_3, x_3, 0, 0, 0 \rangle + \langle -8x_4, 5x_4, 0, x_4, 0, 0 \rangle + \langle -5x_6, 7x_6, 0, 0, -4x_6, x_6 \rangle \\
&= x_3 \big\langle -3, 2, 1, 0, 0, 0 \big\rangle + x_4 \langle -8, 5, 0, 1, 0, 0 \rangle + x_6 \big\langle -5, 7, 0, 0, -4, 1 \big\rangle.
\end{aligned}$$

where the free variables x_3, x_4, and x_6 can be ***any*** real numbers. Thus, the three vectors above Span *nullspace*(A), and we write:

$$\mathit{nullspace}(A) = \mathit{Span}(\{\langle -3, 2, 1, 0, 0, 0 \rangle, \langle -8, 5, 0, 1, 0, 0 \rangle, \langle -5, 7, 0, 0, -4, 1 \rangle\}) \trianglelefteq \mathbb{R}^6.$$

If we were to arrange these four vectors on top of each other in a dependence test equation, just like we did for the basis for *rowspace*(A), we can see that a similar pattern appears: there is a 1 in the component corresponding to the ***free variable*** x_i that produced that vector, and there are only zeroes ***above*** and ***below*** that 1:

$$\begin{array}{cccccccccc}
 & x_3 & < & -3, & 2, & \mathbf{1}, & 0, & 0, & 0 & > \\
+ & x_4 & < & -8, & 5, & 0, & \mathbf{1}, & 0, & 0 & > \\
+ & x_6 & < & -5, & 7, & 0, & 0, & -4, & \mathbf{1} & > & = & \vec{\mathbf{0}}_6.
\end{array}$$

Using the same logic as we applied to the rowspace, x_3 has to be zero, otherwise we get a non-zero entry in the 3rd component of the sum, thanks to the 1 in the 3nd component of the 1st vector. Now that $x_3 = 0$, we must also have $x_4 = 0$ because of the 1 in the 4th component of the 2nd vector. Finally, $x_6 = 0$ thanks to the Zero-Factors Theorem. Thus, this set both Spans *nullspace*(A) and is linearly independent, and so we have a ***basis*** for *nullspace*(A).

Lastly, to find a basis for *nullspace*$(A^\top)$, we will need $A^\top$ and its rref:

$$A^\top = \begin{bmatrix} 5 & 3 & 1 & 2 \\ 7 & 4 & 2 & 4 \\ 1 & 1 & -1 & -2 \\ 5 & 4 & -2 & -4 \\ 6 & 3 & 3 & 5 \\ 0 & -1 & 3 & 2 \end{bmatrix} \quad \text{with rref} \quad R' = \begin{bmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

If we refer to the four variables for the homogeneous system corresponding to $A^\top$ as y_1 through y_4, we can set up the three equations:

$$\begin{array}{rcrcrcl} y_1 & & + & 2y_3 & & = & 0 \\ & y_2 & - & 3y_3 & & = & 0 \\ & & & & y_4 & = & 0 \end{array}$$

The leading variables are y_1, y_2 and y_4, and the only free variable is y_3. We solve for the leading variables in terms of y_3, and we get: $y_1 = -2y_3$, $y_2 = 3y_3$ and $y_4 = 0$.

Thus, our solutions are:

$$\langle y_1, y_2, y_3, y_4 \rangle = \langle -2y_3, 3y_3, y_3, 0 \rangle = y_3 \Big\langle -2, 3, 1, 0 \Big\rangle.$$

Our basis for $nullspace(A^\top)$ therefore consists of a single vector, and we write:

$$nullspace(A^\top) = Span(\{\langle -2, 3, 1, 0 \rangle\}) \trianglelefteq \mathbb{R}^4. \square$$

Rank and Nullity

The dimensions of the Four Fundamental Matrix Spaces go by special names:

Definition/Theorem: Rank and Nullity:

Let A be an $m \times n$ matrix. The dimension of the ***nullspace*** of A is called the ***nullity*** of A.

The dimension of the ***rowspace*** of A is exactly the same as the dimension of the ***columnspace*** of A, and we call this common dimension the ***rank*** of A.

Furthermore, since $rowspace(A) = colspace(A^\top)$, and $colspace(A) = rowspace(A^\top)$, we can conclude that $rank(A) = rank(A^\top)$.

We write these dimensions symbolically as:

$$\begin{aligned} \textbf{rank(A)} &= dim(rowspace(A)) = dim(colspace(A)) = \textbf{rank(A}^\top\textbf{)}, \\ \textbf{nullity(A)} &= dim(nullspace(A)), \quad \text{and} \\ \textbf{nullity(A}^\top\textbf{)} &= dim(nullspace(A^\top)). \end{aligned}$$

Proof: All we need to show is that $dim(rowspace(A)) = dim(colspace(A))$. We saw that the non-zero rows of the rref R of A form a basis for $rowspace(A)$. Thus, the dimension of $rowspace(A)$ is the number of ***leading ones*** of R. However, the Minimizing Theorem says that the columns of A corresponding to the ***leading ones*** of R form a basis for the columnspace of A. Thus, the dimension of $colspace(A)$ is also the number of leading ones of R, and so these two dimensions are equal. ∎

Example: For the matrix in our previous Example, we found that:

$$\begin{aligned} rowspace(A) &= Span(\{\langle 1, 0, 3, 8, 0, 5 \rangle, \langle 0, 1, -2, -5, 0, -7 \rangle, \langle 0, 0, 0, 0, 1, 4 \rangle\}), \\ colspace(A) &= Span(\{\langle 5, 3, 1, 2 \rangle, \langle 7, 4, 2, 4 \rangle, \langle 6, 3, 3, 5 \rangle\}), \\ nullspace(A) &= Span(\{\langle -3, 2, 1, 0, 0, 0 \rangle, \langle -8, 5, 0, 1, 0, 0 \rangle, \langle -5, 7, 0, 0, -4, 1 \rangle\}), \text{ and} \\ nullspace(A^\top) &= Span(\{\langle -2, 3, 1, 0 \rangle\}). \end{aligned}$$

We also showed in this Example that each set of vectors above is a ***basis*** for the corresponding matrix space. Thus: $rank(A) = 3 = rank(A^\top)$, $nullity(A) = 3$ and $nullity(A^\top) = 1$. □

Simply staring at a matrix which is ***not*** in reduced row echelon form will usually not allow you to correctly guess its rank or nullity. However, we can set some ***bounds*** on these dimensions:

> ***Theorem/Definition — Bounds on Rank and Nullity:***
> Suppose A is an $m \times n$ matrix. Then:
> $$0 \leq rank(A) = rank(A^\top) \leq \min(m,n),$$
> $$n - m \leq nullity(A) \leq n, \quad \text{and} \quad m - n \leq nullity(A^\top) \leq m.$$
> We say that A has ***full-rank*** if $rank(A) = \min(m,n)$.

The symbol ***min(m,n)*** means the ***smaller of the two values*** m and n. For example, $\min(4,7) = 4$. We leave the details of the proofs for these inequalities as an Exercise. Notice also that if $m \geq n$, then $n - m \leq 0$, and thus we effectively just get $0 \leq nullity(A) \leq n$, since nullity cannot be negative. Likewise, if $m \leq n$, we just get $0 \leq nullity(A^\top) \leq m$.

Example: If A is a 5×9 matrix, then $0 \leq rank(A) = rank(A^\top) \leq 5 = \min(5,9)$. Since $9 - 5 = 4$, $nullity(A)$ is between 4 and 9, inclusively. However, $5 - 9 = -4$, and since $nullity(A^\top)$ cannot be negative, all we can conclude is that $0 \leq nullity(A^\top) \leq 5$. However, we already knew this since $nullspace(A^\top) \trianglelefteq \mathbb{R}^5$. Thus, our bounds give us no additional useful information about $nullity(A^\top)$. □

The Dimension Theorem for Matrices

We now present one of the central Theorems of Linear Algebra:

> ***Theorem — The Dimension Theorem for Matrices:***
> For any $m \times n$ matrix A:
> $$rank(A) + nullity(A) = n, \quad \text{and similarly,} \quad rank(A^\top) + nullity(A^\top) = m.$$

Proof: If R is the rref of A, the rank of A is the number of leading 1's of R, and the nullity of A is the number of free variables. But since we have n variables, and ***every*** variable is ***either leading or free*** (but ***not both***), the first equation follows. A similar argument applies to $A^\top$. ■

Example: In the previous Example, A is a 4×6 matrix with rank **3** and nullity **3**. The Dimension Theorem is thus verified, since $rank(A) + nullity(A) = 3 + 3 = 6 = n$.
A is ***not*** a full-rank matrix, since $\min(4,6) = 4 \neq 3$. □

Sight-Reading the Nullspace

We will now present a way to find a basis for the nullspace of a matrix ***by inspection***, without having to explicitly solve for our leading variables. Let us bring back the rref R and the basis that we found for $nullspace(A)$ in the previous Example:

$$R = \begin{bmatrix} \mathbf{1} & 0 & 3 & 8 & 0 & 5 \\ 0 & \mathbf{1} & -2 & -5 & 0 & -7 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \begin{aligned} \{ &< -3,\ 2,\ 1,\ 0,\ 0,\ 0 >, \\ &< -8,\ 5,\ 0,\ 1,\ 0,\ 0 >, \\ &< -5,\ 7,\ 0,\ 0,\ -4,\ 1 > \}. \end{aligned}$$

We begin by identifying the ***leading ones*** in the rref, which are highlighted, and the ***free variables***, as usual. Our free variables are x_3, x_4, and x_6. This tells us that we will need ***three*** vectors in our basis, corresponding to each free variable. We create a skeleton for these three vectors and write them on separate lines so we can align the six components properly. For each vector, place a ***one*** on the entry corresponding to the free variable, and put a ***zero*** on the entries ***above*** and ***below*** this one on the other vectors, because free variables do not affect each other. Put a zero as well on all coordinates to the ***right*** of each one, because a free variable only affects a leading variable to its ***left***. For example, x_4 only affects x_1 and x_2, but not x_5. Leave the other entries blank for now:

$$\begin{aligned} \{ &< \_\_,\ \_\_,\ \mathbf{1},\ 0,\ 0,\ 0 >, \\ &< \_\_,\ \_\_,\ 0,\ \mathbf{1},\ 0,\ 0 >, \\ &< \_\_,\ \_\_,\ 0,\ 0,\ \_\_,\ \mathbf{1} > \}. \end{aligned}$$

The entries in the 3rd column of R appear in the 1st basis vector for the nullspace, but with the ***opposite signs***, in the 1st and 2nd components. This happens because we ***solve*** for x_1 and x_2 from the homogeneous system, thus x_3 moves to the ***other side*** of the equation. Similarly, the entries in the 4th column of R appear in the 2nd basis vector, again in the 1st and 2nd components, but with opposite signs. Finally, the entries in the 6th column of R appear in the 3rd basis vector, but with the opposite signs, in the 1st, 2nd and 5th components.

In general, the entries in column i corresponding to the free variable x_i will appear in a single basis vector for *nullspace*(A), in the component corresponding to the ***leading one*** on that row, but with the ***opposite sign***. Thus, the "4" on column 6 appears as −4 in the 5th component of the final basis vector, because the leading 1 on the row containing 4 is in the 5th column. In this way, we complete our basis for *nullspace*(A) without having to solve the corresponding homogeneous system: just remember to ***reverse*** the signs and place the entry in the correct component. As a final check, the numbers appearing vertically in column i of R must now appear horizontally in the basis vector corresponding to x_i.

The General Solution of $A\vec{x} = \vec{b}$

The columnspace and nullspace of A are important because they are intrinsically connected to ***any*** matrix equation $A\vec{x} = \vec{b}$. The following can be proven using the Gauss-Jordan algorithm and two equivalent Theorems in Section 1.5, so we leave its proof as an Exercise:

Theorem — The Columnspace Test for Consistency:

The matrix equation $A\vec{x} = \vec{b}$ is ***consistent*** ***if and only if*** $\vec{b}$ is a member of ***colspace(A)***.

Furthermore, if $A\vec{x} = \vec{b}$ is consistent, suppose $\vec{x}_p$ is a ***fixed*** solution (also called a ***particular solution***) of this system. Then, a vector $\vec{x}$ is a solution of this system ***if and only if*** it can be written in the form: $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_h$ is a member of ***nullspace(A)***.

Consequently, if $\vec{x}$ and $\vec{y}$ are any two solutions to $A\vec{x} = \vec{b}$, then $\vec{x} - \vec{y} \in nullspace(A)$.

Recall that in Section 1.2, we defined the ***translate*** of a Span of vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$. Since all subspaces of $\mathbb{R}^n$ can be expressed as $Span(B)$ for some basis B, we will make the following equivalent definition:

Definition: If $\vec{b}$ is a fixed vector of $\mathbb{R}^n$, and W is a ***subspace*** of $\mathbb{R}^n$, then:
$\vec{b} + W = \left\{ \vec{b} + \vec{w} \mid \vec{w} \in W \right\}$ is called a ***translate*** of the subspace W.

From the Theorem above, the set of all solutions to $A\vec{x} = \vec{b}$ is a ***translate*** of the nullspace of A, and we can rewrite part of the previous Theorem as follows:

Theorem: If $\vec{b}$ is a member of ***colspace(A)***, the set X of all solutions $\vec{x}$ to $A\vec{x} = \vec{b}$ is a translate of ***nullspace(A)***, that is:

$$X = \vec{x}_p + nullspace(A),$$

where $\vec{x}_p$ is a fixed or ***particular solution*** for $A\vec{x} = \vec{b}$.

Example: Let us consider the system given by the augmented matrix:

$$\left[A \middle| \vec{b} \right] = \left[\begin{array}{ccccc|c} 3 & -15 & -5 & 1 & 3 & 2 \\ -2 & 10 & 3 & -2 & -2 & -3 \\ 4 & -20 & -5 & 8 & 3 & 5 \\ 2 & -10 & -4 & -2 & 2 & -2 \end{array} \right], \quad \text{with rref} \quad R = \left[\begin{array}{ccccc|c} \mathbf{1} & -5 & 0 & 7 & 0 & 3 \\ 0 & 0 & \mathbf{1} & 4 & 0 & 5 \\ 0 & 0 & 0 & 0 & \mathbf{1} & 6 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right].$$

The system is consistent, the leading variables are x_1, x_3 and x_5, and the free variables are x_2 and x_4. We solve the equations:

$$\begin{array}{ccccccccc} x_1 & - & 5x_2 & & & + & 7x_4 & & = & 3 \\ & & & & x_3 & + & 4x_4 & & = & 5 \\ & & & & & & & x_5 & = & 6 \end{array}$$

for x_1, x_3 and x_5, and we obtain the solutions:

$$\begin{aligned} \langle x_1, x_2, x_3, x_4, x_5 \rangle &= \langle 3 + 5x_2 - 7x_4, x_2, 5 - 4x_4, x_4, 6 \rangle \\ &= \langle 3, 0, 5, 0, 6 \rangle + x_2 \langle 5, 1, 0, 0, 0 \rangle + x_4 \langle -7, 0, -4, 1, 0 \rangle, \end{aligned}$$

where x_2 and x_4 are free. Thus, our particular solution can be $\vec{x}_p = \langle 3, 0, 5, 0, 6 \rangle$. Notice that 3, 5 and 6 are precisely the numbers on the ***right side*** of the rref. They appear in $\vec{x}_p$, but take note of their ***locations***. They are found exactly in the entry for the corresponding ***leading variable*** on that row.

If we ignore the rightmost column of R, we can sight-read a basis for the nullspace of A:

$$nullspace(A) = Span\left(\left\{ \langle 5, 1, 0, 0, 0 \rangle, \langle -7, 0, -4, 1, 0 \rangle \right\} \right).$$

Notice that these are the ***same vectors*** that appear in the other two terms of our solutions. We verify that they are not parallel, so they are linearly ***independent*** and thus form a ***basis*** for $nullspace(A)$. We can thus write:

$$\vec{x}_h = x_2 \langle 5, 1, 0, 0, 0 \rangle + x_4 \langle -7, 0, -4, 1, 0 \rangle.$$

Therefore, we can write our solution set X to $A\vec{x} = \vec{b}$ as:

$$X = \vec{x}_p + nullspace(A) = \langle 3, 0, 5, 0, 6 \rangle + Span\left(\left\{ \left\langle 5, 1, 0, 0, 0 \right\rangle, \left\langle -7, 0, -4, 1, 0 \right\rangle \right\}\right).$$

As a bonus, we verify the Dimension Theorem for the 4×5 matrix A: $rank(A) = 3$, $nullity(A) = 2$, and $3 + 2 = 5$, the number of columns of A. □

Properties of Full-Rank Matrices

Systems of linear equations involving full-rank matrices have special qualities, depending on their size:

Examples: Let us look at some full-rank matrices from each type according to size. Consider:

$$A_1 = \begin{bmatrix} -3 & -5 & -6 & 2 \\ 2 & 6 & -4 & -3 \\ 4 & 7 & 7 & -5 \end{bmatrix}, \; A_2 = \begin{bmatrix} -3 & -5 & 2 \\ 2 & 6 & -3 \\ 4 & 7 & -5 \end{bmatrix}, \text{ and } \; A_3 = \begin{bmatrix} 3 & 5 & -2 \\ -2 & 0 & 4 \\ 1 & -3 & -3 \\ 5 & 6 & -5 \end{bmatrix}.$$

A_1 is a 3×4 matrix (underdetermined), A_2 is a 3×3 matrix (square), and A_3 is a 4×3 matrix (overdetermined). Thus, any of them would be of full rank ***if and only if*** their rank is 3. Thus, we want to see 3 leading 1's in all the rrefs. Now, let us try to solve the systems:

$$A_1\vec{x} = \begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix}, \; A_2\vec{y} = \begin{bmatrix} -1 \\ -4 \\ 5 \end{bmatrix}, \text{ and } \; A_3\vec{z} = \begin{bmatrix} -1 \\ 4 \\ 2 \\ -3 \end{bmatrix}.$$

For the first system, we form the augmented matrix:

$$\left[\begin{array}{cccc|c} -3 & -5 & -6 & 2 & -1 \\ 2 & 6 & -4 & -3 & -4 \\ 4 & 7 & 7 & -5 & 5 \end{array}\right] \text{ with rref } R_1 = \left[\begin{array}{cccc|c} 1 & 0 & 7 & 0 & 4 \\ 0 & 1 & -3 & 0 & -3 \\ 0 & 0 & 0 & 1 & -2 \end{array}\right].$$

We see three leading 1's, so indeed A_1 has full rank. This system is ***consistent***, and in fact we have an ***infinite*** number of solutions because x_3 is a free variable. We can also conclude in this case that $A_1\vec{x} = \vec{b}$ is solvable for ***any*** $\vec{b} \in \mathbb{R}^3$, likewise with an infinite number of solutions, because the rref of $\left[A_1 \,|\, \vec{b} \right]$ will have exactly the ***same left side*** regardless of which $\vec{b}$ we put on the right side.

Similarly, let us solve the 2nd system using the augmented matrix:

$$\left[\begin{array}{ccc|c} -3 & -5 & 2 & -1 \\ 2 & 6 & -3 & -4 \\ 4 & 7 & -5 & 5 \end{array}\right] \text{ with rref } R_2 = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & -2 \end{array}\right].$$

Notice that A_2 is just A_1 but we deleted the 3rd column of A_1. Since the other three columns are linearly ***independent***, as seen in the columns of R_1, they remain independent when we assemble them in A_2. Again, we have three leading 1's, so A_2 is also of full rank. But this time, we do not have any free variable, and so the solution $\vec{y} = \langle 4, -3, -2 \rangle$ is ***unique***. Notice also that the left side of R_2 is $\boldsymbol{I_3}$, and the rref of $\left[A_2 \,|\, \vec{b} \right]$ will contain $\boldsymbol{I_3}$ on the left side for ***any*** $\vec{b} \in \mathbb{R}^3$. Thus, the system $A_2\vec{y} = \vec{b}$ will have ***exactly one*** solution for ***any*** $\vec{b} \in \mathbb{R}^3$.

Finally, let us form the augmented matrix:

$$\left[\begin{array}{ccc|c} 3 & 5 & -2 & -1 \\ -2 & 0 & 4 & 4 \\ 1 & -3 & -3 & 2 \\ 5 & 6 & -5 & -3 \end{array}\right] \text{ with rref } R_3 = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Again, we have three leading 1's, and so A_3 is also of full rank. There are no free variables in R_3 either, but notice that there is a ***row of zeroes***. The system is still ***consistent***, though, and the solution $\vec{z} = \langle 8, -3, 5 \rangle$ is also ***unique***. The row of zeroes on the left side, though, should tell us that there ***will*** be systems $\left[A_3 \,|\, \vec{b} \right]$ which are ***inconsistent***. For example, let us change just one component on the final column (say, change 4 to 5, which we show boxed below), and we get the augmented matrix:

$$\left[\begin{array}{ccc|c} 3 & 5 & -2 & -1 \\ -2 & 0 & 4 & \boxed{5} \\ 1 & -3 & -3 & 2 \\ 5 & 6 & -5 & -3 \end{array}\right] \text{ with rref } R_4 = \left[\begin{array}{ccc|c} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right].$$

Notice that the right-hand column of R_4 is now completely different from that of R_3. More significantly, the bottom row now tells us that this new system is ***inconsistent***.

In fact, since $colspace(A_3) \trianglelefteq \mathbb{R}^4$ and $rank(A_3) = 3 < 4$, there will be an ***infinite*** number of vectors $\vec{b} \in \mathbb{R}^4$ which are ***not*** in $colspace(A_3)$, and for these vectors, $\left[A_3 \,|\, \vec{b} \right]$ is ***inconsistent***.□

Let us generalize our observations above in the following, whose proof is an Exercise:

Theorem — Linear Systems with a Full-Rank Coefficient Matrix:

Suppose that $\left[A \,|\, \vec{b} \right]$ is an augmented matrix, where A is an $m \times n$ ***full-rank*** matrix. Then:

1. If $m < n$ (the system is ***underdetermined***), then the system is ***consistent*** for ***any*** $\vec{b} \in \mathbb{R}^m$, and furthermore, the system always has an ***infinite*** number of solutions.
2. If $m = n$ (the system is ***square***), then the rref of A is $\boldsymbol{I_n}$, and the system is ***consistent*** for ***any*** $\vec{b} \in \mathbb{R}^n$.
   Furthermore, the system has ***exactly one*** solution for every $\vec{b} \in \mathbb{R}^n$.
3. If $m > n$ (the system is ***overdetermined***), ***and*** the system is ***consistent***, then it has ***exactly one*** solution. However, there are an ***infinite number of vectors*** $\vec{b} \in \mathbb{R}^m$ for which the system is ***inconsistent***.
   Thus, we can also say that an overdetermined full-rank system has ***at most one*** solution.

1.8 Section Summary

The Four Fundamental Matrix Spaces: Let A be an $m \times n$ matrix. We define the subspaces:

$$\begin{aligned} \textbf{\textit{rowspace(A)}} &= Span(\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m\}) \trianglelefteq \mathbb{R}^{\boldsymbol{n}}, \\ \textbf{\textit{colspace(A)}} &= Span(\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}) \trianglelefteq \mathbb{R}^{\boldsymbol{m}}, \text{ and} \\ \textbf{\textit{nullspace(A)}} &= \left\{ \vec{x} \in \mathbb{R}^{\boldsymbol{n}} \mid A\vec{x} = \vec{\mathbf{0}}_{\boldsymbol{m}} \right\} \trianglelefteq \mathbb{R}^{\boldsymbol{n}}, \end{aligned}$$

where $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m$ are the rows of A (considered as vectors from $\mathbb{R}^{\boldsymbol{n}}$), and $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ are the columns of A (considered as vectors from $\mathbb{R}^{\boldsymbol{m}}$).

$A^{\top}$ (pronounced "*A* ***transpose***") is the $n \times m$ matrix obtained from A by writing row 1 of A as column 1 of $A^{\top}$, writing row 2 of A as column 2 of $A^{\top}$, and so on. The fourth fundamental matrix space is:

$$\textbf{\textit{nullspace}}(\boldsymbol{A}^{\top}) = \left\{ \vec{y} \in \mathbb{R}^{\boldsymbol{m}} \mid A^{\top}\vec{y} = \vec{\mathbf{0}}_{\boldsymbol{n}} \right\} \trianglelefteq \mathbb{R}^{\boldsymbol{m}}.$$

Let R be the rref of A. The ***non-zero rows*** of R form a basis for the rowspace of A.
The columns of A corresponding to the ***leading columns*** of R form a basis for the columnspace of A.
We can find a basis for $nullspace(A)$ from R, by solving for leading variables in terms of free variables, and expressing each member of $nullspace(A)$ as a linear combination of vectors, one for each ***free variable***.

The dimensions of these spaces are known by:

$$\begin{aligned} \textbf{\textit{rank(A)}} &= dim(rowspace(A)) = dim(colspace(A)) = \textbf{\textit{rank}}(\boldsymbol{A}^{\top}), \\ \textbf{\textit{nullity(A)}} &= dim(nullspace(A)), \text{ and } \textbf{\textit{nullity}}(\boldsymbol{A}^{\top}) = dim(nullspace(A^{\top})). \end{aligned}$$

The Dimension Theorem for Matrices:

For any $m \times n$ matrix A: $rank(A) + nullity(A) = n$, and $rank(A^{\top}) + nullity(A^{\top}) = m$.
We also have the bounds: $0 \le rank(A) = rank(A^{\top}) \le \min(m,n)$, $n - m \le nullity(A) \le n$, and $m - n \le nullity(A^{\top}) \le m$. We say that A has ***full-rank*** if $rank(A) = \min(m,n)$.

The Columnspace Test for Consistency:

The matrix equation $A\vec{x} = \vec{b}$ is ***consistent*** ***if and only if*** $\vec{b}$ is a member of ***colspace(A)***..
Furthermore, if $A\vec{x} = \vec{b}$ is consistent, suppose $\vec{x}_p$ is a ***fixed*** solution (also called a ***particular solution***) of this system. Then, a vector $\vec{x}$ is a solution of this system ***if and only if*** it can be written as: $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_h$ is a member of ***nullspace(A)***. In particular, if $\vec{x}$ and $\vec{y}$ are any two solutions to $A\vec{x} = \vec{b}$, then $\vec{x} - \vec{y}$ is a member of $nullspace(A)$. In other words, the set X of all solutions $\vec{x}$ of a consistent matrix equation of $A\vec{x} = \vec{b}$ is a translate of the nullspace, that is, $X = \vec{x}_p + nullspace(A)$, where $\vec{x}_p$ is a particular solution for $A\vec{x} = \vec{b}$.

Linear Systems with a Full-Rank Coefficient Matrix:

Suppose that $\left[A \,|\, \vec{b} \right]$ is an augmented matrix, where A is an $m \times n$ ***full-rank*** matrix. Then:

1. If $m < n$ (the system is ***underdetermined***), then the system is ***consistent*** for ***any*** $\vec{b} \in \mathbb{R}^{\boldsymbol{m}}$, and furthermore, the system always has an ***infinite*** number of solutions.
2. If $m = n$ (the system is ***square***), then the system is ***consistent*** for ***any*** $\vec{b} \in \mathbb{R}^{\boldsymbol{m}}$, and furthermore, the system has ***exactly one*** solution.
3. If $m > n$ (the system is ***overdetermined***), ***and*** the system is ***consistent***, then it has ***exactly one*** solution. However, there are ***an infinite number of vectors*** $\vec{b} \in \mathbb{R}^{\boldsymbol{m}}$ for which the system is ***inconsistent***. Thus, an overdetermined full-rank system has ***at most one*** solution.

1.8 Exercises

For Exercises (1) to (16): (a) Find a basis for *rowspace*(A), *colspace*(A), *nullspace*(A), and *nullspace*$(A^\top)$, if possible, for the matrix A that appears in the following Exercises from Section 1.6. You were given the rref R of A in Section 1.6, so you need only to compute $A^\top$ and its rref. Use technology if allowed by your instructor. The bases should consist of vectors with integer entries (we suggest that you also practice sight-reading a basis for the nullspaces). (b) State the rank and nullity of A and $A^\top$, and verify that the Dimension Theorem is satisfied by both A and $A^\top$. (c) Express the original rows of A in terms of the basis for *rowspace*(A) that you found in (a).

1. Exercise 10.
2. Exercise 12.
3. Exercise 14.
4. Exercise 16.
5. Exercise 17.
6. Exercise 19.
7. Exercise 20.
8. Exercise 22.
9. Exercise 23.
10. Exercise 25.
11. Exercise 28.
12. Exercise 29.
13. Exercise 30.
14. Exercise 31.
15. Exercise 33.
16. Exercise 34.

For Exercises (17) to (36): Suppose that $\left[A\,|\,\vec{b}\right]$ is a consistent system of equations whose rref is shown in the following Exercises from Section 1.4. Express all the solutions of the system, in the form $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_p$ is a fixed solution, and $\vec{x}_h$ is a member of the nullspace of A. Again, practice sight-reading the nullspace.

17. Exercise 7.
18. Exercise 8.
19. Exercise 10.
20. Exercise 11.
21. Exercise 14.
22. Exercise 15.
23. Exercise 17.
24. Exercise 19.
25. Exercise 20.
26. Exercise 22.
27. Exercise 23.
28. Exercise 25.
29. Exercise 27.
30. Exercise 29.
31. Exercise 30.
32. Exercise 32.
33. Exercise 33.
34. Exercise 34.
35. Exercise 35.
36. Exercise 36.

For Exercises (37) to (45): Express all the solutions of the system $A\vec{x} = \vec{b}$, in the form $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_p$ is a fixed solution, and $\vec{x}_h$ is a member of the nullspace of A. Use technology to find the rref, if permitted by your instructor.

37. $$\begin{array}{rcrcrcr} 2x & + & 3y & + & 7z & = & -10 \\ 3x & - & 7y & - & 47z & = & 77 \end{array}$$

38. $$\begin{array}{rcrcrcrcr} 2x_1 & - & x_2 & + & 4x_3 & - & 12x_4 & = & -8 \\ -3x_1 & + & 5x_2 & - & 2x_3 & - & 9x_4 & = & 17 \\ 7x_1 & - & 4x_2 & + & 3x_3 & + & 14x_4 & = & 13 \end{array}$$

39. $$\begin{array}{rcrcrcrcr} -3x_1 & + & 2x_2 & + & 20x_3 & + & 6x_4 & = & 36 \\ -2x_1 & + & 5x_2 & + & 39x_3 & + & 3x_4 & = & 4 \\ 3x_1 & - & 2x_2 & - & 20x_3 & + & 4x_4 & = & 54 \end{array}$$

40. $$\begin{array}{rcrcrcrcr} 3x_1 & - & 12x_2 & - & 2x_3 & + & x_4 & = & 13 \\ 2x_1 & - & 8x_2 & + & x_3 & + & 17x_4 & = & 4 \\ -6x_1 & + & 24x_2 & + & 5x_3 & + & 5x_4 & = & -28 \\ 4x_1 & - & 16x_2 & - & 3x_3 & - & x_4 & = & 18 \end{array}$$

41. $$\begin{aligned} 5x_1 + 6x_2 - 22x_3 - 27x_4 &= -3 \\ -x_1 - 3x_2 + 17x_3 + 9x_5 &= -3 \\ 7x_1 + 4x_2 - 51x_4 + 22x_5 &= -13 \end{aligned}$$

42. $$\begin{aligned} 2x_1 + 8x_2 - 12x_3 - 5x_4 + 29x_5 &= -10 \\ 5x_1 + 20x_2 - 30x_3 + 2x_4 + 29x_5 &= 33 \\ -3x_1 - 12x_2 + 18x_3 - x_4 - 18x_5 &= -19 \end{aligned}$$

43. $$\begin{aligned} -4x_1 + 20x_2 + 2x_3 - 20x_4 + 36x_5 &= -46 \\ 2x_1 - 10x_2 + 5x_3 - 2x_4 + 6x_5 &= -43 \\ 3x_1 - 15x_2 + 4x_3 + 4x_4 - 5x_5 &= -26 \\ -x_1 + 5x_2 - 2x_3 - x_5 &= 16 \end{aligned}$$

44. $$\begin{aligned} 7x_1 + 9x_2 - 4x_3 - 2x_4 - 30x_5 &= 4 \\ -4x_1 - 3x_2 + 2x_3 + x_4 + 6x_5 &= 5 \\ 6x_1 + 5x_2 + 3x_3 + 3x_4 + 8x_5 &= -21 \\ -2x_1 - 7x_2 + 4x_3 - 3x_4 &= -33 \end{aligned}$$

45. $$\begin{aligned} 9x_1 + 3x_2 + 36x_3 + 7x_4 - 2x_5 + 31x_6 &= 23 \\ 7x_1 + 4x_2 + 23x_3 + 2x_4 - 5x_5 + 31x_6 &= 14 \\ -2x_1 + 4x_2 - 22x_3 + 10x_4 - 5x_5 + 72x_6 &= -29 \\ 5x_1 + 2x_2 + 19x_3 - 7x_4 - 2x_5 - 21x_6 &= 31 \\ 3x_1 + 6x_2 - 3x_3 - 2x_4 - 5x_5 + 19x_6 &= 26 \end{aligned}$$

Assisted Computation: For Exercises (46) to (56): You are given a matrix A, its rref R, and the rref R' of $A^\top$. (a) Find a basis for *rowspace*(A), *colspace*(A), and *nullspace*(A), if possible, using only the information found in A and R. (b) Find a basis for *nullspace*$(A^\top)$ using only the information found in R'. Again, practice sight-reading the nullspaces. (c) State the rank and nullity of A and $A^\top$, and verify that the Dimension Theorem is satisfied by both A and $A^\top$.

46. $A = \begin{bmatrix} 3 & 2 & 8 & 9 \\ 5 & 7 & 6 & 4 \\ 16 & 29 & 6 & -7 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & 5 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$; $R' = \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

47. $A = \begin{bmatrix} 5 & 6 & 2 & 1 \\ -4 & -7 & 5 & 2 \\ 3 & 2 & 6 & 3 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & 0 \\ 0 & 1 & -3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$; $R' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

48. $A = \begin{bmatrix} 5 & 6 & 2 & 1 & 8 \\ 4 & 7 & -5 & -2 & 11 \\ 3 & 2 & 6 & 3 & 0 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & 0 & 6 \\ 0 & 1 & -3 & 0 & -3 \\ 0 & 0 & 0 & 1 & -4 \end{bmatrix}$; $R' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

49. $A = \begin{bmatrix} 3 & 4 & 0 & 3 & 7 \\ 5 & 7 & -1 & 4 & 3 \\ 1 & 2 & -2 & -1 & -15 \\ 4 & 3 & 7 & 2 & 7 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & 0 & 2 \\ 0 & 1 & -3 & 0 & -5 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$; $R' = \begin{bmatrix} 1 & 0 & -3 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

50. $A = \begin{bmatrix} 4 & 2 & 6 & 5 & 9 \\ 6 & 3 & 9 & 7 & 10 \\ 17 & 8 & 29 & 20 & 28 \\ 28 & 13 & 49 & 30 & 25 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 5 & 0 & -8 \\ 0 & 1 & -7 & 0 & 3 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$; $R' = \begin{bmatrix} 1 & 0 & 0 & -9 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

51. $A = \begin{bmatrix} 4 & 11 & 16 & 9 & 15 \\ 2 & 5 & 6 & 4 & 5 \\ 5 & 12 & 13 & 10 & 11 \\ 7 & 9 & -13 & 8 & -20 \\ 10 & 19 & 6 & 17 & 1 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & -7 & 0 & -9 \\ 0 & 1 & 4 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$; $R' = \begin{bmatrix} 1 & 0 & 0 & -6 & -3 \\ 0 & 1 & 0 & 3 & -4 \\ 0 & 0 & 1 & 5 & 6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

52. $A = \begin{bmatrix} 3 & 2 & 2 & -1 & -1 \\ 1 & 1 & -1 & -1 & 0 \\ 0 & 2 & -10 & -3 & 4 \\ -1 & -6 & 26 & 7 & -13 \\ -4 & -7 & 19 & 8 & -6 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & 0 & 0 \\ 0 & 1 & -5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$; $R' = \begin{bmatrix} 1 & 0 & 0 & -3 & 0 \\ 0 & 1 & 0 & 8 & 0 \\ 0 & 0 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

53. $A = \begin{bmatrix} 4 & 2 & 4 & -2 & 1 \\ 9 & 4 & 6 & -7 & 2 \\ 0 & 2 & 12 & 10 & 1 \\ 11 & 5 & 8 & -8 & 2 \\ -6 & -2 & 0 & 8 & 1 \\ -9 & -4 & -6 & 7 & 1 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & -2 & -3 & 0 \\ 0 & 1 & 6 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$ $R' = \begin{bmatrix} 1 & 0 & 9 & 0 & 5 & 3 \\ 0 & 1 & -4 & 0 & 2 & 5 \\ 0 & 0 & 0 & 1 & -4 & -6 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

54. $A = \begin{bmatrix} 3 & 1 & -4 & -2 & 0 \\ 0 & -1 & 7 & -2 & 5 \\ 12 & 1 & 5 & -14 & 15 \\ -1 & -1 & 6 & -1 & 4 \\ 12 & 0 & 12 & -17 & 22 \\ -1 & 0 & -1 & 4 & -6 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -7 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$; $R^{/} = \begin{bmatrix} 1 & 0 & 4 & 0 & 5 & 0 \\ 0 & 1 & 3 & 0 & 2 & 0 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

55. $A = \begin{bmatrix} 3 & 1 & 9 & 0 & -4 & 1 \\ 4 & -3 & -1 & 2 & -5 & 8 \\ 1 & -2 & -4 & 1 & -1 & 4 \\ -6 & 1 & -9 & -2 & 9 & -6 \\ -1 & 1 & 1 & 2 & -3 & -1 \\ 9 & 2 & 24 & -18 & 18 & -5 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 2 & 0 & 0 & 5 \\ 0 & 1 & 3 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$;

$R^{/} = \begin{bmatrix} 1 & 0 & 0 & 2 & 0 & 4 \\ 0 & 1 & 0 & -5 & 0 & -3 \\ 0 & 0 & 1 & 8 & 0 & 2 \\ 0 & 0 & 0 & 0 & 1 & -7 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

56. $A = \begin{bmatrix} 3 & -1 & 3 & 7 & 2 & 2 & 15 \\ -4 & 3 & 11 & 4 & 2 & 3 & -17 \\ -3 & 2 & 6 & 1 & 1 & 1 & -16 \\ 1 & 4 & 40 & 37 & 12 & 17 & 24 \\ -5 & 3 & 7 & -1 & 0 & 1 & -22 \end{bmatrix}$; $R = \begin{bmatrix} 1 & 0 & 4 & 5 & 0 & 0 & 3 \\ 0 & 1 & 9 & 8 & 0 & 0 & -4 \\ 0 & 0 & 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$;

$R^{/} = \begin{bmatrix} 1 & 0 & 0 & 3 & 0 \\ 0 & 1 & 0 & 5 & 0 \\ 0 & 0 & 1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

57. Use ***The Equality of Spans Theorem*** to show that if B is obtained from A using a single row operation, then $rowspace(A) = rowspace(B)$. You will need to consider the three types of elementary row operations. Hint: the Proof will be very similar to that of the Invariance of Solutions Theorem in Section 1.4, so review this Proof first.

58. Use the previous Exercise to show that the rowspace of A is the Span of the non-zero rows of the reduced row echelon form R of A.

59. Suppose that A is an $m \times n$ matrix. Prove that:

$$0 \leq rank(A) \leq \min(m,n),\ n - m \leq nullity(A) \leq n, \text{ and } m - n \leq nullity(A^{\top}) \leq m.$$

60. Prove that the matrix equation $A\vec{x} = \vec{b}$ is ***consistent*** ***if and only if*** $\vec{b} \in colspace(A)$.

61. Prove that every solution $\vec{x}$ of a consistent matrix equation $A\vec{x} = \vec{b}$ can be written as $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_p$ is a ***fixed*** solution of $A\vec{x} = b$ and $\vec{x}_h \in nullspace(A)$. Hint: compute $A(\vec{x} - \vec{x}_p)$. What does this imply?

62. Suppose that the matrix equation $A\vec{x} = \vec{b}$ is consistent, where A is an $m \times n$ matrix. Use the previous Exercise to prove that the equation has exactly one solution ***if and only if*** $nullspace(A) = \left\{\vec{\mathbf{0}}_n\right\}$.

63. Suppose that you are told that the rank of a matrix A is 5, the nullity is 8, and the columnspace of A is a subspace of $\mathbb{R}^6$. How big is A?

64. Suppose that A is an $a \times n$ matrix, and B is a $b \times n$ matrix, where a and b could possibly be different positive integers. Prove that $rank(A) \leq rank(B)$ ***if and only if*** $nullity(A) \geq nullity(B)$.

65. Suppose that A is an $n \times n$ matrix. Prove that $nullity(A) = nullity(A^{\top})$.

66. Suppose that $\left[A|\vec{b}\right]$ is an augmented matrix, where A is an $m \times n$ ***full-rank*** matrix.

    a. If $m < n$ (the system is ***underdetermined***), prove that the system is ***consistent*** for ***any*** $\vec{b} \in \mathbb{R}^m$, and furthermore, the system always has an ***infinite*** number of solutions.

    b. If $m = n$ (the system is ***square***), prove that the system is ***consistent*** for ***any*** $\vec{b} \in \mathbb{R}^m$, and furthermore, the system has ***exactly one*** solution.

    c. If $m > n$ (the system is ***overdetermined***), ***and*** the system is ***consistent***, prove that the system has ***exactly one*** solution.

    d. Prove that in the third case ($m > n$), there are ***an infinite number of vectors*** $\vec{b} \in \mathbb{R}^m$ for which the system is ***inconsistent***. Hint: start with $colspace(A)$, and apply the Extension Theorem to obtain a basis for $\mathbb{R}^m$. Why do we get an infinite number of vectors $\vec{b}$ for which $\left[A|\vec{b}\right]$ is inconsistent?

67. ***True or False:*** Determine whether each statement is true or false, and briefly explain your answer by citing a Theorem, providing a counterexample, or a convincing argument.

    a. If A is a 7×4 matrix, then A can have rank 5.

    b. If A is a 4×7 matrix, then A can have nullity 5.

    c. If A is a 7×4 matrix, then A can have nullity 5.

    d. If A is a 7×4 matrix, then $rank(A) + nullity(A) = 7$.

    e. If A is a 6×8 full rank matrix, then $nullity(A) = 2$.

    f. If A is a 5×8 full-rank matrix, then $\left[A|\vec{b}\right]$ is always consistent for any $\vec{b} \in \mathbb{R}^5$.

    g. If A is a 5×8 full-rank matrix, then $\left[A|\vec{b}\right]$ always has a unique solution for any $\vec{b} \in \mathbb{R}^5$.

h. If A is a 7×5 full-rank matrix, and $\left[A \,|\, \vec{b} \right]$ is consistent for some $\vec{b} \in \mathbb{R}^7$, then this system has a unique solution .

i. If A is a 7×5 full-rank matrix, and $\left[A \,|\, \vec{b} \right]$ is consistent for some $\vec{b} \in \mathbb{R}^7$, then $\left[A \,|\, \vec{d} \right]$ is consistent for ***any other*** $\vec{d} \in \mathbb{R}^7$.

j. If A is a 7×5 full-rank matrix, then $\left[A \,|\, \vec{b} \right]$ always has a unique solution for any $\vec{b} \in \mathbb{R}^7$.

k. If A is a 7×7 full-rank matrix, then $\left[A \,|\, \vec{b} \right]$ always has a unique solution for any $\vec{b} \in \mathbb{R}^7$.

l. If A is a 6×8 full-rank matrix, then $\left[A \,|\, \vec{b} \right]$ always has an infinite number of solutions for any $\vec{b} \in \mathbb{R}^6$.

m. If A is a 5×8 matrix, then $\left[A \,|\, \vec{b} \right]$ is always consistent for any $\vec{b} \in \mathbb{R}^5$.

n. If A is a 5×8 matrix, and $\left[A \,|\, \vec{b} \right]$ is consistent for some $\vec{b} \in \mathbb{R}^5$, then $\left[A \,|\, \vec{d} \right]$ is consistent for ***any other*** $\vec{d} \in \mathbb{R}^5$.

o. If A is a 7×5 matrix, then $\left[A \,|\, \vec{b} \right]$ always has a unique solution for any $\vec{b} \in \mathbb{R}^7$.

68. ***The Uniqueness of the Reduced Row Echelon Form:*** We are now in a position to prove that if A is an $m \times n$ matrix, and we obtain two matrices H and J from A using a finite sequence of elementary row operations, and both H and J are in reduced row echelon form, then $H = J$. Thus, the rref of A is ***unique***. We will use the Principle of Mathematical Induction.

a. First let us take care of the trivial case: If A consists entirely of zeroes, prove that $H = A = J$.
Thus we can assume for the rest of the Exercise that A is a ***non-zero*** matrix.

b. Explain why $rowspace(H) = rowspace(A) = rowspace(J)$.

c. Explain why the number of non-zero rows of H must be the same as the number of non-zero rows of J. Hint: what does this number represent?

Thus we can conclude that both H and J have r non-zero rows, for some positive number r. We must now show that every pair of corresponding rows are equal.

d. Numerical Warm-Up: both H and J below are in rref, and both of them have rank 3:

$$H = \begin{bmatrix} 1 & 5 & 0 & -4 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \quad J = \begin{bmatrix} 1 & 0 & -4 & 0 & 7 \\ 0 & 1 & 3 & 0 & 9 \\ 0 & 0 & 0 & 1 & -8 \end{bmatrix}.$$

Explain why $rowspace(H)$ is ***not*** the same as $rowspace(J)$, even though they have the same dimension. Use the Equality of Spans Theorem. Start with comparing the first rows.

e. The Basis Step for Induction: Let $\vec{h}_1$ be the first row of H and $\vec{j}_1$ the first row of J. Prove that the leading 1 of $\vec{h}_1$ is exactly in the same place as the leading 1 of $\vec{j}_1$. Hint: Keep in mind for this entire Exercise that the leading 1's below each row are to the ***right*** of those from the previous rows, and use The Equality of Spans Theorem.

f. (continued) Next, prove that $\vec{h}_1 = \vec{j}_1$, that is, the ***rest*** of the entries in the first row must also be exactly the same. Hint: use the same ideas as in part (e).

g. If $r = 1$, then we are done. Otherwise, assume by the Inductive Hypothesis that we have already shown that $\vec{h}_1 = \vec{j}_1$, $\vec{h}_2 = \vec{j}_2, \ldots,$ $\vec{h}_k = \vec{j}_k$ for some $k < r$, where h_i is row i of H and j_i is row i of J.
Complete the proof with the Inductive Step: show that $\vec{h}_{k+1} = \vec{j}_{k+1}$.

Section 1.9 Orthogonal Complements

A plane Π through the origin in $\mathbb{R}^3$ has equation $ax + by + cz = 0$, and we saw that this can be written as a dot product:

$$\langle a, b, c\rangle \circ \langle x, y, z\rangle = 0.$$

Thus, every vector $\langle x, y, z\rangle$ on Π is ***orthogonal*** to any vector on the normal line $L = Span(\{\langle a, b, c\rangle\})$. We will now generalize this idea:

> ***Definition/Theorem:*** If W is a subspace of $\mathbb{R}^n$, then $W^\perp$ (pronounced ***"W perp"***), the ***orthogonal complement*** of W, defined as:
>
> $$W^\perp = \left\{\vec{v} \in \mathbb{R}^n \mid \vec{v} \circ \vec{w} = 0 \text{ for } \textbf{\textit{all }} \vec{w} \in W\right\}$$
>
> is also a ***subspace*** of $\mathbb{R}^n$.

Proof: First, $W^\perp$ contains $\vec{\mathbf{0}}_n$, since $\vec{\mathbf{0}}_n \circ \vec{w} = 0$ for all $\vec{w} \in W$ (in fact, for all $\vec{w} \in \mathbb{R}^n$). Thus, $W^\perp$ is ***non-empty***. Next, suppose that $\vec{v}$ and $\vec{u}$ are vectors in $W^\perp$. Thus:

$$\vec{v} \circ \vec{w} = 0 \text{ and } \vec{u} \circ \vec{w} = 0 \text{ for all } \vec{w} \in W.$$

We must show that $\vec{v} + \vec{u}$ and $r\vec{v}$ are also vectors in $W^\perp$, for any $r \in \mathbb{R}$. Thus:

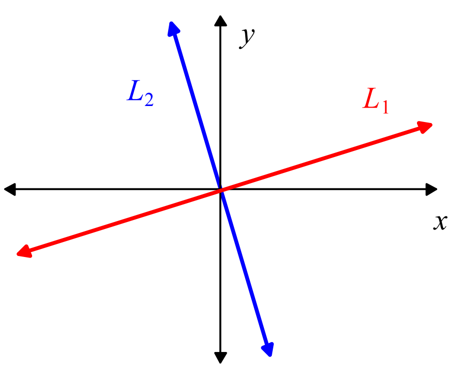
$$(\vec{v} + \vec{u}) \circ \vec{w} = \vec{v} \circ \vec{w} + \vec{u} \circ \vec{w} = 0 + 0 = 0 \quad \text{for all } \vec{w} \in W,$$

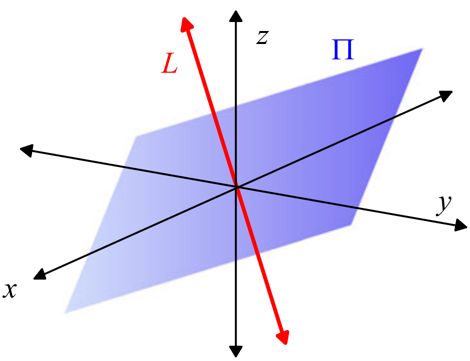
so again $\vec{v} + \vec{u}$ is a vector in $W^\perp$. Similarly:

$$(r\vec{v}) \circ \vec{w} = r(\vec{v} \circ \vec{w}) = r(0) = 0 \quad \text{for all } \vec{w} \in W.$$

Thus $W^\perp$ is ***closed*** under addition and scalar multiplication, and is a subspace of $\mathbb{R}^n$ ■

Examples: Let L_1 be any line in $\mathbb{R}^2$ passing through the origin. If $L_1 = Span(\{\vec{e}_1\})$, the x-axis, then clearly every vector on $L_2 = Span(\{\vec{e}_2\})$, the y-axis, is orthogonal to L_1, and vice versa. Thus $L_1^\perp = L_2$ and $L_2^\perp = L_1$. More generally, if L_1 has non-zero slope m, then the line L_2 with slope $m' = -1/m$ and passing through the origin is perpendicular to L_1, as we know from basic algebra:





$L_1 : y = mx,\ m \neq 0$, and $L_2 : y = -\dfrac{1}{m}x$ $\quad$ $\Pi : ax + by + cz = 0$, and $L = Span(\{\langle a,b,c\rangle\})$

Orthogonal Complements in $\mathbb{R}^2$ and $\mathbb{R}^3$

Similarly, let Π be any plane in $\mathbb{R}^3$ passing through the origin, with equation $ax + by + cz = 0$. Its normal vector is $\vec{n} = \langle a, b, c \rangle$. All the vectors on the normal line $L = Span(\{\vec{n}\})$ are perpendicular to all the vectors on Π, and vice versa, as represented in the diagram above on the right. Thus we can say that $\Pi^\perp = L$, and $L^\perp = \Pi$. $\square$

More generally, checking that a given vector $\vec{v}$ is orthogonal to ***all*** vectors in a subspace W seems like a monumental task. However, the following Theorem gives us a significant shortcut. Its proof will be left as an Exercise:

Theorem: If $W = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}) \trianglelefteq \mathbb{R}^n$, then:

$$W^\perp = \left\{ \vec{v} \in \mathbb{R}^n \,|\, \vec{v} \circ \vec{w}_i = 0 \text{ for all } i = 1 \ldots k \right\}.$$

Example: Suppose $S = \{\langle 1, 3, -2, 5 \rangle, \langle -2, 5, 7, -8 \rangle\}$ and $W = Span(S) \trianglelefteq \mathbb{R}^4$. Let us find a basis for $W^\perp$. We want to find all vectors $\langle x_1, x_2, x_3, x_4 \rangle$ so that:

$$\langle 1, 3, -2, 5 \rangle \circ \langle x_1, x_2, x_3, x_4 \rangle = 0, \text{ and}$$
$$\langle -2, 5, 7, -8 \rangle \circ \langle x_1, x_2, x_3, x_4 \rangle = 0.$$

In other words, we want to solve:

$$\begin{aligned} x_1 + 3x_2 - 2x_3 + 5x_4 &= 0 \quad \text{and} \\ -2x_1 + 5x_2 + 7x_3 - 8x_4 &= 0. \end{aligned}$$

This is a ***homogeneous*** system of equations, and the vectors that we want are precisely the vectors in $nullspace(A)$, where A is the coefficient matrix:

$$A = \begin{bmatrix} 1 & 3 & -2 & 5 \\ -2 & 5 & 7 & -8 \end{bmatrix} \text{ with rref } R = \begin{bmatrix} 1 & 0 & -\frac{31}{11} & \frac{49}{11} \\ 0 & 1 & \frac{3}{11} & \frac{2}{11} \end{bmatrix}.$$

The leading variables are x_1 and x_2, and the free variables are x_3 and x_4. This is a good opportunity to practice sight-reading the nullspace when ***fractions*** are involved. We need two vectors, one for x_3 and one for x_4, and we get:

$$\left\langle \frac{31}{11}, -\frac{3}{11}, \mathbf{1}, 0 \right\rangle \text{ and } \left\langle -\frac{49}{11}, -\frac{2}{11}, 0, \mathbf{1} \right\rangle.$$

Clearing the denominators from these two vectors, we can write:

$$W^\perp = Span(\{\langle 31, -3, 11, 0 \rangle, \langle -49, -2, 0, 11 \rangle\}).$$

We can easily check by taking dot products that every vector in our Spanning set is orthogonal to every vector in $S = \{\langle 1, 3, -2, 5 \rangle, \langle -2, 5, 7, -8 \rangle\}$, so we have confidence that our description is correct. For example, taking the dot product of the first vectors of each set, we get:

$$\langle 31, -3, 11, 0 \rangle \circ \langle 1, 3, -2, 5 \rangle = 31 - 9 - 22 = 0,$$

and similarly for the three other pairs. $\square$

We were able to describe $W^\perp$ by finding the ***nullspace*** of a coefficient matrix A in this Example. This will be true in general, but before we can prove this, we need a new point of view:

A Dot Product Perspective of Matrix Multiplication

One of the beauties of Mathematics is that we can sometimes look at the same object in ***different*** ways. The matrix product $A\vec{x}$ is a good example. We first defined $A\vec{x}$ as a ***linear combination*** of the ***columns*** of A using the ***coefficients*** from $\vec{x}$. Let us spell it out, entry by entry:

$$A\vec{x} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix} + x_2 \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{bmatrix} + \cdots + x_n \begin{bmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{bmatrix}$$

However, notice that the top entry of $A\vec{x}$ can be written as:

$$x_1 a_{1,1} + x_2 a_{1,2} + \cdots + x_n a_{1,n} = \langle x_1, x_2, \ldots, x_n \rangle \circ \langle a_{1,1}, a_{1,2}, \ldots a_{1,n} \rangle = \vec{x} \circ \vec{r}_1,$$

where $\vec{r}_1$ is the first row of A. Similarly, we can see that the second entry is $\vec{x} \circ \vec{r}_2$. Continuing thus, we get:

$$A\vec{x} = \begin{bmatrix} \vec{r}_1 \\ \vec{r}_2 \\ \vdots \\ \vdots \\ \vec{r}_m \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \vec{x} \circ \vec{r}_1 \\ \vec{x} \circ \vec{r}_2 \\ \vdots \\ \vdots \\ \vec{x} \circ \vec{r}_m \end{bmatrix} = \begin{bmatrix} \vec{r}_1 \circ \vec{x} \\ \vec{r}_2 \circ \vec{x} \\ \vdots \\ \vdots \\ \vec{r}_m \circ \vec{x} \end{bmatrix}.$$

From this, we can see that:

Theorem: A vector $\vec{x} \in \mathbb{R}^n$ is a solution to $A\vec{x} = \vec{\mathbf{0}}_m$ ***if and only if*** $\vec{x} \circ \vec{r}_i = 0$ for all the rows $\vec{r}_i$ of A. In other words, $\vec{x}$ is in the ***nullspace*** of A ***if and only if*** $\vec{x}$ is ***orthogonal*** to ***all*** the ***rows*** of A. Thus:

$$\text{If } W = rowspace(A), \text{ then } W^{\perp} = nullspace(A).$$

$$\text{Similarly, if } U = nullspace(A), \text{ then } U^{\perp} = rowspace(A).$$

This last Theorem shows us the relationship between ***rowspace(A)*** and ***nullspace(A)***, and also gives us an efficient ***algorithm*** to describe the orthogonal complement of a subspace:

Theorem: Suppose $W = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}) \trianglelefteq \mathbb{R}^n$. If we form the matrix A with ***rows*** $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$, then:

$$W = rowspace(A) \quad \text{and} \quad W^{\perp} = nullspace(A).$$

Thus, the ***non-zero rows*** of the ***rref*** of A form a ***basis*** for W, and we can obtain a ***basis*** for $W^{\perp}$ exactly as we would find a basis for ***nullspace(A)*** using the rref of A.

Note: This is the ***only*** place in this book where we assemble vectors into the ***rows*** of a matrix. The rest of the time, we will assemble vectors into the columns of a matrix.

Example: Let us consider the subspace $W = Span(S) \trianglelefteq \mathbb{R}^5$, where:

$$S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\} = \{\langle 3, -2, 1, 5, 0\rangle, \langle 5, -3, 2, 6, 1\rangle, \langle -8, 3, -5, 3, -7\rangle, \langle 4, 1, 0, -2, 3\rangle\}.$$

Notice that there are ***four*** vectors in S. Our main objective is to find a basis for $W^{\perp}$. To do this, we assemble the four vectors that generate W into the ***rows*** of a matrix, as prescribed by our Theorem:

$$A = \begin{bmatrix} 3 & -2 & 1 & 5 & 0 \\ 5 & -3 & 2 & 6 & 1 \\ -8 & 3 & -5 & 3 & -7 \\ 4 & 1 & 0 & -2 & 3 \end{bmatrix}, \quad \text{with rref:} \quad R = \begin{bmatrix} 1 & 0 & 0 & \frac{2}{5} & \frac{2}{5} \\ 0 & 1 & 0 & -\frac{18}{5} & \frac{7}{5} \\ 0 & 0 & 1 & -\frac{17}{5} & \frac{8}{5} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The leading variables are x_1, x_2 and x_3, and the free variables are x_4 and x_5. Again, let us sight-read the nullspace. We will need two vectors, corresponding to x_4 and x_5:

$$W^{\perp} = nullspace(A) = Span\left(\left\{\left\langle -\frac{2}{5}, \frac{18}{5}, \frac{17}{5}, \mathbf{1}, 0\right\rangle, \left\langle -\frac{2}{5}, -\frac{7}{5}, -\frac{8}{5}, 0, \mathbf{1}\right\rangle\right\}\right), \quad \text{or}$$
$$W^{\perp} = Span(\{\langle -2, 18, 17, 5, 0\rangle, \langle -2, -7, -8, 0, 5\rangle\}),$$

by clearing the denominators, as we did in the previous Example. Thus, $W^{\perp}$ is ***2-dimensional***.

There is, however, a bonus outcome from the rref. Since we assembled the vectors from S into the rows of A, then $W = Span(S) = rowspace(A)$. But we saw in the previous Section that the non-zero rows of R form a ***basis*** for $rowspace(A)$. But notice that there are only ***three*** non-zero rows of R. Thus, we can more efficiently say that:

$$W = Span\left(\left\{\left\langle 1, 0, 0, \frac{2}{5}, \frac{2}{5}\right\rangle, \left\langle 0, 1, 0, -\frac{18}{5}, \frac{7}{5}\right\rangle, \left\langle 0, 0, 1, -\frac{17}{5}, \frac{8}{5}\right\rangle\right\}\right), \quad \text{or}$$
$$W = Span(\{\langle 5, 0, 0, 2, 2\rangle, \langle 0, 5, 0, -18, 7\rangle, \langle 0, 0, 5, -17, 8\rangle\}),$$

again, by clearing denominators. Thus, W is only ***3-dimensional***.

Let us think about this some more: Since W is 3-dimensional, this means that the original Spanning set S which consists of ***four*** vectors has to be ***dependent***, by the Dependent Sets from Spanning Sets Theorem. It is far from obvious, though, how the four vectors are related to each other. The rref R only tells us the dependency relationships of the ***columns*** of A, but unfortunately, ***not*** its rows. If we really want to know how the four original vectors depend on each other, we would need to use ***The Minimizing Theorem*** and assemble the vectors in S into the columns of a matrix, say:

$$B = A^{\top} = \begin{bmatrix} 3 & 5 & -8 & 4 \\ -2 & -3 & 3 & 1 \\ 1 & 2 & -5 & 0 \\ 5 & 6 & 3 & -2 \\ 0 & 1 & -7 & 3 \end{bmatrix} \quad \text{with rref} \quad R' = \begin{bmatrix} 1 & 0 & 9 & 0 \\ 0 & 1 & -7 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Now we can see that $\vec{w}_3$ is a linear combination of the first two vectors: $\vec{w}_3 = 9\vec{w}_1 - 7\vec{w}_2$.
Thus, by the Minimizing Theorem, $S' = \{\vec{w}_1, \vec{w}_2, \vec{w}_4\}$ is another basis for W. □

You will prove the following properties of orthogonal complements in the Exercises.

Theorem — Properties of Orthogonal Complements:
For any subspace $W \trianglelefteq \mathbb{R}^n$:
a) $W \cap W^\perp = \{\vec{\mathbf{0}}_n\}$
b) $(W^\perp)^\perp = W$. Thus, we can say that W and $W^\perp$ are orthogonal complements of ***each other***, or that W and $W^\perp$ form an ***orthogonal pair*** of subspaces.

The Dimension Theorem for Matrices can be used to prove an analogous statement concerning a subspace W and its orthogonal complement $W^\perp$. Its proof is also left as an Exercise:

Theorem — The Dimension Theorem for Orthogonal Complements:
If W is a subspace of $\mathbb{R}^n$ with orthogonal complement $W^\perp$, then: $dim(W) + dim(W^\perp) = n$.

Example: Suppose that:

$$W = Span(\{\langle 15, -10, 5, 25, 0\rangle, \langle -9, 6, -3, -15, 0\rangle, \langle 1, -2, -5, 11, 8\rangle, \langle 5, -3, 3, 6, -2\rangle\}) \trianglelefteq \mathbb{R}^5.$$

For now, all we can say is that W is ***at most*** 4-dimensional. As before, we find $W^\perp$ by assembling the vectors into the ***rows*** of a matrix and finding its nullspace:

$$A = \begin{bmatrix} 15 & -10 & 5 & 25 & 0 \\ -9 & 6 & -3 & -15 & 0 \\ 1 & -2 & -5 & 11 & 8 \\ 5 & -3 & 3 & 6 & -2 \end{bmatrix} \quad \text{with rref:} \quad R = \begin{bmatrix} 1 & 0 & 3 & -3 & -4 \\ 0 & 1 & 4 & -7 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The leading variables are x_1 and x_2, and the free variables are x_3, x_4 and x_5. Since there are two non-zero rows in R and three free variables, $dim(W) = 2$ and $dim(W^\perp) = 3$. This verifies that $dim(W) + dim(W^\perp) = 2 + 3 = 5$, the dimension of the ambient space $\mathbb{R}^5$. The non-zero rows of R will be a basis for W, and by sight-reading the nullspace, we get a basis for $W^\perp$:

$$W = Span(\{\langle 1, 0, 3, -3, -4\rangle, \langle 0, 1, 4, -7, -6\rangle\}), \quad \text{and}$$
$$W^\perp = Span(\{\langle -3, -4, 1, 0, 0\rangle, \langle 3, 7, 0, 1, 0\rangle, \langle 4, 6, 0, 0, 1\rangle\}).$$

We can check that any vector in our basis for W is orthogonal to any vector in our basis for $W^\perp$. □

Using dim(W) to Find Other Bases for W

Knowing the dimension of a subspace W allows us to more efficiently check whether or not a subset B of W is a basis for W:

Theorem — The "Two for the Price of One" or "Two-for-One" Theorem:
Suppose W is a subspace of $\mathbb{R}^n$, and $dim(W) = k$. Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ be any subset of k vectors from W. Then: B is a ***basis*** for W ***if and only if*** either B is ***linearly independent*** ***or*** B ***Spans*** W. In other words, it is necessary and sufficient to check B for ***only one condition*** without checking the other, if B already contains the ***correct number of vectors***.

Proof: $(\Rightarrow)$ This direction is obvious because by definition, a basis B is linearly independent ***and*** Spans W.

$(\Leftarrow)$ Suppose that $dim(W) = k$ and $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is ***linearly independent***. We must show that B also ***Spans*** W.

Let us use Proof by Contradiction. Suppose that B does ***not*** Span W, so we can find some vector $\vec{w} \in W$ so that $\vec{w} \notin Span(B)$. But then $B \cup \{\vec{w}\}$ will still be linearly ***independent***, according to ***The Extension Theorem***.

However, if B' is any basis for W, then B' will have exactly k members, and so $B \cup \{\vec{w}\}$ is an independent set from $W = Span(B')$ that has more elements than B'. This is impossible by the Dependent Sets from Spanning Sets Theorem.

Similarly, if B ***Spans*** W, let us show that B must also be ***linearly independent***. If B were ***dependent***, then according to ***The Minimizing Theorem***, we can find a subset B' of B which still ***Spans*** W but is also linearly ***independent***. Thus, B' is a ***basis*** for W. Since B is dependent, the subset B' has strictly ***less*** than k vectors. This contradicts the fact that $dim(W) = k$. ■

In practice, once we know that B is a subset of W with $dim(W)$ members, it is much easier to ***check linear independence*** rather than Spanning, because it always involves a homogeneous system.

Example: In the previous Example, we investigated $W = Span(S)$, where:

$$\begin{aligned} S &= \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\} \\ &= \{\langle 15, -10, 5, 25, 0\rangle, \langle -9, 6, -3, -15, 0\rangle, \langle 1, -2, -5, 11, 8\rangle, \langle 5, -3, 3, 6, -2\rangle\}. \end{aligned}$$

We found that $\dim(W) = 2$, and a basis for W is the set:

$$B = \{\langle 1, 0, 3, -3, -4\rangle, \langle 0, 1, 4, -7, -6\rangle\}.$$

Notice, though, that these two vectors bear no resemblance whatsoever to the original set of four vectors. However, since $\dim(W) = 2$, we only need ***two*** vectors from S which are ***not parallel*** to each other in order to form a basis for W. We can see that $\vec{w}_1$ and $\vec{w}_2$ are actually parallel to each other, but there are no other pairs in S which are parallel. Therefore, ***each*** of the following sets of vectors:

$$\{\vec{w}_1, \vec{w}_3\}, \{\vec{w}_1, \vec{w}_4\}, \{\vec{w}_2, \vec{w}_3\}, \{\vec{w}_2, \vec{w}_4\}, \text{ and } \{\vec{w}_3, \vec{w}_4\}$$

are also bases for W. □

We must warn that if $dim(W) = k > 2$, it is not enough that we make sure that no two vectors in a subset $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset W$ are parallel to each other before declaring that S is a basis for W, even if S has exactly the correct number of vectors. We must test, as usual, that S is ***linearly independent*** by checking that the rref of $A = [\, \vec{w}_1 \; \vec{w}_2 \ldots \vec{w}_k \,]$ does not contain any free variables.

Example: Suppose we have $W = Span(S) \trianglelefteq \mathbb{R}^5$, where:

$$\begin{aligned} S &= \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5\} \\ &= \{\langle 3, -2, 4, 1, -3\rangle, \langle 1, -1, 2, -1, 0\rangle, \langle 7, -4, 8, 5, -9\rangle, \langle 1, 2, -1, 0, 3\rangle, \langle 5, -1, 2, 11, -12\rangle\}. \end{aligned}$$

Let us find a basis for W and a basis for $W^{\perp}$. We assemble the five vectors into the ***rows*** of a matrix:

$$A = \begin{bmatrix} 3 & -2 & 4 & 1 & -3 \\ 1 & -1 & 2 & -1 & 0 \\ 7 & -4 & 8 & 5 & -9 \\ 1 & 2 & -1 & 0 & 3 \\ 5 & -1 & 2 & 11 & -12 \end{bmatrix}, \quad \text{with rref:} \quad R = \begin{bmatrix} 1 & 0 & 0 & 3 & -3 \\ 0 & 1 & 0 & -\frac{10}{3} & 5 \\ 0 & 0 & 1 & -\frac{11}{3} & 4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

There are three leading variables (x_1, x_2 and x_3) and two free variables (x_4 and x_5). Thus, W is 3-dimensional and $W^\perp$ is 2-dimensional. Our basis for W will be the non-zero rows of R, which we scale to clear denominators:

$$W = Span(\{\langle 1, 0, 0, 3, -3\rangle, \langle 0, 3, 0, -10, 15\rangle, \langle 0, 0, 3, -11, 12\rangle\}).$$

We find a basis for $W^\perp$ by sight-reading the nullspace of R, taking note that our free variables are x_4 and x_5:

$$\left\langle -3, \frac{10}{3}, \frac{11}{3}, \mathbf{1}, 0 \right\rangle, \langle 3, -5, -4, 0, \mathbf{1}\rangle.$$

Clearing fractions, we get a basis consisting of vectors with integer coordinates, and write:

$$W^\perp = Span(\{\langle -9, 10, 11, 3, 0\rangle, \langle 3, -5, -4, 0, 1\rangle\}).$$

Let us think some more about W. Notice that there are ***five*** vectors in S, but W is only 3-dimensional. However, none of the vectors in S are parallel to each other. Thus, it is far from obvious if a random set of three vectors from S will be dependent or independent. For example, we can decide if $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is independent by assembling these vectors into the ***columns*** of a matrix and applying the Minimizing Theorem. But since there are only 5 vectors in S, we may as well assemble all 5 vectors into the columns of a matrix (which would be $A^\top$) and apply the Minimizing Theorem:

$$B = A^\top = \begin{bmatrix} 3 & 1 & 7 & 1 & 5 \\ -2 & -1 & -4 & 2 & -1 \\ 4 & 2 & 8 & -1 & 2 \\ 1 & -1 & 5 & 0 & 11 \\ -3 & 0 & -9 & 3 & -12 \end{bmatrix}, \quad \text{with rref} \quad R' = \begin{bmatrix} 1 & 0 & 3 & 0 & 4 \\ 0 & 1 & -2 & 0 & -7 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

From R', we see that $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$ is in fact ***dependent***, with $\vec{w}_3 = 3\vec{w}_1 - 2\vec{w}_2$. However, we also see from R' that $\{\vec{w}_1, \vec{w}_2, \vec{w}_4\}$ is ***independent***, since these correspond to the leading columns. Thus, we have found another basis for W:

$$W = Span(\{\vec{w}_1, \vec{w}_2, \vec{w}_4\}).$$

We can ask, though: is it possible to use a ***different*** combination of three columns from B? Indeed, we can, as long as these columns are also independent. Thus, we can also write:

$$W = Span(\{\vec{w}_2, \vec{w}_3, \vec{w}_4\}), \quad \text{and}$$
$$W = Span(\{\vec{w}_3, \vec{w}_4, \vec{w}_5\}).$$

In fact, there are three more possible combinations that will also produce a basis for W. On the other hand, $\{\vec{w}_1, \vec{w}_2, \vec{w}_5\}$ would ***not*** be a basis either, because $\vec{w}_5 = 4\vec{w}_1 - 7\vec{w}_2$. □

1.9 Section Summary

If W is a subspace of $\mathbb{R}^n$, then the ***orthogonal complement*** of W, defined as:

$$W^{\perp} = \left\{ \vec{v} \in \mathbb{R}^n \,|\, \vec{v} \circ \vec{w} = 0 \text{ for all } \vec{w} \in W \right\},$$

is also a subspace of $\mathbb{R}^n$. If $W = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}) \trianglelefteq \mathbb{R}^n$, then:

$$W^{\perp} = \left\{ \vec{v} \in \mathbb{R}^n \,|\, \vec{v} \circ \vec{w}_i = 0 \text{ for all } i = 1 \ldots k \right\}.$$

Furthermore, if we form the matrix A with ***rows*** $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$, then:

$$W = rowspace(A) \quad \text{and} \quad W^{\perp} = nullspace(A).$$

Thus, $nullspace(A)$ and $rowspace(A)$ are ***orthogonal complements*** of ***each other***.

For any subspace $W \trianglelefteq \mathbb{R}^n$: (a) $W \cap W^{\perp} = \left\{ \vec{\mathbf{0}}_n \right\}$, and (b) $(W^{\perp})^{\perp} = W$.

The Dimension Theorem for Orthogonal Complements:

If W is a subspace of $\mathbb{R}^n$ with orthogonal complement $W^{\perp}$, then: $dim(W) + dim(W^{\perp}) = n$.

The "Two for the Price of One" or "Two-for-One" Theorem:

Suppose W is a subspace of $\mathbb{R}^n$, and $dim(W) = k$.

Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ be any subset of k vectors from W.

Then: B is a ***basis*** for W ***if and only if*** either B is ***linearly independent or*** B ***Spans*** W.

1.9 Exercises

For Exercises (1) to (10): (a) Assemble the vectors in each Exercise into the ***rows*** of a matrix A, and find the rref R of A. (b) Use R to find a basis for each subspace W, and find a basis for $W^{\perp}$ as well. Both bases should consist of vectors with integer entries. (c) State the dimensions of W and $W^{\perp}$ and verify that the Dimension Theorem is true for these subspaces.

1. $W = Span(\{\langle 1, 5, -2\rangle, \langle 3, 4, 7\rangle\})$
2. $W = Span(\{\langle 2, 6, 5, -4\rangle, \langle 5, -2, 7, 1\rangle\})$
3. $W = Span(\{\langle 3, -2, 5, 7, 0\rangle, \langle 4, 1, 0, -3, 6\rangle\})$
4. $W = Span(\{\langle 2, -5, 6, -3\rangle\})$
5. $W = Span(\{\langle 3, -1, 5, 2, 6\rangle\})$
6. $W = Span(\{\langle 2, 6, 5, -4\rangle, \langle 5, -2, 7, 1\rangle, \langle 3, -8, 2, 6\rangle\})$
7. $W = Span(\{\langle 2, 6, 5, -4\rangle, \langle 5, -2, 7, 1\rangle, \langle 3, -8, 2, 5\rangle\})$. Compare to Exercises 2 and 6.
8. $W = Span(\{\langle -2, 4, 5, -4, 9\rangle, \langle -6, 2, 4, 2, 0\rangle, \langle 5, 5, -3, -7, 11\rangle\})$
9. $W = Span(\{\langle 3, -2, 5, 7, 0\rangle, \langle 4, 1, 0, -3, 6\rangle, \langle -2, 5, 4, 0, -2\rangle, \langle 1, 2, 5, -2, 3\rangle\})$
10. $W = Span(\{\langle -2, 1, 3, -1, 4\rangle, \langle -4, 2, 5, 0, 6\rangle, \langle -8, 4, 7, 6, 6\rangle, \langle -5, 3, -1, 0, 1\rangle, \langle 2, -1, 0, -5, 2\rangle\})$
11. In the final Example of this Section, we asked if certain subsets of $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5\}$ were also a basis for the 3-dimensional subspace $W = Span(S)$. Decide whether or not the following subsets also form a basis for W:

    a. $\{\vec{w}_1, \vec{w}_3, \vec{w}_4\}$ b. $\{\vec{w}_1, \vec{w}_4, \vec{w}_5\}$ c. $\{\vec{w}_1, \vec{w}_2, \vec{w}_5\}$

    d. $\{\vec{w}_2, \vec{w}_4, \vec{w}_5\}$ e. $\{\vec{w}_1, \vec{w}_3, \vec{w}_5\}$ f. $\{\vec{w}_2, \vec{w}_3, \vec{w}_5\}$

12. Suppose $W = Span(S) \trianglelefteq \mathbb{R}^5$, where:

$$S = \{\langle 2,-4, 7, 5, 3\rangle, \langle -6, 12,-21,-15,-9\rangle, \langle 3,-2, 6, 1,-4\rangle, \langle -11, 2,-16, 5, 26\rangle\}.$$

a. Form the matrix A whose ***rows*** are the vectors of S and find its rref R.
b. Use R to find a basis for $W^\perp$ with integer coefficients.
c. Use R to find a basis for W with integer coefficients.
d. Find the dimensions of W and $W^\perp$, and verify the Dimension Theorem.
e. Decide which of the following sets of vectors are also bases for W (and briefly explain how you arrived at your conclusion):

$$B^{(1)} = \{\langle 2,-4, 7, 5, 3\rangle, \langle -6, 12,-21,-15,-9\rangle\};$$
$$B^{(2)} = \{\langle 2,-4, 7, 5, 3\rangle, \langle -11, 2,-16, 5, 26\rangle\};$$
$$B^{(3)} = \{\langle 3,-2,6,1,-4\rangle, \langle -11, 2,-16, 5, 26\rangle\}.$$

13. Suppose $W = Span(S) \trianglelefteq \mathbb{R}^5$, where:

$$S = \{\langle 3,-2, 2, 1,-2\rangle, \langle -3, 2, 10, 11, 6\rangle, \langle 3,-2, 6, 1,-4\rangle, \langle 6,-4, 7, 5,-3\rangle\}.$$

a. Form the matrix A whose ***rows*** are the vectors of S and find its rref R.
b. Use R to find a basis for $W^\perp$ with integer coefficients.
c. Use R to find a basis for W with integer coefficients.
d. Find the dimensions of W and $W^\perp$, and verify the Dimension Theorem.
e. Now, form the matrix B whose ***columns*** are the vectors of S and find its rref R'.
f. Use R' to decide which of the following sets of vectors are also bases for W (and briefly explain how you arrived at your conclusion):

$$B^{(1)} = \{\langle 3,-2, 2, 1,-2\rangle, \langle -3, 2, 10, 11, 6\rangle, \langle 3,-2, 6, 1,-4\rangle\};$$
$$B^{(2)} = \{\langle 3,-2, 2, 1,-2\rangle, \langle -3, 2, 10, 11, 6\rangle, \langle 6,-4, 7, 5,-3\rangle\};$$
$$B^{(3)} = \{\langle 3,-2, 2, 1,-2\rangle, \langle 3,-2, 6, 1,-4\rangle, \langle 6,-4, 7, 5,-3\rangle\};$$
$$B^{(4)} = \{\langle -3, 2, 10, 11, 6\rangle, \langle 3,-2, 6, 1,-4\rangle, \langle 6,-4, 7, 5,-3\rangle\}.$$

14. Suppose $W = Span(S) \trianglelefteq \mathbb{R}^4$, where:

$$S = \{\langle 3,-2, 1, 4\rangle, \langle 5,-4, 3, 7\rangle, \langle -4,6,-8,-7\rangle, \langle -5, 6,-4,-6\rangle, \langle 2, 6,-8, 3\rangle\}.$$

a. Form the matrix A whose ***rows*** are the vectors of S and find its rref R.
b. Use R to find a basis for $W^\perp$ with integer coefficients.
c. Use R to find a basis for W with integer coefficients.
d. Find the dimensions of W and $W^\perp$, and verify the Dimension Theorem.
e. Now, form the matrix B whose ***columns*** are the vectors of S and find its rref R'.
f. Use R' to decide which of the following sets of vectors are also bases for W (and briefly explain how you arrived at your conclusion):

$$B^{(1)} = \{\langle 3,-2, 1, 4\rangle, \langle 5,-4, 3, 7\rangle, \langle -4,6,-8,-7\rangle\};$$
$$B^{(2)} = \{\langle 3,-2, 1, 4\rangle, \langle 5,-4, 3, 7\rangle, \langle -5, 6,-4,-6\rangle\};$$
$$B^{(3)} = \{\langle 5,-4, 3, 7\rangle, \langle -4,6,-8,-7\rangle, \langle -5, 6,-4,-6\rangle\};$$
$$B^{(4)} = \{\langle -4,6,-8,-7\rangle, \langle -5, 6,-4,-6\rangle, \langle 2, 6,-8, 3\rangle\}.$$

Assisted Computation: For Exercises (15) to (18): the following Exercises are similar in spirit to the previous three. You are given a set S of m vectors from some $\mathbb{R}^n$. These vectors were assembled in an $m \times n$ matrix A whose ***rows*** are the m vectors. The rref R of A is given, as well as the rref R' of $A^\top$. Suppose that $W = Span(S)$. (a) Find a basis for W using R. (b) Find a basis for W using R'. (c) Find a basis for $W^\perp$. The bases in (a) through (c) should have integer components. (d) Use ***The Two-For-One Theorem*** to decide whether or not each set of vectors $B^{(i)}$ forms a basis for W. There are no further computations necessary.

15. $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\} \subset \mathbb{R}^5$,
$S = \{\langle 4, 3, 8, 5, -5\rangle, \langle 5, 7, -3, 6, 5\rangle, \langle 3, 4, -1, 4, 0\rangle, \langle 6, 5, 10, 7, -3\rangle\}$;
$$R = \begin{bmatrix} 1 & 0 & 5 & 0 & 4 \\ 0 & 1 & -4 & 0 & 3 \\ 0 & 0 & 0 & 1 & -6 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \; R' = \begin{bmatrix} 1 & 0 & 0 & 7/5 \\ 0 & 1 & 0 & 4/5 \\ 0 & 0 & 1 & -6/5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$
$B^{(1)} = \{\vec{w}_1, \vec{w}_2\}$; $B^{(2)} = \{\vec{w}_1, \vec{w}_2, \vec{w}_4\}$; $B^{(3)} = \{\vec{w}_2, \vec{w}_3, \vec{w}_4\}$.

16. $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5\} \subset \mathbb{R}^5$,
$S = \{\langle 5, 8, -3, 7, -4\rangle, \langle 3, 4, -2, 4, -2\rangle, \langle 6, 4, -5, 7, -2\rangle, \langle -8, -9, 5, -7, 5\rangle, \langle -3, 6, 3, 5, -2\rangle\}$;
$$R = \begin{bmatrix} 1 & 0 & 0 & -10/3 & -2/3 \\ 0 & 1 & 0 & 4/3 & -1/3 \\ 0 & 0 & 1 & -13/3 & -2/3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \; R' = \begin{bmatrix} 1 & 0 & -3 & 0 & 5 \\ 0 & 1 & 7 & 0 & -4 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$
$B^{(1)} = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$; $B^{(2)} = \{\vec{w}_1, \vec{w}_2, \vec{w}_5\}$; $B^{(3)} = \{\vec{w}_3, \vec{w}_4, \vec{w}_5\}$.

17. $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5\} \subset \mathbb{R}^5$,
$S = \{\langle 7, 4, -3, 2, 9\rangle, \langle 3, -5, 2, -1, 6\rangle, \langle 6, 9, -7, 3, 8\rangle, \langle 5, 3, 8, 3, -6\rangle, \langle 4, -2, 5, -1, 7\rangle\}$;
$$R = \begin{bmatrix} 1 & 0 & 0 & 0 & 81/44 \\ 0 & 1 & 0 & 0 & 25/66 \\ 0 & 0 & 1 & 0 & -29/44 \\ 0 & 0 & 0 & 1 & -487/132 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \; R' = \begin{bmatrix} 1 & 0 & 0 & 8 & 0 \\ 0 & 1 & 0 & -5 & 0 \\ 0 & 0 & 1 & -6 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$
$B^{(1)} = \{\vec{w}_1, \vec{w}_2, \vec{w}_4, \vec{w}_5\}$; $B^{(2)} = \{\vec{w}_1, \vec{w}_3, \vec{w}_4, \vec{w}_5\}$; $B^{(3)} = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\}$.

18. $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5, \vec{w}_6\} = \{\langle 3, 5, -4, 3, 3, 5\rangle, \langle -2, -4, 6, -4, -2, 0\rangle,$
$\langle 1, -1, 12, -7, -1, -1\rangle, \langle 5, 6, 5, -2, 3, 4\rangle, \langle 0, -1, 5, -3, -1, -3\rangle, \langle 4, 6, -2, 2, 4, 10\rangle\} \subset \mathbb{R}^6$;
R and R' are found on the next page.
$B^{(1)} = \{\vec{w}_1, \vec{w}_2, \vec{w}_4\}$; $B^{(2)} = \{\vec{w}_2, \vec{w}_3, \vec{w}_5\}$; $B^{(3)} = \{\vec{w}_1, \vec{w}_2, \vec{w}_6\}$; $B^{(4)} = \{\vec{w}_4, \vec{w}_5, \vec{w}_6\}$.
Note/Hint: Determining whether $B^{(4)}$ is independent or not might require a minute of computation.

$$R = \begin{bmatrix} 1 & 0 & 7 & -4 & 0 & 2 \\ 0 & 1 & -5 & 3 & 0 & -5 \\ 0 & 0 & 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}; \; R' = \begin{bmatrix} 1 & 0 & 0 & 1 & -1/2 & 2 \\ 0 & 1 & 0 & -1/2 & -1/2 & 1 \\ 0 & 0 & 1 & 1 & 1/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For Exercises (19) to (27): Each problem refers to an $m \times n$ matrix A from Section 1.8, where you will also find the rref of A (you will not need the rref of $A^\top$). Let $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n$ be the columns of A. Use the Two-For-One Theorem to decide which of the given sets form a basis for *colspace*(A). Begin by reviewing the basis for *colspace*(A) that you obtained in Section 1.8 using the Minimizing Theorem. These can be found in the Answer Key.

19. Exercise 47.
    a. $\{\vec{c}_1, \vec{c}_3\}$ b. $\{\vec{c}_1, \vec{c}_3, \vec{c}_4\}$ c. $\{\vec{c}_2, \vec{c}_3, \vec{c}_4\}$
    d. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3\}$ e. $\{\vec{c}_3, \vec{c}_4\}$

20. Exercise 49.
    a. $\{\vec{c}_1, \vec{c}_2, \vec{c}_5\}$ b. $\{\vec{c}_2, \vec{c}_3, \vec{c}_5\}$ c. $\{\vec{c}_3, \vec{c}_4, \vec{c}_5\}$
    d. $\{\vec{c}_2, \vec{c}_4, \vec{c}_5\}$ e. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3\}$

21. Exercise 50.
    a. $\{\vec{c}_1, \vec{c}_2, \vec{c}_5\}$ b. $\{\vec{c}_1, \vec{c}_3, \vec{c}_4\}$ c. $\{\vec{c}_2, \vec{c}_4, \vec{c}_5\}$
    d. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3\}$ e. $\{\vec{c}_2, \vec{c}_3, \vec{c}_5\}$

22. Exercise 51.
    a. $\{\vec{c}_1, \vec{c}_3, \vec{c}_4\}$ b. $\{\vec{c}_2, \vec{c}_3, \vec{c}_5\}$ c. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3\}$
    d. $\{\vec{c}_3, \vec{c}_4, \vec{c}_5\}$ e. $\{\vec{c}_2, \vec{c}_3, \vec{c}_4\}$

23. Exercise 52.
    a. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3, \vec{c}_4\}$ b. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3, \vec{c}_5\}$ c. $\{\vec{c}_1, \vec{c}_3, \vec{c}_4, \vec{c}_5\}$
    d. $\{\vec{c}_2, \vec{c}_3, \vec{c}_4, \vec{c}_5\}$ e. $\{\vec{c}_1, \vec{c}_4, \vec{c}_5\}$

24. Exercise 53.
    a. $\{\vec{c}_1, \vec{c}_2, \vec{c}_4\}$ b. $\{\vec{c}_1, \vec{c}_3, \vec{c}_5\}$ c. $\{\vec{c}_1, \vec{c}_3, \vec{c}_4\}$
    d. $\{\vec{c}_3, \vec{c}_4, \vec{c}_5\}$ e. $\{\vec{c}_2, \vec{c}_3, \vec{c}_4\}$

25. Exercise 54.
    a. $\{\vec{c}_1, \vec{c}_3, \vec{c}_4, \vec{c}_5\}$ b. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3, \vec{c}_4\}$ c. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3, \vec{c}_5\}$
    d. $\{\vec{c}_2, \vec{c}_3, \vec{c}_4, \vec{c}_5\}$ e. $\{\vec{c}_3, \vec{c}_4, \vec{c}_5\}$

26. Exercise 55.
    a. $\{\vec{c}_1, \vec{c}_3, \vec{c}_4, \vec{c}_5\}$ b. $\{\vec{c}_2, \vec{c}_3, \vec{c}_4, \vec{c}_5\}$ c. $\{\vec{c}_3, \vec{c}_4, \vec{c}_5, \vec{c}_6\}$
    d. $\{\vec{c}_1, \vec{c}_3, \vec{c}_5, \vec{c}_6\}$ e. $\{\vec{c}_1, \vec{c}_2, \vec{c}_3, \vec{c}_6\}$

27. Exercise 56.
    a. $\{\vec{c}_1, \vec{c}_2, \vec{c}_4, \vec{c}_5\}$ b. $\{\vec{c}_2, \vec{c}_3, \vec{c}_4, \vec{c}_5\}$ c. $\{\vec{c}_2, \vec{c}_3, \vec{c}_5, \vec{c}_6\}$
    d. $\{\vec{c}_1, \vec{c}_2, \vec{c}_5, \vec{c}_7\}$ e. $\{\vec{c}_3, \vec{c}_4, \vec{c}_5, \vec{c}_7\}$

28. Prove that if $W = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}) \trianglelefteq \mathbb{R}^n$, then:

$$W^\perp = \left\{\vec{v} \in \mathbb{R}^n \,|\, \vec{v} \circ \vec{w}_i = 0 \text{ for all } i = 1 \ldots k\right\}.$$

Hint: start by writing the complete (original) definition of $W^\perp$ for any subspace W of $\mathbb{R}^n$.

29. Prove that in any $\mathbb{R}^n$: (a) $\left\{\vec{\mathbf{0}}_n\right\}^\perp = \mathbb{R}^n$, and (b) $(\mathbb{R}^n)^\perp = \left\{\vec{\mathbf{0}}_n\right\}$.

Hint: for (a) write the definition of $\left\{\vec{\mathbf{0}}_n\right\}^\perp$ and explain why every vector in $\mathbb{R}^n$ satisfies this definition, and for (b) use the previous Exercise and the fact that $\mathbb{R}^n$ is Spanned by $\vec{e}_1$ through $\vec{e}_n$.

30. Prove ***The Dimension Theorem for Orthogonal Complements:*** If W is a subspace of $\mathbb{R}^n$ with orthogonal complement $W^\perp$, then: $dim(W) + dim(W^\perp) = n$. Hint: Assemble a basis for W in the ***rows*** of a matrix and apply ***The Dimension Theorem for Matrices***.

31. Let $W \trianglelefteq \mathbb{R}^n$. Prove that $W \cap W^\perp = \left\{\vec{\mathbf{0}}_n\right\}$. Hint: Suppose $\vec{w} \in W$ and $\vec{w} \in W^\perp$. What can you say about the dot product of $\vec{w}$ with itself?

32. Use the idea behind the previous Exercise to give another proof that $(\mathbb{R}^n)^\perp = \left\{\vec{\mathbf{0}}_n\right\}$.

33. Let $W \trianglelefteq \mathbb{R}^n$. Our goal in this Exercise is to prove that $(W^\perp)^\perp = W$.

   a. Explain why Exercise 29 takes care of the cases when W is one of the trivial subspaces.

   For the rest of this Exercise, we can therefore assume that W is a non-trivial subspace.

   b. Let us use the symbol U for $W^\perp$. Prove that $W \subset U^\perp$. Hint: write down the definition of U (which is $W^\perp$) as well as $U^\perp$. Stare at these two definitions until you can clearly explain in writing why every vector in W also satisfies the definition of $U^\perp$.

   c. In Exercise 39 of Section 1.7, we discussed the concept of ***nested subspaces***. Part (b) tells us that $W \trianglelefteq U^\perp \trianglelefteq \mathbb{R}^n$ is a nesting of subspaces. Use The Dimension Theorem for Orthogonal Complements as well as Exercise 39 of Section 1.7 to show that in fact, $W = U^\perp$. This completes the proof that for all subspaces W of $\mathbb{R}^n$: $(W^\perp)^\perp = W$.

34. Let $W \trianglelefteq \mathbb{R}^n$, and suppose $dim(W) = 2$, and $B = \{\vec{w}_1, \vec{w}_2\} \subset W$. Use the Two-for-One Theorem to prove that B is a ***basis*** for W ***if and only if*** $\vec{w}_1$ and $\vec{w}_2$ are ***not parallel*** to each other.

35. Let A be an $m \times n$ matrix. Prove that $nullspace(A^\top)$ is the orthogonal complement of $colspace(A)$.

36. ***So You Think You Know the Zero Vector?*** In the course of this Chapter, we made several statements involving $\vec{\mathbf{0}}_n$. Decide whether each of the following statements is True or False. Cite a Definition, Theorem or counterexample in your explanation.

   a. $\vec{\mathbf{0}}_n$ is parallel to all vectors $\vec{v} \in \mathbb{R}^n$.

   b. $\vec{\mathbf{0}}_n$ is orthogonal to all vectors $\vec{v} \in \mathbb{R}^n$.

   c. All zero vectors are the same, so for example, $\vec{\mathbf{0}}_4 = \vec{\mathbf{0}}_7$.

   d. If $3 \cdot \vec{v} = \vec{\mathbf{0}}_n$, then $\vec{v} = \vec{\mathbf{0}}_n$.

   e. For all $\vec{v} \in \mathbb{R}^n$ and $k \in \mathbb{R}$: If $k \cdot \vec{v} = \vec{\mathbf{0}}_n$, then $\vec{v} = \vec{\mathbf{0}}_n$.

   f. $\vec{\mathbf{0}}_n$ is always a member of $Span(S)$, where S is a non-empty subset of $\mathbb{R}^n$.

   g. For all $\vec{v} \in \mathbb{R}^n$: $0 \cdot \vec{v} = \vec{\mathbf{0}}_n$.

   h. For all $\vec{v} \in \mathbb{R}^n$ and $k \in \mathbb{R}$: if $k \cdot \vec{v} = \vec{\mathbf{0}}_n$, and $\vec{v} \neq \vec{\mathbf{0}}_n$, then $k = 0$.

   i. If $\vec{\mathbf{0}}_n \in S$, a set of vectors from $\mathbb{R}^n$, then $Span(S)$ is undefined.

   j. The matrix equation $A\vec{x} = \vec{\mathbf{0}}_n$ is called a homogeneous equation.

k. The solution $\vec{x} = \vec{\mathbf{0}}_n$ to the matrix equation $A\vec{x} = \vec{\mathbf{0}}_n$ is called a non-trivial solution.

l. If $\vec{\mathbf{0}}_n \in S$, then S is automatically linearly dependent.

m. If $\vec{\mathbf{0}}_n \in S$, then S is automatically linearly independent.

n. If S does ***not*** contain $\vec{\mathbf{0}}_n$, then S is automatically linearly independent.

o. $\vec{\mathbf{0}}_n$ is a member of every subspace W of $\mathbb{R}^n$.

p. If $\vec{\mathbf{0}}_n \in S$, then S cannot be a basis for any subspace W of $\mathbb{R}^n$.

q. The set $\left\{\vec{\mathbf{0}}_n\right\}$ is a basis for the trivial subspace $W = \left\{\vec{\mathbf{0}}_n\right\}$.

r. The trivial subspace $\left\{\vec{\mathbf{0}}_n\right\}$ is 1-dimensional.

s. The trivial subspace $\left\{\vec{\mathbf{0}}_n\right\}$ does not have a basis.

t. $\vec{\mathbf{0}}_n$ is the only vector in $\mathbb{R}^n$ which is orthogonal to ***itself***.

u. For any subspace W of $\mathbb{R}^n$: $W \cap W^{\perp} = \left\{\vec{\mathbf{0}}_n\right\}$.

37. ***More True or False:*** Determine whether each statement is true or false, and briefly explain your answer by citing a Theorem, providing a counterexample, or a convincing argument.

a. All subspaces of $\mathbb{R}^n$ have a basis.

b. All nontrivial subspaces of $\mathbb{R}^n$ have a basis.

c. If $Span(\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}) = Span(\{\vec{w}_1, \vec{w}_2\}) = W$, then W is 2-dimensional.

d. If $Span(\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}) = Span(\{\vec{w}_1, \vec{w}_2\}) = W$, then W is at least 3-dimensional.

e. If $Span(\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}) = Span(\{\vec{w}_1, \vec{w}_2\}) = W$, then W is at most 2-dimensional.

f. If $W = Span(\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\})$, then W is 4-dimensional.

g. If $W = Span(\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\})$, then W is at least 4-dimensional.

h. If $W = Span(\{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\})$, then W is at most 4-dimensional.

i. If $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\} \subset W$ is a linearly independent set, then W is 3-dimensional.

j. If $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\} \subset W$ is a linearly independent set, then W is at least 3-dimensional.

k. If $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\} \subset W$ is a linearly independent set, then W is at most 3-dimensional.

l. If $dim(W) = 3$ and $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\} \subset W$, then S is dependent.

m. If $dim(W) = 5$ and $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\} \subset W$, then S must be independent.

n. If $dim(W) = 5$ and $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\} \subset W$, then S cannot Span W.

o. If $dim(W) = 5$ and $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5, \vec{w}_6\} \subset W$, then S must Span W.

p. If $dim(W) = 5$ and $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5\} \subset W$, and no two vectors in S are parallel to each other, then S must be a basis for W.

q. If $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5, \vec{w}_6, \vec{w}_7, \vec{w}_8\} \subset \mathbb{R}^5$, and $W = Span(S)$, then W could be 7-dimensional.

r. If $S = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5, \vec{w}_6, \vec{w}_7, \vec{w}_8\} \subset \mathbb{R}^5$, and $W = Span(S)$, then W could be 5-dimensional.

s. If S is a basis for W, then S cannot contain the zero vector.

t. In order to apply the Dimension Theorem to a matrix A, we need to know how many rows A has. (Note that only A is involved in the first equation of the Dimension Theorem).

A Summary of Chapter 1

Linear Algebra is the study of ***vector spaces***, which are generalizations of numbers, and functions with special properties called ***linear transformations*** that map one vector space to another. We will introduce linear transformations in Chapter 2.

The main examples of vector spaces are the ***Euclidean n-spaces***, denoted $\mathbb{R}^n$, which consist of all n-tuples or ***vectors*** $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$. These vectors can be viewed also as column matrices or row matrices. Operations on vectors include ***vector addition, scalar multiplication,*** and ***linear combinations***. These operations enjoy many properties analogous to those of real numbers.

The ***Span*** of a set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ from $\mathbb{R}^m$ is the set of ***all possible linear combinations*** of the vectors in S: $Span(S) = \{x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n \mid x_1, x_2, \ldots, x_n \in \mathbb{R}\}$.
In $\mathbb{R}^2$, the Span of a set of vectors is either $\left\{\vec{\mathbf{0}}_2\right\}$, a ***line*** through the origin, or all of $\mathbb{R}^2$. In $\mathbb{R}^3$, the Span of a set of vectors is either $\left\{\vec{\mathbf{0}}_3\right\}$, a ***line*** through the origin, a ***plane*** through the origin, or all of $\mathbb{R}^3$. The Span of a set of vectors is the fundamental example of a ***subspace*** of $\mathbb{R}^n$ (defined below), and so these geometric objects describe ***all*** the subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$.

If $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle \in \mathbb{R}^n$, define $\vec{u} \circ \vec{v} = u_1v_1 + u_2v_2 + \cdots + u_nv_n$, their ***dot product***. The ***length*** of $\vec{u}$ is $\|\vec{u}\| = \sqrt{\vec{u} \circ \vec{u}} = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2}$. We say $\vec{u}$ is a unit vector if $\|\vec{u}\| = 1$. Two vectors $\vec{u}$ and $\vec{v}$ are ***orthogonal*** if $\vec{u} \circ \vec{v} = 0$.

The Cauchy-Schwarz Inequality says that for all $\vec{u}, \vec{v} \in \mathbb{R}^n$: $|\vec{u} \circ \vec{v}| \leq \|\vec{u}\|\,\|\vec{v}\|$, and
The Triangle Inequality says that $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$.

An $m \times n$ ***matrix*** is a rectangular array organized into m ***rows*** and n ***columns***.

We can test if $\vec{b} \in \mathbb{R}^m$ is a member of $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\})$ by solving a linear system of m equations in the n unknowns $x_1, x_2, \ldots, x_n$. The ***Gauss-Jordan Algorithm*** does this efficiently: We assemble $A = \left[\vec{v}_1\ \vec{v}_2\ \ldots\ \vec{v}_n \mid \vec{b}\right]$ and transform it into its ***reduced row echelon form*** or ***rref***, using a sequence of ***elementary row operations***. From the rref, we can read off all the solutions, if any exist.

A set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$ is ***linearly independent*** if the only solution to the ***dependence test equation*** $x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_n\vec{v}_n = \vec{\mathbf{0}}_m$ is the ***trivial solution***: $x_1 = 0,\ x_2 = 0,\ \ldots,\ x_n = 0$. An equation of this form with at least one non-zero coefficient is called a ***dependence equation***, and we say that S is ***linearly dependent***.

The Equality of Spans Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^k$ and $S' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\} \subset \mathbb{R}^k$. Then: $Span(S) = Span(S')$ ***if and only if*** every $\vec{v}_i$ can be written as a linear combination of the $\vec{w}_1$ through $\vec{w}_m$, ***and*** every $\vec{w}_j$ can also be written as a linear combination of the $\vec{v}_1$ through $\vec{v}_n$.

The Elimination Theorem: Suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a ***linearly dependent*** set of vectors from $\mathbb{R}^m$, and $c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n = \vec{\mathbf{0}}_m$, where ***none*** of the coefficients c_i is 0, then $Span(S) = Span(S - \{\vec{v}_i\})$, for all $i = 1..n$.

The Minimizing Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset \mathbb{R}^m$, and let $A = \left[\ \vec{v}_1\ \vec{v}_2\ \ldots\ \vec{v}_n\ \right]$. Suppose that R is the rref of A, and $i_1, i_2, \ldots i_k$ are the columns of R that contain the leading 1's. Then the set $S' = \{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}$, that is, the subset of vectors of S consisting of the corresponding ***columns*** of A, is a ***linearly independent*** set, and $Span(S) = Span(S')$. Furthermore, every $\vec{v}_i \in S - S'$, that is, the vectors of S corresponding to the ***free variables*** of R, can be expressed as linear combinations of the vectors of S', using the ***coefficients*** found in the corresponding column of R.

The In/dependent Sets from Spanning Sets Theorem: Suppose $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\} \subset \mathbb{R}^{\boldsymbol{k}}$, and we form $Span(S)$. Suppose now we form a new set $L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$ consisting of m randomly chosen vectors from $Span(S)$. We can then conclude that ***if*** $m > n$, ***then*** L is automatically ***linearly dependent***. Consequently, ***if*** L is ***linearly independent***, ***then*** $m \leq n$.

The Extension Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***linearly independent*** set of vectors from $\mathbb{R}^{\boldsymbol{m}}$, and suppose $\vec{v}_{n+1} \notin Span(S)$. Then: $S' = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{v}_{n+1}\}$ is ***still linearly independent***.

A ***subspace*** W of $\mathbb{R}^{\boldsymbol{n}}$, written as $W \trianglelefteq \mathbb{R}^{\boldsymbol{n}}$, is a non-empty subset of vectors of $\mathbb{R}^{\boldsymbol{n}}$ (the ***ambient space*** of W) which is closed under vector addition and scalar multiplication: if $\vec{u}, \vec{v} \in W$, and $k \in \mathbb{R}$, then $\vec{u} + \vec{v} \in W$, and $k \cdot \vec{v} \in W$. The Span of a set of vectors is the fundamental example of a subspace.

A ***basis*** for a ***non-zero*** subspace W of $\mathbb{R}^{\boldsymbol{n}}$ is a ***linearly independent*** set of vectors $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\} \subset W$ which ***Spans*** W. Any two ***bases*** for W contain the same number of vectors, called the ***dimension*** of W, and we write ***dim(W) = k***. Any non-zero subspace $W \trianglelefteq \mathbb{R}^{\boldsymbol{n}}$ has a basis consisting of k vectors, where $k \leq n$. By convention, $W = \left\{\vec{\mathbf{0}}_{\boldsymbol{n}}\right\}$ has no basis, and $dim\left(\left\{\vec{\mathbf{0}}_{\boldsymbol{n}}\right\}\right) = 0$.

The Four Fundamental Matrix Spaces: Let A be an $m \times n$ matrix. We define the subspaces:

$$\textbf{\textit{rowspace(A)}} = Span(\{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_m\}) \trianglelefteq \mathbb{R}^{\boldsymbol{n}}, \quad \textbf{\textit{colspace(A)}} = Span(\{\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_n\}) \trianglelefteq \mathbb{R}^{\boldsymbol{m}},$$

$$\textbf{\textit{nullspace(A)}} = \left\{\vec{x} \in \mathbb{R}^{\boldsymbol{n}} \mid A\vec{x} = \vec{\mathbf{0}}_{\boldsymbol{m}}\right\} \trianglelefteq \mathbb{R}^{\boldsymbol{n}}, \text{ and } \textbf{\textit{nullspace}}(\boldsymbol{A^{\top}}) = \left\{\vec{y} \in \mathbb{R}^{\boldsymbol{m}} \mid A^{\top}\vec{y} = \vec{\mathbf{0}}_{\boldsymbol{n}}\right\} \trianglelefteq \mathbb{R}^{\boldsymbol{m}}.$$

where the $\vec{r}_i$ are the ***rows*** of A, and the $\vec{c}_j$ are the ***columns*** of A. The ***transpose*** of A, $A^{\top}$, is the $n \times m$ matrix obtained by writing row 1 of A as column 1 of $A^{\top}$, row 2 of A as column 2 of $A^{\top}$, and so on.

The rref R of A can be used to find a basis for the nullspace of A, by expressing each member of the nullspace as a linear combination of vectors, one for each free variable. The ***non-zero rows*** of R form a basis for the rowspace of A. The columns of A corresponding to the ***leading columns*** of R form a basis for the columnspace of A. The ***dimensions*** of these spaces are known as:

$$\textbf{\textit{rank(A)}} = dim(rowspace(A)) = dim(colspace(A)) = \textbf{\textit{rank}}(\boldsymbol{A^{\top}}),$$

$$\textbf{\textit{nullity(A)}} = dim(nullspace(A)), \text{ and } \textbf{\textit{nullity}}(\boldsymbol{A^{\top}}) = dim(nullspace(A^{\top})).$$

The Dimension Theorem for Matrices: For any $m \times n$ matrix A:

$$rank(A) + nullity(A) = n, \text{ and } rank(A^{\top}) + nullity(A^{\top}) = m.$$

The Columnspace Test for Consistency: The matrix equation $A\vec{x} = \vec{b}$ is ***consistent*** ***if and only if*** $\vec{b}$ is a member of ***colspace(A)***. Furthermore, if $A\vec{x} = \vec{b}$ is consistent, suppose $\vec{x}_p$ is a ***fixed*** solution (also called a ***particular solution***) of this system. Then, a vector $\vec{x}$ is a solution of this system ***if and only if*** it can be written as: $\vec{x} = \vec{x}_p + \vec{x}_h$, where $\vec{x}_h$ is a member of the ***nullspace(A)***.

If $W \trianglelefteq \mathbb{R}^{\boldsymbol{n}}$, we define: $W^{\perp} = \left\{\vec{v} \in \mathbb{R}^{\boldsymbol{n}} \,|\, \vec{v} \circ \vec{w} = 0 \text{ for } \textbf{\textit{all}}\ \vec{w} \in W\right\}$, the ***orthogonal complement*** of W, which is also a subspace of $\mathbb{R}^{\boldsymbol{n}}$. If $W = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}) \trianglelefteq \mathbb{R}^{\boldsymbol{n}}$, then:

$$W^{\perp} = \left\{\vec{v} \in \mathbb{R}^{\boldsymbol{n}} \,|\, \vec{v} \circ \vec{w}_i = 0 \text{ for all } i = 1 \ldots k\right\}.$$

If we form the matrix A with ***rows*** $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$, then: $W = rowspace(A)$ and $W^{\perp} = nullspace(A)$. Thus, $nullspace(A)$ and $rowspace(A)$ are ***orthogonal complements*** of each other.

For any subspace $W \trianglelefteq \mathbb{R}^{\boldsymbol{n}}$: (a) $W \cap W^{\perp} = \left\{\vec{\mathbf{0}}_{\boldsymbol{n}}\right\}$, and (b) $(W^{\perp})^{\perp} = W$.

The Dimension Theorem for Orthogonal Complements: If W is a subspace of $\mathbb{R}^{\boldsymbol{n}}$ with orthogonal complement $W^{\perp}$, then: $dim(W) + dim(W^{\perp}) = n$.

The "Two-for-One" Theorem: If $\dim(W) = k$, and $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a subset of vectors from W, then B is a ***basis*** for W ***if and only if*** either B is linearly ***independent*** ***or*** B ***Spans*** W.

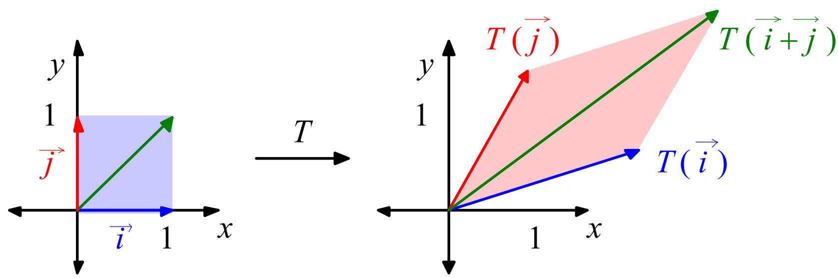
Chapter 2

Adding Movement and Colors:

Linear Transformations on Euclidean Spaces

We defined ***Linear Algebra*** as the study of objects called ***vector spaces***, which are generalizations of numbers, their ***structure***, and functions with special properties called ***linear transformations*** that map one space to another. In this Chapter, we will begin the study of linear transformations from one Euclidean space to another.

We will first see that these linear transformations are encoded by ***matrices***, that is, they can be computed by performing a matrix product. In $\mathbb{R}^2$ and $\mathbb{R}^3$, we will see that linear transformations have a ***geometric*** interpretation:



We will study the properties of these linear transformations, and construct important subspaces that are related to them, namely the ***kernel*** and the ***range***. We will see that unlike functions that we see in Calculus, where plotting a finite number of points often does not show us the whole picture, linear transformations can be completely described by knowing how they behave on a ***basis***.

We will see how to combine linear transformations using ***addition***, ***scalar multiplication***, and ***compositions***. These will lead us to a similar set of arithmetic operations on matrices, with compositions in particular corresponding to a general ***matrix product***.

We will study special kinds of linear transformations which are ***one-to-one*** or ***onto***, and give some easy conditions under which a transformation can be tested for these properties. Most importantly, we will see how to construct the ***inverse*** of a ***linear operator***, when it is ***both*** one-to-one and onto, in the same manner that we create the cube root function, $y = \sqrt[3]{x}$ from $y = x^3$.

Analogously, we will see how to use the Gauss-Jordan Algorithm to find the ***inverse*** of a square matrix, when it exists, and see that an invertible square matrix can be ***factored*** into simple matrices called ***elementary matrices***.

Finally, we will look at some special families of linear operators, called ***diagonal***, ***triangular***, and ***symmetric*** operators, and study their properties.

2.1 Mapping Spaces: Introduction to Linear Transformations

In basic Calculus, we study functions with real values that are defined on an interval I. Let us begin by generalizing this idea.

> ***Definition:*** Let X and Y be any two sets. A ***function*** $f : X \to Y$ is a rule (or a recipe, or a formula) that receives as its input an element x of X, and assigns to x as its output a ***unique*** element y of Y.
>
> We write $y = f(x)$, as usual, and also call y the ***image*** of x under f.
>
> We call X the ***domain*** of f, and call Y the ***codomain*** of f. We can also call f a ***map*** (because it tells us where to go), and say that f ***maps*** x ***to*** y, and more generally, f ***maps*** X ***into*** Y.

Example: Let us consider the two sets:

$$X = \{\text{all human beings}\}, \text{ and}$$

$$Y = \{\text{all 366 days between January 1 and December 31 (including February 29)}\}.$$

For every person $x \in X$ we can define the function:

$$b : X \to Y, \text{ where}$$

$$b(x) = \text{ the \textbf{birthday} of the person } x.$$

Then, b is a function from X into Y because every human being has a unique birthday.□

Linear Transformations

Now, let us focus our attention on a special kind of function that maps vector spaces into each other:

> ***Definition:*** A ***linear transformation*** $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function that satisfies:
> for all $\vec{u}, \vec{v} \in \mathbb{R}^n$, and for all $k \in \mathbb{R}$:
>
> ***The Additivity Property:*** $T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v})$
>
> ***The Homogeneity Property:*** $T(k \cdot \vec{u}) = k \cdot T(\vec{u})$
>
> In the special case when $T : \mathbb{R}^n \to \mathbb{R}^n$, that is, the domain is the same space as the codomain, we call T a ***linear operator***.
> If the codomain is $\mathbb{R} = \mathbb{R}^1$, we call T a ***linear functional***.

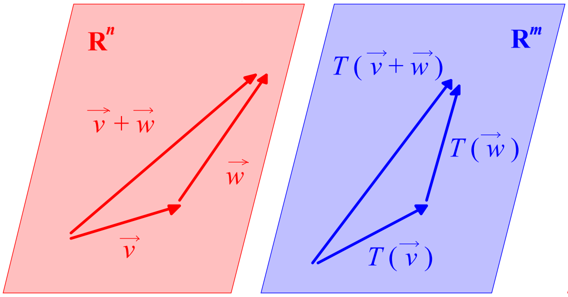
Notice that the domain of a linear transformation is always ***all*** of $\mathbb{R}^n$. Furthermore, most functions that we see in Calculus ***do not*** possess either of these two properties. For example, in general:

$$\sin(u + v) \neq \sin(u) + \sin(v),$$

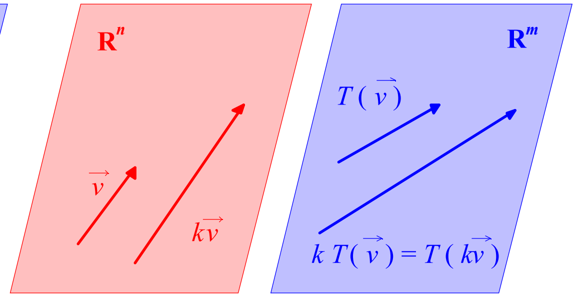
$$\sin(2u) \neq 2\sin(u),$$

$$(x + y)^2 \neq x^2 + y^2.$$

The Additivity Property essentially says that ***triangles are preserved***, and the Homogeneity Property says that ***proportionalities*** and ***parallel vectors are preserved***. These two linearity properties can be visualized as follows:



The Additivity Property



The Homogeneity Property

Example: Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be defined by:

$$T(\langle x, y\rangle) = \langle 2x - 3y, x + 2y, -x + y\rangle.$$

As a warm-up, let us compute $T(\langle -5, 7\rangle)$:

$$T(\langle -5, 7\rangle) = \langle 2(-5) - 3(7), -5 + 2(7), -(-5) + 7\rangle = \langle -31, 9, 12\rangle.$$

Notice that we ***input*** a vector from $\mathbb{R}^2$ into T, which then ***outputs*** a vector from $\mathbb{R}^3$.

Next, let us show that the two properties in the Definition are satisfied. Suppose $\langle x_1, y_1\rangle$ and $\langle x_2, y_2\rangle \in \mathbb{R}^2$. Then:

$$\begin{aligned}
&T(\langle x_1, y_1\rangle + \langle x_2, y_2\rangle)\\
&= T(\langle x_1 + x_2, y_1 + y_2\rangle)\\
&= \langle 2(x_1 + x_2) - 3(y_1 + y_2), (x_1 + x_2) + 2(y_1 + y_2), -(x_1 + x_2) + (y_1 + y_2)\rangle\\
&= \langle 2x_1 + 2x_2 - 3y_1 - 3y_2, x_1 + x_2 + 2y_1 + 2y_2, -x_1 - x_2 + y_1 + y_2\rangle\\
&= \langle 2x_1 - 3y_1, x_1 + 2y_1, -x_1 + y_1\rangle + \langle 2x_2 - 3y_2, x_2 + 2y_2, -x_2 + y_2\rangle\\
&= T(\langle x_1, y_1\rangle) + T(\langle x_2, y_2\rangle),
\end{aligned}$$

and similarly:

$$\begin{aligned}
T(k\langle x, y\rangle) &= T(\langle kx, ky\rangle)\\
&= \langle 2kx - 3ky, kx + 2ky, -kx + ky\rangle\\
&= k\langle 2x - 3y, x + 2y, -x + y\rangle\\
&= kT(\langle x, y\rangle).
\end{aligned}$$

Thus, T is a linear transformation.□

Now, if we identify the vector $\langle x, y\rangle$ and $T(\langle x, y\rangle) = \langle 2x - 3y, x + 2y, -x + y\rangle$ with their corresponding ***column matrices***, as usual, we can rewrite the action of T above as:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x-3y \\ x+2y \\ -x+y \end{bmatrix} = \begin{bmatrix} 2x \\ x \\ -x \end{bmatrix} + \begin{bmatrix} -3y \\ 2y \\ y \end{bmatrix}$$

$$= x\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} + y\begin{bmatrix} -3 \\ 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & -3 \\ 1 & 2 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix},$$

where in the last step, we recall that a ***linear combination*** of column matrices can be written as a ***matrix product***. Let us show in general that every linear transformation is indeed ***equivalent*** to a matrix product in a natural way:

Theorem: A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a ***linear transformation*** ***if and only if*** we can find an $m \times n$ matrix A so that the action of T can be performed by ***matrix multiplication:***

$$T(\vec{x}) = A\vec{x},$$

where, on the right side, we view $\vec{x} \in \mathbb{R}^n$ as an $n \times 1$ matrix, and $T(\vec{x}) \in \mathbb{R}^m$ as an $m \times 1$ matrix. We refer to A as the ***standard matrix*** of T, and we write:

$$[T] = A = \left[\, T(\vec{e}_1) \mid T(\vec{e}_2) \mid \ldots \mid T(\vec{e}_n) \,\right].$$

Consequently, the standard matrix of T is ***unique:*** if B is another matrix such that $T(\vec{x}) = B\vec{x}$ for ***all*** $\vec{x} \in \mathbb{R}^n$, then $A = B = [T]$.

We also say that A ***represents*** T, or A is a ***representation*** of T. In particular, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is an ***operator***, $[T]$ is an $n \times n$ or ***square*** matrix, and if $T : \mathbb{R}^n \to \mathbb{R}$ is a ***linear functional***, then $[T]$ is an $1 \times n$ or ***row*** matrix.

Proof: ($\Leftarrow$) Suppose that the matrix A exists so that $T(\vec{x}) = A\vec{x}$.

Then for all $\vec{x}$, $\vec{y} \in \mathbb{R}^n$ and $k \in \mathbb{R}$: $T(\vec{x}+\vec{y}) = A(\vec{x}+\vec{y}) = A\vec{x}+A\vec{y}$, and $T(k\vec{x}) = A(k\vec{x}) = k(A\vec{x})$, by the properties of matrix multiplication that we saw in Chapter 1. Thus, T is both additive and homogeneous, and so T is a linear transformation.

($\Rightarrow$) Now, suppose that T satisfies the two conditions of a linear transformation.

If $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle \in \mathbb{R}^n$, then:

$$\begin{aligned} T(\vec{x}) &= T(\langle x_1, x_2, \ldots, x_n \rangle) \\ &= T(x_1\vec{e}_1 + x_2\vec{e}_2 + \cdots + x_n\vec{e}_n) \\ &= T(x_1\vec{e}_1) + T(x_2\vec{e}_2) + \cdots + T(x_n\vec{e}_n) \quad \text{by Additivity} \\ &= x_1T(\vec{e}_1) + x_2T(\vec{e}_2) + \cdots + x_nT(\vec{e}_n) \quad \text{by Homogeneity.} \end{aligned}$$

Now for the magic: if we assemble the $T(\vec{e}_i)$ into the ***columns*** of a matrix A:

$$A = [T] = \left[\, T(\vec{e}_1) \mid T(\vec{e}_2) \mid \ldots \mid T(\vec{e}_n) \,\right],$$

then we can view the last line for $T(\vec{x})$ above as the matrix product:

$$T(\vec{x}) = \left[\ T(\vec{e}_1)\ |\ T(\vec{e}_2)\ |\ \ldots\ |\ T(\vec{e}_n)\ \right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A\vec{x}.$$

The uniqueness of $[T]$ will be proven in the Exercises.■

Example: If $T : \mathbb{R}^4 \to \mathbb{R}^3$ is a linear transformation and we are told that:

$$\begin{aligned} T(\vec{e}_1) &= \langle 3,7,-2\rangle, \\ T(\vec{e}_2) &= \langle -4,0,9\rangle, \\ T(\vec{e}_3) &= \langle -1,1,5\rangle, \quad \text{and} \\ T(\vec{e}_4) &= \langle 2,-1,-7\rangle. \end{aligned}$$

Then:

$$[T] = A = \begin{bmatrix} 3 & -4 & -1 & 2 \\ 7 & 0 & 1 & -1 \\ -2 & 9 & 5 & -7 \end{bmatrix}.$$

If we wanted to compute, say, $T(\langle 5,-3,4,8\rangle)$, we multiply:

$$T(\langle 5,-3,4,8\rangle) = \begin{bmatrix} 3 & -4 & -1 & 2 \\ 7 & 0 & 1 & -1 \\ -2 & 9 & 5 & -7 \end{bmatrix} \begin{bmatrix} 5 \\ -3 \\ 4 \\ 8 \end{bmatrix} = \begin{bmatrix} 39 \\ 31 \\ -73 \end{bmatrix},$$

and thus $T(\langle 5,-3,4,8\rangle) = \langle 39,31,-73\rangle$.□

Some Basic Examples

Let us take a look at some easy yet important linear transformations. The function:

$$\begin{aligned} Z_{n,m} &: \mathbb{R}^n \to \mathbb{R}^m, && \text{given by} \\ Z_{n,m}(\vec{x}) &= \vec{\mathbf{0}}_m && \text{for } \textbf{\textit{all}}\ \vec{x} \in \mathbb{R}^n, \end{aligned}$$

is a linear transformation, called the ***zero transformation*** of $\mathbb{R}^n$ into $\mathbb{R}^m$.

Since $Z_{n,m}(\vec{e}_i) = \vec{\mathbf{0}}_m$ for all i, the matrix of $Z_{n,m}$ is the $m \times n$ matrix with 0 in all the entries, and it is called the ***zero m × n matrix***, denoted by:

$$[Z_{n,m}] = \mathbf{0}_{m\times n} = \begin{bmatrix} 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{bmatrix}.$$

By construction, for all $\vec{x} \in \mathbb{R}^n$: $\mathbf{0}_{m \times n}\vec{x} = \vec{\mathbf{0}}_m$, so the zero matrices behave like the number 0 under ***multiplication*** of real numbers.

For any $\mathbb{R}^n$, we have the ***identity operator*** on $\mathbb{R}^n$:

$$\boldsymbol{I}_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n \quad \text{defined by}$$
$$\boldsymbol{I}_{\mathbb{R}^n}(\vec{x}) = \vec{x} \quad \text{for all } x \in \mathbb{R}^n.$$

Since $\boldsymbol{I}_{\mathbb{R}^n}(\vec{e}_i) = \vec{e}_i$, for all $i = 1..n$, we get as its matrix the ***identity matrix*** $\boldsymbol{I}_n$:

$$[\boldsymbol{I}_{\mathbb{R}^n}] = \boldsymbol{I}_n = \left[\, \vec{e}_1\, \vec{e}_2 \cdots \vec{e}_n \,\right] = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}.$$

By performing a matrix product, we get:

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \text{in other words:}$$

$$\boldsymbol{I}_n \vec{x} = \vec{x}.$$

Thus, $\boldsymbol{I}_n$ behaves like the number 1 under ***multiplication*** of real numbers. More generally, for any $\mathbb{R}^n$ and any $k \in \mathbb{R}$, we have the ***scaling operator:***

$$\boldsymbol{S}_k : \mathbb{R}^n \to \mathbb{R}^n, \quad \text{given by}$$
$$\boldsymbol{S}_k(\vec{x}) = k\vec{x} \quad \text{for all } \vec{x} \in \mathbb{R}^n.$$

As an Exercise, you will find the matrix of the scaling operators, but we give an Example below as a hint. For obvious reasons, they are also called ***scalar product operators***.□

Examples: The zero transformation $Z_{3,2} : \mathbb{R}^3 \to \mathbb{R}^2$ has matrix:

$$[Z_{3,2}] = \mathbf{0}_{2\times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

The identity operator $\boldsymbol{I}_{\mathbb{R}^4} : \mathbb{R}^4 \to \mathbb{R}^4$ and the scaling operator $\boldsymbol{S}_5 : \mathbb{R}^3 \to \mathbb{R}^3$ have matrices:

$$[\boldsymbol{I}_{\mathbb{R}^4}] = \boldsymbol{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad [\boldsymbol{S}_5] = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{bmatrix}.\,\square$$

Elementary Matrices

Let us review elementary row operations and construct special matrices using them:

> ***Definition:*** An $n \times n$ matrix E is called an ***elementary matrix*** if it is obtained by performing a ***single*** elementary row operation on the identity matrix $\boldsymbol{I_n}$.

Recall that there are three types of elementary row operations:

Type:	***Notation:***
1. Multiply row i by a ***nonzero*** scalar c	$R_i \to cR_i$
2. Exchange row i and row j	$R_i \leftrightarrow R_j$
3. Add c times row j to row i	$R_i \to R_i + cR_j$

Examples: The following are elementary matrices of Type 1, 2 and 3, respectively:

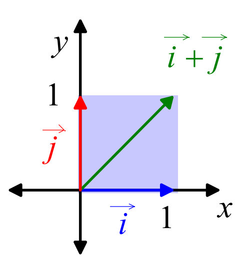
$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix}.$$

Each is obtained from $\boldsymbol{I_3}$, respectively, by multiplying row 2 by 3, exchanging row 1 and row 3, and adding 5 times row 2 to row 3. However, the following are ***not*** elementary matrices:

$$F_1 = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \quad F_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad F_3 = \begin{bmatrix} 1 & 0 & -5 \\ 0 & 1 & 7 \\ 0 & 0 & 1 \end{bmatrix}.$$

F_1 is not elementary because ***two rows*** (1 and 3) have been multiplied by a non-zero constant. F_2 is not elementary because ***two*** pairs of rows (row 1 and row 2, followed by row 2 and row 3) of $\boldsymbol{I_3}$ have been exchanged. F_3 is not elementary because a multiple of row 3 was added to row 1 ***and*** row 2.□

The linear operators that correspond to these matrices have geometric properties. For now, we will focus on 2×2 elementary matrices of Type 1 and 3. We will see how these operators transform $\mathbb{R}^2$ by looking at their effects on the three vectors $\left\{\vec{i}, \vec{j}, \vec{i}+\vec{j}\right\}$, that form what we call the ***basic box:***



The Basic Box

Horizontal and Vertical Dilations and Contractions

A 2×2 Type 1 elementary matrix has the form:

$$\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix} \text{ or } \begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}.$$

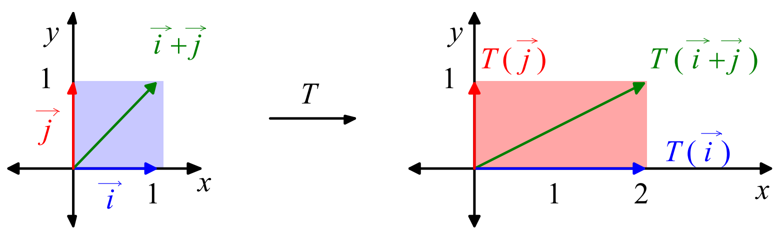
In the first case, the effect is on the ***horizontal*** or x-axis, and in the second case, the effect is on the ***vertical*** or y-axis. If $|c| > 1$, this operator corresponds to a ***dilation*** operator, and if $|c| < 1$, to a ***contraction*** operator. Furthermore, if $c < 0$, the operator also produces a ***reflection*** across the y-axis in the first case, and across the x-axis in the second case.

Example: Let T be the operator with standard matrix: $[T] = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$.

Recalling that $T(\vec{i})$ is in column 1, and $T(\vec{j})$ is in column 2, we get:

$$T(\vec{i}) = \langle 2, 0 \rangle = 2\vec{i}, \quad T(\vec{j}) = \langle 0, 1 \rangle = \vec{j}, \quad \text{and} \quad T(\vec{i}+\vec{j}) = 2\vec{i}+\vec{j} = \langle 2, 1 \rangle,$$

by applying the Additivity Property. The effect on the basic box is shown below:

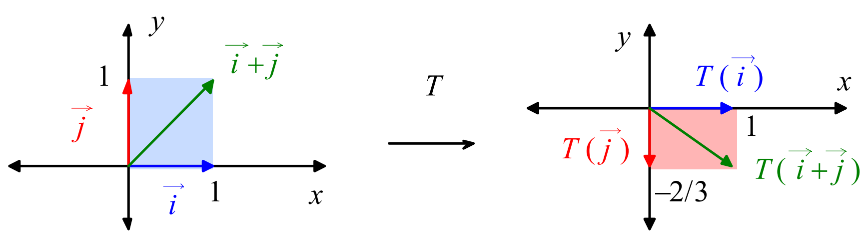


The Action of A Horizontal Dilation

Notice that the box has been ***dilated*** or "stretched" ***horizontally*** by a factor of 2. The vertical unit vector $\vec{j}$ is not affected by T.□

Example: Similarly, suppose T is the Type 1 operator with $[T] = \begin{bmatrix} 1 & 0 \\ 0 & -2/3 \end{bmatrix}$.

This time, $T(\vec{i}) = \vec{i}$, $T(\vec{j}) = -\frac{2}{3}\vec{j}$, and the basic box is ***contracted*** or "shrunk" ***vertically*** by a factor of 2/3 and ***reflected*** across the x –axis:



The Action of A Vertical Contraction Combined with a Reflection. □

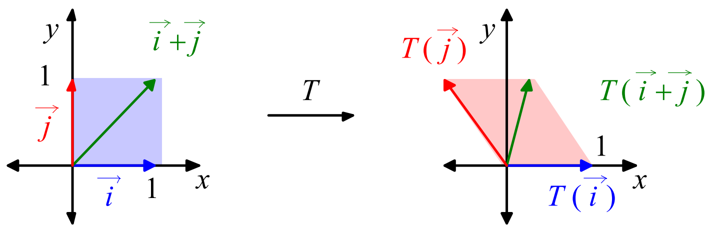
Shear Operators

A 2×2 Type 3 elementary matrix has the form: $\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}$ or $\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$.

In the first case, the unit vector $\vec{i}$ is not affected, but $T\left(\vec{j}\right) = c\vec{i} + \vec{j}$, so the image of $\vec{j}$ is now leaning to the right or left, depending on whether c is positive or negative. Because of this, the first kind is called a ***horizontal shear operator***. In the second case, $\vec{j}$ is not affected, but the image of $\vec{i}$ is now tilting up or down, so the second kind is called a ***vertical shear operator***.

Example: Let T be the operator with $[T] = \begin{bmatrix} 1 & -3/4 \\ 0 & 1 \end{bmatrix}$.

Again, we see that $T\left(\vec{i}\right) = \vec{i}$, $T\left(\vec{j}\right) = \langle -3/4, 1 \rangle$, and $T\left(\vec{i}+\vec{j}\right) = \langle 1, 0 \rangle + \langle -3/4, 1 \rangle = \langle 1/4, 1 \rangle$. The effect on the basic box is shown below:



The Action of A Horizontal Shear Transformation. □

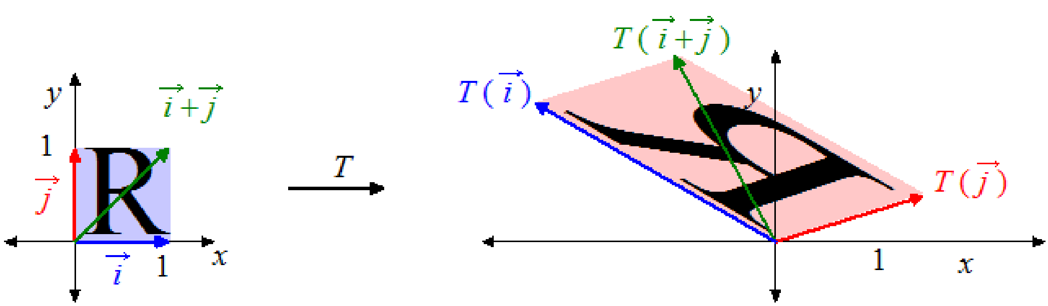
Example: The action of ***any*** operator on $\mathbb{R}^2$ can be visualized by its action on the basic box.

Suppose T is the operator with: $[T] = \begin{bmatrix} -5/2 & 3/2 \\ 3/2 & 1/2 \end{bmatrix}$.

As in the previous Example, we get:

$$T\left(\vec{i}\right) = \langle -5/2, 3/2 \rangle, \quad T\left(\vec{j}\right) = \langle 3/2, 1/2 \rangle, \quad \text{and} \quad T\left(\vec{i}+\vec{j}\right) = T\left(\vec{i}\right) + T\left(\vec{j}\right) = \langle -1, 2 \rangle.$$

The effect on the basic box is shown below, with its effect on the letter “R” inside the box shown for dramatic effect.



The Action of An Arbitrary Operator on $\mathbb{R}^2$. □

2.1 Section Summary

Let X and Y be any two sets. A ***function*** $f : X \to Y$ is a rule (or a recipe, or a formula) that receives as input an element x of X, and assigns to x as its output a ***unique*** element y of Y. We call X the ***domain*** of f, and call Y the ***codomain*** of f. We can also call f a ***map***, and say that f ***maps*** X ***into*** Y.

A ***linear transformation*** $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function that satisfies, for all $\vec{u}, \vec{v} \in \mathbb{R}^n$, and for all $k \in \mathbb{R}$:

The Additivity Property: $\quad T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v})$.

The Homogeneity Property: $\quad T(k\vec{u}) = kT(\vec{u})$.

In the special case when $T : \mathbb{R}^n \to \mathbb{R}^n$, we call T a ***linear operator***. If the codomain is $\mathbb{R} = \mathbb{R}^1$, we call T a ***linear functional***.

A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation ***if and only if*** we can find an $m \times n$ matrix A so that the action of T can be performed by matrix multiplication, that is, $T(\vec{x}) = A\vec{x}$, where we view $\vec{x} \in \mathbb{R}^n$ as an $n \times 1$ matrix. We refer to A as the ***standard matrix*** of T, and we write $[T] = A = \left[\, T(\vec{e}_1) \mid T(\vec{e}_2) \mid \ldots \mid T(\vec{e}_n) \,\right]$.

Some basic transformations include (where $\vec{x} \in \mathbb{R}^n$):

- The ***zero transformations*** $Z_{n,m} : \mathbb{R}^n \to \mathbb{R}^m$, given by $Z_{n,m}(\vec{x}) = \vec{\mathbf{0}}_m$.
- The ***identity operators*** $\boldsymbol{I}_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n$ defined by $\boldsymbol{I}_{\mathbb{R}^n}(\vec{x}) = \vec{x}$.
- The ***scaling operators*** $\boldsymbol{S}_k : \mathbb{R}^n \to \mathbb{R}^n$, given by $\boldsymbol{S}_k(\vec{x}) = k\vec{x}$.

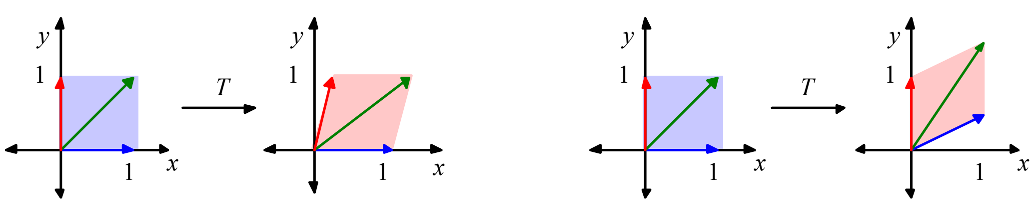
An $n \times n$ matrix E is called an ***elementary matrix*** if it is obtained by performing a ***single*** elementary row operation on the identity matrix $\boldsymbol{I}_n$.

A 2×2 Type 1 elementary matrix has the form: $\begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}$ or $\begin{bmatrix} 1 & 0 \\ 0 & c \end{bmatrix}$.

In the first case, the effect is on the ***horizontal*** or x-axis, and in the second case, the effect is on the ***vertical*** or y-axis. If $|c| > 1$, this operator corresponds to a ***dilation*** operator, and if $|c| < 1$, to a ***contraction*** operator. Furthermore, if $c < 0$, the operator also produces a ***reflection*** across the y-axis in the first case, and across the x-axis in the second case.

A 2×2 Type 3 elementary matrix has the form: $\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}$ or $\begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$.

In the first case, the unit vector $\vec{i}$ is not affected, but $T\left(\vec{j}\right) = c\vec{i} + \vec{j}$, so the image of $\vec{j}$ is now leaning to the right or left, depending on whether c is positive or negative. In the second case, $\vec{j}$ is not affected, but the image of $\vec{i}$ is now tilting up or down. Both are examples of ***shear*** operators.



A Horizontal Shear Operator and A Vertical Shear Operator

2.1 Exercises

1. Suppose that we have the two sets:

$$X = \{\text{the set of } \textbf{\textit{all parents}}, \text{ dead or alive}\}, \text{ and}$$

$$Y = \{\text{the set of } \textbf{\textit{all people}}, \text{ dead or alive}\}.$$

Decide which of the following are functions, and explain your decision briefly:

a. $f : X \to Y$, where $f(x)$ = the oldest child of x.
b. $g : X \to Y$, where $g(x)$ = the oldest daughter of x.
c. $h : Y \to X$, where $h(y)$ = the mother of y.
d. $k : Y \to Y$, where $k(y)$ = the youngest brother of y.
e. $p : X \to Y$, where $p(x)$ = the oldest grandchild of x.
f. $q : Y \to X$, where $q(y)$ = the paternal grandmother of y.

2. Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ be the function given by:

$$T(\langle x, y\rangle) = \langle 2x + 3y, x - 5y, 4x + y\rangle.$$

a. Compute $T(\langle 3, -7\rangle)$.
b. Show explicitly, using the two properties in the definition, that T is a linear transformation.
c. Find $[T]$.
d. Verify that your answer to (c) is correct by recomputing $T(\langle 3, -7\rangle)$ as a matrix product.

3. Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be the function given by:

$$T(\langle x_1, x_2, x_3, x_4\rangle) = \langle 2x_1 - 5x_3, 3x_2 + x_3 - 2x_4, 3x_1 + 8x_2\rangle.$$

a. Compute $T(\langle 5, -3, 7, 2\rangle)$.
b. Show explicitly, using the two properties in the definition, that T is a linear transformation.
c. Find $[T]$.
d. Verify that your answer to (c) is correct by recomputing $T(\langle 5, -3, 7, 2\rangle)$ as a matrix product.

4. Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be the function given by:

$$T(\langle x, y, z\rangle) = \langle 3x + 2y - 5z, x + 4z, 2y - 7z, 4x + 9y\rangle.$$

a. Compute $T(\langle 3, 8, -6\rangle)$.
b. Show explicitly, using the two properties in the definition, that T is a linear transformation.
c. Find $[T]$.
d. Verify that your answer to (c) is correct by recomputing $T(\langle 3, 8, -6\rangle)$ as a matrix product.

5. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be the function given by:

$$T(\langle x, y, z\rangle) = \langle 5x - 3y - 2z, 4x - 6y + 3z, 2x + 2y\rangle.$$

a. Compute $T(\langle 2, -7, 4\rangle)$.
b. Show explicitly, using the two properties in the definition, that T is a linear transformation (in this case, T is an operator).

c. Find $[T]$.

d. Verify that your answer to (c) is correct by recomputing $T(\langle 2,-7,4\rangle)$ as a matrix product.

6. Is $T : \mathbb{R}^2 \to \mathbb{R}^2$, given by $T(\langle x,y\rangle) = \langle y-3, x+5\rangle$ a linear transformation? Why or why not?

7. Is $T : \mathbb{R}^2 \to \mathbb{R}^2$, given by $T(\langle x,y\rangle) = \langle x+2y, 5xy\rangle$ a linear transformation? Why or why not?

8. Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation given by:

$$T(\vec{i}) = \langle 0,-5,3\rangle \text{ and } T(\vec{j}) = \langle 2,4,-7\rangle.$$

a. Find $[T]$.

b. Give a general formula for $T(\langle x,y\rangle)$

c. Use $[T]$ to compute $T(\langle 7,-2\rangle)$.

9. Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^2$ is a linear transformation given by:

$$T(\vec{i}) = \langle -3,5\rangle,\ T(\vec{j}) = \langle 2,7\rangle, \text{ and } T(\vec{k}) = \langle 0,4\rangle.$$

a. Find $[T]$.

b. Give a general formula for $T(\langle x,y,z\rangle)$

c. Use $[T]$ to compute $T(\langle 5,-2,6\rangle)$.

10. Suppose that $T : \mathbb{R}^5 \to \mathbb{R}^5$ is an operator given by:

$$T(\vec{e}_1) = \vec{e}_3,\ T(\vec{e}_2) = \vec{e}_5,\ T(\vec{e}_3) = \vec{e}_2,\ T(\vec{e}_4) = \vec{e}_4, \text{ and } T(\vec{e}_5) = \vec{e}_1.$$

a. Find $[T]$.

b. Give a general formula for $T(\langle x_1,x_2,x_3,x_4,x_5\rangle)$

c. Use $[T]$ to compute $T(\langle 3,0,-5,2,9\rangle)$.

d. Write a sentence explaining exactly what T does to any vector from $\mathbb{R}^5$.

11. Suppose we know that T is a linear transformation with codomain $\mathbb{R}^3$, and:

$$T(3\vec{v}_1 + \vec{v}_2) = \langle 5,-2,7\rangle, \text{ and}$$
$$T(5\vec{v}_1 + 2\vec{v}_2) = \langle 4,0,-3\rangle,$$

for some vectors $\vec{v}_1$ and $\vec{v}_2$ from the domain. Find $T(\vec{v}_1)$ and $T(\vec{v}_2)$.

For Exercise (12) to (23): Show the effect on the basic box of the following operators T, where $[T]$ is the indicated matrix. Note that (12) to (19) are elementary matrices, but (20) to (23) are not.

12. $\begin{bmatrix} 5/2 & 0 \\ 0 & 1 \end{bmatrix}$ 13. $\begin{bmatrix} -3/5 & 0 \\ 0 & 1 \end{bmatrix}$ 14. $\begin{bmatrix} 1 & 0 \\ 0 & 5/3 \end{bmatrix}$ 15. $\begin{bmatrix} 1 & 0 \\ 0 & -4/5 \end{bmatrix}$

16. $\begin{bmatrix} 1 & 3/4 \\ 0 & 1 \end{bmatrix}$ 17. $\begin{bmatrix} 1 & -2/3 \\ 0 & 1 \end{bmatrix}$ 18. $\begin{bmatrix} 1 & 0 \\ -7/4 & 1 \end{bmatrix}$ 19. $\begin{bmatrix} 1 & 0 \\ 4/3 & 1 \end{bmatrix}$

20. $\begin{bmatrix} 3 & -2 \\ 5 & 1 \end{bmatrix}$ 21. $\begin{bmatrix} -3 & 1 \\ 2 & -4 \end{bmatrix}$ 22. $\begin{bmatrix} 5 & 3 \\ -4 & 1 \end{bmatrix}$ 23. $\begin{bmatrix} 3 & -2 \\ -6 & 4 \end{bmatrix}$

Explain what happened to the basic box in Exercise 23, and why this happened.

For Exercises (24) and (25): We can also visualize the action of an operator on $\mathbb{R}^2$ by its action on any two vectors $\vec{u}$ and $\vec{v}$ that form a ***basis*** for $\mathbb{R}^2$, that is, two ***non-parallel*** vectors. For the following Exercises, compute $T(\vec{u})$, $T(\vec{v})$ and $T(\vec{u}+\vec{v})$, then sketch $\vec{u}$, $\vec{v}$ and $\vec{u}+\vec{v}$ in one coordinate system and $T(\vec{u})$, $T(\vec{v})$ and $T(\vec{u}+\vec{v})$ in another. What do you notice about the vectors?

24. $[T] = \begin{bmatrix} 5/13 & 12/13 \\ 12/13 & -5/13 \end{bmatrix}$, $\vec{u} = \langle 3, 2\rangle$, $\vec{v} = \langle -2, 3\rangle$.

25. $[T] = \begin{bmatrix} -2 & 0 \\ 3 & 2 \end{bmatrix}$, $\vec{u} = \langle 2, 1\rangle$, $\vec{v} = \langle 1, -2\rangle$.

26. Decide which of the following matrices are elementary, and which are not. If the matrix is elementary, decide if it is of Type 1, 2 or 3.

a. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -3 & 1 \end{bmatrix}$ b. $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 5 \\ 0 & 0 & 1 \end{bmatrix}$ c. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -3 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ d. $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

e. $\begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ f. $\begin{bmatrix} 0 & 0 & 3 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ g. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \sqrt{2} \end{bmatrix}$ h. $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$

i. $\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 2 & 0 & 1 \end{bmatrix}$ j. $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$ k. $\begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ l. $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

27. Prove ***directly*** that the standard matrix $A = [T] = \left[\, T(\vec{e}_1) \mid T(\vec{e}_2) \mid \ldots \mid T(\vec{e}_n) \,\right]$ is ***unique:*** if B is another matrix such that $T(\vec{x}) = A\vec{x} = B\vec{x}$ for ***all*** $\vec{x} \in \mathbb{R}^n$, then $A = B = [T]$. Hint: rewrite $A\vec{x} = B\vec{x}$ into $(A - B)\vec{x} = \vec{\mathbf{0}}_m$. If any of the entries of $A - B$ is non-zero, think of a ***specific*** $\vec{x} \in \mathbb{R}^n$ which would make $(A - B)\vec{x}$ a ***non-zero*** vector.

28. Show that the matrix of T from Exercise 10 can be obtained from the identity matrix $\boldsymbol{I}_5$ by a sequence of Type 2 row operations. For this reason, this is an example of what is called a ***permutation matrix***, because it is a ***rearrangement*** of the columns of $\boldsymbol{I}_n$.

29. Find the standard matrix of the scaling operator $\boldsymbol{S}_k$ on $\mathbb{R}^n$, where $k \in \mathbb{R}$, given by $\boldsymbol{S}_k(\vec{x}) = k\vec{x}$.

30. Starting with the two properties of a linear transformation T, we found from our only Theorem in this Section that T can be computed using a matrix product: $T(\vec{x}) = A\vec{x}$. Use this to prove that for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we must have: $T\left(\vec{\mathbf{0}}_n\right) = \vec{\mathbf{0}}_m$.

31. Now, using the Additivity Property and the property of the zero vector, prove ***directly*** that for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we must have $T\left(\vec{\mathbf{0}}_n\right) = \vec{\mathbf{0}}_m$.

    Hint: compute $T\left(\vec{\mathbf{0}}_n + \vec{\mathbf{0}}_n\right)$ in two different ways.

32. Now, using the Homogeneity Property and the multiplicative property of the scalar zero, prove ***directly*** that for any linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$, we must have $T\left(\vec{\mathbf{0}}_n\right) = \vec{\mathbf{0}}_m$. How should the Hint from the previous Exercise be modified?

2.2 Rotations, Projections and Reflections

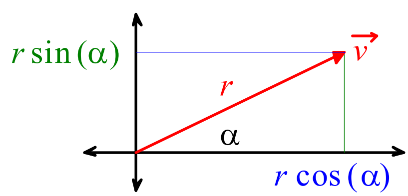
We saw in the previous Section that linear operators corresponding to Type 1 and 3 elementary matrices have geometric interpretations, namely as contraction, dilation and shear operators. We will see in this Section other geometric operators, namely ***rotations*** in $\mathbb{R}^2$, ***reflections*** in $\mathbb{R}^2$ and $\mathbb{R}^3$ (of which Type 2 elementary matrices are examples), and ***projections*** in $\mathbb{R}^2$ and $\mathbb{R}^3$. First, let's go out for a spin:

Rotations in $\mathbb{R}^2$

We can use the idea of ***polar coordinates*** to write a vector $\vec{v} = \langle x, y \rangle$ in $\mathbb{R}^2$ as:

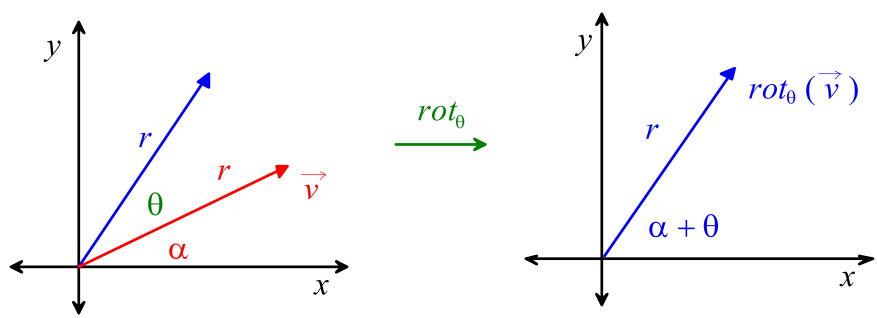
$$\vec{v} = \langle x, y \rangle = \langle r\cos(\alpha), r\sin(\alpha) \rangle,$$

for some non-negative number $r = \|\vec{v}\|$ and some angle $\alpha \in [0, 2\pi)$, where $\vec{v}$ makes an angle of α with respect to the positive x-axis when drawn in standard position.



A Vector in Polar Coordinates

If we rotate $\vec{v}$ counterclockwise by an angle θ about the origin and call this new vector $rot_\theta(\vec{v})$, then the length of $rot_\theta(\vec{v})$ will still be r, but it will now make an angle of $\alpha + \theta$ with respect to the positive x-axis, as seen in the diagram below:



A Vector $\vec{v}$ and $rot_\theta(\vec{v})$, its Counterclockwise Rotation by θ

Instead of directly verifying that rot_θ satisfies the two linearity properties, we will find a ***matrix*** for rot_θ, thus showing that it is indeed a linear transformation.

The coordinates of the rotated vector $rot_\theta(\vec{v})$ are:

$$rot_\theta(\vec{v}) = \Big\langle r\cos(\alpha + \theta), r\sin(\alpha + \theta) \Big\rangle.$$

But by using ***addition formulas*** from trigonometry, we have:

$$\begin{bmatrix} r\cos(\alpha+\theta) \\ r\sin(\alpha+\theta) \end{bmatrix} = \begin{bmatrix} r(\cos(\alpha)\cos(\theta) - \sin(\alpha)\sin(\theta)) \\ r(\sin(\alpha)\cos(\theta) + \cos(\alpha)\sin(\theta)) \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\theta)\cdot r\cos(\alpha) - \sin(\theta)\cdot r\sin(\alpha) \\ \sin(\theta)\cdot r\cos(\alpha) + \cos(\theta)\cdot r\sin(\alpha) \end{bmatrix}$$

$$= \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} r\cos(\alpha) \\ r\sin(\alpha) \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix},$$

expressed as a matrix product. Thus, we see that indeed:

Theorem: The ***rotation transformation*** $rot_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that takes a vector $\vec{v}$ in standard position and rotates $\vec{v}$ counterclockwise by an angle of $\theta > 0$ about the origin is a ***linear transformation***, with:

$$[rot_\theta] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Example: Let us find the matrix of the counterclockwise rotation by the angle $\theta = \cos^{-1}(3/5) \approx 53^0$ about the origin. We have:

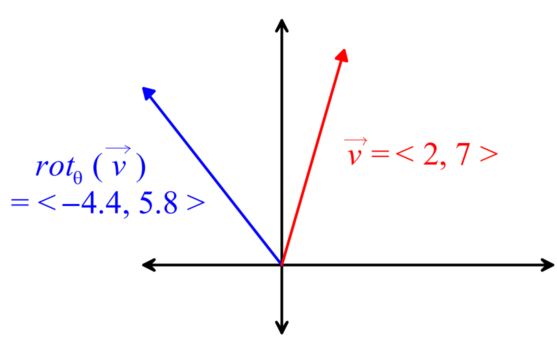
$$\cos(\theta) = 3/5 \quad \text{and} \quad \sin(\theta) = 4/5, \quad \text{thus:}$$

$$[rot_\theta] = \begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}.$$

To demonstrate its action, let us find $rot_\theta(\langle 2, 7\rangle)$:

$$rot_\theta(\langle 2,7\rangle) = \begin{bmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{bmatrix}\begin{bmatrix} 2 \\ 7 \end{bmatrix} = \begin{bmatrix} -4.4 \\ 5.8 \end{bmatrix}.$$

We graph $\langle 2, 7\rangle$ and $rot_\theta(\langle 2, 7\rangle) = \langle -4.4, 5.8\rangle$ below, and observe that their lengths are the same but $rot_\theta(\langle 2, 7\rangle)$ is rotated counterclockwise from $\langle 2, 7\rangle$ by about 53^0:



The Vector $\vec{v} = \langle 2, 7\rangle$ and its Rotation $rot_\theta(\vec{v})$ by $\theta = \cos^{-1}(3/5)$. □

Basic Projections and Reflections in $\mathbb{R}^2$

We can perform a variety of projection and reflection operators on a vector $\langle x,y\rangle \in \mathbb{R}^2$. These operators, and their counterparts in $\mathbb{R}^3$ which we will see later, have important applications in ***computer graphics***.

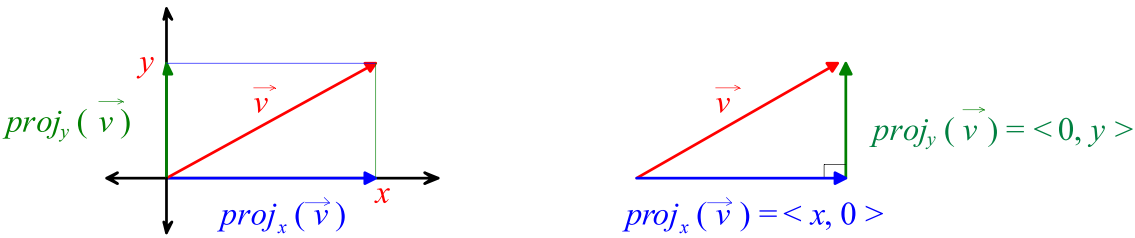
The projection of $\vec{v} = \langle x, y\rangle$ onto the x-axis, denoted $proj_x(\vec{v})$, is the vector $\langle x, 0\rangle$:

$$proj_x(\langle x, y\rangle) = \langle x, 0\rangle,$$

and similarly, the projection of $\vec{v}$ onto the y-axis is the vector $\langle 0, y\rangle$:

$$proj_y(\langle x, y\rangle) = \langle 0, y\rangle.$$

We see their geometric interpretation below:



The Relationships Among $\vec{v}$, $proj_x(\vec{v})$, and $proj_y(\vec{v})$

The key relationship among these vectors is seen in the right triangle that they form:

$$\vec{v} = \langle x, y\rangle = \langle x, 0\rangle + \langle 0, y\rangle = proj_x(\vec{v}) + proj_y(\vec{v}),$$

where $proj_x(\vec{v})$ is orthogonal to $proj_y(\vec{v})$. This is an example of what is called an ***orthogonal decomposition***, and the rest of our examples will involve this concept as well.

As with rotations, we will show that these are indeed linear transformations by finding their standard matrices. But since their actions are very simple, it is easy to check by direct multiplication that:

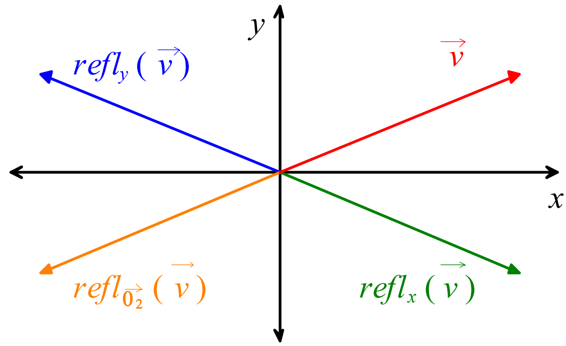
$$proj_x(\vec{v}) = \begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix} \text{ and}$$

$$proj_y(\vec{v}) = \begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

Thus $proj_x$ and $proj_y$ are linear transformations, with:

$$[proj_x] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \text{ and } [proj_y] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Similarly, we can take $\vec{v}$ and ***reflect*** it across the x-axis, the y-axis, or the origin (in the same way that we reflect graphs of functions):



A Vector $\vec{v}$ and its Three Basic Reflections in $\mathbb{R}^2$.

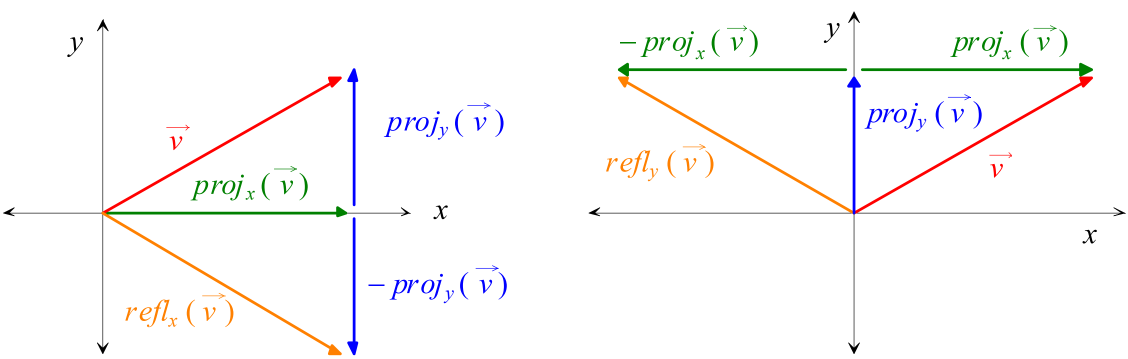
We compute these operators, and see their standard matrices, via:

$$refl_x\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x \\ -y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix},$$

$$refl_y\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -x \\ y \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}, \text{ and}$$

$$refl_{\vec{0}_2}\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -x \\ -y \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}.$$

Notice that $[refl_x]$ and $[refl_y]$ are both 2×2 Type 1 elementary matrices, with $c = -1$, and $\left[refl_{\vec{0}_2}\right] = \boldsymbol{S}_{-1}$, which represents scalar multiplication by -1. The reflection operators across the x- and y-axes can be related to the projection operators through the following diagrams:



The Geometric Relationships Among $\vec{v}$, $proj_x(\vec{v})$, $proj_y(\vec{v})$, $refl_x(\vec{v})$ and $refl_y(\vec{v})$

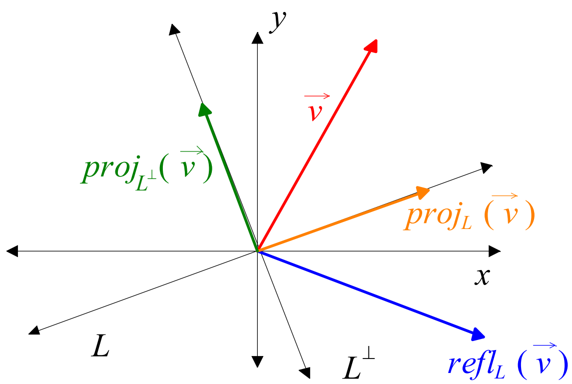
From these, we see that:

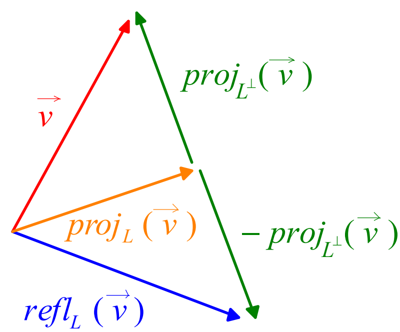
$$refl_x(\vec{v}) = proj_x(\vec{v}) - proj_y(\vec{v}), \text{ and}$$

$$refl_y(\vec{v}) = proj_y(\vec{v}) - proj_x(\vec{v}).$$

General Projections and Reflections in $\mathbb{R}^2$

The x-axis and y-axis are two orthogonal lines that pass through the origin. More generally, if L is ***any*** line through the origin in $\mathbb{R}^2$, there is a unique line $L^{\perp}$, also passing through the origin, that is ***orthogonal*** to L (recall from Chapter 1 that we call $L^{\perp}$ the ***orthogonal complement*** of L). We can define the projection operators onto L and $L^{/}$ and the reflection operator across L by the following vector diagrams:





The Projections of $\vec{v}$ Onto a Line L and its Orthogonal Complement $L^{\perp}$, and the Reflection of $\vec{v}$ Across L.

It was easy to find $proj_x(\vec{v})$ and $proj_y(\vec{v})$ if we knew $\vec{v} = \langle x, y\rangle$, but for a random line L and its orthogonal complement $L^{\perp}$, these projections are not that obvious. However, our goal is to satisfy the equation:

$$\vec{v} = proj_L(\vec{v}) + proj_{L^{\perp}}(\vec{v}),$$

where $proj_L(\vec{v})$ is parallel to L, and $proj_{L^{\perp}}(\vec{v})$ is parallel to $L^{\perp}$. Once we find these two projections, we can find the reflection across L via:

$$refl_L(\vec{v}) = proj_L(\vec{v}) - proj_{L^{\perp}}(\vec{v}),$$

as seen from the diagram. Let us demonstrate how to find these three vectors:

Example: Let L be the line in $\mathbb{R}^2$ with Cartesian equation $y = \dfrac{2}{3}x$.

The vector $\langle 3, 2\rangle$ is parallel to L, and since the vector $\langle -2, 3\rangle$ is orthogonal to $\langle 3, 2\rangle$ as we easily check with the dot product, $\langle -2, 3\rangle$ must be parallel to $L^{\perp}$. From the diagram, we want $proj_L(\vec{v})$ to be parallel to L, and $proj_{L^{\perp}}(\vec{v})$ parallel to $L^{\perp}$. Thus:

$$proj_L(\vec{v}) = a\langle 3, 2\rangle, \text{ and}$$
$$proj_{L^{\perp}}(\vec{v}) = b\langle -2, 3\rangle,$$

for some scalar multiples by a and b. However, we want:

$$\vec{v} = proj_L(\vec{v}) + proj_{L^{\perp}}(\vec{v}).$$

Using $\vec{v} = \langle x, y\rangle$, we get:

$$\langle x, y\rangle = a\langle 3, 2\rangle + b\langle -2, 3\rangle.$$

This vector equation is equivalent to the linear system:

$$\begin{aligned} 3a - 2b &= x \\ 2a + 3b &= y. \end{aligned}$$

By eliminating b, we can solve for a:

$$\begin{aligned} 9a - 6b &= 3x \\ 4a + 6b &= 2y \end{aligned} \Rightarrow 13a = 3x + 2y \quad \text{or} \quad a = \frac{3}{13}x + \frac{2}{13}y.$$

Similarly, by eliminating a, we get $b = -\frac{2}{13}x + \frac{3}{13}y$. Thus, we get:

$$\begin{aligned} proj_L(\vec{v}) &= a\langle 3, 2\rangle = \left(\frac{3}{13}x + \frac{2}{13}y\right)\langle 3, 2\rangle \\ &= \left\langle \frac{9}{13}x + \frac{6}{13}y,\ \frac{6}{13}x + \frac{4}{13}y\right\rangle, \text{ and} \\ proj_{L^\perp}(\vec{v}) &= b\langle -2, 3\rangle = \left(-\frac{2}{13}x + \frac{3}{13}y\right)\langle -2, 3\rangle \\ &= \left\langle \frac{4}{13}x - \frac{6}{13}y,\ -\frac{6}{13}x + \frac{9}{13}y\right\rangle. \end{aligned}$$

From these, we can see that these projections are indeed operators, with matrices:

$$[proj_L] = \begin{bmatrix} \frac{9}{13} & \frac{6}{13} \\ \frac{6}{13} & \frac{4}{13} \end{bmatrix} \quad \text{and} \quad [proj_{L^\perp}] = \begin{bmatrix} \frac{4}{13} & -\frac{6}{13} \\ -\frac{6}{13} & \frac{9}{13} \end{bmatrix}.$$

Finally, we can find the reflection across L via:

$$\begin{aligned} refl_L(\vec{v}) &= proj_L(\vec{v}) - proj_{L^\perp}(\vec{v}) \\ &= \left\langle \frac{9}{13}x + \frac{6}{13}y,\ \frac{6}{13}x + \frac{4}{13}y\right\rangle - \left\langle \frac{4}{13}x - \frac{6}{13}y,\ -\frac{6}{13}x + \frac{9}{13}y\right\rangle \\ &= \left\langle \frac{5}{13}x + \frac{12}{13}y,\ \frac{12}{13}x - \frac{5}{13}y\right\rangle, \end{aligned}$$

and thus reflection across L is indeed an operator, with:

$$[refl_L] = \begin{bmatrix} \frac{5}{13} & \frac{12}{13} \\ \frac{12}{13} & -\frac{5}{13} \end{bmatrix}.$$

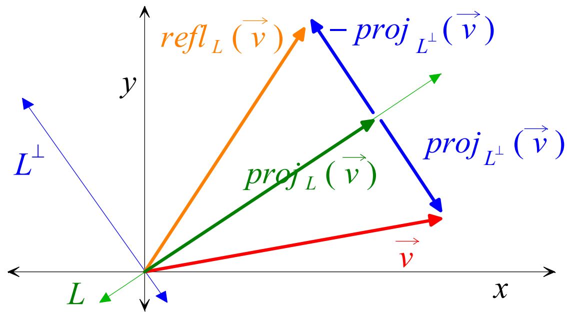
Let us demonstrate the result of these three operators on $\vec{v} = \langle 4, 1\rangle$. We get:

$$proj_L(\vec{v}) = \begin{bmatrix} \frac{9}{13} & \frac{6}{13} \\ \frac{6}{13} & \frac{4}{13} \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{42}{13} \\ \frac{28}{13} \end{bmatrix} \approx \begin{bmatrix} 3.23 \\ 2.15 \end{bmatrix},$$

$$proj_{L^\perp}(\vec{v}) = \begin{bmatrix} \frac{4}{13} & -\frac{6}{13} \\ -\frac{6}{13} & \frac{9}{13} \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{10}{13} \\ -\frac{15}{13} \end{bmatrix} \approx \begin{bmatrix} 0.77 \\ -1.15 \end{bmatrix}, \text{ and}$$

$$refl_L(\vec{v}) = \begin{bmatrix} \frac{5}{13} & \frac{12}{13} \\ \frac{12}{13} & -\frac{5}{13} \end{bmatrix}\begin{bmatrix} 4 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{32}{13} \\ \frac{43}{13} \end{bmatrix} \approx \begin{bmatrix} 2.46 \\ 3.31 \end{bmatrix}.$$

Let us put these all together in the following diagram:



Projections and Reflections with respect to $L : y = \frac{2}{3}x$

Projections and Reflections in $\mathbb{R}^3$

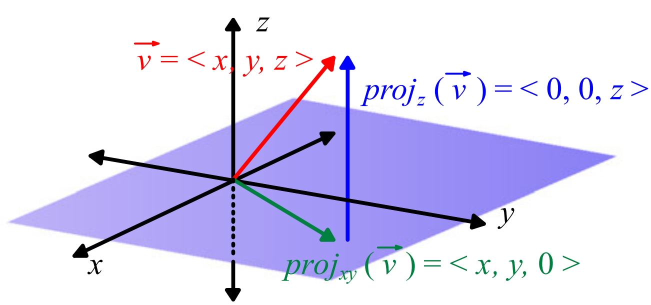
We can also define basic projection and reflection operators in $\mathbb{R}^3$, but we now have more varieties. Notice that in $\mathbb{R}^2$, the lines L through the origin are the non-trivial ***subspaces*** of $\mathbb{R}^2$. But for $\mathbb{R}^3$, the non-trivial subspaces are lines L through the origin as well as planes Π passing through the origin. The simplest subspaces are thus the x-, y- and z-axes, and the xy-, yz-, and xz-planes, and so for $\vec{v} = \langle x, y, z \rangle$, we can define their ***projection operators:***

$$proj_x(\langle x,y,z\rangle) = \langle x,0,0\rangle, \qquad proj_{xy}(\langle x,y,z\rangle) = \langle x,y,0\rangle,$$
$$proj_y(\langle x,y,z\rangle) = \langle 0,y,0\rangle, \qquad proj_{xz}(\langle x,y,z\rangle) = \langle x,0,z\rangle, \quad \text{and}$$
$$proj_z(\langle x,y,z\rangle) = \langle 0,0,z\rangle, \qquad proj_{yz}(\langle x,y,z\rangle) = \langle 0,y,z\rangle.$$

These projections are connected by the following relationships:

$$\vec{v} = \langle x,y,z\rangle = \langle x,0,0\rangle + \langle 0,y,z\rangle = proj_x(\langle x,y,z\rangle) + proj_{yz}(\langle x,y,z\rangle),$$
$$\vec{v} = \langle x,y,z\rangle = \langle 0,y,0\rangle + \langle x,0,z\rangle = proj_y(\langle x,y,z\rangle) + proj_{xz}(\langle x,y,z\rangle), \quad \text{and}$$
$$\vec{v} = \langle x,y,z\rangle = \langle 0,0,z\rangle + \langle x,y,0\rangle = proj_z(\langle x,y,z\rangle) + proj_{xy}(\langle x,y,z\rangle).$$

It is important to note in the pairings, for example, that the z-axis is the ***orthogonal complement*** of the xy-plane, and so on. Once again, we obtain an ***orthogonal decomposition***, as we verify that $\langle x,y,0\rangle \circ \langle 0,0,z\rangle = 0$, and so on. The relationships among $\vec{v}$, $proj_{xy}(\vec{v})$ and $proj_z(\vec{v})$ can be visualized by imagining the sun to be directly overhead at high noon: If $\vec{v}$ is an actual arrow anchored to the origin, then $proj_{xy}(\vec{v})$ would be the ***shadow*** that $\vec{v}$ makes on the ground:



A Basic Orthogonal Decomposition in $\mathbb{R}^3$

Let us look next at the reflections across the coordinate planes. To do this, pretend, for instance, that the xy-plane is a ***mirror***. The reflection of $\langle x, y, z\rangle$ across the xy-plane is thus $\langle x, y, -z\rangle$. But notice that we can write:

$$\begin{aligned} & refl_{xy}(\langle x, y, z\rangle) \\ &= \langle x, y, -z\rangle \\ &= \langle x, y, 0\rangle - \langle 0, 0, z\rangle \\ &= proj_{xy}(\langle x, y, z\rangle) - proj_z(\langle x, y, z\rangle). \end{aligned}$$

Again, since the z-axis is the ***orthogonal complement*** of the xy-plane, this equation is analogous to the equation from $\mathbb{R}^2$:

$$refl_L(\vec{v}) = proj_L(\vec{v}) - proj_{L^\perp}(\vec{v}).$$

Now, by ***reversing the roles*** of the xy-plane and the z-axis, we can define the reflection across the z-axis in $\mathbb{R}^3$ via:

$$\begin{aligned} & refl_z(\langle x, y, z\rangle) \\ &= proj_z(\langle x, y, z\rangle) - proj_{xy}(\langle x, y, z\rangle) \\ &= \langle 0, 0, z\rangle - \langle x, y, 0\rangle \\ &= \langle -x, -y, z\rangle. \end{aligned}$$

We can now summarize the six basic ***reflection operators:***

$$\begin{aligned} refl_x(\langle x, y, z\rangle) &= \langle x, -y, -z\rangle, & refl_{xy}(\langle x, y, z\rangle) &= \langle x, y, -z\rangle, \\ refl_y(\langle x, y, z\rangle) &= \langle -x, y, -z\rangle, & refl_{xz}(\langle x, y, z\rangle) &= \langle x, -y, z\rangle, \quad \text{and} \\ refl_z(\langle x, y, z\rangle) &= \langle -x, -y, z\rangle, & refl_{yz}(\langle x, y, z\rangle) &= \langle -x, y, z\rangle. \end{aligned}$$

We note that the matrices of the reflections across the coordinate planes are 3×3 Type 1 matrices with $c = -1$, analogous to what we saw in $\mathbb{R}^2$. We will see in the Exercises that Type 2 elementary matrices represent reflection operators in $\mathbb{R}^2$ and $\mathbb{R}^3$.

From our discussion above, it makes sense to investigate projections and reflections as they relate to an arbitrary plane Π together with its normal line L. If $\vec{v} = \langle x, y, z\rangle \in \mathbb{R}^3$, we want to produce the ***projection operators*** onto Π and L in order to express $\vec{v}$ as an ***orthogonal decomposition:***

$$\vec{v} = proj_\Pi(\vec{v}) + proj_L(\vec{v}), \quad \text{where } proj_\Pi(\vec{v}) \in \Pi \ \text{ and } \ proj_L(\vec{v}) \in L.$$

We must show that this sum can be constructed in exactly one way. From this, we get the ***reflection operators:***

$$refl_\Pi(\vec{v}) = proj_\Pi(\vec{v}) - proj_L(\vec{v}), \quad \text{and} \quad refl_L(\vec{v}) = proj_L(\vec{v}) - proj_\Pi(\vec{v}).$$

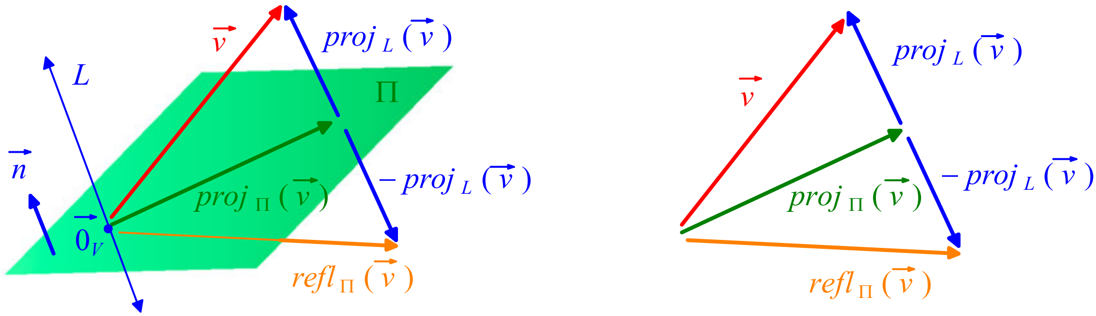
Let us illustrate these computations:

Example: Let Π be the plane in $\mathbb{R}^3$ with Cartesian equation:

$$3x - 5y + 2z = 0.$$

Π has normal vector $\vec{n} = \langle 3, -5, 2\rangle$, and normal line $L = Span(\{\vec{n}\})$. We will show that $proj_\Pi$, $proj_L$ and $refl_\Pi$, are all operators by finding their standard matrices (leaving $refl_L$ as an Exercise).

Let $\vec{v} = \langle x, y, z\rangle$ be any vector in $\mathbb{R}^3$. We see below the vectors that we are looking for:



The Relationships Among $\vec{v}$, $proj_\Pi(\vec{v})$, $proj_L(\vec{v})$ and $refl_\Pi(\vec{v})$

We will use a different strategy from that in our Example in $\mathbb{R}^2$. Let us begin with $proj_L(\vec{v})$. This vector is parallel to $\vec{n}$, so:

$$proj_L(\vec{v}) = k\langle 3, -5, 2\rangle = \langle 3k, -5k, 2k\rangle,$$

for some scalar multiple k. For now, let us assume that $proj_\Pi(\vec{v})$ actually exists, but we will verify later on that this assumption is justified.

Since $\vec{v} = proj_\Pi(\vec{v}) + proj_L(\vec{v})$, we must have:

$$\begin{aligned} & proj_\Pi(\vec{v}) \\ &= \vec{v} - proj_L(\vec{v}) \\ &= \langle x, y, z\rangle - \langle 3k, -5k, 2k\rangle \\ &= \langle x - 3k, y + 5k, z - 2k\rangle. \end{aligned}$$

But since $proj_\Pi(\vec{v}) \in \Pi$, this vector must be ***orthogonal*** to $\vec{n}$, so the correct value of k must satisfy the equation:

$$\begin{aligned} 0 &= \vec{n} \circ proj_\Pi(\vec{v}) \\ &= \langle 3, -5, 2\rangle \circ \langle x - 3k, y + 5k, z - 2k\rangle \\ &= 3(x - 3k) - 5(y + 5k) + 2(z - 2k) \\ &= 3x - 5y + 2z - (9 + 25 + 4)k. \end{aligned}$$

From this, we find:

$$k = \frac{3x - 5y + 2z}{38}$$

as the only possible solution. Thus, we get:

$$\begin{aligned} & proj_L(\vec{v}) \\ &= \langle 3k, -5k, 2k\rangle \\ &= \left\langle 3\left(\frac{3x - 5y + 2z}{38}\right), -5\left(\frac{3x - 5y + 2z}{38}\right), 2\left(\frac{3x - 5y + 2z}{38}\right)\right\rangle \\ &= \left\langle \frac{9x - 15y + 6z}{38}, \frac{-15x + 25y - 10z}{38}, \frac{6x - 10y + 4z}{38}\right\rangle. \end{aligned}$$

Consequently, we also get:

$$\begin{aligned}
&proj_\Pi(\vec{v})\\
&= \vec{v} - proj_L(\vec{v})\\
&= \langle x, y, z\rangle - \left\langle \frac{9x-15y+6z}{38}, \frac{-15x+25y-10z}{38}, \frac{6x-10y+4z}{38} \right\rangle\\
&= \left\langle \frac{29x+15y-6z}{38}, \frac{15x+13y+10z}{38}, \frac{-6x+10y+34z}{38} \right\rangle.
\end{aligned}$$

We will now check that we were justified in assuming that $proj_\Pi(\vec{v})$ exists by checking that the final vector above is on Π, that is:

$$\begin{aligned}
&\vec{n} \circ proj_\Pi(\vec{v})\\
&= 3\left(\frac{29x+15y-6z}{38}\right) - 5\left(\frac{15x+13y+10z}{38}\right) + 2\left(\frac{-6x+10y+34z}{38}\right)\\
&= \frac{1}{38}(87x+45y-18z-75x-65y-50z-12x+20y+68z) = 0,
\end{aligned}$$

for all x, y, and z. Thus, our solution is indeed correct. Next, we find the reflection operator across Π using our two projections:

$$\begin{aligned}
refl_\Pi(\vec{v}) &= proj_\Pi(\vec{v}) - proj_L(\vec{v})\\
&= \left\langle \frac{29x+15y-6z}{38}, \frac{15x+13y+10z}{38}, \frac{-6x+10y+34z}{38} \right\rangle -\\
&\quad \left\langle \frac{9x-15y+6z}{38}, \frac{-15x+25y-10z}{38}, \frac{6x-10y+4z}{38} \right\rangle\\
&= \left\langle \frac{10x+15y-6z}{19}, \frac{15x-6y+10z}{19}, \frac{-6x+10y+15z}{19} \right\rangle.
\end{aligned}$$

Finally, we can assemble the matrices of the three operators:

$$[proj_L] = \begin{bmatrix} \frac{9}{38} & -\frac{15}{38} & \frac{6}{38} \\ -\frac{15}{38} & \frac{25}{38} & -\frac{10}{38} \\ \frac{6}{38} & -\frac{10}{38} & \frac{4}{38} \end{bmatrix}; \quad [proj_\Pi] = \begin{bmatrix} \frac{29}{38} & \frac{15}{38} & -\frac{6}{38} \\ \frac{15}{38} & \frac{13}{38} & \frac{10}{38} \\ -\frac{6}{38} & \frac{10}{38} & \frac{34}{38} \end{bmatrix}; \quad [refl_\Pi] = \begin{bmatrix} \frac{10}{19} & \frac{15}{19} & -\frac{6}{19} \\ \frac{15}{19} & -\frac{6}{19} & \frac{10}{19} \\ -\frac{6}{19} & \frac{10}{19} & \frac{15}{19} \end{bmatrix}.$$

Notice that these three matrices all share a very special property: certain entries in each matrix appear ***in equal pairs***. For example, within each matrix, the entry in row 1, column 2, and the entry in row 2, column 1, are exactly the same, as well as the entries in row 2, column 3, and row 3, column 2. These are all examples of a special family of matrices called ***symmetric matrices***. They will play a very important role in Chapters 8 and 9.□

2.2 Section Summary

The ***rotation transformation*** $rot_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ that takes a vector $\vec{v}$ in standard position and rotates $\vec{v}$ counterclockwise by an angle of θ is a ***linear transformation***, with:

$$[rot_\theta] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

In $\mathbb{R}^2$, we can define the ***projections*** of a vector $\vec{v}$ onto the x-axis and y-axis, its ***reflections*** across the x-axis and y-axis, and its reflection across the origin, as:

$$proj_x(\langle x, y\rangle) = \langle x, 0\rangle, \qquad proj_y(\langle x, y\rangle) = \langle 0, y\rangle,$$
$$refl_x(\langle x, y\rangle) = \langle x, -y\rangle, \qquad refl_y(\langle x, y\rangle) = \langle -x, y\rangle, \text{ and}$$
$$refl_{\vec{\mathbf{0}}_2}(\langle x, y\rangle) = \langle -x, -y\rangle.$$

More generally, given any line L through the origin in $\mathbb{R}^2$ and its orthogonal complement $L^\perp$, it is possible to take any vector $\vec{v} \in \mathbb{R}^2$ and find its ***orthogonal decomposition*** $\vec{v} = proj_L(\vec{v}) + proj_{L^\perp}(\vec{v})$, where $proj_L(\vec{v})$ is parallel to L, and $proj_{L^\perp}(\vec{v})$ is parallel to $L^\perp$. The reflection across L can be defined using these projections as:

$$refl_L(\vec{v}) = proj_L(\vec{v}) - proj_{L^\perp}(\vec{v}).$$

In $\mathbb{R}^3$, we can define the six basic ***projection operators:***

$$proj_x(\langle x, y, z\rangle) = \langle x, 0, 0\rangle, \qquad proj_{xy}(\langle x, y, z\rangle) = \langle x, y, 0\rangle,$$
$$proj_y(\langle x, y, z\rangle) = \langle 0, y, 0\rangle, \qquad proj_{xz}(\langle x, y, z\rangle) = \langle x, 0, z\rangle, \quad \text{and}$$
$$proj_z(\langle x, y, z\rangle) = \langle 0, 0, z\rangle, \qquad proj_{yz}(\langle x, y, z\rangle) = \langle 0, y, z\rangle.$$

Similarly, we can define the six basic ***reflection operators:***

$$refl_x(\langle x, y, z\rangle) = \langle x, -y, -z\rangle, \qquad refl_{xy}(\langle x, y, z\rangle) = \langle x, y, -z\rangle,$$
$$refl_y(\langle x, y, z\rangle) = \langle -x, y, -z\rangle, \qquad refl_{xz}(\langle x, y, z\rangle) = \langle x, -y, z\rangle, \quad \text{and}$$
$$refl_z(\langle x, y, z\rangle) = \langle -x, -y, z\rangle, \qquad refl_{yz}(\langle x, y, z\rangle) = \langle -x, y, z\rangle.$$

More generally, given a random plane Π through the origin in $\mathbb{R}^3$ and its orthogonal complement L, it is possible to take any vector $\vec{v} \in \mathbb{R}^3$ and find its ***orthogonal decomposition:***

$$\vec{v} = proj_\Pi(\vec{v}) + proj_L(\vec{v}), \text{ where } proj_\Pi(\vec{v}) \in \Pi \text{ and } proj_L(\vec{v}) \in L.$$

From this, we get the ***reflection operators:***

$$refl_\Pi(\vec{v}) = proj_\Pi(\vec{v}) - proj_L(\vec{v}), \text{ and } refl_L(\vec{v}) = proj_L(\vec{v}) - proj_\Pi(\vec{v}).$$

2.2 Exercises

For Exercises (1) to (5): For the following angles θ: (a) Find the standard matrix of rot_θ, the counterclockwise rotation by θ in $\mathbb{R}^2$; (b) Compute $rot_\theta(\vec{v})$ for $\vec{v} = \langle 5, 3 \rangle$, providing both exact and approximate answers; and (c) Sketch $\vec{v}$ and $rot_\theta(\vec{v})$ and check with a ruler and protractor that $rot_\theta(\vec{v})$ has the same length as $\vec{v}$ but is rotated counterclockwise by θ.

1. $\pi/6$
2. $\sin^{-1}(3/5)$
3. $\cos^{-1}(-5/13)$
4. $\tan^{-1}(5/12)$
5. $5\pi/8$. Hint: Use the Half-Angle Formulas.

For Exercises (6) to (10): If $\theta < 0$, the formula for $[rot_\theta]$ is exactly the same, but the geometric effect is a ***clockwise*** rotation by $|\theta|$. For the following θ: (a) Find the matrix of rot_θ; (b) Compute $rot_\theta(\vec{v})$ for $\vec{v} = \langle 5, 3 \rangle$, providing both exact and approximate answers, and (c) Sketch $\vec{v}$ and $rot_\theta(\vec{v})$ and check with a ruler and protractor that $rot_\theta(\vec{v})$ has the same length as $\vec{v}$ but is rotated clockwise by $|\theta|$.

6. $-2\pi/3$
7. $\sin^{-1}(-20/29)$
8. $\tan^{-1}(-4/3)$
9. $\sin^{-1}(15/17) - \pi$
10. $-2\cos^{-1}(20/29)$. Hint: Use the Double Angle Formulas.

For Exercises (11) to (15): (a) Find the matrices of $proj_L$, $proj_{L^\perp}$ and $refl_L$ for the lines L in $\mathbb{R}^2$ given by the following Cartesian equations; (b) Compute the values of these three operators on $\vec{v} = \langle 3, 2 \rangle$, providing both exact and approximate answers; and (c) Sketch the graphs of L, $L^\perp$, $\vec{v}$ and the images of $\vec{v}$ under these three operators, as shown in the Example in this Section.

11. $y = \dfrac{3}{5}x$
12. $y = \dfrac{4}{7}x$
13. $y = -\dfrac{4}{5}x$
14. $y = -\dfrac{7}{3}x$
15. $y = 3x$

For Exercises (16) to (20): Find the standard matrix of $proj_\Pi$, $proj_L$ and $refl_\Pi$ for the planes Π in $\mathbb{R}^3$ and corresponding normal line L, where Π is given by the following Cartesian equations:

16. $4x + 2y - 3z = 0$
17. $2x - 5y + 6z = 0$
18. $7x - 4y - 5z = 0$
19. $3x + 5z = 0$

20. $2y - 7z = 0$

21. Find the matrix of the counterclockwise rotation in $\mathbb{R}^2$ by $\theta = \pi/2$. Is this a 2×2 elementary matrix? Why or why not?

22. Find the matrix of $refl_L$, the reflection across L for the normal line to the plane Π with Cartesian equation $3x - 5y + 2z = 0$ as seen in the final Example of this Section. How is $[refl_L]$ related to $[refl_\Pi]$?

23. Find the matrix of $refl_L$, the reflection across L for the normal line to the plane Π with Cartesian equation $2x - 5y + 6z = 0$ from Exercise 17.

24. ***Type 2 Elementary Matrices:*** We will see in this Exercise that Type 2 elementary matrices indeed correspond to ***reflections*** in $\mathbb{R}^2$ or $\mathbb{R}^3$.

 Let T be the operator with $[T] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

 Note that this is the ***only*** 2×2 Type 2 elementary matrix.

 a. Find $T(\vec{v})$ and $T(\vec{w})$ for $\vec{v} = \langle 5, 2 \rangle$ and $\vec{w} = \langle -3, 4 \rangle$.

 b. Sketch the four vectors involved in (a), and the line $y = x$. Convince yourself that $T(\vec{v})$ is a mirror-image of $\vec{v}$ for each vector in (a).

 c. Now, find the standard matrices of $proj_L$, $proj_{L^\perp}$ and $refl_L$ for the line in $\mathbb{R}^2$:

$$y = x.$$

 Which of these matrices correspond to $[T]$?

 d. Consider the 3×3 Type 2 elementary matrix $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$.

 Show that this is the matrix of the reflection in $\mathbb{R}^3$, $refl_\Pi$, across the plane Π with Cartesian equation $y = x$, i.e. $x - y = 0$.

 e. There are exactly two other 3×3 Type 2 elementary matrices. Find them, and for each, find the corresponding plane Π such that the reflection across Π has this matrix as its standard matrix.

 f. Consider the 4×4 Type 2 elementary matrix $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$.

 If T is the corresponding operator, find an explicit formula for $T(\langle x_1, x_2, x_3, x_4 \rangle)$. Write a sentence describing in words what T does to any vector $\vec{v} \in \mathbb{R}^4$. (Since we cannot visualize $\mathbb{R}^4$, we cannot see the effect of T on $\mathbb{R}^4$, but we can still explain what T ***does*** to a vector.)

25. Suppose that $\vec{v} = \langle a, b\rangle$ is a ***unit vector*** in $\mathbb{R}^2$, L is the line $Span(\{\vec{v}\})$, and $L^\perp$ is the orthogonal complement of L. Prove that the matrices of $proj_L$, $proj_{L^\perp}$ and $refl_L$ are given by:

$$[proj_L] = \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix},$$

$$[proj_{L^\perp}] = \begin{bmatrix} b^2 & -ab \\ -ab & a^2 \end{bmatrix}, \text{ and}$$

$$[refl_L] = \begin{bmatrix} a^2 - b^2 & 2ab \\ 2ab & b^2 - a^2 \end{bmatrix}.$$

26. Suppose that $\vec{n} = \langle a, b, c\rangle$ is a ***unit vector*** in $\mathbb{R}^3$, Π is the plane in $\mathbb{R}^3$ with equation $ax + by + cz = 0$, and L is the normal line $Span(\{\vec{n}\})$ to Π. Prove that the matrices of $proj_L$, $proj_\Pi$ and $refl_\Pi$ are given by:

$$[proj_L] = \begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix},$$

$$[proj_\Pi] = \begin{bmatrix} 1-a^2 & -ab & -ac \\ -ab & 1-b^2 & -bc \\ -ac & -bc & 1-c^2 \end{bmatrix} = \begin{bmatrix} b^2+c^2 & -ab & -ac \\ -ab & a^2+c^2 & -bc \\ -ac & -bc & a^2+b^2 \end{bmatrix}, \text{ and}$$

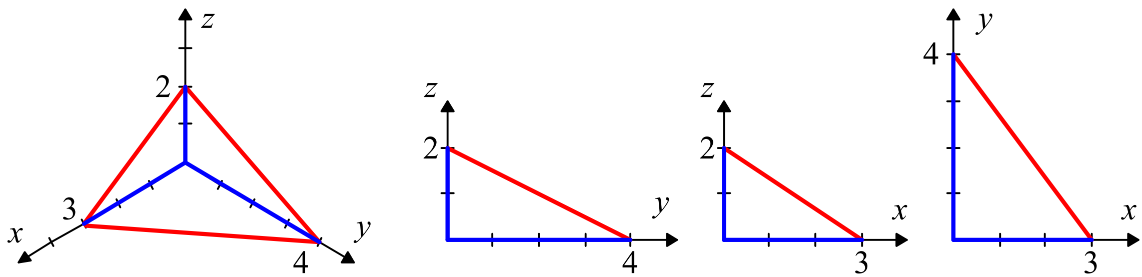
$$[refl_\Pi] = \begin{bmatrix} 1-2a^2 & -2ab & -2ac \\ -2ab & 1-2b^2 & -2bc \\ -2ac & -2bc & 1-2c^2 \end{bmatrix}$$
$$= \begin{bmatrix} b^2+c^2-a^2 & -2ab & -2ac \\ -2ab & a^2+c^2-b^2 & -2bc \\ -2ac & -2bc & a^2+b^2-c^2 \end{bmatrix}.$$

27. Let us go backwards. Suppose you were told that:

$$[proj_\Pi] = \frac{1}{109}\begin{bmatrix} 73 & 18 & -48 \\ 18 & 100 & 24 \\ -48 & 24 & 45 \end{bmatrix},$$

for some plane Π. Find a Cartesian equation for Π.

28. ***Application — Drawing Three-Dimensional Objects:*** We can use the idea of a projection in order to precisely draw a 3-dimensional object from any perspective that we choose. Traditionally, we draw the positive coordinate axes using three half lines that make an angle of 120^0 with each other. Let us call this the standard perspective. Our instincts would tell us that we would see the standard perspective if we look at the origin from the direction $\langle 1,1,1\rangle$. Let us illustrate this by drawing the tetrahedron with vertices $(0,0,0)$, $(3,0,0)$, $(0,4,0)$ and $(0,0,2)$:



A Tetrahedron in Standard Perspective, and its Front, Side and Top Views

The ***front view*** is what we see from the $\vec{i}$ direction, the ***side view*** is what we see from the $-\vec{j}$ direction (so that the positive x-axis goes to the ***right*** instead of left), and the ***top view*** is what we see from the $\vec{k}$ direction.

More generally, let $\vec{n} = \langle a, b, c\rangle$ be a ***unit vector*** in $\mathbb{R}^3$, and let Π be the plane $ax + by + cz = 0$. Imagine that you are standing in the direction of $\vec{n}$ and facing an object near the origin. To avoid getting dizzy, orient your head in such a manner that the z-axis still looks like a ***vertical*** line from our vantage point (this is possible as long as $\vec{n} \neq \pm\vec{k}$). Let us call the projections of the edges of the solid on Π, as seen from this perspective, the ***view*** of the solid from the direction $\vec{n}$. However, the edges of the object could now appear ***shorter*** than their true lengths, and two edges that are supposed to be perpendicular to each other may no longer appear to be perpendicular.

The purpose of this Exercise is to perform the calculations to find the ***perceived lengths*** of $\vec{i}$, $\vec{j}$ and $\vec{k}$ and the ***perceived angles*** between any two of these unit vectors. With these calculations, we can precisely draw on a plane (such as a computer screen or a page in a book) the view of the solid from the direction $\vec{n}$.

Let us use the final Example in the text for motivation. For the plane Π with equation $3x - 5y + 2z = 0$:

$$[proj_{\Pi}] = \begin{bmatrix} \frac{29}{38} & \frac{15}{38} & -\frac{6}{38} \\ \frac{15}{38} & \frac{13}{38} & \frac{10}{38} \\ -\frac{6}{38} & \frac{10}{38} & \frac{34}{38} \end{bmatrix}.$$

The normal vector $\langle 3,-5,2\rangle$ is not a unit vector, though, so we normalize it to:

$$\vec{n} = \frac{1}{\sqrt{38}}\langle 3,-5,2\rangle.$$

Now, recall that for any linear transformation T, the ***columns*** of T are $T(\vec{e}_1)$ through $T(\vec{e}_n)$, respectively. Thus:

$$proj_\Pi\left(\vec{i}\right) = \left\langle \frac{29}{38}, \frac{15}{38}, -\frac{6}{38} \right\rangle,$$
$$proj_\Pi\left(\vec{j}\right) = \left\langle \frac{15}{38}, \frac{13}{38}, \frac{10}{38} \right\rangle, \text{ and}$$
$$proj_\Pi\left(\vec{k}\right) = \left\langle -\frac{6}{38}, \frac{10}{38}, \frac{34}{38} \right\rangle.$$

a. Find the lengths of these three vectors. Do you notice a pattern? Find decimal approximations for them.

b. Find the dot products of all three pairs of these three vectors.

c. Find the angle $\alpha_{i,j}$ between $proj_\Pi\left(\vec{i}\right)$ and $proj_\Pi\left(\vec{j}\right)$ (review Section 1.3).

Approximate $\alpha_{i,j}$ to 2 decimal places, and express your answer in degrees.

d. Repeat (c) for the angle $\alpha_{j,k}$ between $proj_\Pi\left(\vec{j}\right)$ and $proj_\Pi\left(\vec{k}\right)$ and similarly for $\alpha_{i,k}$.

e. Now, we draw our coordinate system based on the view from $\vec{n}$. Draw a vertical line which will be the z-axis. Use your protractor to measure the three angles you found in (b) and (c). Use the three lengths that you found in (a) to measure off tick-marks on each of the corresponding axes, using 1 unit = 5 centimeters. Use these tick-marks to sketch the view of the tetrahedron found in the introduction from the perspective of $\vec{n}$.

Let us generalize the construction above. We saw in a previous Exercise that:

$$[proj_\Pi] = \begin{bmatrix} 1-a^2 & -ab & -ac \\ -ab & 1-b^2 & -bc \\ -ac & -bc & 1-c^2 \end{bmatrix}$$

where $\vec{n} = \langle a, b, c\rangle$ is a ***unit*** normal vector to Π.

f. Show that $\left\| proj_\Pi\left(\vec{i}\right) \right\| = \sqrt{1-a^2}$.

Hint: for (f) and (g), you will need the fact that $a^2 + b^2 + c^2 = 1$.
State and prove analogous formulas for $\left\| proj_\Pi\left(\vec{j}\right) \right\|$ and $\left\| proj_\Pi\left(\vec{k}\right) \right\|$.
Again, these are the ***perceived lengths*** of the three unit vectors, as seen from $\vec{n}$.

g. Show that $proj_\Pi\left(\vec{i}\right) \circ proj_\Pi\left(\vec{j}\right) = -ab$.

State and prove analogous formulas for $proj_\Pi\left(\vec{i}\right) \circ proj_\Pi\left(\vec{k}\right)$ and $proj_\Pi\left(\vec{j}\right) \circ proj_\Pi\left(\vec{k}\right)$.

h. Show that $\cos(\alpha_{i,j}) = \dfrac{-ab}{\sqrt{1-a^2}\sqrt{1-b^2}}$.

State and prove analogous formulas for $\cos(\alpha_{j,k})$ and $\cos(\alpha_{i,k})$.

Again, $\alpha_{i,j}$, $\alpha_{i,k}$, and $\alpha_{j,k}$ are the ***perceived angles*** between $\vec{i}$ and $\vec{j}$, $\vec{i}$ and $\vec{k}$, and $\vec{j}$ and $\vec{k}$, respectively, as seen from $\vec{n}$.

i. Bonus: If a, b and c are all ***positive***, that is, we are viewing the origin from the ***first octant***, prove that the three angles must satisfy:

$$\cos^{-1}\left(\frac{-ab}{\sqrt{1-a^2}\sqrt{1-b^2}}\right) + \cos^{-1}\left(\frac{-ac}{\sqrt{1-a^2}\sqrt{1-c^2}}\right) + \cos^{-1}\left(\frac{-bc}{\sqrt{1-b^2}\sqrt{1-c^2}}\right) = 2\pi.$$

Hint: interpret this equation geometrically.
How are the three positive axes oriented in relation to each other?

2.3 Operations on Linear Transformations and Matrices

In Algebra, we can take two functions, say $f(x)$ and $g(x)$, and combine them using addition, subtraction, multiplication and division. In Linear Algebra, though, the only arithmetic operations that we can perform on vectors are ***addition*** and ***scalar multiplication***, so we begin with the following:

Definitions: If $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^n \to \mathbb{R}^m$ are ***linear transformations***, and $k \in \mathbb{R}$, then we can define the ***sum***, ***difference*** and ***scalar product*** of these transformations as:

$$T_1 + T_2 : \mathbb{R}^n \to \mathbb{R}^m,$$

$$T_1 - T_2 : \mathbb{R}^n \to \mathbb{R}^m, \text{ and}$$

$$kT_1 : \mathbb{R}^n \to \mathbb{R}^m,$$

the functions with actions given, respectively, by: for any $\vec{v} \in \mathbb{R}^n$:

$$(T_1 + T_2)(\vec{v}) = T_1(\vec{v}) + T_2(\vec{v}),$$

$$(T_1 - T_2)(\vec{v}) = T_1(\vec{v}) - T_2(\vec{v}), \text{ and}$$

$$(kT_1)(\vec{v}) = kT_1(\vec{v}).$$

We will prove later that these functions are actually ***linear transformations*** by finding their matrices using $[T_1]$ and $[T_2]$. We will combine these two matrices using arithmetic operations on matrices that we will also be defining in this section, thus producing matrices for these new linear transformations.

Example: Suppose that:

$$T_1, T_2 : \mathbb{R}^3 \to \mathbb{R}^2, \text{ given by:}$$

$$T_1(\langle x, y, z \rangle) = \langle 3x - 2y + 5z, 2x + 4y - 3z \rangle, \text{ and}$$

$$T_2(\langle x, y, z \rangle) = \langle 4x + 7y - 2z, x - y + 4z \rangle.$$

Their matrices are:

$$[T_1] = \begin{bmatrix} 3 & -2 & 5 \\ 2 & 1 & -3 \end{bmatrix} \text{ and } [T_2] = \begin{bmatrix} 4 & 7 & -2 \\ 1 & -1 & 4 \end{bmatrix}.$$

The sum of these two linear transformations is:

$$\begin{aligned} &(T_1 + T_2)(\langle x, y, z \rangle) \\ &= \langle 3x - 2y + 5z, 2x + y - 3z \rangle + \langle 4x + 7y - 2z, x - y + 4z \rangle \\ &= \langle 7x + 5y + 3z, 3x + z \rangle. \end{aligned}$$

Notice that this can be written as a matrix product:

$$(T_1 + T_2)(\langle x, y, z \rangle) = \begin{bmatrix} 7 & 5 & 3 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Thus, at least for this example, $T_1 + T_2$ is indeed a ***linear transformation***, and:

$$[T_1 + T_2] = \begin{bmatrix} 7 & 5 & 3 \\ 3 & 0 & 1 \end{bmatrix}.$$

Notice that the matrix that we obtained for $T_1 + T_2$ has entries that are the sum of the corresponding entries for the matrices for T_1 and T_2.

Similarly, let us look at a scalar product:

$$\begin{aligned} & 7T_1(\langle x, y, z\rangle) \\ &= 7\langle 3x - 2y + 5z, 2x + y - 3z\rangle \\ &= \langle 21x - 14y + 35z, 14x + 7y - 21z\rangle. \end{aligned}$$

Its matrix is thus:

$$[7T_1] = \begin{bmatrix} 21 & -14 & 35 \\ 14 & 7 & -21 \end{bmatrix}.$$

The entries of this matrix are 7 times each corresponding entry from the matrix for T_1. These observations are hardly coincidences, and it is therefore natural to define next the addition, subtraction, and scalar multiplication of matrices.□

The Arithmetic of Matrices

We will use the notation from Chapter 1, where we denote by $(A)_{i,j}$ the ***entry*** of the matrix A in row i, column j. This allows us to make the following:

Definitions: If A and B are both $m \times n$ matrices, and $k \in \mathbb{R}$ is any scalar, then we can define the ***sum***, ***difference*** and ***scalar product*** of these matrices, denoted:

$$A + B,\ A - B, \text{ and } kA.$$

These are also $m \times n$ matrices with entries given by:

$$\begin{aligned} (A + B)_{i,j} &= (A)_{i,j} + (B)_{i,j}, \\ (A - B)_{i,j} &= (A)_{i,j} - (B)_{i,j}, \quad \text{and} \\ (kA)_{i,j} &= k(A)_{i,j}. \end{aligned}$$

In particular, we can define the ***negative*** of a matrix, $-A$, to be:

$$-A = (-1)A,$$

with the property that:

$$A + (-A) = (-A) + A = \mathbf{0}_{m\times n}.$$

Since we are adding or subtracting corresponding pairs of entries from A and B when we compute $A + B$ and $A - B$, the same is true if we partition A and B into columns or rows. Thus, we have the following:

Theorem: If A and B are both $m \times n$ matrices, and we partition A and B into ***columns*** as:

$$A = \begin{bmatrix} \vec{a}_1 \, \vec{a}_2 \ldots \vec{a}_n \end{bmatrix}, \text{ and } B = \begin{bmatrix} \vec{b}_1 \, \vec{b}_2 \ldots \vec{b} \end{bmatrix},$$

then we have:

$$\begin{aligned} A + B &= \left[\left(\vec{a}_1 + \vec{b}_1\right) \left(\vec{a}_2 + \vec{b}_2\right) \ldots \left(\vec{a}_n + \vec{b}_n\right) \right], \\ A - B &= \left[\left(\vec{a}_1 - \vec{b}_1\right) \left(\vec{a}_2 - \vec{b}_2\right) \ldots \left(\vec{a}_n - \vec{b}_n\right) \right], \text{ and} \\ kA &= \left[k\vec{a}_1 \, k\vec{a}_2 \ldots k\vec{a}_n \right], \end{aligned}$$

for any scalar $k \in \mathbb{R}$. Similarly, if we partition A and B into ***rows*** as:

$$A = \begin{bmatrix} \vec{r}_1 \\ \vec{r}_2 \\ \vdots \\ \vec{r}_m \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} \vec{s}_1 \\ \vec{s}_2 \\ \vdots \\ \vec{s}_m \end{bmatrix},$$

then we have:

$$A + B = \begin{bmatrix} \vec{r}_1 + \vec{s}_1 \\ \vec{r}_2 + \vec{s}_2 \\ \vdots \\ \vec{r}_m + \vec{s}_m \end{bmatrix}, \quad A - B = \begin{bmatrix} \vec{r}_1 - \vec{s}_1 \\ \vec{r}_2 - \vec{s}_2 \\ \vdots \\ \vec{r}_m - \vec{s}_m \end{bmatrix}, \quad \text{and} \quad kA = \begin{bmatrix} k\vec{r}_1 \\ k\vec{r}_2 \\ \vdots \\ k\vec{r}_m \end{bmatrix}.$$

Now we are ready to construct the matrices for our combined linear transformations.

Theorem: If $T_1 : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{m}}$ and $T_2 : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{m}}$ are ***linear transformations***, with matrices $[T_1]$ and $[T_2]$ respectively, and k is any scalar, then for any $\vec{v} \in \mathbb{R}^{\boldsymbol{n}}$:

$$\begin{aligned} (T_1 + T_2)(\vec{v}) &= ([T_1] + [T_2])\vec{v}, \\ (T_1 - T_2)(\vec{v}) &= ([T_1] - [T_2])\vec{v}, \text{ and} \\ (kT_1)(\vec{v}) &= (k[T_1])\vec{v}. \end{aligned}$$

Consequently, $T_1 + T_2$, $T_1 - T_2$ and kT_1 are ***linear transformations*** with matrices given by, respectively:

$$\begin{aligned} [T_1 + T_2] &= [T_1] + [T_2], \\ [T_1 - T_2] &= [T_1] - [T_2], \text{ and} \\ [kT_1] &= k[T_1]. \end{aligned}$$

Proof: We will prove the first property and leave the other two as Exercises. For convenience, suppose that $[T_1] = A$ and $[T_2] = B$. We want to show that $(T_1 + T_2)(\vec{v}) = (A + B)\vec{v}$, and therefore $T_1 + T_2$ is indeed a linear transformation with $[T_1 + T_2] = A + B$. By definition:

$$
\begin{aligned}
(T_1 + T_2)(\vec{v}) &= T_1(\vec{v}) + T_2(\vec{v}) \\
&= A\vec{v} + B\vec{v}.
\end{aligned}
$$

Now, as we did in the previous Theorem, let us partition A and B into ***columns:***

$$A = \left[\, \vec{a}_1\ \vec{a}_2\ \ldots\ \vec{a}_n \,\right], \text{ and } B = \left[\, \vec{b}_1\ \vec{b}_2\ \ldots\ \vec{b}_n \,\right].$$

Thus we have:

$$
\begin{aligned}
&(T_1 + T_2)(\vec{v}) \\
&= A\vec{v} + B\vec{v} \\
&= \left[\, \vec{a}_1\ \vec{a}_2\ \ldots\ \vec{a}_n \,\right]\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \left[\, \vec{b}_1\ \vec{b}_2\ \ldots\ \vec{b}_n \,\right]\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \\
&= v_1\vec{a}_1 + v_2\vec{a}_2 + \cdots + v_n\vec{a}_n + v_1\vec{b}_1 + v_2\vec{b}_2 + \cdots + v_n\vec{b}_n \\
&= v_1\vec{a}_1 + v_1\vec{b}_1 + v_2\vec{a}_2 + v_2\vec{b}_2 + \cdots + v_n\vec{a}_n + v_n\vec{b}_n \\
&= v_1\left(\vec{a}_1 + \vec{b}_1\right) + v_2\left(\vec{a}_2 + \vec{b}_2\right) + \cdots + v_n\left(\vec{a}_n + \vec{b}_n\right) \\
&\quad \text{(by the “Right” Distributive Property from Section 1.1)} \\
&= \left[\, \left(\vec{a}_1 + \vec{b}_1\right)\ \left(\vec{a}_2 + \vec{b}_2\right)\ \ldots\ \left(\vec{a}_n + \vec{b}_n\right) \,\right]\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = (A + B)\vec{v}. \blacksquare
\end{aligned}
$$

Compositions of Linear Transformations

Now let us see a more sophisticated way to combine linear transformations. In Algebra, we can form the ***composition*** of two functions $f(x)$ and $g(x)$ by:

$$(f \circ g)(x) = f(g(x)),$$

We must therefore be careful that the ***value*** of $g(x)$ is a member of the ***domain*** of $f(x)$ in order to successfully perform this computation. We can compose linear transformations as well, but we likewise have to be aware of the domains and codomains:

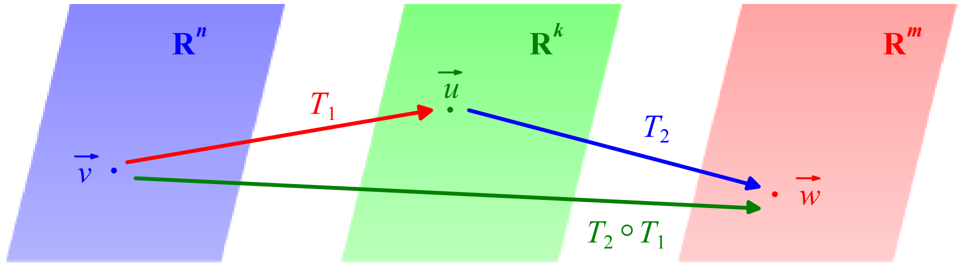
Definition/Theorem: If $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ are ***linear transformations***, then we can define their ***composition:***

$$T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^m,$$

which is also a ***linear transformation***, whose action is given as follows:

Suppose $\vec{u} \in \mathbb{R}^n$, $T_1(\vec{u}) = \vec{v} \in \mathbb{R}^k$, and $T_2(\vec{v}) = \vec{w} \in \mathbb{R}^m$. Then:

$$(T_2 \circ T_1)(\vec{u}) = T_2(T_1(\vec{u})) = T_2(\vec{v}) = \vec{w}.$$



The Composition of T_1 with T_2

Note that the ***domain*** of T_2 must be the ***codomain*** of T_1 in order for $T_2 \circ T_1$ to be defined. This is called a ***compatibility requirement***. The fact that the composition of two linear transformations is also a linear transformation will follow directly from the linearity properties that the two transformations enjoy:

Proof of Linearity: We need to show Additivity and Homogeneity:

$$(T_2 \circ T_1)(\vec{u}_1 + \vec{u}_2) = (T_2 \circ T_1)(\vec{u}_1) + (T_2 \circ T_1)(\vec{u}_2) \quad \text{and}$$

$$(T_2 \circ T_1)(k\vec{u}_1) = k(T_2 \circ T_1)(\vec{u}_1).$$

We will prove the first equation and leave the second as an Exercise:

$$\begin{aligned}
(T_2 \circ T_1)(\vec{u}_1 + \vec{u}_2) &= T_2(T_1(\vec{u}_1 + \vec{u}_2)) && \text{by the definition of } T_2 \circ T_1, \\
&= T_2(T_1(\vec{u}_1) + T_1(\vec{u}_2)) && \text{by the additivity of } T_1, \\
&= T_2(T_1(\vec{u}_1)) + T_2(T_1(\vec{u}_2)) && \text{by the additivity of } T_2, \\
&= (T_2 \circ T_1)(\vec{u}_1) + (T_2 \circ T_1)(\vec{u}_2) && \text{by the definition of } T_2 \circ T_1. \blacksquare
\end{aligned}$$

Example: Suppose that $T_1 : \mathbb{R}^3 \to \mathbb{R}^2$ and $T_2 : \mathbb{R}^2 \to \mathbb{R}^4$ are given by:

$$T_1(\langle x, y, z \rangle) = \langle x - y + 2z,\, 3x + y - z \rangle, \quad \text{and}$$

$$T_2(\langle u, v \rangle) = \langle u + 2v,\, 5u - v,\, 3u,\, u + v \rangle.$$

Then $(T_2 \circ T_1) : \mathbb{R}^3 \to \mathbb{R}^4$ is given by:

$$\begin{aligned}
&(T_2 \circ T_1)(\langle x, y, z \rangle) \\
&= T_2(T_1(\langle x, y, z \rangle)) \\
&= T_2(\langle x - y + 2z,\, 3x + y - z \rangle) \\
&= \langle x - y + 2z + 2(3x + y - z),\, 5(x - y + 2z) - (3x + y - z), \\
&\quad\ 3(x - y + 2z),\, x - y + 2z + 3x + y - z \rangle \\
&= \langle 7x + y,\, 2x - 6y + 11z,\, 3x - 3y + 6z,\, 4x + z \rangle.
\end{aligned}$$

Thus $T_2 \circ T_1$ is indeed a linear transformation, with matrix:

$$[T_2 \circ T_1] = \begin{bmatrix} 7 & 1 & 0 \\ 2 & -6 & 11 \\ 3 & -3 & 6 \\ 4 & 0 & 1 \end{bmatrix}. \square$$

Clearly, finding the matrix of the composition is not an obvious process. But now that we have seen the addition, subtraction and scalar products of matrices, it would probably be no surprise that we need to ***multiply*** $[T_2]$ and $[T_1]$ to get $[T_2 \circ T_1]$. Let us now see how to generalize the construction of a ***matrix product***.

General Matrix Products

In the previous Chapter, we defined the matrix product $A\vec{x}$, where A is an $m \times n$ matrix and $\vec{x}$ is an $n \times 1$ column matrix, as the $m \times 1$ column matrix obtained by forming the linear combination of the columns of A with corresponding coefficients from $\vec{x}$. We will now extend this definition in a natural manner:

> ***Definition — Matrix Product:*** If A is an $m \times k$ matrix, and B is a $k \times n$, then we can construct the $m \times n$ ***matrix product*** AB, where:
>
> $$\text{column } i \text{ of } AB = A \times (\text{column } i \text{ of } B).$$
>
> In other words, if we partition B into ***columns***, and write:
>
> $$B = \left[\, \vec{b}_1 \,|\, \vec{b}_2 \,|\, \cdots \,|\, \vec{b}_n \,\right],$$
>
> then:
>
> $$AB = \left[\, A\vec{b}_1 \,|\, A\vec{b}_2 \,|\, \cdots \,|\, A\vec{b}_n \,\right].$$

Notice that we require that the number of columns of A must equal the number of rows of B. This is again known as a ***compatibility requirement***. Also, notice that from the definition of AB, we can write the individual entries of AB as:

$$(AB)_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \cdots + a_{i,k}b_{k,j}.$$

But recall that a matrix product $A\vec{c}$ is a column vector consisting of the ***dot product*** of the rows of A with $\vec{c}$, and in fact we can see that the formula above looks just like a dot product. If we partition A into its ***rows*** $\vec{r}_1, \vec{r}_2 \ldots \vec{r}_m$ and B into its ***columns*** $\vec{b}_1, \vec{b}_2 \ldots \vec{b}_n$, then, indeed:

$$AB = \begin{bmatrix} \vec{r}_1 \\ \vec{r}_2 \\ \vdots \\ \vec{r}_m \end{bmatrix} \begin{bmatrix} \vec{b}_1 & \vec{b}_2 & \cdots & \vec{b}_n \end{bmatrix} = \begin{bmatrix} \vec{r}_1 \circ \vec{b}_1 & \vec{r}_1 \circ \vec{b}_2 & \cdots & \vec{r}_1 \circ \vec{b}_n \\ \vec{r}_2 \circ \vec{b}_1 & \vec{r}_2 \circ \vec{b}_2 & \cdots & \vec{r}_2 \circ \vec{b}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{r}_m \circ \vec{b}_1 & \vec{r}_m \circ \vec{b}_2 & \cdots & \vec{r}_m \circ \vec{b}_n \end{bmatrix}.$$

Let us summarize this observation in the following:

Theorem: If A is an $m \times k$ matrix, and B is a $k \times n$, then AB is an $m \times n$ matrix, and:

$$(AB)_{i,j} = \textbf{\textit{row i}} \text{ of } A \circ \textbf{\textit{column j}} \text{ of } B.$$

Example: Let us continue with our previous Example, and study:

$$T_1 : \mathbb{R}^3 \to \mathbb{R}^2, \text{ given by:}$$
$$T_1(\langle x, y, z\rangle) = \langle x - y + 2z, 3x + y - z\rangle, \text{ and}$$
$$T_2 : \mathbb{R}^2 \to \mathbb{R}^4, \text{ given by:}$$
$$T_2(\langle u, v\rangle) = \langle u + 2v, 5u - v, 3u, u + v\rangle.$$

Their matrices are:

$$[T_1] = A = \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \end{bmatrix}, \text{ and } [T_2] = B = \begin{bmatrix} 1 & 2 \\ 5 & -1 \\ 3 & 0 \\ 1 & 1 \end{bmatrix}.$$

Since B is a 4×2 matrix and A is a 2×3, the product BA is defined and should be a 4×3 matrix. Multiplying these two matrices, we get:

$$BA = \begin{bmatrix} 1 & 2 \\ 5 & -1 \\ 3 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ 3 & 1 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \cdot 1 + 2 \cdot 3 & 1 \cdot -1 + 2 \cdot 1 & 1 \cdot 2 + 2 \cdot -1 \\ 5 \cdot 1 + (-1) \cdot 3 & 5 \cdot -1 + (-1) \cdot 1 & 5 \cdot 2 + (-1) \cdot (-1) \\ 3 \cdot 1 + 0 \cdot 3 & 3 \cdot -1 + 0 \cdot 1 & 3 \cdot 2 + 0 \cdot -1 \\ 1 \cdot 1 + 1 \cdot 3 & 1 \cdot -1 + 1 \cdot 1 & 1 \cdot 2 + 1 \cdot -1 \end{bmatrix} = \begin{bmatrix} 7 & 1 & 0 \\ 2 & -6 & 11 \\ 3 & -3 & 6 \\ 4 & 0 & 1 \end{bmatrix}. \square$$

This shows that, at least for this example:

$$[T_2 \circ T_1] = BA = [T_2][T_1].$$

Before we can prove this property in general, though, we need to learn more about the properties of matrix products, which we will see in the next Section. In the meantime, we can generalize a construction from Chapter 1:

Linear Combinations of Linear Transformations and Matrices

If $T_1, T_2, \ldots, T_k$ are all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, and $c_1, c_2, \ldots, c_k$ are any scalars, then we can construct the ***linear combination*** of these linear transformations with the corresponding coefficients in the natural manner, by:

$$(c_1T_1 + c_2T_2 + \cdots + c_kT_k)(\vec{v}) = c_1T_1(\vec{v}) + c_2T_2(\vec{v}) + \cdots + c_kT_k(\vec{v}).$$

Analogously, if $A_1, A_2, \ldots, A_k$ are all $m \times n$ matrices, then we can also construct the ***linear combination*** of these matrices with the corresponding coefficients by:

$$c_1A_1 + c_2A_2 + \cdots + c_kA_k.$$

From our Theorem above concerning the matrices for $T_1 + T_2$ and kT, we naturally see by Induction (as you will prove in the Exercises) that:

$$[c_1T_1 + c_2T_2 + \cdots + c_kT_k] = c_1[T_1] + c_2[T_2] + \cdots + c_k[T_k].$$

Example: Let $T_1, T_2, T_3 : \mathbb{R}^4 \to \mathbb{R}^3$, given by:

$$\begin{aligned} T_1(\langle x_1, x_2, x_3, x_4\rangle) &= \langle 5x_3 - x_1, 4x_2 - 7x_3 + 2x_4, 8x_3 - 5x_1\rangle, \\ T_2(\langle x_1, x_2, x_3, x_4\rangle) &= \langle 7x_2 + x_4, x_1 + 5x_2 - 3x_4, x_2 - x_3 - 7x_4\rangle, \quad \text{and} \\ T_3(\langle x_1, x_2, x_3, x_4\rangle) &= \langle x_2 + x_3 - 5x_4, x_1 - 9x_4, 2x_3 + 3x_4\rangle. \end{aligned}$$

Let us find the matrix of $4T_1 - 5T_2 + 7T_3$ in two ways: by using the definition of the linear combination of linear transformations, and by computing the corresponding linear combination of the matrices of the three linear transformations.

First, we have:

$$\begin{aligned} &(4T_1 - 5T_2 + 7T_3)(\langle x_1, x_2, x_3, x_4\rangle) \\ &= 4\langle 5x_3 - x_1, 4x_2 - 7x_3 + 2x_4, 8x_3 - 5x_1\rangle - 5\langle 7x_2 + x_4, x_1 + 5x_2 - 3x_4, x_2 - x_3 - 7x_4\rangle \\ &\quad + 7\langle x_2 + x_3 - 5x_4, x_1 - 9x_4, 2x_3 + 3x_4\rangle \\ &= \langle 20x_3 - 4x_1, 16x_2 - 28x_3 + 8x_4, 32x_3 - 20x_1\rangle \\ &\quad + \langle -35x_2 - 5x_4, -5x_1 - 25x_2 + 15x_4, -5x_2 + 5x_3 + 35x_4\rangle \\ &\quad + \langle 7x_2 + 7x_3 - 35x_4, 7x_1 - 63x_4, 14x_3 + 21x_4\rangle \\ &= \langle -4x_1 - 28x_2 + 27x_3 - 40x_4, 2x_1 - 9x_2 - 28x_3 - 40x_4, -20x_1 - 5x_2 + 51x_3 + 56x_4\rangle. \end{aligned}$$

Thus:

$$[4T_1 - 5T_2 + 7T_3] = \begin{bmatrix} -4 & -28 & 27 & -40 \\ 2 & -9 & -28 & -40 \\ -20 & -5 & 51 & 56 \end{bmatrix}.$$

Next, we find the linear combination of the matrices:

$$4[T_1] - 5[T_2] + 7[T_3]$$

$$= 4\begin{bmatrix} -1 & 0 & 5 & 0 \\ 0 & 4 & -7 & 2 \\ -5 & 0 & 8 & 0 \end{bmatrix} - 5\begin{bmatrix} 0 & 7 & 0 & 1 \\ 1 & 5 & 0 & -3 \\ 0 & 1 & -1 & -7 \end{bmatrix} + 7\begin{bmatrix} 0 & 1 & 1 & -5 \\ 1 & 0 & 0 & -9 \\ 0 & 0 & 2 & 3 \end{bmatrix}$$

$$= \begin{bmatrix} -4 & 0 & 20 & 0 \\ 0 & 16 & -28 & 8 \\ -20 & 0 & 32 & 0 \end{bmatrix} + \begin{bmatrix} 0 & -35 & 0 & -5 \\ -5 & -25 & 0 & 15 \\ 0 & -5 & 5 & 35 \end{bmatrix} + \begin{bmatrix} 0 & 7 & 7 & -35 \\ 7 & 0 & 0 & -63 \\ 0 & 0 & 14 & 21 \end{bmatrix}$$

$$= \begin{bmatrix} -4 & -28 & 27 & -40 \\ 2 & -9 & -28 & -40 \\ -20 & -5 & 51 & 56 \end{bmatrix}.$$

This verifies our previous computation.□

2.3 Section Summary

If $T_1,\ T_2 : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations, and k is any scalar, then we can define $T_1 + T_2$, $T_1 - T_2$, and $kT_1 : \mathbb{R}^n \to \mathbb{R}^m$, as linear transformations, by:

$$\begin{aligned} (T_1 + T_2)(\vec{v}) &= T_1(\vec{v}) + T_2(\vec{v}), \\ (T_1 - T_2)(\vec{v}) &= T_1(\vec{v}) - T_2(\vec{v}), \text{ and} \\ (kT_1)(\vec{v}) &= kT_1(\vec{v}). \end{aligned}$$

If A and B are both $m \times n$ matrices, and k is any scalar, then we can define: $A + B$, $A - B$, and kA as $m \times n$ matrices with the entry in row i, column j, given by:

$$\begin{aligned} (A + B)_{i,j} &= (A)_{i,j} + (B)_{i,j}, \\ (A - B)_{i,j} &= (A)_{i,j} - (B)_{i,j}, \text{ and} \\ (kA)_{i,j} &= k(A)_{i,j}. \end{aligned}$$

If $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations, with matrices $[T_1]$ and $[T_2]$ respectively, and k is any scalar, then:

$$\begin{aligned} [T_1 + T_2] &= [T_1] + [T_2], \\ [T_1 - T_2] &= [T_1] - [T_2], \text{ and} \\ [kT_1] &= k[T_1]. \end{aligned}$$

If $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ are linear transformations, then we can define their ***composition:***

$$T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^m,$$

as a linear transformation, with action given as follows: Suppose $\vec{u} \in \mathbb{R}^n$, $T_1(\vec{u}) = \vec{v} \in \mathbb{R}^k$, and $T_2(\vec{v}) = \vec{w} \in \mathbb{R}^m$. Then:

$$(T_2 \circ T_1)(\vec{u}) = T_2(T_1(\vec{u})) = T_2(\vec{v}) = \vec{w}.$$

If A is an $m \times k$ matrix, and B is a $k \times n$, then we can construct the $m \times n$ ***matrix product AB***, where column i of AB is $A \cdot$ (column i of B). Thus, if we write B in terms of its columns as $B = \left[\, \vec{b}_1 \,|\, \vec{b}_2 \,|\, \cdots \,|\, \vec{b}_n \,\right]$, then $AB = \left[\, A\vec{b}_1 \,|\, A\vec{b}_2 \,|\, \cdots \,|\, A\vec{b}_n \,\right]$.

We can also view the entries of AB as ***dot products***, via:

$$(AB)_{i,j} = \textbf{\textit{row i}} \text{ of } A \circ \textbf{\textit{column j}} \text{ of } B.$$

If $T_1, T_2, \ldots, T_k$ are all linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, and $c_1, c_2, \ldots, c_k$ are any scalars, then we can construct the ***linear combination*** of these linear transformations with the corresponding coefficients by:

$$(c_1T_1 + c_2T_2 + \cdots + c_kT_k)(\vec{v}) = c_1T_1(\vec{v}) + c_2T_2(\vec{v}) + \cdots + c_kT_k(\vec{v}).$$

Analogously, if $A_1, A_2, \ldots, A_k$ are all $m \times n$ matrices, then we can also construct the ***linear combination*** of these matrices with the corresponding coefficients by:

$$c_1A_1 + c_2A_2 + \cdots + c_kA_k, \text{ and thus:}$$

$$[c_1T_1 + c_2T_2 + \cdots + c_kT_k] = c_1[T_1] + c_2[T_2] + \cdots + c_k[T_k].$$

2.3 Exercises

1. Let $T_1, T_2 : \mathbb{R}^3 \to \mathbb{R}^2$ be given by:

$$T_1(\langle x, y, z \rangle) = \langle 3x - 2y + 5z, x + 4y - 7z \rangle, \text{ and}$$

$$T_2(\langle x, y, z \rangle) = \langle 2x + 9z, x - y + 3z \rangle.$$

   a. Use the definition of the sum of two linear transformations directly to compute $(T_1 + T_2)(\langle x, y, z \rangle)$.
   b. Use your answer in (a) to find the matrix of $T_1 + T_2$.
   c. Find $[T_1]$ and $[T_2]$, and use these to compute $[T_1] + [T_2]$.
   d. Verify that $[T_1] + [T_2] = [T_1 + T_2]$.
   e. Similarly, use the definition to find $(-4T_1)(\langle x, y, z \rangle)$ directly, and find $[-4T_1]$. Show that $[-4T_1] = -4[T_1]$.

2. Let $T_1, T_2 : \mathbb{R}^3 \to \mathbb{R}^4$ be given by:

$$T_1(\langle x, y, z \rangle) = \langle x - 2y + 3z, x - 4z, 2y, x - y + z \rangle, \text{ and}$$

$$T_2(\langle x, y, z \rangle) = \langle 2x + z, x - y, x + 3z, -4x \rangle.$$

   a. Use the definition of the sum of two linear transformations directly to compute $(T_1 + T_2)(\langle x, y, z \rangle)$.
   b. Use your answer in (a) to find the matrix of $T_1 + T_2$.
   c. Find $[T_1]$ and $[T_2]$, and use these to compute $[T_1] + [T_2]$.
   d. Verify that $[T_1] + [T_2] = [T_1 + T_2]$.
   e. Similarly, use the definition to find $(3T_1)(\langle x, y, z \rangle)$ directly, and find $[3T_1]$. Show that $[3T_1] = 3[T_1]$.

3. Consider the following matrices:

$$A = \begin{bmatrix} 3 & -7 & 4 \\ 2 & 8 & -3 \end{bmatrix}, \; B = \begin{bmatrix} -4 & 2 \\ 7 & 3 \\ -2 & -5 \end{bmatrix}, \; C = \begin{bmatrix} 7 & -4 \\ -2 & 3 \end{bmatrix},$$

$$D = \begin{bmatrix} -5 & 3 & -7 \\ 4 & -1 & -2 \end{bmatrix}, \text{ and } E = \begin{bmatrix} -1 & 3 \\ 2 & -4 \\ 5 & 1 \end{bmatrix}.$$

Compute, if possible, the following matrices, and state their dimensions:

a. $A + D$ b. $A - B$ c. $2B - 5E$ d. $7C + 4A$ e. BC

f. CB g. CC h. BE i. EA j. DB

k. $D(BC)$ l. $(DB)C$ m. $(EC)A$ n. $E(CA)$ o. $C(DB)$

You may use the answers to (e) and (j) for parts (k), (l) and (o). Compare the answers to parts (k) and (l) together, and parts (m) and (n) together. What property does this remind you of?

4. Consider the following matrices:

$$A = \begin{bmatrix} 5 & -3 & 2 & 4 \\ 7 & 1 & -6 & 8 \\ -4 & -5 & 3 & 1 \end{bmatrix}, \; B = \begin{bmatrix} -3 & 7 & 2 & 1 & -5 \\ 0 & 4 & 3 & -6 & 2 \\ 6 & -1 & 0 & -8 & 1 \end{bmatrix}, \; C = \begin{bmatrix} 2 & 6 & -1 \\ 3 & 0 & 4 \\ -2 & 5 & 7 \\ 1 & -8 & -3 \end{bmatrix},$$

$$D = \begin{bmatrix} 6 & 0 & -3 & 1 & 7 \\ 3 & 5 & 2 & -6 & -3 \\ -8 & 4 & -7 & 3 & 1 \\ 0 & -2 & 5 & -4 & 6 \end{bmatrix}, \; E = \begin{bmatrix} 4 & -6 & 7 \\ 3 & 0 & -2 \\ 5 & 1 & -8 \\ -3 & 4 & 6 \\ 1 & -3 & 0 \end{bmatrix}, \; F = \begin{bmatrix} 7 & 1 & -9 & 6 \\ -4 & 0 & 4 & -3 \\ 3 & -2 & -7 & 0 \\ 0 & 5 & 2 & 1 \\ 8 & 6 & -4 & -1 \end{bmatrix}.$$

Compute, if possible, the following matrix products, and state their dimensions:

a. AC b. CA c. AD d. DA e. DE

f. EA g. FC h. FA i. DF j. FB

k. $(CA)D$ l. $C(AD)$ m. $(DF)C$ n. $D(FC)$ o. $(AC)(AC)$

You may use previous computations to perform (k) through (o).

5. Let $T_1 : \mathbb{R}^2 \to \mathbb{R}^4$, and $T_2 : \mathbb{R}^4 \to \mathbb{R}^3$ be given by:

$$T_1(\langle x, y \rangle) = \langle 3x - 2y, 5x + y, -x + 3y, 4y \rangle, \quad \text{and}$$

$$T_2(\langle x_1, x_2, x_3, x_4 \rangle) = \langle 3x_1 - 5x_4, 7x_2 + 2x_3 - x_4, 6x_3 + 9x_4 \rangle.$$

a. Explain why the composition $T_2 \circ T_1$ is well defined. State the domain and the codomain of $T_2 \circ T_1$.

b. Is the composition $T_1 \circ T_2$ also well defined? Why or why not?
c. Use the definition of a composition directly to find $(T_2 \circ T_1)(\langle x, y\rangle)$.
d. Use (c) to find $[T_2 \circ T_1]$.
e. Form $[T_2]$ and $[T_1]$ and compute the matrix product $[T_2][T_1]$. Verify that it equals $[T_2 \circ T_1]$.

6. Let $T_1 : \mathbb{R}^3 \to \mathbb{R}^4$, and $T_2 : \mathbb{R}^4 \to \mathbb{R}^3$ be given by:

$$T_1(\langle x, y, z\rangle) = \langle 3x + 5y - z,\ 2x - y + 4z,\ x + z,\ y - 2z\rangle, \quad \text{and}$$
$$T_2(\langle x_1, x_2, x_3, x_4\rangle) = \langle 3x_1 - 5x_4,\ 7x_2 + 2x_3 - x_4,\ 6x_3 + 9x_4\rangle.$$

Note that we are using the same T_2 as the previous Exercise.

a. Explain why ***both*** compositions $T_2 \circ T_1$ and $T_1 \circ T_2$ are well defined. State the domain and the codomain of each.
b. Use directly the definition of a composition to find $(T_2 \circ T_1)(\langle x, y, z\rangle)$ and $(T_1 \circ T_2)(\langle x_1, x_2, x_3, x_4\rangle)$.
c. Use (b) to find $[T_2 \circ T_1]$ and $[T_1 \circ T_2]$.
d. Form $[T_2]$ and $[T_1]$ and compute the matrix products $[T_2][T_1]$ and $[T_1][T_2]$. Verify that they equal $[T_2 \circ T_1]$ and $[T_1 \circ T_2]$ respectively.

7. Let $T_1 : \mathbb{R}^2 \to \mathbb{R}^5$, and $T_2 : \mathbb{R}^5 \to \mathbb{R}^2$ be given by:

$$T_1(\langle x, y\rangle) = \langle x + 2y,\ 3x - y,\ 5x + 7y,\ 6y, -2x\rangle, \quad \text{and}$$
$$T_2(\langle x_1, x_2, x_3, x_4, x_5\rangle) = \langle 3x_1 + 7x_2 - 6x_3 + 5x_4 - 8x_5,\ 9x_1 + 2x_3 - x_4 + x_5\rangle.$$

a. Explain why ***both*** compositions $T_2 \circ T_1$ and $T_1 \circ T_2$ are well defined. State the domain and the codomain of each.
b. Use directly the definition of a composition to find $(T_2 \circ T_1)(\langle x, y\rangle)$ and $(T_1 \circ T_2)(\langle x_1, x_2, x_3, x_4, x_5\rangle)$.
c. Use (b) to find $[T_2 \circ T_1]$ and $[T_1 \circ T_2]$.
d. Form $[T_2]$ and $[T_1]$ and compute the matrix products $[T_2][T_1]$ and $[T_1][T_2]$. Verify that they equal $[T_2 \circ T_1]$ and $[T_1 \circ T_2]$ respectively.

8. Let $T_1 : \mathbb{R}^n \to \mathbb{R}^m$ and $T_2 : \mathbb{R}^n \to \mathbb{R}^m$ be linear transformations, with $m \times n$ matrices $[T_1]$ and $[T_2]$ respectively, $k \in \mathbb{R}$, and $\vec{v} \in \mathbb{R}^n$. Prove that:

$$(T_1 - T_2)(\vec{v}) = ([T_1] - [T_2])\vec{v}, \quad \text{and}$$
$$(kT_1)(\vec{v}) = (k[T_1])\vec{v}.$$

Consequently, prove that:

$$[T_1 - T_2] = [T_1] - [T_2], \quad \text{and}$$
$$[kT_1] = k[T_1].$$

9. Let $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ be linear transformations. Prove that for all $k \in \mathbb{R}$, and $\vec{v} \in \mathbb{R}^n$: $(T_2 \circ T_1)(k\vec{v}) = k(T_2 \circ T_1)(\vec{v})$, that is, $T_2 \circ T_1$ satisfies the Homogeneity Property.

10. Use induction on k to prove that: if $c_1, c_2, \ldots, c_k$ are scalars and $T_1, T_2, \ldots, T_k$ are linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^m$, then the linear combination:

$$c_1T_1 + c_2T_2 + \cdots + c_kT_k$$

is also a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. Start with $k = 2$ as your "basis step." At the same time, prove that:

$$[c_1T_1 + c_2T_2 + \cdots + c_kT_k] = c_1[T_1] + c_2[T_2] + \cdots + c_k[T_k].$$

11. Show that if Π is a plane in $\mathbb{R}^3$ through $\vec{\mathbf{0}}_3$ with normal line L, then the reflection operator across Π can be written in terms of the projection onto L as the linear combination:

$$refl_\Pi = \boldsymbol{I}_{\mathbb{R}^3} - 2proj_L.$$

Hint: $\boldsymbol{I}_{\mathbb{R}^3}(\vec{v}) = \vec{v}$ for all $\vec{v} \in \mathbb{R}^3$.

12. ***Multiplication by Identity and Zero Matrices:*** Show that if A is an $k \times n$ matrix and B is an $n \times k$ matrix, then $A\boldsymbol{I}_n = A$ and $\boldsymbol{I}_nB = B$. State and prove corresponding properties for $A\mathbf{0}_{n\times r}$ and $\mathbf{0}_{r\times n}B$.

13. What can you say about the dimensions of A and B if ***both*** products AB and BA are defined, although not necessarily of the same size? Analogously, let $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ be linear transformations. What can you say about n, k and m if both compositions $T_2 \circ T_1$ and $T_1 \circ T_2$ are well defined?

14. Prove that if A is an $m \times k$ matrix and B is a $k \times n$ matrix, then we can define the matrix product AB by partitioning A into ***rows*** as follows:

$$\text{If } A = \begin{bmatrix} \vec{r}_1 \\ \vec{r}_2 \\ \vdots \\ \vec{r}_m \end{bmatrix} \text{ then } AB = \begin{bmatrix} \vec{r}_1B \\ \vec{r}_2B \\ \vdots \\ \vec{r}_mB \end{bmatrix}.$$

Hint: use the dot product formula for a matrix product. Compare this to the definition of AB using the ***columns*** of B.

15. In Chapter 1, we showed that if A is an $m \times n$ matrix and $\vec{b}$ is an $m \times 1$ matrix, then the matrix equation $A\vec{x} = \vec{b}$ is solvable for $\vec{x}$ ***if and only if*** $\vec{b}$ is in the ***columnspace*** of A. Analogously, prove that if $\vec{d}$ is a $1 \times n$ matrix, then the matrix equation $\vec{y}A = \vec{d}$ is solvable for $\vec{y}$ ***if and only if*** $\vec{d}$ is in the ***rowspace*** of A.

16. Suppose that Π is a plane in $\mathbb{R}^3$ passing through the origin with Cartesian equation $ax + by + cz = 0$, where $\vec{n} = \langle a, b, c\rangle$ is a ***unit*** vector. Show that we can write $[proj_\Pi]$ as the matrix product:

$$[proj_\Pi] = \begin{bmatrix} a \\ b \\ c \end{bmatrix}\begin{bmatrix} a & b & c \end{bmatrix}.$$

Hint: review the general formula for $[proj_\Pi]$ at the end of the Exercises in Section 2.2.

2.4 Properties of Operations on Linear Transformations and Matrices

In this Section, we will see that linear transformations and their arithmetic, as well as the arithmetic of matrix operations, enjoy some analogous properties that the arithmetic of vectors and real numbers possesses. However, we will also point out that some properties are not always possessed by these matrix operations.

Properties of Matrix Addition and Scalar Multiplication

Many of the properties of vector arithmetic are inherited by matrix addition and scalar multiplication:

Theorem — Properties of Matrix Addition and Scalar Multiplication:
If A, B and C are $m \times n$ matrices, and r and s are scalars, then the following properties hold:

1. ***The Commutative Property of Addition*** $\quad A + B = B + A$
2. ***The Associative Property of Addition*** $\quad A + (B + C) = (A + B) + C$
3. ***The "Left" Distributive Property*** $\quad (r + s)A = rA + sA$
4. ***The "Right" Distributive Property*** $\quad r(A + B) = rA + rB$
5. ***The Associative Property of Scalar Multiplication*** $\quad r(sA) = (rs)A = s(rA)$

Proof: We will show that the first property is true, and leave the rest as Exercises. First of all, A and B are both $m \times n$ matrices, so $A + B$ and $B + A$ are both $m \times n$ matrices. Let us use the notation $(A)_{i,j}$ instead of $a_{i,j}$, and similarly for $(A + B)_{i,j}$ and so on, to look at each entry:

$$\begin{aligned} (A + B)_{i,j} &= (A)_{i,j} + (B)_{i,j} \\ &= (B)_{i,j} + (A)_{i,j} \quad \text{(since both are numbers)} \\ &= (B + A)_{i,j}. \end{aligned}$$

Since all corresponding pairs of entries are equal, $A + B = B + A$. ■

The rest of the properties are proved similarly by examining the sizes and the entries of the matrices and their combinations on both sides of each equation.

Basic Properties of Matrix Multiplication

Unlike matrix addition and scalar multiplication, matrix multiplication has fewer nice properties, even when combined with the relatively simple operations of matrix addition and scalar multiplication:

Theorem: If A and B are $m \times k$ matrices, C and D are $k \times n$ matrices, and r is a scalar, then the following properties hold:

1. The "Left" Distributive Property $\quad (A+B)C = AC + BC$

2. The "Right" Distributive Property $\quad A(C+D) = AC + AD$

3. The Associative Property of Mixed (Scalar and Matrix) Products $\quad r(BC) = (rB)C = B(rC)$

Proof: Again, we will prove the first property and leave the other two as Exercises. We will prove it using the dot product formula for matrix products. First, since A and B are on the left side of the product, we partition them into their ***rows:***

$$A = \begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_m \end{bmatrix} \text{ and } B = \begin{bmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \vdots \\ \vec{b}_m \end{bmatrix}.$$

We saw in the previous Section that with this partitioning:

$$A + B = \begin{bmatrix} \vec{a}_1 \\ \vec{a}_2 \\ \vdots \\ \vec{a}_m \end{bmatrix} + \begin{bmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \vdots \\ \vec{b}_m \end{bmatrix} = \begin{bmatrix} \vec{a}_1 + \vec{b}_1 \\ \vec{a}_2 + \vec{b}_2 \\ \vdots \\ \vec{a}_m + \vec{b}_m \end{bmatrix}.$$

Next, since C is on the right side of the product, we partition C into ***columns:***

$$C = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix}.$$

We now have:

$$(A+B)C = \begin{bmatrix} \vec{a}_1 + \vec{b}_1 \\ \vec{a}_2 + \vec{b}_2 \\ \vdots \\ \vec{a}_m + \vec{b}_m \end{bmatrix} \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix}$$

$$= \begin{bmatrix} \left(\vec{a}_1 + \vec{b}_1\right) \circ \vec{c}_1 & \left(\vec{a}_1 + \vec{b}_1\right) \circ \vec{c}_2 & \cdots & \left(\vec{a}_1 + \vec{b}_1\right) \circ \vec{c}_n \\ \left(\vec{a}_2 + \vec{b}_2\right) \circ \vec{c}_1 & \left(\vec{a}_2 + \vec{b}_2\right) \circ \vec{c}_2 & \cdots & \left(\vec{a}_2 + \vec{b}_2\right) \circ \vec{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \left(\vec{a}_m + \vec{b}_m\right) \circ \vec{c}_1 & \left(\vec{a}_m + \vec{b}_m\right) \circ \vec{c}_2 & \cdots & \left(\vec{a}_m + \vec{b}_m\right) \circ \vec{c}_n \end{bmatrix}$$

(by the dot product formula for the matrix product)

$$= \begin{bmatrix} \vec{a}_1 \circ \vec{c}_1 + \vec{b}_1 \circ \vec{c}_1 & \vec{a}_1 \circ \vec{c}_2 + \vec{b}_1 \circ \vec{c}_2 & \cdots & \vec{a}_1 \circ \vec{c}_n + \vec{b}_1 \circ \vec{c}_n \\ \vec{a}_2 \circ \vec{c}_1 + \vec{b}_2 \circ \vec{c}_1 & \vec{a}_2 \circ \vec{c}_2 + \vec{b}_2 \circ \vec{c}_2 & \cdots & \vec{a}_2 \circ \vec{c}_n + \vec{b}_2 \circ \vec{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_m \circ \vec{c}_1 + \vec{b}_m \circ \vec{c}_1 & \vec{a}_m \circ \vec{c}_2 + \vec{b}_m \circ \vec{c}_2 & \cdots & \vec{a}_m \circ \vec{c}_n + \vec{b}_m \circ \vec{c}_n \end{bmatrix}$$

(by the Distributive Property of Dot Products)

$$= \begin{bmatrix} \vec{a}_1 \circ \vec{c}_1 & \vec{a}_1 \circ \vec{c}_2 & \cdots & \vec{a}_1 \circ \vec{c}_n \\ \vec{a}_2 \circ \vec{c}_1 & \vec{a}_2 \circ \vec{c}_2 & \cdots & \vec{a}_2 \circ \vec{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{a}_m \circ \vec{c}_1 & \vec{a}_m \circ \vec{c}_2 & \cdots & \vec{a}_m \circ \vec{c}_n \end{bmatrix} + \begin{bmatrix} \vec{b}_1 \circ \vec{c}_1 & \vec{b}_1 \circ \vec{c}_2 & \cdots & \vec{b}_1 \circ \vec{c}_n \\ \vec{b}_2 \circ \vec{c}_1 & \vec{b}_2 \circ \vec{c}_2 & \cdots & \vec{b}_2 \circ \vec{c}_n \\ \vdots & \vdots & \ddots & \vdots \\ \vec{b}_m \circ \vec{c}_1 & \vec{b}_m \circ \vec{c}_2 & \cdots & \vec{b}_m \circ \vec{c}_n \end{bmatrix}$$

(by the definition of matrix addition)

$= AC + BC$ (again, by the dot product formula for the matrix product).■

The Associative Property of Matrix Multiplication

The most difficult property to prove regarding matrix multiplication is its ***associative*** nature, and thus we now focus on it:

Theorem: If A is an $m \times p$ matrix, B is a $p \times q$ matrix, and C is a $q \times n$ matrix, then:

$$\boldsymbol{A(BC) = (AB)C}.$$

Proof: First, let us check that both sides of the equation are well defined, and the resulting matrices are of the ***same size***. Since B is a $p \times q$ matrix, and C is a $q \times n$ matrix, the product BC is a $p \times n$ matrix. Since A is an $m \times p$ matrix, the product $A(BC)$ is an $m \times n$ matrix. Similarly, AB is $m \times q$, and $(AB)C$ is $m \times n$. Thus, both sides are $m \times n$ matrices.

Now, we have to show that both sides, pair-wise, have exactly the ***same entries***. First, let us assume that $C = \vec{x}$, a $q \times 1$ matrix. Let us write the middle matrix B in terms of its ***column*** vectors:

$$B = \left[\ \vec{b}_1\ \vec{b}_2\ \ldots\ \vec{b}_q\ \right].$$

From the definition of the matrix product:

$$AB = \left[\ A\vec{b}_1\ A\vec{b}_2\ \ldots\ A\vec{b}_q\ \right].$$

Thus, when we multiply on the right by $\vec{x}$, we get:

$$(AB)\vec{x} = \left[\ A\vec{b}_1\ A\vec{b}_2\ \ldots\ A\vec{b}_q\ \right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{bmatrix} = x_1\left(A\vec{b}_1\right) + x_2\left(A\vec{b}_2\right) + \cdots + x_q\left(A\vec{b}_q\right)$$

by the basic definition of the product of a matrix with a column vector. Now, let us work on $A(B\vec{x})$.

First, we find $B\vec{x}$:

$$B\vec{x} = \left[\, \vec{b}_1\ \vec{b}_2\ \dots\ \vec{b}_q \,\right] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{bmatrix} = x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_q\vec{b}_q.$$

Thus:

$$\begin{aligned} A(B\vec{x}) &= A\left(x_1\vec{b}_1 + x_2\vec{b}_2 + \cdots + x_q\vec{b}_q\right) \\ &= A\left(x_1\vec{b}_1\right) + A\left(x_2\vec{b}_2\right) + \cdots + A\left(x_q\vec{b}_q\right) \\ &\quad \text{(by the “Right” Distributive Property)} \\ &= x_1\left(A\vec{b}_1\right) + x_2\left(A\vec{b}_2\right) + \cdots + x_q\left(A\vec{b}_q\right) \end{aligned}$$

by the Associative Property of Mixed Products. Thus the two sides are the same.

Now, if C is an arbitrary $q \times n$ matrix, then we can write C in terms of its ***columns:***

$$C = \left[\, \vec{c}_1\ \ \vec{c}_2\ \dots\ \vec{c}_n \,\right],$$

and from the previous analysis:

$$(AB)\vec{c}_i = A(B\vec{c}_i)$$

for every column $\vec{c}_i$. Thus, column i of $(AB)C$ is exactly the same as that of $A(BC)$, and therefore $(AB)C = A(BC)$. ■

Example: Consider the matrices:

$$A = \begin{bmatrix} 5 & -3 \\ 4 & 0 \\ 2 & 6 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & -2 & 3 & 1 \\ -4 & 8 & 0 & -5 \end{bmatrix}, \quad \text{and} \quad C = \begin{bmatrix} 3 & -7 \\ 0 & 5 \\ 4 & -2 \\ 9 & 6 \end{bmatrix}.$$

Note that A is 3×2, B is 2×4, and C is 4×2. Thus, AB is 3×4 and BC is 2×2. We compute these matrix products below:

$$AB = \begin{bmatrix} 5 & -3 \\ 4 & 0 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 7 & -2 & 3 & 1 \\ -4 & 8 & 0 & -5 \end{bmatrix} = \begin{bmatrix} 47 & -34 & 15 & 20 \\ 28 & -8 & 12 & 4 \\ -10 & 44 & 6 & -28 \end{bmatrix}, \quad \text{and}$$

$$BC = \begin{bmatrix} 7 & -2 & 3 & 1 \\ -4 & 8 & 0 & -5 \end{bmatrix} \begin{bmatrix} 3 & -7 \\ 0 & 5 \\ 4 & -2 \\ 9 & 6 \end{bmatrix} = \begin{bmatrix} 42 & -59 \\ -57 & 38 \end{bmatrix}.$$

From these, we see that the product $(AB)C$ is a 3×2 matrix, and $A(BC)$ is also a 3×2 matrix. By the Associative Property, these two products should be the same, and in fact we can verify this by directly computing these products and checking that they are equal:

$$(AB)C = \begin{bmatrix} 47 & -34 & 15 & 20 \\ 28 & -8 & 12 & 4 \\ -10 & 44 & 6 & -28 \end{bmatrix} \begin{bmatrix} 3 & -7 \\ 0 & 5 \\ 4 & -2 \\ 9 & 6 \end{bmatrix} = \begin{bmatrix} 381 & -409 \\ 168 & -236 \\ -258 & 110 \end{bmatrix}, \quad \text{and}$$

$$A(BC) = \begin{bmatrix} 5 & -3 \\ 4 & 0 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 42 & -59 \\ -57 & 38 \end{bmatrix} = \begin{bmatrix} 381 & -409 \\ 168 & -236 \\ -258 & 110 \end{bmatrix}. \square$$

We note that it is almost a mathematical ***miracle*** that matrix multiplication is associative, considering that this operation is defined in a rather strange way. This just goes to show that creative ideas often lead to interesting consequences.

The Matrix of a Composition

Now we are ready to prove that in general:

> ***Theorem:*** If $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ are linear transformations, then:
> $$[T_2 \circ T_1] = [T_2][T_1].$$

Proof: For convenience, as before, let $[T_1] = A$ and $[T_2] = B$. We must show that $[T_2 \circ T_1] = BA$. By the uniqueness property of the standard matrix, we must show that for any $\vec{v} \in \mathbb{R}^n$:

$$(T_2 \circ T_1)(\vec{v}) = (BA)\vec{v}.$$

Now, by definition:

$$(T_2 \circ T_1)(\vec{v}) = T_2(T_1(\vec{v})) = T_2(A\vec{v}) = B(A\vec{v}) = (BA)\vec{v},$$

where the last equation follows from the Associative Property of Matrix Multiplication.■

k-fold Compositions

If $T_1, T_2, \ldots, T_{k-1}, T_k$ are all linear transformations with the property that ***the codomain of*** T_i ***is the domain of*** T_{i+1}, for all $i = 1..k-1$, then we can inductively construct the ***k-fold composition*** of these linear transformations by:

$$(T_k \circ T_{k-1} \circ \cdots \circ T_2 \circ T_1)(\vec{v}) = T_k((T_{k-1} \circ \cdots \circ T_2 \circ T_1)(\vec{v})).$$

The above formula says that we first compute the $(k-1)$-fold composition $T_{k-1} \circ \cdots \circ T_2 \circ T_1$, evaluated at the vector $\vec{v}$, and use T_k to evaluate the final vector.

The matrix of the k-fold composition is the product of the matrices in the composition, and conveniently, they are in the same order:

$$[T_k \circ T_{k-1} \circ \cdots \circ T_2 \circ T_1] = [T_k][T_{k-1}]\cdots[T_2][T_1].$$

Since matrix multiplication is ***associative***, it does not matter how the long product on the right is computed, as long as the intermediate products are kept in the ***same order***.

Example: Let $T_1, T_2, T_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be given by:

$$\begin{aligned} T_1 &= \text{rotation counterclockwise by } \pi/3, \\ T_2 &= \text{projection onto the } y\text{-axis, and} \\ T_3 &= \text{reflection across the } x\text{-axis.} \end{aligned}$$

Let us find the matrix of $T_3 \circ T_2 \circ T_1$. First, the individual matrices are:

$$[T_1] = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}; \quad [T_2] = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}; \quad [T_3] = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

By the Associative Property of Matrix Multiplication, it doesn't matter whether we compute $([T_3][T_2])[T_1]$ or $[T_3]([T_2][T_1])$. Just for fun, let's do both:

$$\begin{aligned} ([T_3][T_2])[T_1] &= \left(\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right) \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -\sqrt{3}/2 & -1/2 \end{bmatrix}, \text{ and similarly:} \\ [T_3]([T_2][T_1]) &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \right) \\ &= \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -\sqrt{3}/2 & -1/2 \end{bmatrix}. \end{aligned}$$

As expected, we get the same answer.□

Powers of Square Matrices and Linear Operators

Now that we understand better the operations of matrix multiplication and composition of transformations, let us see when we can multiply a matrix A by ***itself***. The compatibility requirements of matrix multiplication and composition of linear transformations easily tell us the following:

> ***Theorem:*** The matrix product AA can be formed ***if and only if*** A is an $n \times n$ or ***square*** matrix. Analogously, the composition $T \circ T$ can be formed ***if and only if*** the domain and codomain of T are the ***same*** Euclidean space $\mathbb{R}^n$, i.e., T is an ***operator***.

It is more natural to write the product AA as A^2 and the composition $T \circ T$ as T^2.

Furthermore, now that we know that matrix multiplication is ***associative***, we have:

$$A(AA) = (AA)A$$

so the expression A^3 is well defined. Similarly, by induction, we will write:

$$A^k = A \bullet A^{k-1} = A \bullet A \bullet \cdots \bullet A, \text{ and}$$

$$T^k(\vec{v}) = T(T^{k-1}(\vec{v})) = T(T(\ldots T(\vec{v})))$$

where there are k factors of A appearing in the product, and k evaluations of T appearing in the k-fold composition.

Example: If $A = \begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix}$, then:

$$A^2 = \begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix} = \begin{bmatrix} -21 & -25 \\ 30 & -26 \end{bmatrix}, \quad \text{and}$$

$$A^3 = A \bullet A^2 = \begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} -21 & -25 \\ 30 & -26 \end{bmatrix} = \begin{bmatrix} -213 & 55 \\ -66 & -202 \end{bmatrix}. \square$$

We can combine the operations of computing the powers of a square matrix, scalar multiplication and matrix addition into a familiar expression:

Definition: If $p(x) = c_0 + c_1x + c_2x^2 + \cdots + c_kx^k$ is a polynomial with real coefficients, and A is any $n \times n$ matrix, then we define the ***polynomial evaluation***, $p(A)$, by:

$$p(A) = c_0\boldsymbol{I_n} + c_1A + c_2A^2 + \cdots + c_kA^k.$$

Warning: Do not forget to multiply c_0 by $\boldsymbol{I_n}$, otherwise you will have a hard time adding a scalar to a matrix. This is a reasonable convention, because in Algebra, $x^0 = 1$, as long as $x \neq 0$. It is therefore natural to ***define*** A^0 as $\boldsymbol{I_n}$, as long as A is not a zero matrix.

Example: Let A be the same matrix as the previous Example, and suppose $p(x) = 3 - 5x + 7x^2 - 4x^3$. Using the powers of A computed above, we get:

$$p(A) = 3\boldsymbol{I_2} - 5A + 7A^2 - 4A^3$$

$$= 3\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - 5\begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix} + 7\begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix}^2 - 4\begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix}^3$$

$$= \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} - 5\begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix} + 7\begin{bmatrix} -21 & -25 \\ 30 & -26 \end{bmatrix} - 4\begin{bmatrix} -213 & 55 \\ -66 & -202 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix} + \begin{bmatrix} -15 & 25 \\ -30 & -10 \end{bmatrix} + \begin{bmatrix} -147 & -175 \\ 210 & -182 \end{bmatrix} + \begin{bmatrix} 852 & -220 \\ 264 & 808 \end{bmatrix} = \begin{bmatrix} 693 & -370 \\ 444 & 619 \end{bmatrix}.$$

Multiplication by Identity and Zero Matrices

In Section 2.1, we saw the zero matrices $\mathbf{0}_{m\times n}$ and the $n \times n$ identity matrices $\boldsymbol{I_n}$. Multiplication by these special matrices are easily performed. The proof of the following properties were Exercises in Section 2.3:

> ***Theorem:*** If A is any $m \times n$ matrix, then:
>
> $$A\boldsymbol{I_n} = A, \;\; \boldsymbol{I_m}A = A, \;\; A\mathbf{0}_{n\times p} = \mathbf{0}_{m\times p}, \;\text{ and }\; \mathbf{0}_{p\times m}A = \mathbf{0}_{p\times n}.$$

T is Uniquely Determined by a Basis

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, we can assemble $[T]$ using $T(\vec{e}_1)$ through $T(\vec{e}_n)$. However, the standard basis is not that special. The next Theorem, whose proof we leave as an Exercise, says that if we know how T behaves on ***any*** basis for $\mathbb{R}^n$, we can still compute T for ***any*** $\vec{v} \in \mathbb{R}^n$:

> ***Theorem:*** If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is any ***basis*** for $\mathbb{R}^n$, then the action of T is uniquely determined by the vectors $T(\vec{v}_1)$, $T(\vec{v}_2)$, $\ldots$, $T(\vec{v}_n)$. More specifically, if $\vec{v} \in \mathbb{R}^n$ and $\vec{v}$ is expressed (uniquely) as:
>
> $$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n, \quad \text{then:} \quad T(\vec{v}) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n).$$

Example: Suppose $T : \mathbb{R}^3 \to \mathbb{R}^4$ is a linear transformation and we are given that:

$$T(\langle 1, 2, 1\rangle) = \langle 3, 5, 2, -4\rangle,$$

$$T(\langle -1, 1, 0\rangle) = \langle 7, 4, -3, 8\rangle, \;\text{ and}$$

$$T(\langle 1, 1, 1\rangle) = \langle 2, 1, 5, 6\rangle.$$

Let us use this information to compute $T(\langle 3, -5, 8\rangle)$. We are supposed to check first that the set $B = \{\langle 1, 2, 1\rangle, \langle -1, 1, 0\rangle, \langle 1, 1, 1\rangle\}$ is a ***basis*** for $\mathbb{R}^3$. However, let us see if we can kill two birds with one stone. If we try to determine if $\langle 3, -5, 8\rangle$ is a member of $Span(B)$ by assembling the vectors in an augmented matrix as usual:

$$\begin{bmatrix} 1 & -1 & 1 & | & 3 \\ 2 & 1 & 1 & | & -5 \\ 1 & 0 & 1 & | & 8 \end{bmatrix}, \quad \text{we get the rref} \quad \begin{bmatrix} 1 & 0 & 0 & | & -18 \\ 0 & 1 & 0 & | & 5 \\ 0 & 0 & 1 & | & 26 \end{bmatrix}.$$

We find 3 leading 1's in the rref, and this proves that B is ***linearly independent*** and therefore a basis, as we need 3 vectors to form a basis for $\mathbb{R}^3$. It also tells us that:

$$\langle 3, -5, 8\rangle = -18\langle 1, 2, 1\rangle + 5\langle -1, 1, 0\rangle + 26\langle 1, 1, 1\rangle.$$

Thus, by the linearity properties:

$$\begin{aligned} T(\langle 3,-5,8\rangle) &= T(-18\langle 1,2,1\rangle + 5\langle -1,1,0\rangle + 26\langle 1,1,1\rangle) \\ &= T(-18\langle 1,2,1\rangle) + T(5\langle -1,1,0\rangle) + T(26\langle 1,1,1\rangle) \\ &= -18\,T(\langle 1,2,1\rangle) + 5\,T(\langle -1,1,0\rangle) + 26\,T(\langle 1,1,1\rangle) \\ &= -18\langle 3,5,2,-4\rangle + 5\langle 7,4,-3,8\rangle + 26\langle 2,1,5,6\rangle \\ &= \langle 33,-44,79,268\rangle. \square \end{aligned}$$

The Existence of Zero Divisors

Because the set of $n \times n$ matrices have an addition and multiplication operation with nice properties, it possesses what is called a ***ring structure***. In such a structure, the product of the zero element with any other element is again the zero element. This is certainly true for matrices, as we stated above. However, the ***converse*** is not always true, and we refer to these exceptions with a special term:

> ***Definition:*** Suppose that A and B are two ***non-zero*** $n \times n$ matrices, with the property that $AB = \mathbf{0}_{n\times n}$. Then, A and B are both called ***zero divisors***.

Example: Let $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$, and $B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$. We can easily check that:

$$AB = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

and thus A and B are both zero divisors. Notice, however, that:

$$BA = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},$$

which is ***not*** the zero matrix! This does not disqualify A or B from being zero divisors, though. $\square$

Reversing the order of multiplication gave us a different answer. This is worth pointing out in general:

AB ≠ BA Most of the Time!

Despite all the wonderful, natural properties of matrix arithmetic, there is an important exception. It is analogous to the well-known property regarding the composition of functions that in general, $f \circ g \neq g \circ f$:

> ***Matrix multiplication, in general, is NOT commutative!***

Example: Let $A = \begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 7 \\ 9 & -4 \end{bmatrix}$. Then:

$$AB = \begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix}\begin{bmatrix} 1 & 7 \\ 9 & -4 \end{bmatrix} = \begin{bmatrix} -42 & 41 \\ 24 & 34 \end{bmatrix}, \text{ but:}$$

$$BA = \begin{bmatrix} 1 & 7 \\ 9 & -4 \end{bmatrix}\begin{bmatrix} 3 & -5 \\ 6 & 2 \end{bmatrix} = \begin{bmatrix} 45 & 9 \\ 3 & -53 \end{bmatrix}. \square$$

There are, of course, exceptions to this rule. In other words, there ***do*** exist pairs of matrices A and B where $AB = BA$. This will be explored further in the last Exercise of this Section.

2.4 Section Summary

Under compatible conditions (that is, when all operations in all the expression are well defined), matrix arithmetic enjoys the following properties:

$$\begin{array}{ll} A + B = B + A & (A+B)C = AC + BC \\ A + (B + C) = (A + B) + C & A(C + D) = AC + AD \\ (r+s)A = rA + sA & r(BC) = (rB)C = B(rC) \\ r(A+B) = rA + rB & A(BC) = (AB)C \\ r(sA) = (rs)A = s(rA) & \\ A\boldsymbol{I_n} = A & \boldsymbol{I_m}A = A \\ A\boldsymbol{0_{n\times p}} = \boldsymbol{0_{m\times p}} & \boldsymbol{0_{p\times m}}A = \boldsymbol{0_{p\times n}} \end{array}$$

However, matrix multiplication, in general, is ***not commutative***.

If $T_1 : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{k}}$ and $T_2 : \mathbb{R}^{\boldsymbol{k}} \to \mathbb{R}^{\boldsymbol{m}}$ are linear transformations, then: $[T_2 \circ T_1] = [T_2][T_1]$.

The matrix product AA can be formed ***if and only if*** A is an $n \times n$ matrix. Analogously, the composition $T \circ T$ can be formed ***if and only if*** the domain and codomain of T are the same Euclidean space $\mathbb{R}^{\boldsymbol{n}}$, i.e., T is an ***operator***. More generally, we can construct A^k for a positive integer k.

If $p(x) = c_0 + c_1x + c_2x^2 + \cdots + c_kx^k$ is a polynomial with real coefficients, and A is any $n \times n$ matrix, then we define the ***polynomial evaluation***, $p(A)$, by: $p(A) = c_0\boldsymbol{I_n} + c_1A + c_2A^2 + \cdots + c_kA^k$.

If $T : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{m}}$ is a linear transformation, and $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a ***basis*** for $\mathbb{R}^{\boldsymbol{n}}$, then the action of T is uniquely determined by the vectors $T(\vec{v}_1),\ T(\vec{v}_2),\ \ldots,\ T(\vec{v}_n)$.
More specifically, if $\vec{v} \in \mathbb{R}^{\boldsymbol{n}}$ and $\vec{v}$ is expressed (uniquely) as:

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n, \quad \text{then:}$$
$$T(\vec{v}) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n).$$

Two $n \times n$ matrices A and B with the property that $AB = \boldsymbol{0_{n\times n}}$, but ***neither*** A nor B is $\boldsymbol{0_{n\times n}}$ are called ***zero divisors***.

2.4 Exercises

1. Consider the matrices: $A = \begin{bmatrix} 5 & -3 & 6 & 8 \\ 2 & 4 & -1 & 7 \\ -4 & 6 & 3 & 0 \end{bmatrix}$, and $B = \begin{bmatrix} 6 & -4 & -7 & 3 \\ -8 & 0 & 1 & -6 \\ -3 & 7 & 5 & 4 \end{bmatrix}$.

   $C = \begin{bmatrix} 7 & 9 \\ -3 & 4 \\ 2 & -1 \\ 0 & 6 \end{bmatrix}$, and $D = \begin{bmatrix} -5 & 3 \\ 0 & -6 \\ -2 & 7 \\ 8 & 1 \end{bmatrix}$. Compute the following:

   a. $A+B$ b. $(A+B)C$ c. AC d. BC e. $AC+BC$

   f. $C+D$ g. $B(C+D)$ h. BD i. $BC+BD$ j. $(A+B)(C+D)$

   Verify that (b) and (e) are equal, and that (g) and (i) are equal.

2. Let $T_1 : \mathbb{R}^3 \to \mathbb{R}^4$, $T_2 : \mathbb{R}^4 \to \mathbb{R}^2$ and $T_3 : \mathbb{R}^2 \to \mathbb{R}^5$ be given by:

$$T_1(\langle x, y, z\rangle) = \langle 2x - 3y,\, 5y - 7z,\, x - y + 4z,\, 6x + y - z\rangle, \text{ and}$$

$$T_2(\langle x_1, x_2, x_3, x_4\rangle) = \langle 5x_1 + 2x_3 - x_4,\, 2x_1 + 8x_2 - 6x_3 + 7x_4\rangle$$

$$T_3(\langle x, y\rangle) = \langle x + 2y,\, x - y,\, 7x + 3y,\, 4x + y,\, x + 5y\rangle$$

   a. Find $[T_1]$, $[T_2]$ and $[T_3]$ and state their dimensions.

   b. Use the definition from Section 2.3 to directly get a formula for $(T_2 \circ T_1)(\langle x, y, z\rangle)$.

   c. Use (b) to get a formula for $[T_2 \circ T_1]$. State its dimension.

   d. Compute the matrix product $[T_2][T_1]$. Check that you get the same answer as (c).

   e. Repeat parts (b), (c) and (d) for the composition $T_3 \circ T_2$.

   f. Find the matrix of $T_3 \circ T_2 \circ T_1$ using any method. You may use some of your computations above. State its dimension.

3. Let $T_1, T_2, T_3 : \mathbb{R}^2 \to \mathbb{R}^2$ be:

$$T_1 = \text{reflection across the line } y = \tfrac{3}{5}x,$$

$$T_2 = \text{rotation clockwise by } \tan^{-1}\left(-\tfrac{4}{3}\right), \text{ and}$$

$$T_3 = \text{projection onto the line } y = -\tfrac{7}{3}x.$$

   a. Use your answers from the Exercises of Section 2.2 to find the matrices of these three linear transformations.

   b. Find the matrix of the compositions $T_2 \circ T_1$ and $T_1 \circ T_3$.

   c. Find the matrix of the compositions $T_3 \circ T_2 \circ T_1$ and $T_1 \circ T_3 \circ T_2$. You may use your answers from (b). Do you get the same matrix?

4. Let $T_1 : \mathbb{R}^3 \to \mathbb{R}^2$ and $T_2 : \mathbb{R}^2 \to \mathbb{R}^3$ be given by:

$$T_1(\langle x, y, z\rangle) = \langle 2x - 3y + z,\ 4x - 5y - 7z\rangle, \text{ and}$$

$$T_2(\langle x, y\rangle) = \langle 5x - 4y,\ x - 3y,\ 7x + 2y\rangle.$$

Find the matrices of T_1, T_2, $T_1 \circ T_2$ and $T_2 \circ T_1$, and state their dimensions.

5. Let $A = \begin{bmatrix} -3 & 7 \\ 5 & -2 \end{bmatrix}$, and let $p(x) = 4 - 6x + 5x^2 - 2x^3 + 7x^4$.

Compute A^2, A^3 and A^4, and use these to compute $p(A)$.

6. Let $A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & -1 & 0 \\ -1 & 4 & 7 \end{bmatrix}$, and let $p(x) = 3 + 7x - 5x^2 + 2x^3$.

Compute A^2 and A^3, and use these to compute $p(A)$.

7. Suppose that you are told that $T : \mathbb{R}^2 \to \mathbb{R}^3$ is a linear transformation, and that:

$$T(\langle 3, 10\rangle) = \langle 3, -7, 2\rangle, \text{ and}$$

$$T(\langle 2, 7\rangle) = \langle 5, 3, -8\rangle.$$

a. Explain why the set $\{\langle 3, 10\rangle, \langle 2, 7\rangle\}$ is a basis for $\mathbb{R}^2$.
b. Use the given information to find $T(\langle -5, 8\rangle)$.

8. Suppose that you are told that $T : \mathbb{R}^3 \to \mathbb{R}^5$ is a linear transformation, and that:

$$T(\langle 1, 0, 1\rangle) = \langle 3, -7, 5, 2, -4\rangle,$$

$$T(\langle 1, 1, -1\rangle) = \langle 2, 5, 4, -3, 8\rangle, \text{ and}$$

$$T(\langle -1, 1, 1\rangle) = \langle 6, 2, 1, 5, 3\rangle.$$

a. Show that $\{\langle 1, 0, 1\rangle, \langle 1, 1, -1\rangle, \langle -1, 1, 1\rangle\}$ form a basis for $\mathbb{R}^3$. Note that these are the three vectors whose images in $\mathbb{R}^5$ are provided.
b. Use the given information to find $T(\langle 6, 5, -2\rangle)$.

For Exercises (9) to (12): Prove the following properties of matrix addition and scalar multiplication, where A, B, and C are all $m \times n$ matrices and r and s are scalars:

9. $A + (B + C) = (A + B) + C$

10. $(r + s)A = rA + sA$

11. $r(A + B) = rA + rB$

12. $r(sA) = (rs)A = s(rA)$

For Exercises (13) and (14): Prove the following properties of matrix multiplication, where A and B are $m \times k$ matrices, C and D are $k \times n$ matrices, and r is a scalar:

13. $A(C + D) = AC + AD$

14. $r(BC) = (rB)C = B(rC)$

15. Prove in general that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a ***basis*** for $\mathbb{R}^n$, then the action of T is uniquely determined by the vectors $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_n)\}$ from $\mathbb{R}^m$.
More specifically, if $\vec{v} \in \mathbb{R}^n$ and $\vec{v}$ is expressed (uniquely) as:
$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n, \quad \text{then:}$$
$$T(\vec{v}) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n).$$

16. In Section 2.2, we saw the rotation matrices:
$$[rot_\theta] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$
Express each matrix below as the product of a rotation matrix with a reflection matrix (across either the x-axis or the y-axis), then describe in words the action of each matrix on $\mathbb{R}^2$. Be sure that factors are in the correct order.
   a. $\begin{bmatrix} -\cos(\theta) & \sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$
   b. $\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}$

17. Find two zero divisors of the ring of 2×2 matrices where ***none*** of the entries are zero.

18. Find two zero divisors of the ring of 3×3 matrices, where each matrix has ***at least two*** non-zero entries.

19. We know in general that AB and BA could be different matrices. However: prove that the rotation matrices ***commute*** with each other. In other words:
$$[rot_\alpha] \bullet [rot_\beta] = [rot_\beta] \bullet [rot_\alpha] = [rot_{\alpha+\beta}].$$
Write a sentence or two explaining the meaning of these equations. Hint: you will need some famous trigonometric identities.

20. ***The Center of Mat(n, n):*** In this Exercise, we will denote the set of all $n \times n$ matrices by the symbol $Mat(n,n)$. We have demonstrated that in general, matrix multiplication is ***not*** commutative. However, we know that for the identity matrix $\boldsymbol{I_n}$:
$$\boldsymbol{I_n}A = A = A\boldsymbol{I_n}$$
for all $n \times n$ matrices A. Thus, we can say that $\boldsymbol{I_n}$ ***commutes*** with ***all*** $n \times n$ matrices A.
   a. Warm-up: Show that $\mathbf{0}_{n\times n}$ also commutes with all $n \times n$ matrices A. Hint: Use a Property that you proved in one of the previous Exercises.
   b. For any $k \in \mathbb{R}$, show that $k\boldsymbol{I_n}$ also commutes with all $n \times n$ matrices A.
   We will now define the ***Center*** of $Mat(n,n)$ to be:
$$Center(Mat(n,n)) = \left\{ C \in Mat(n,n) \,|\, AC = CA \text{ for } \textbf{\textit{all}}\; n \times n \text{ matrices } A. \right\}$$
   Thus, the matrices we found in (a) and (b) all belong to $Center(M(n,n))$. The goal of the rest of this Exercise is to show that these are the ***only*** members of $Center(M(n,n))$, that is, to prove that:

$$Center(M(n,n)) = \{k\boldsymbol{I_n} \mid k \in \mathbb{R}\}.$$

Notice that this includes $\mathbf{0}_{n\times n}$ if we let $k = 0$.

c. Consider the 3×3 matrix:

$$A_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let C be any 3×3 matrix. Write the entries of C in $[c_{i,j}]$ notation as usual. Find the products A_1C and CA_1.

d. If we want A_1C and CA_1 to be ***equal***, which entries of C have to be zero?

e. Repeat parts (c) and (d) for the matrix:

$$A_2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

f. Put the last three parts together to show that if we want C to commute with ***both*** A_1 and A_2, then C has to be of the form:

$$C = \begin{bmatrix} c_{11} & 0 & 0 \\ 0 & c_{22} & 0 \\ 0 & 0 & c_{33} \end{bmatrix}.$$

Now, let A_3 and A_4 be the Type 2 elementary 3×3 matrices:

$$A_3 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ and } A_4 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

g. Show that if we want to force the equalities: $A_3C = CA_3$ and $A_4C = CA_4$, then we must have $c_{11} = c_{22} = c_{33}$. Explain why this proves that:

$$Center(Mat(3,3)) = \{k\boldsymbol{I_3} \mid k \in \mathbb{R}\}.$$

h. Create analogous matrices A_1, A_2, $\ldots$ etc. for 4×4 matrices, and show that $Center(Mat(4,4)) = \{k\boldsymbol{I_4} \mid k \in \mathbb{R}\}$. What is the smallest number of these matrices A_i that will accomplish the proof?

i. Generalize the argument above to show that $Center(Mat(n,n)) = \{k\boldsymbol{I_n} \mid k \in \mathbb{R}\}$ for any n. What is the smallest number of these matrices A_i that will accomplish the proof? The answer in general will depend on n.

2.5 The Kernel and Range; One-to-One and Onto Transformations

In this Section, we will investigate two special subspaces associated to T, namely the ***kernel*** and the ***range*** of T. This is probably the first time in your life that you are encountering the first word in a technical way. It usually goes with the phrase "of corn," or perhaps "of truth." You have seen the second word in Precalculus, though, where the range of a function $y = f(x)$ is usually defined as:

$$range(f) = \left\{ y \in \mathbb{R} \,|\, y = f(x) \text{ for some } x \in domain(f) \right\}.$$

An analogous definition will appear below in the context of Linear Algebra. The kernel and range of T will allow us to decide if T has certain desirable properties, namely whether or not T is ***one-to-one*** or T is ***onto***.

The Kernel and Range of a Linear Transformation

Definition: If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, we define the ***kernel*** of T as the set:

$$ker(T) = \left\{ \vec{v} \in \mathbb{R}^n \,|\, T(\vec{v}) = \vec{\mathbf{0}}_m \right\} \subset \mathbb{R}^n.$$

Similarly, we define the ***range*** of T as the set:

$$range(T) = \left\{ \vec{w} \in \mathbb{R}^m \,|\, \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in \mathbb{R}^n \right\} \subset \mathbb{R}^m.$$

We emphasize that $ker(T)$ is from $\mathbb{R}^n$, and $range(T)$ is from $\mathbb{R}^m$

Now, since we want $T(\vec{v}) = [T]\vec{v} = \vec{\mathbf{0}}_m$ in order for $\vec{v}$ to be in $ker(T)$, we can immediately see that $\vec{v} \in ker(T)$ ***if and only if*** $\vec{v} \in nullspace([T])$. We already know that $nullspace([T]) \trianglelefteq \mathbb{R}^n$ from Chapter 1, and so $ker(T)$ is a subspace of $\mathbb{R}^n$. Notice also that we write $ker(T)$ and not $ker([T])$, and we emphasize that $ker(T)$ is a subspace of the ***domain*** $\mathbb{R}^n$ of T.

Similarly, if any member $\vec{w}$ of $range(T)$ must be of the form $\vec{w} = T(\vec{v})$ for some vector $\vec{v} \in \mathbb{R}^n$. Again, $T(\vec{v}) = [T]\vec{v}$, and let us remind ourselves from Chapter 1 that $[T]\vec{v}$ is a ***linear combination*** of the ***columns*** of $[T]$. In other words, $\vec{w} \in range(T)$ ***if and only if*** $\vec{w} \in colspace([T])$. We know from Chapter 1 that $colspace([T]) \trianglelefteq \mathbb{R}^m$, and so $range(T)$ is a subspace of $\mathbb{R}^m$. We write $range(T)$ and not $range([T])$.

Let us summarize these results in the following:

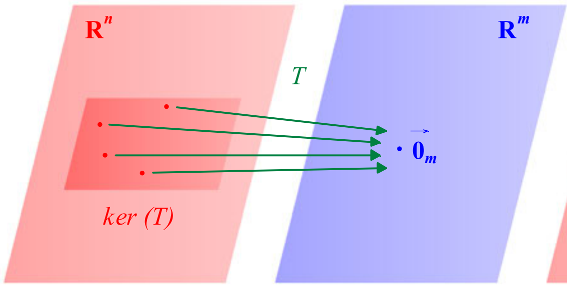
Theorem: If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a ***linear transformation***, then:

$$ker(T) = nullspace([T]) \trianglelefteq \mathbb{R}^n, \text{ and}$$
$$range(T) = colspace([T]) \trianglelefteq \mathbb{R}^m.$$

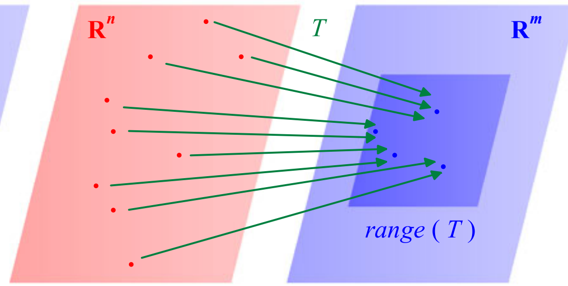
We call the dimension of $ker(T)$ the ***nullity*** of T, written $nullity(T)$. Similarly, we call the dimension of $range(T)$ the ***rank*** of T, written $rank(T)$. Thus:

$$nullity(T) = dim(nullspace([T])) = nullity([T]), \text{ and}$$
$$rank(T) = dim(colspace([T])) = rank([T]).$$

We summarize the concepts of kernel and the range below:



The Kernel of T



The Range of T

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be given by:

$$[T] = \begin{bmatrix} 2 & -1 & 1 \\ -3 & 5 & 9 \\ -1 & -2 & -8 \\ 6 & -4 & 0 \end{bmatrix}, \quad \text{with rref } R = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The only free variable is x_3, and we solve for the ***nullspace*** of this matrix, as before:

$$\langle x_1, x_2, x_3 \rangle = \langle -2x_3, -3x_3, x_3 \rangle = x_3 \langle -2, -3, 1 \rangle, \text{ and thus:}$$
$$ker(T) = Span(\{\langle -2, -3, 1 \rangle\}) \trianglelefteq \mathbb{R}^3, \text{ and } nullity(T) = 1.$$

The leading 1's are in the first and second columns of R, and thus the first and second columns of $[T]$ form a basis for the ***columnspace*** of this matrix. Thus:

$$range(T) = Span(\{\langle 2, -3, -1, 6 \rangle, \langle -1, 5, -2, -4 \rangle\}) \trianglelefteq \mathbb{R}^4, \text{ and } rank(T) = 2.$$

Notice that $ker(T)$ is a subspace of the ***domain*** $\mathbb{R}^3$ and $range(T)$ is a subspace of the ***codomain*** $\mathbb{R}^4$, as they should be. $\square$

The Dimension Theorem for Linear Transformations

Because the kernel and the range of a linear transformation T are the nullspace and columnspace of the matrix $[T]$, we can rephrase the Dimension Theorem for Matrices as follows:

Theorem — The Dimension Theorem for Linear Transformations:
Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then:

$$rank(T) + nullity(T) = n = dim(domain(T))$$

Example: Continuing with the previous Example, we saw that $nullity(T) = 1$ and $rank(T) = 2$. The ***domain*** of T is $\mathbb{R}^3$, and thus we verify that:

$$rank(T) + nullity(T) = 2 + 1 = 3. \square$$

One-to-One Transformations

We will now investigate linear transformations that possess a quality similar to that of some functions that we see in ordinary Algebra:

Definition: We say that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** or ***injective*** if the image of two different vectors from the domain are different vectors of the codomain:

$$\text{If } \vec{v}_1 \neq \vec{v}_2 \text{ then } T(\vec{v}_1) \neq T(\vec{v}_2).$$

We also say that T is an ***injection*** or an ***embedding***.

Since it is easier to solve an equation rather than an inequality, let us rephrase this definition in a better way by using the ***contrapositive*** of the implication above, as we saw in Chapter Zero. An implication is true ***if and only if*** its contrapositive is also true, and so we obtain the following:

Theorem: A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** ***if and only if*** the only way two vectors from the domain have the same image in the codomain is for them to be the same vector to begin with:

$$\text{If } T(\vec{v}_1) = T(\vec{v}_2) \text{ then } \vec{v}_1 = \vec{v}_2.$$

In other words, the ***only solution*** to $T(\vec{v}_1) = T(\vec{v}_2)$ is $\vec{v}_1 = \vec{v}_2$.

This condition, unfortunately, is still too awkward to verify. We will use it, though, to prove the following Theorem, which will give us an easy way to determine whether or not a linear transformation is one-to-one:

Theorem — The Kernel Test for Injectivity:

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** ***if and only if:***

$$ker(T) = \left\{\vec{\mathbf{0}}_n\right\}.$$

Proof: ($\Rightarrow$) If T is one-to-one, we must show that $ker(T)$ is only $\left\{\vec{\mathbf{0}}_n\right\}$.

Let $\vec{v} \in ker(T)$. This means that $T(\vec{v}) = \vec{\mathbf{0}}_m$. However, we know from the final Exercises in Section 2.1 that $T\left(\vec{\mathbf{0}}_n\right) = \vec{\mathbf{0}}_m$. Thus $T(\vec{v}) = T\left(\vec{\mathbf{0}}_n\right)$. But by the one-to-one property, the only solution to $T(\vec{v}_1) = T(\vec{v}_2)$ is $\vec{v}_1 = \vec{v}_2$, for any two vectors $\vec{v}_1$ and $\vec{v}_2 \in \mathbb{R}^n$.

Thus $\vec{v} = \vec{\mathbf{0}}_n$, so the only possible member of $ker(T)$ is $\vec{\mathbf{0}}_n$. Thus $ker(T) = \left\{\vec{\mathbf{0}}_n\right\}$.

($\Leftarrow$) Conversely, suppose that $ker(T)$ is only $\left\{\vec{\mathbf{0}}_n\right\}$. We must prove that T is one-to-one, that is, the only solution to an equation $T(\vec{v}_1) = T(\vec{v}_2)$ is $\vec{v}_1 = \vec{v}_2$.

Let us use Proof by Contradiction. Suppose that we were lucky enough to find two vectors $\vec{v}_1$ and $\vec{v}_2$ such that $T(\vec{v}_1) = T(\vec{v}_2)$. Then $T(\vec{v}_1) - T(\vec{v}_2) = \vec{\mathbf{0}}_m$, or $T(\vec{v}_1 - \vec{v}_2) = \vec{\mathbf{0}}_m$ by the linearity properties. But this equation tells us that $\vec{v}_1 - \vec{v}_2 \in ker(T)$. Thus $\vec{v}_1 - \vec{v}_2 = \vec{\mathbf{0}}_n$, or $\vec{v}_1 = \vec{v}_2$. Thus, T is one-to-one and we weren't so lucky after all. ■

Example: Suppose $T_1 : \mathbb{R}^3 \to \mathbb{R}^4$ is given by:

$$[T_1] = \begin{bmatrix} 1 & -3 & 4 \\ 2 & -6 & 9 \\ 5 & -15 & 4 \\ -3 & 9 & -7 \end{bmatrix}, \quad \text{with rref} \quad R_1 = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, $ker(T_1) = Span(\{\langle 3, 1, 0\rangle\})$, and thus T_1 is ***not*** one-to-one.

However, let us change one entry on the top row ever so slightly.

Suppose $T_2 : \mathbb{R}^3 \to \mathbb{R}^4$ with matrix given by:

$$[T_2] = \begin{bmatrix} 1 & -2 & 4 \\ 2 & -6 & 9 \\ 5 & -15 & 4 \\ -3 & 9 & -7 \end{bmatrix}, \quad \text{with rref} \quad R_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

This time, $ker(T_2) = \left\{ \vec{\mathbf{0}}_3 \right\}$ and thus T_2 ***is*** one-to-one.□

Knowledge of dimensions of the domain and codomain of T can sometimes tell us that T is automatically ***not*** one-to-one.

Theorem: A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***not*** one-to-one if $n > m$.

Proof: We know that $[T]$ is an $m \times n$ matrix. If $n > m$, the homogeneous system corresponding to this matrix will be ***underdetermined***, and thus it will have an infinite number of solutions. Thus $ker(T) \neq \left\{ \vec{\mathbf{0}}_n \right\}$ and T is not one-to-one.■

Example: Any linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^3$ is automatically not one-to-one. However, a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^5$ could be either one-to-one or not one-to-one. For instance, the zero transformation $Z_{3,5}$ is ***not*** one-to-one.□

Onto Linear Transformations

Now we come to another special type of linear transformation:

Definition: We say that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***onto*** or ***surjective*** if:

$$range(T) = \mathbb{R}^m.$$

We also say that T is a ***surjection*** or a ***covering*** (because T hits all the vectors of $\mathbb{R}^m$).

Since $rank(T) = dim(range(T))$, the following Theorem follows directly:

Theorem: A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***onto*** ***if and only if*** $rank(T) = m$.

Example: Suppose $T_1 : \mathbb{R}^3 \to \mathbb{R}^2$ is given by:

$$[T_1] = \begin{bmatrix} -2 & -8 & 6 \\ 1 & 4 & -3 \end{bmatrix}, \quad \text{with rref} \quad R_1 = \begin{bmatrix} 1 & 4 & -3 \\ 0 & 0 & 0 \end{bmatrix}.$$

There is only one leading 1 in R_1, thus $rank(T_1) = 1 < 2$, and thus T_1 is ***not*** onto. Again, let us change one entry on the top row. Suppose $T_2 : \mathbb{R}^3 \to \mathbb{R}^2$ is given by:

$$[T_2] = \begin{bmatrix} -2 & -8 & 7 \\ 1 & 4 & -3 \end{bmatrix}, \quad \text{with rref} \quad R_2 = \begin{bmatrix} 1 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

This time, there are two leading 1's in R_2, and thus $rank(T_2) = 2$, and thus $range(T_2)$ must be all of $\mathbb{R}^2$. Thus T_2 ***is*** onto. Notice that neither T_1 nor T_2 is one-to-one, since $3 > 2$. □

Like one-to-one linear transformations, knowledge of the dimensions of the domain and codomain can tell us when T is definitely ***not*** onto:

Theorem: A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***not*** onto if $n < m$.

Proof: Once again, $[T]$ is an $m \times n$ matrix. If $n < m$, there are more rows than columns, in other words, any system with $[T]$ as coefficient matrix is overdetermined. Thus, the maximum number of leading 1's that can be found in the rref of $[T]$ is $\boldsymbol{n}$, so, $dim(colspace([T])) \leq n$. But since $n < m$, the dimension of the columnspace is strictly ***less*** than the dimension of the codomain $\mathbb{R}^m$. Thus, $rank(T) < m$, so the range of T ***cannot*** be all of $\mathbb{R}^m$. ■

Example: A linear transformation $T_1 : \mathbb{R}^4 \to \mathbb{R}^7$ cannot be onto, but it could be one-to-one. On the other hand, a linear transformation $T_2 : \mathbb{R}^7 \to \mathbb{R}^4$ cannot be one-to-one, but it could be onto. □

Notice that $T : \mathbb{R}^n \to \mathbb{R}^m$ ***cannot be one-to-one*** if $n > m$, and T ***cannot be onto*** if $n < m$. Therefore, when $n = m$, that is, when T is an ***operator***, we cannot tell in advance whether or not T will be one-to-one, onto, both, or neither.

Example: Suppose that $T_1 : \mathbb{R}^3 \to \mathbb{R}^3$ is given by:

$$[T_1] = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & -2 \\ -3 & 2 & -19 \end{bmatrix}, \quad \text{with rref} \quad R_1 = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -5 \\ 0 & 0 & 0 \end{bmatrix}.$$

Since R_1 has a free variable, namely z, we see that $ker(T_1) = Span(\{\langle -3, 5, 1\rangle\})$, a ***line*** through the origin. Thus, T_1 is ***not*** one-to-one. Also, since there are only two leading 1's, $rank(T_1) = 2 < 3$, and so T_1 is ***not*** onto either. Since the leading 1's are found in the 1st and 2nd columns, we can say that:

$$range(T_1) = Span(\{\langle 2, 1, -3\rangle, \langle 1, 1, 2\rangle\})$$

Thus $range(T_1)$ is a ***plane*** through the origin and not all of $\mathbb{R}^3$. Thus, T_1 is ***neither*** one-to-one, nor onto. However, we can take this opportunity to verify that the Dimension Theorem is still true:

$$rank(T_1) + nullity(T_1) = 2 + 1 = 3.$$

Now, this time, let us change the bottom row of T_1 by just a little bit. Let $T_2 : \mathbb{R}^3 \to \mathbb{R}^3$ be given by:

$$[T_2] = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 1 & -2 \\ -3 & 2 & -18 \end{bmatrix}, \quad \text{with rref } R_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Note that R_2 is the identity matrix $\boldsymbol{I_3}$. There are no free variables, so $ker(T_2) = \left\{ \vec{\mathbf{0}}_3 \right\}$ and T_2 is ***one-to-one***. There are 3 leading 1's in R, so $rank(T_2) = 3$, which means that $range(T_2) = \mathbb{R}^3$. Thus T_2 is also ***onto***. Thus we conclude that T_2 is ***both*** one-to-one and onto.□

We have seen that just glancing at the rref of $[T]$ immediately tells us whether or not T is one-to-one and/or onto. It is worth summarizing these observations in the following:

Theorem: Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and R is the rref of $[T]$. Then:

1. T is ***one-to-one*** ***if and only if*** R does ***not*** have any free variables.
2. T is ***onto*** ***if and only if*** R does ***not*** have any row consisting only of zeroes.

In Section 1.8, we encountered the concept of a ***full-rank*** $m \times n$ matrix A, that is, where $rank(A) = \min(m,n)$. We saw in that Section that full-rank matrices have special properties. We will also call a linear transformation T full-rank if $[T]$ is a full-rank matrix. Let us translate the properties that we saw for full-rank matrices in Section 1.8 in the language of linear transformations. We leave the Proof as an Exercise.

Theorem — Equivalent Properties for Full-Rank Linear Transformations:
Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then:

1. if $m < n$: T is full-rank ***if only only if*** T is ***onto***.
2. if $m = n$: T is full-rank ***if and only if*** T is ***both one-to-one and onto***.
3. if $m > n$: T is full-rank ***if and only if*** T is ***one-to-one***.

Example: Suppose that $T_1 : \mathbb{R}^4 \to \mathbb{R}^3$ is the linear transformation defined by:

$$[T_1] = \begin{bmatrix} 3 & 7 & 2 & 7 \\ -2 & 1 & 3 & 6 \\ 5 & -4 & -7 & -3 \end{bmatrix} \quad \text{with rref } R_1 = \begin{bmatrix} 1 & 0 & 0 & 8 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 1 & 9 \end{bmatrix}.$$

R_1 does not have a row of zeroes, so T_1 is ***onto***. But note that $rank(T_1) = 3 = min(3,4)$, which verifies that T_1 is full-rank.

Now, suppose that $T_2 : \mathbb{R}^4 \to \mathbb{R}^4$ is the operator defined by:

$$[T_2] = \begin{bmatrix} 3 & 7 & 2 & 7 \\ -2 & 1 & 3 & 6 \\ 5 & -4 & -7 & -3 \\ 2 & 6 & 9 & -1 \end{bmatrix} \quad \text{with rref } R_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that R_2 is $\boldsymbol{I_4}$, which has no free variables nor rows of zeroes. Thus $rank(T_2) = 4 = min(4,4)$, and we can conclude that T_2 is full-rank and is ***both*** ***one-to-one*** and ***onto***.

Finally, suppose that $T_3 : \mathbb{R}^3 \to \mathbb{R}^4$ is the linear transformation defined by:

$$[T_3] = \begin{bmatrix} 3 & 7 & 2 \\ -2 & 1 & 3 \\ 5 & -4 & -7 \\ 2 & 6 & 9 \end{bmatrix} \text{ with rref } R_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Notice that $[T_3]$ contains the first three columns of $[T_2]$, and we know from R_2 that ***all*** the columns of $[T_2]$ are linearly independent, and so the first three columns remain independent on their own in $[T_3]$. Thus $[T_3]$ has no free variables and so T_3 is ***one-to-one***.

Again, we see that $rank(T_3) = 3 = min(3,4)$, and so T_3 is also full-rank.□

A Recap of The One-to-One and Onto Properties

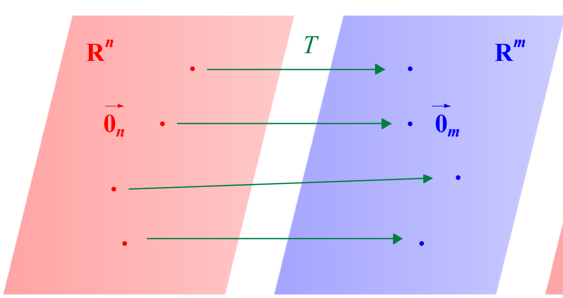
The two properties of being one-to-one or onto are of such fundamental importance in Linear Algebra (and in many other areas in Mathematics) that we will now summarize the basic tests for these properties.

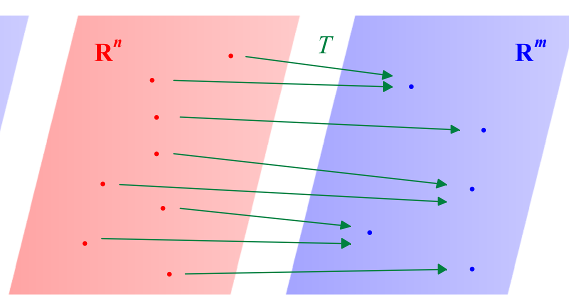
A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one ***if and only*** if $ker(T) = \left\{\vec{\mathbf{0}}_n\right\}$.

This means that if $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{\mathbf{0}}_m$, then $\vec{v} = \vec{\mathbf{0}}_n$. This also means that $nullity(T) = 0$.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is onto ***if and only if*** $range(T) = \mathbb{R}^m$.

This means that for any vector $\vec{w} \in \mathbb{R}^m$, we can find at least one vector $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{w}$ (although more than one such vector $\vec{v}$ could exist for a given $\vec{w}$). This also means that $rank(T) = m$.

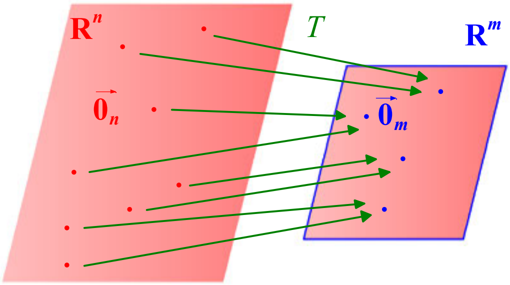


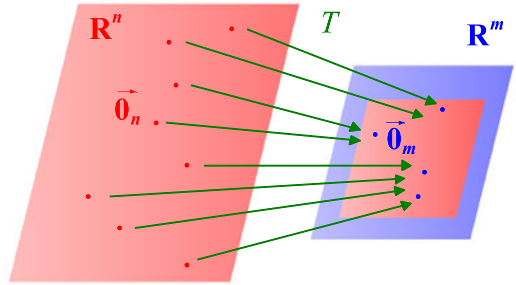


T is ***one-to-one*** $\Leftrightarrow$ $ker(T) = \left\{\vec{\mathbf{0}}_n\right\}$ $\qquad$ T is ***onto*** $\Leftrightarrow$ $range(T) = \mathbb{R}^m$

T can be one-to-one but not onto, onto but not one-to-one, ***neither*** one-to-one nor onto, or ***both*** one-to-one and onto.

However, if $T : \mathbb{R}^n \to \mathbb{R}^m$ with $n > m$, then T is automatically ***not*** one-to-one. In this case, T could be onto, or neither onto nor one-to-one.



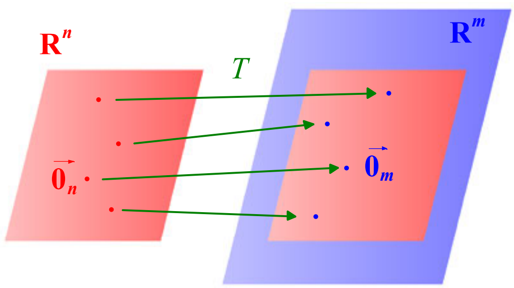


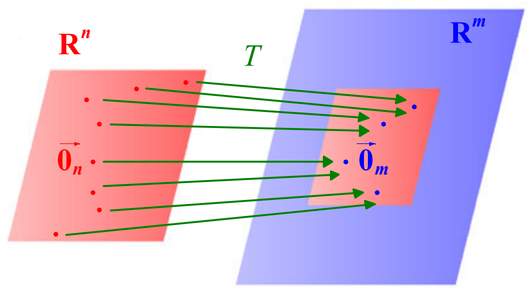
T is onto, but not one-to-one
$rank\,(T) = m$; T is full rank

$n > m$

T is neither onto nor one-to-one
$rank\,(T) < m$

Similarly, if $n < m$, then T is automatically ***not*** onto. In this case, T could be one-to-one, or neither one-to-one nor onto.





T is one-to-one, but not onto
$rank\,(T) = n$; T is full rank

$n < m$

T is neither one-to-one nor onto
$rank\,(T) < n$

2.5 Section Summary

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, we define the ***kernel*** of T as the subspace:

$$ker(T) = \left\{ \vec{v} \in \mathbb{R}^n \mid T(\vec{v}) = \vec{\mathbf{0}}_m \right\} = nullspace([T]) \trianglelefteq \mathbb{R}^n.$$

We call the dimension of $ker(T)$ the ***nullity*** of T, written $nullity(T)$.

The ***range*** of T is the subspace:

$$range(T) = \left\{ \vec{w} \in \mathbb{R}^m \,|\, \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in \mathbb{R}^n \right\} = colspace([T]) \trianglelefteq \mathbb{R}^m.$$

We call the dimension of $range(T)$ the ***rank*** of T, written $rank(T)$.

The Dimension Theorem for Linear Transformations states that:

$$rank(T) + nullity(T) = n = dim\big(\text{domain of } T\big).$$

We say that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** or ***injective*** if the image of different vectors from the domain are different vectors from the codomain: If $\vec{v}_1 \neq \vec{v}_2$ then $T(\vec{v}_1) \neq T(\vec{v}_2)$.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is one-to-one ***if and only if*** $ker(T) = \left\{ \vec{\mathbf{0}}_n \right\}$.

The same linear transformation T is ***onto*** or ***surjective*** if $range(T) = \mathbb{R}^m$, which is true ***if and only if*** $rank(T) = m$.

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is automatically ***not*** one-to-one if $n > m$, and T is automatically ***not*** onto if $n < m$.

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then:

1. if $m < n$: T is full-rank ***if only only if*** T is ***onto***.
2. if $m = n$: T is full-rank ***if and only if*** T is ***both one-to-one and onto***.
3. if $m > n$: T is full-rank ***if and only if*** T is ***one-to-one***.

2.5 Exercises

1. Let $T_1 : \mathbb{R}^4 \to \mathbb{R}^3$ be given by:

   $$T_1(\langle x_1, x_2, x_3, x_4 \rangle) = \langle 3x_1 + x_2 - 7x_3 + 8x_4, 2x_1 + 2x_2 - 2x_3 - 4x_4, -2x_1 + x_2 + 8x_3 - 17x_4 \rangle.$$

   a. Find $[T_1]$.
   b. Find the rref R_1 of $[T_1]$
   c. Use R_1 to find a basis for the kernel of T_1.
   d. Find the nullity of T_1.
   e. Is T_1 one-to-one?
   f. Use R_1 to find a basis for the range of T_1.
   g. Find the rank of T_1.
   h. Is T_1 onto?
   i. Verify The Dimension Theorem for T_1.
2. Let $T_2 : \mathbb{R}^3 \to \mathbb{R}^4$ be given by:

   $$T_2(\langle x_1, x_2, x_3 \rangle) = \langle 3x_1 - 6x_2 + 5x_3, 2x_1 - 4x_2 + 7x_3, -5x_1 + 10x_2 + 3x_3, -x_1 + 2x_2 + 8x_3 \rangle.$$

   a. Find $[T_2]$.
   b. Find the rref R_2 of $[T_2]$.
   c. Use R_2 to find a basis for the kernel of T_2.
   d. Find the nullity of T_2.
   e. Is T_2 one-to-one?
   f. Use R_2 to find a basis for the range of T_2.
   g. Find the rank of T_2.
   h. Is T_2 onto?
   i. Verify The Dimension Theorem for T_2.
3. Let $T_3 : \mathbb{R}^3 \to \mathbb{R}^3$ be the operator given by:

   $$T_3(\langle x_1, x_2, x_3 \rangle) = \langle -5x_1 - 7x_2 + 2x_3, -2x_1 + x_2 + 16x_3, 3x_1 - 2x_2 - 26x_3 \rangle$$

   a. Find $[T_3]$.
   b. Find the rref R_3 of $[T_3]$
   c. Use R to find a basis for the kernel of T_3.
   d. Find the nullity of T_3.
   e. Is T_3 one-to-one?
   f. Use R_3 to find a basis for the range of T_3.

g. Find the rank of T_3.

h. Is T_3 onto?

i. Verify The Dimension Theorem for T_3.

j. Describe geometrically the ***kernel*** and the ***range*** of T_3 as a subspace of $\mathbb{R}^3$. If the subspace is a ***line***, give a direction vector for it, and if it is a ***plane***, give an equation for it in the standard form $ax + by + cz = 0$.

k. Is the range ***orthogonal*** to the kernel? Explain why this does not contradict anything we have learned so far.

For Exercises (4) to (20): Given the following linear transformations T, their standard matrices $[T]$, and the rref R of each matrix: (a) find a basis for the kernel of T (if possible), (b) state the nullity of T, (c) decide if T is one-to-one or not, (d) find a basis for the range of T, (e) state the rank of T, (f) decide if T is onto or not, (g) decide if T is full-rank, and (h) verify the Dimension Theorem for T.

4. $T : \mathbb{R}^3 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 7 \\ 3 & 5 & 5 \\ -3 & 2 & -19 \\ 3 & 10 & -5 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$

5. $T : \mathbb{R}^3 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} 2 & 3 & 4 \\ 3 & 4 & 7 \\ 3 & 5 & 5 \\ -3 & 2 & -18 \\ 3 & 10 & -5 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$ (Compare to Exercise 4.)

6. $T : \mathbb{R}^5 \to \mathbb{R}^3,\ [T] = \begin{bmatrix} 3 & -2 & -6 & -9 & 8 \\ -5 & 3 & 7 & 16 & -13 \\ -8 & 5 & 13 & 25 & -21 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 4 & -5 & 2 \\ 0 & 1 & 9 & -3 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

7. $T : \mathbb{R}^5 \to \mathbb{R}^3,\ [T] = \begin{bmatrix} 3 & -2 & -6 & -9 & 8 \\ -5 & 3 & 7 & 16 & -13 \\ -8 & 5 & 13 & 25 & -20 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 4 & -5 & 0 \\ 0 & 1 & 9 & -3 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

8. $T : \mathbb{R}^5 \to \mathbb{R}^3,\ [T] = \begin{bmatrix} 3 & -2 & -2 & 2 & 3 \\ -5 & 3 & 9 & -10 & -34 \\ -8 & 5 & 4 & -5 & -2 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 0 & 2 & -1 \\ 0 & 1 & 0 & 3 & 2 \\ 0 & 0 & 1 & -1 & -5 \end{bmatrix}$

(Compare Exercises 6 through 8 to each other.)

9. $T:\mathbb{R}^5 \to \mathbb{R}^4,\ [T] = \begin{bmatrix} 3 & -2 & -6 & -2 & 15 \\ -5 & 3 & 7 & 9 & 6 \\ -8 & 5 & 13 & 10 & -14 \\ 6 & -3 & -3 & -8 & 2 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 4 & 0 & 3 \\ 0 & 1 & 9 & 0 & -8 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

10. $T:\mathbb{R}^5 \to \mathbb{R}^4,\ [T] = \begin{bmatrix} 3 & -9 & 6 & 9 & -2 \\ -5 & 15 & -7 & 0 & 9 \\ -2 & 6 & -3 & -1 & 7 \\ 2 & -6 & 5 & 11 & -8 \end{bmatrix},\ R = \begin{bmatrix} 1 & -3 & 0 & -7 & 0 \\ 0 & 0 & 1 & 5 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

11. $T:\mathbb{R}^5 \to \mathbb{R}^4,\ [T] = \begin{bmatrix} 3 & 6 & -2 & -1 & -11 \\ -5 & -7 & 3 & -4 & -14 \\ -2 & -3 & 4 & 3 & 26 \\ 2 & 5 & 7 & -2 & 10 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 0 & 0 & 2 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 1 & 5 \end{bmatrix}$

12. $T:\mathbb{R}^5 \to \mathbb{R}^4,\ [T] = \begin{bmatrix} 3 & 6 & 3 & -2 & -1 \\ -5 & -7 & -11 & 3 & -4 \\ -2 & -3 & -4 & 4 & 3 \\ 2 & 5 & 0 & 7 & -2 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 5 & 0 & 0 \\ 0 & 1 & -2 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

13. $T:\mathbb{R}^4 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} 15 & 72 & -16 & -44 \\ 30 & 63 & 13 & -25 \\ -10 & 27 & -31 & -29 \\ -5 & -54 & 22 & 38 \\ -15 & 0 & -24 & -12 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & \frac{8}{5} & \frac{4}{5} \\ 0 & 1 & -\frac{5}{9} & -\frac{7}{9} \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

14. $T:\mathbb{R}^4 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} 1 & 2 & -6 & -4 \\ 3 & 6 & 3 & -5 \\ -1 & 7 & -3 & -2 \\ -5 & -4 & 2 & 3 \\ 5 & 0 & -4 & 1 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

15. $T:\mathbb{R}^4 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} 5 & 7 & 2 & 3 \\ 2 & -1 & 3 & 17 \\ -6 & -3 & -5 & -25 \\ -2 & 3 & 1 & -15 \\ 1 & 0 & -1 & 2 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

16. $T:\mathbb{R}^4 \to \mathbb{R}^4,\ [T] = \begin{bmatrix} 2 & -6 & 3 & 2 \\ 3 & -9 & 1 & 10 \\ 2 & -6 & 5 & -2 \\ 5 & -15 & 4 & 12 \end{bmatrix},\ R = \begin{bmatrix} 1 & -3 & 0 & 4 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

17. $T:\mathbb{R}^4 \to \mathbb{R}^4,\ [T] = \begin{bmatrix} 4 & 2 & 1 & -6 \\ 5 & 9 & -2 & -14 \\ -6 & -7 & -1 & 1 \\ 5 & 6 & 3 & 17 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 0 \end{bmatrix}$

18. $T:\mathbb{R}^5 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} -3 & 15 & -5 & 8 & 12 \\ 2 & -10 & -1 & 25 & -4 \\ 5 & -25 & 2 & 31 & 0 \\ 0 & 0 & -3 & 21 & -25 \\ -4 & 20 & -7 & 13 & 37 \end{bmatrix},\ R = \begin{bmatrix} 1 & -5 & 0 & 9 & 0 \\ 0 & 0 & 1 & -7 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

19. $T:\mathbb{R}^5 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} -3 & -5 & 4 & 2 & -5 \\ 2 & 1 & 9 & -4 & -1 \\ 4 & 6 & -2 & -5 & 2 \\ 0 & -1 & 5 & -5 & -3 \\ -3 & -4 & -1 & 3 & -7 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 7 & 0 & 0 \\ 0 & 1 & -5 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

20. $T:\mathbb{R}^5 \to \mathbb{R}^5,\ [T] = \begin{bmatrix} -3 & -5 & 2 & -5 & -4 \\ 2 & -1 & -4 & 4 & -5 \\ 4 & 2 & -5 & 9 & -4 \\ 0 & -3 & -5 & -9 & 1 \\ -3 & -7 & 3 & 2 & -12 \end{bmatrix},\ R = \begin{bmatrix} 1 & 0 & 0 & 7 & -5 \\ 0 & 1 & 0 & -2 & 3 \\ 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$

21. If $T:\mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, prove ***using only the definitions*** that:

a. T is ***one-to-one*** ***if and only if*** every $\vec{w} \in \mathbb{R}^m$ can be expressed as $\vec{w} = T(\vec{v})$ for ***at most one*** vector $\vec{v} \in \mathbb{R}^n$. Hint: for (a) and (b) you can use Proof by Contradiction.

b. T is ***onto*** ***if and only if*** every $\vec{w} \in \mathbb{R}^m$ can be expressed as $\vec{w} = T(\vec{v})$ for ***at least one*** vector $\vec{v} \in \mathbb{R}^n$.

c. T is ***both one-to-one and onto*** ***if and only if*** every $\vec{w} \in \mathbb{R}^m$ can be expressed as $\vec{w} = T(\vec{v})$ for ***exactly one*** vector $\vec{v} \in \mathbb{R}^n$.

22. Suppose that Π is a plane in $\mathbb{R}^3$ passing through the origin, and $L = \Pi^{\perp}$, its normal line. In Section 2.2, we considered the operators $proj_L$, $proj_\Pi$ and $refl_\Pi$. Find the following subspaces (Hint: the answers are either $\left\{\vec{\mathbf{0}}_3\right\}$, L, Π or $\mathbb{R}^3$). Explain your reasoning.

a. $ker(proj_L)$ b. $range(proj_L)$ c. $ker(proj_\Pi)$

d. $range(proj_\Pi)$ e. $ker(refl_\Pi)$ f. $range(refl_\Pi)$

23. ***The Kernel and Range of a Composition:*** The purpose of this Exercise is to investigate the kernel and range of the ***composition*** of two linear transformations. Suppose that:

$$T_1 : \mathbb{R}^n \to \mathbb{R}^k, \quad \text{and} \quad T_2 : \mathbb{R}^k \to \mathbb{R}^m$$

are linear transformations.

a. Write down the general definition of the ***kernel*** of ***any*** linear transformation $T : \mathbb{R}^a \to \mathbb{R}^b$. Use the symbol $\vec{\mathbf{0}}_a$ or $\vec{\mathbf{0}}_b$, whichever is appropriate.

b. Adapt the definition in part (a) to write down the definition of $ker(T_1)$, $ker(T_2)$ and $ker(T_2 \circ T_1)$ as set up above. There should be ***three*** separate definitions. Make sure that you precisely use the symbols $\mathbb{R}^n$, $\mathbb{R}^k$, $\mathbb{R}^m$, $\vec{\mathbf{0}}_n$, $\vec{\mathbf{0}}_k$ and $\vec{\mathbf{0}}_m$, where appropriate.

c. Two out of the three subspaces that you defined in (b) are subspaces of the same Euclidean space. Which of the two kernels live in which same Euclidean space?

d. Use your definitions to prove that $ker(T_1)$ is a subset of $ker(T_2 \circ T_1)$, that is:

$$ker(T_1) \subseteq ker(T_2 \circ T_1).$$

Hint: This means that you must show that every vector $\vec{v}$ that satisfies the definition of $ker(T_1)$ also satisfies the definition of $ker(T_2 \circ T_1)$.

e. Use part (d) to prove that if $T_2 \circ T_1$ is one-to-one, then T_1 is also one-to-one.

f. Write down the contrapositive of the statement in (e).

Now, in a similar way, we will investigate the ranges:

g. Write down the general definition of the ***range*** of ***any*** linear transformation $T : \mathbb{R}^a \to \mathbb{R}^b$.

h. Adapt the definition in part (g) to write down the definition of $range(T_1)$, $range(T_2)$ and $range(T_2 \circ T_1)$ as set up above. There should be ***three*** separate definitions. Make sure that you precisely use the symbols $\mathbb{R}^n$, $\mathbb{R}^k$ and $\mathbb{R}^m$, where appropriate.

i. Two out of the three subspaces that you defined in (h) are subspaces of the same Euclidean space. Which of the two ranges live in which same Euclidean space?

j. Use your definitions to prove that $range(T_2 \circ T_1)$ is a subset of $range(T_2)$, that is:

$$range(T_2 \circ T_1) \subseteq range(T_2).$$

Hint: This means that you must show that every member $\vec{w}$ of $range(T_2 \circ T_1)$ is also a member of $range(T_2)$.

k. Use part (j) to prove that if $T_2 \circ T_1$ is onto, then T_2 is also onto. Do you notice the difference with part (e)?

l. Write down the contrapositive of the statement in (k).

24. Prove that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is a ***one-to-one*** linear transformation and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a set of ***linearly independent*** vectors from $\mathbb{R}^n$, then $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}$ is a set of ***linearly independent*** vectors from $\mathbb{R}^m$.

    Hint: Begin the proof by considering the dependence test equation:

    $$c_1 T(\vec{v}_1) + c_2 T(\vec{v}_2) + \cdots + c_k T(\vec{v}_k) = \vec{\mathbf{0}}_m.$$

    Rewrite the left side using the linearity properties of T and use the Kernel Test for Injectivity.

25. Use the previous Exercise to prove that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is ***any basis*** for $\mathbb{R}^n$, then $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_n)\}$ is a **basis** for $range(T)$.

26. Let $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$.

    a. What is the ambient space of $range(T_1)$?

    b. What is the ambient space of $ker(T_2)$?

    c. Prove that $T_2 \circ T_1 = Z_{n,m}$ ***if and only if*** $range(T_1) \subseteq ker(T_2)$.

       Recall that $Z_{n,m} : \mathbb{R}^n \to \mathbb{R}^m$ is the zero transformation, where $Z_{n,m}(\vec{v}) = \vec{\mathbf{0}}_m$ for all $\vec{v} \in \mathbb{R}^n$. Notice that by (a) and (b), both subspaces $range(T_1)$ and $ker(T_2)$ are in $\mathbb{R}^k$, so requiring one to be a subset of the other is a possibility.

    d. Suppose that $m = n$. State and prove an analogous statement for $T_1 \circ T_2$.

27. We know that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is automatically ***not*** one-to-one if $n > m$. Thus, suppose that $n \leq m$. Prove that $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** ***if and only if*** the rref of $[T]$ has the form:

    $$\begin{bmatrix} \boldsymbol{I}_n \\ \mathbf{0}_{m-n,n} \end{bmatrix}.$$

    This means that we can divide $[T]$ into two parts: the first n rows will contain $\boldsymbol{I}_n$, and the bottom $m - n$ rows will contain all zeroes. In the case that $n = m$, T will be one-to-one if and only if the rref of $[T]$ is $\boldsymbol{I}_n$.

28. We know that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is automatically ***not*** onto if $n < m$. Thus, suppose that $n \geq m$. Prove that $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***onto*** ***if and only if*** the rref of $[T]$ does not contain a row of zeroes.

29. Prove the last Theorem in this Section: Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Then:

    a. If $m < n$: prove that T is full-rank ***if only only if*** T is ***onto***.

    b. If $m = n$: prove that T is full-rank ***if and only if*** T is ***both one-to-one and onto***.

    c. If $m > n$: prove that T is full-rank ***if and only if*** T is ***one-to-one***.

30. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation. Prove the following statements:

    a. If $n \leq m$, then: T is ***one-to-one*** ***if and only if*** for ***any*** basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$, the image set $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_n)\}$ is linearly independent. Hint: think of $ker(T)$.

b. If $n \geq m$, then: T is ***onto*** ***if and only if*** there exists a linearly independent subset $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ from $\mathbb{R}^n$, such that the image set $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_m)\}$ is linearly independent. Note that there are only m vectors in these sets. Hint: T is onto ***if and only if*** $rank(T) = m$.

c. Bonus: show that (a) is still true if the phrase "for any basis" is replaced with "for at least one basis."

31. Suppose that $T : \mathbb{R}^3 \to \mathbb{R}^5$ is a linear transformation. Notice that we can automatically conclude that T is not onto, but T ***could*** be one-to-one. However, suppose we were also told that:

$$T(\langle 1,-2, 1\rangle) = \langle 2,-3, 4,-1, 7\rangle,$$

$$T(\langle 0,-1, 3\rangle) = \langle -3, 2,-1, 4, 2\rangle, \text{ and}$$

$$T(\langle 0,-2, 5\rangle) = \langle 5,-6, 7,-4, 10\rangle.$$

Show that T is ***not*** one-to-one. Hint: find a non-zero vector in $ker(T)$.

32. ***True or False:*** Determine whether each statement is true or false, and briefly explain your answer by either applying a Theorem or providing a counterexample or a convincing argument.

a. A linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^3$ cannot be one-to-one

b. A linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^3$ cannot be onto.

c. A linear transformation $T : \mathbb{R}^5 \to \mathbb{R}^3$ could be neither one-to-one nor onto.

d. A linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^5$ must be one-to-one.

e. A linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^5$ must be onto.

f. A linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^5$ could be neither one-to-one nor onto.

g. An operator $T : \mathbb{R}^5 \to \mathbb{R}^5$ which is not one-to-one is also not onto.

h. An operator $T : \mathbb{R}^5 \to \mathbb{R}^5$ which is not onto is also not one-to-one.

i. An operator $T : \mathbb{R}^5 \to \mathbb{R}^5$ could be onto but not one-to-one.

j. If $T : \mathbb{R}^3 \to \mathbb{R}^5$ is a full-rank linear transformation, then T is one-to-one.

k. If $T : \mathbb{R}^5 \to \mathbb{R}^3$ is a full-rank linear transformation, then T is one-to-one.

l. If $T : \mathbb{R}^3 \to \mathbb{R}^5$ is a full-rank linear transformation, then T is onto.

m. If $T : \mathbb{R}^5 \to \mathbb{R}^3$ is a full-rank linear transformation, then T is onto.

n. If $T : \mathbb{R}^5 \to \mathbb{R}^5$ is a full-rank operator, then T is both one-to-one and onto.

2.6 Invertible Operators and Matrices

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a ***function***, and so we can attack the problem of determining if T can be ***inverted***. In Algebra, we first require that a function:

$$f : D \to R,$$

with domain D and range R, is ***one-to-one*** on D before we find its inverse. If this is indeed the case, then we can construct: $f^{-1} : R \to D$, where we ***reverse*** the roles of the domain and range, with the property that:

$$f^{-1}(y) = x \textit{ if and only if } f(x) = y.$$

We know that ***at least one*** such x exists because R is the ***range*** of f, and so f is ***onto*** R. However, we also know that ***at most one*** such x exists, because f is ***one-to-one***. Hence, there is ***exactly one*** such x for every y. Thus, f^{-1} is a ***function***. In this case, f^{-1} and f also possess the ***cancellation properties:***

$$f^{-1}(f(x)) = f^{-1}(y) = x \text{ for all } x \in D, \textit{ and } f(f^{-1}(y)) = f(x) = y \text{ for all } y \in R.$$

Going back to our linear transformation, it is therefore only natural that we first require that $T : \mathbb{R}^n \to \mathbb{R}^m$ be ***one-to-one*** before we can even think of inverting T. For now, we also want the inverse transformation T^{-1} to be defined on the ***entire codomain***, and so we require T to be ***onto*** (we will update this in Chapter 4). We will therefore agree on the following:

Definition: We say that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***invertible*** if T is both ***one-to-one*** and ***onto***.
We also say equivalently that T is ***bijective***, T is a ***bijection*** or T is an ***isomorphism*.**

This definition consequently requires that the domain and codomain of an invertible transformation be the ***same*** Euclidean space, that is, T must first be an ***operator:***

Theorem: If $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***invertible***, then $n = m$.

Proof: This is another classic example of Proof by Contradiction. Suppose that T is invertible, that is, it is ***both*** one-to-one and onto. Let us assume the opposite of the conclusion, that is, $n \neq m$. But if $n > m$, then T ***cannot*** be one-to-one, and if $n < m$, then T ***cannot*** be onto, both yielding contradictions. Thus, we must have $n = m$. ■

Because of this Theorem, we only have to ask if T is invertible when T is an operator, that is:

$$T : \mathbb{R}^n \to \mathbb{R}^n,$$

for some Euclidean space $\mathbb{R}^n$. Let us rewrite our introduction above in the language and notation of Linear Algebra: If $T : \mathbb{R}^n \to \mathbb{R}^n$ is both one-to-one and onto, then, we can define an ***inverse function:*** $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, with the property that:

$$\text{for every } \vec{w} \in \mathbb{R}^n\text{: } T^{-1}(\vec{w}) = \vec{v} \textit{ if and only if } T(\vec{v}) = \vec{w}.$$

We know that ***at least one*** such $\vec{v}$ exists because T is onto, and ***at most one*** such $\vec{v}$ exists because T is one-to-one. Thus, there is ***exactly one*** such $\vec{v}$ that will satisfy the equation above, and so T^{-1} is indeed a ***function***.

Let us not forget, though, that $T : \mathbb{R}^n \to \mathbb{R}^n$ is not just any ordinary function, but is rather an ***operator***. In other words, it is also a ***linear transformation:*** it is ***additive*** and ***homogeneous***. It would be very nice if T^{-1} were ***also*** an ***operator***. In fact, it turns out that not only is this true, it also gives us an ***equivalent*** condition for an operator to be ***invertible:***

Theorem: A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is ***invertible*** ***if and only if*** we can find another ***unique*** linear operator, $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, the ***inverse operator*** for T, such that if $\vec{v} \in \mathbb{R}^n$ and $T(\vec{v}) = \vec{w}$, then we define:

$$T^{-1}(\vec{w}) = \vec{v},$$

and thus:

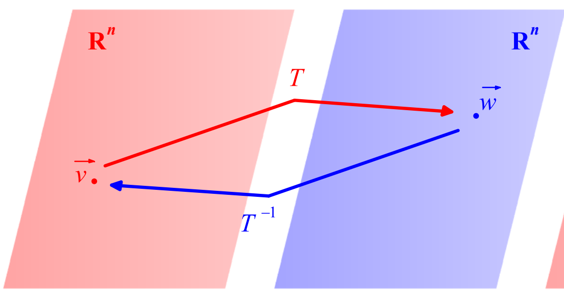
$$(T^{-1} \circ T)(\vec{v}) = \vec{v} \quad \text{and} \quad (T \circ T^{-1})(\vec{w}) = \vec{w}.$$

In other words:

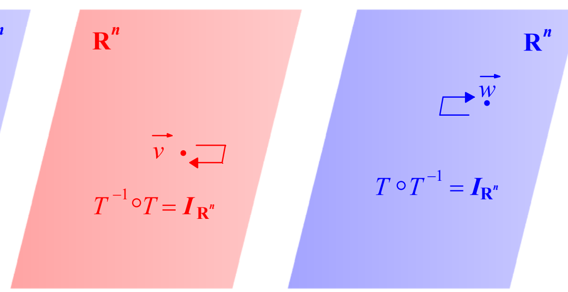
$$T^{-1} \circ T = T \circ T^{-1} = \boldsymbol{I}_{\mathbb{R}^n},$$

the ***identity operator*** on $\mathbb{R}^n$.
Furthermore, if T is invertible, then T^{-1} is also invertible, and $(T^{-1})^{-1} = T$. Thus, we can say that T and T^{-1} are ***inverses of each other***.

We can visualize these equations through the following diagrams:



The Composition of T with T^{-1}



$T^{-1} \circ T = \boldsymbol{I}_{\mathbb{R}^n} = T \circ T^{-1}$

Proof: ($\Rightarrow$) The forward direction is our introduction above: Suppose T is invertible, meaning, T is both one-to-one and onto. Let us explicitly construct the inverse operator T^{-1}. Let $\vec{w} \in \mathbb{R}^n$. We want to define $T^{-1}(\vec{w})$. Since T is onto, there is ***at least one*** vector $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{w}$. But since T is also one-to-one, there is ***at most one*** such $\vec{v}$. In other words, there is ***exactly one*** $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{w}$. Thus, let $T^{-1}(\vec{w}) = \vec{v}$. We immediately get that:

$$T(T^{-1}(\vec{w})) = T(\vec{v}) = \vec{w} = \boldsymbol{I}_{\mathbb{R}^n}(\vec{w}), \quad \text{and}$$

$$T^{-1}(T(\vec{v})) = T^{-1}(\vec{w}) = \vec{v} = \boldsymbol{I}_{\mathbb{R}^n}(\vec{v}).$$

Thus, $T \circ T^{-1}$ and $T^{-1} \circ T$ are both the identity operators on $\mathbb{R}^n$. Notice from the construction above that there was ***exactly one*** possible choice for $T^{-1}(\vec{w})$ in order to make both compositions the identity operator, and therefore as a function, T^{-1} is ***unique***.

We also need to show that T^{-1} is itself a ***linear transformation***, that is, it possesses the properties of ***additivity*** and ***homogeneity***:

$$T^{-1}(\vec{w}_1 + \vec{w}_2) = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2), \quad \text{and}$$
$$T^{-1}(k \cdot \vec{w}_1) = k \cdot T^{-1}(\vec{w}_1),$$

for all $\vec{w}_1, \vec{w}_2 \in \mathbb{R}^n$ and $k \in \mathbb{R}$. By construction, we have $T^{-1}(\vec{w}_1) = \vec{v}_1$ and $T^{-1}(\vec{w}_2) = \vec{v}_2$, where $T(\vec{v}_1) = \vec{w}_1$ and $T(\vec{v}_2) = \vec{w}_2$. Thus:

$$T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2) = \vec{w}_1 + \vec{w}_2, \quad \text{and so:}$$
$$T^{-1}(\vec{w}_1 + \vec{w}_2) = \vec{v}_1 + \vec{v}_2 = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2).$$

Similarly, $T(k \cdot \vec{v}_1) = k \cdot T(\vec{v}_1) = k \cdot \vec{w}_1$, and so $T^{-1}(k \cdot \vec{w}_1) = k \cdot \vec{v}_1 = k \cdot T^{-1}(\vec{w}_1)$.

($\Leftarrow$) Conversely, suppose now we can construct another linear operator, $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, such that $T \circ T^{-1}$ and $T^{-1} \circ T$ are both the identity operators on $\mathbb{R}^n$. We must show that T is invertible, in other words, T is both ***one-to-one*** and ***onto***.

Suppose that $T(\vec{v}_1) = T(\vec{v}_2)$. Then $T^{-1}(T(\vec{v}_1)) = T^{-1}(T(\vec{v}_2))$, and since $T^{-1} \circ T$ is the identity operator on $\mathbb{R}^n$, we get $\vec{v}_1 = \vec{v}_2$. Thus, T is ***one-to-one***.

Now, suppose $\vec{w} \in \mathbb{R}^n$. Let us find another vector $\vec{v} \in \mathbb{R}^n$ such that $T(\vec{v}) = \vec{w}$. A natural choice (in fact, the only choice), would be $\vec{v} = T^{-1}(\vec{w})$, because we shall be rewarded with:

$$T(\vec{v}) = T(T^{-1}(\vec{w})) = T \circ T^{-1}(\vec{w}) = \vec{w},$$

since $T \circ T^{-1}$ is the identity operator on $\mathbb{R}^n$. Thus T is ***onto***. Notice that we used the fact that ***both compositions*** $T \circ T^{-1}$ and $T^{-1} \circ T$ are the identity operators on $\mathbb{R}^n$. This proves the converse.

Before we go to the last part of the Theorem, let us rewrite the first part of the Theorem using more ***neutral*** notation:

Theorem: (restatement) A linear operator $T_1 : \mathbb{R}^n \to \mathbb{R}^n$ is ***invertible*** ***if and only if*** we can find another linear operator, $T_2 : \mathbb{R}^n \to \mathbb{R}^n$, the ***inverse operator*** for T_1, such that if $\vec{v} \in \mathbb{R}^n$ and $\vec{w} = T_1(\vec{v})$, then $T_2(\vec{w}) = \vec{v}$, and thus:

$$(T_2 \circ T_1)(\vec{v}) = \vec{v} \quad \text{and} \quad (T_1 \circ T_2)(\vec{w}) = \vec{w},$$

for all $\vec{v}, \vec{w} \in \mathbb{R}^n$, that is, $T_2 \circ T_1 = \boldsymbol{I}_{\mathbb{R}^n} = T_1 \circ T_2$.

Now, suppose T is invertible with inverse T^{-1}. We want to show that T^{-1} is also invertible. We already know that these two operators satisfy the equation:

$$T^{-1} \circ T = \boldsymbol{I}_{\mathbb{R}^n} = T \circ T^{-1}.$$

But now, let us interpret this from the point of view of T^{-1}. In the notation above, let:

$$T_1 = T^{-1} \quad \text{and} \quad T_2 = T.$$

Thus, the compositions can be written as:

$$T^{-1} \circ T = T_1 \circ T_2 \quad \text{and} \quad T \circ T^{-1} = T_2 \circ T_1.$$

But we know that both of these compositions are $\boldsymbol{I}_{\mathbb{R}^n}$, and so by the rewritten version of our Theorem, $T_1 = T^{-1}$ must be invertible, with inverse $T_2 = T$. This completes the proof.■

Invertible Matrices

We defined parallel operations of addition and subtraction for matrices and for linear transformations, and related the composition of two linear transformations to the product of their matrices. In the same way, we will now define the concept of an ***invertible matrix*** using the concept of invertible linear operators. Since the minimum requirement for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ to be invertible is that $n = m$, by the same token, we will require a matrix to be ***square*** before we determine whether or not it is invertible:

> ***Definition:*** An $n \times n$ matrix A is ***invertible*** if the linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ corresponding to $A = [T]$ is an ***invertible operator***. In other words, the operator defined by:
>
> $$T(\vec{v}) = A\vec{v},$$
>
> for all $\vec{v} \in \mathbb{R}^n$, is an invertible operator on $\mathbb{R}^n$.

Now, if we know that an operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, then its composition with its inverse operator T^{-1} must be the identity operator, that is:

$$T^{-1} \circ T = \boldsymbol{I}_{\mathbb{R}^n} = T \circ T^{-1}.$$

However, we know that when a composition $T_2 \circ T_1$ is defined, then $[T_2 \circ T_1] = [T_2][T_1]$, that is the product of the two individual matrices. Applying this idea above, we get:

$$[T^{-1} \circ T] = [\boldsymbol{I}_{\mathbb{R}^n}] = [T \circ T^{-1}], \text{ in other words:}$$
$$[T^{-1}][T] = \boldsymbol{I}_n = [T][T^{-1}].$$

But recall that from our definition above, $[T] = A$. Since $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, its matrix $[T^{-1}]$ is likewise an $n \times n$ matrix. Let us simply call it B. Thus, we can rewrite the last equation as: $AB = \boldsymbol{I}_n = BA$.

We will show below that this construction is ***reversible***, and so we have the following:

> ***Theorem/Definition:*** An $n \times n$ matrix A is ***invertible*** ***if and only if*** we can find another $n \times n$ matrix B such that: $AB = \boldsymbol{I}_n = BA$.
>
> We call B the ***inverse matrix*** of A, and denote it by A^{-1}.
>
> If A is invertible, then the inverse matrix A^{-1} is likewise invertible, and: $(A^{-1})^{-1} = A$.
>
> In other words, $B^{-1} = A$. Thus, we can say that A and A^{-1} are ***inverses of each other***.

Proof: $(\Rightarrow)$ The forward direction is our discussion above: If A is invertible, then the operator $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(\vec{v}) = A\vec{v}$, for all $\vec{v} \in \mathbb{R}^n$, is an invertible operator. Thus there exists an inverse operator $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ such that $T^{-1} \circ T = \boldsymbol{I}_{\mathbb{R}^n} = T \circ T^{-1}$. If we now call $[T^{-1}]$ the matrix B, then we get $AB = \boldsymbol{I}_n = BA$ as above.

$(\Leftarrow)$ Now for the converse: Suppose we are given that we can find another $n \times n$ matrix B such that $AB = \boldsymbol{I}_n = BA$. We can define another linear operator $T_2 : \mathbb{R}^n \to \mathbb{R}^n$ via:

$$T_2(\vec{v}) = B\vec{v},$$

for all $\vec{v} \in \mathbb{R}^n$. But then, we get:

$$(T \circ T_2)(\vec{v}) = T(T_2(\vec{v})) = T(B\vec{v}) = A(B\vec{v}) = (AB)\vec{v} = \boldsymbol{I}_n\vec{v} = \vec{v} = \boldsymbol{I}_{\mathbb{R}^n}(\vec{v}),$$

for all $\vec{v} \in \mathbb{R}^n$, and similarly:

$$(T_2 \circ T)(\vec{v}) = T_2(T(\vec{v})) = T_2(A\vec{v}) = B(A\vec{v}) = (BA)\vec{v} = \boldsymbol{I_n}\vec{v} = \vec{v} = \boldsymbol{I}_{\mathbb{R}^n}(\vec{v}),$$

for all $\vec{v} \in \mathbb{R}^n$. Notice that we used the Associative Property of Matrix Multiplication above. Thus, both $T \circ T_2$ and $T_2 \circ T$ are the identity operator on $\mathbb{R}^n$. Therefore, T is an invertible operator with inverse operator T_2. Thus, A is invertible.■

Notice that in the statement of the Theorem above, we called B ***the*** inverse of A. This presumes that B, that is, A^{-1}, is ***unique***, and indeed it is:

Theorem — The Uniqueness of the Matrix Inverse:
If an $n \times n$ matrix A is ***invertible***, then its inverse matrix B is ***unique***. This means that if B and C both satisfy the equations: $AB = \boldsymbol{I_n} = BA$ and $AC = \boldsymbol{I_n} = CA$, then $B = C$.

Proof: If we multiply both sides of the given equation $AC = \boldsymbol{I_n}$ on the left by B, then we get:

$$B(AC) = B\boldsymbol{I_n}, \text{ and thus, } (BA)C = B.$$

However, since we are also given that $BA = \boldsymbol{I_n}$, then we get: $\boldsymbol{I_n}C = B$, which gives us $C = B$.■

The general process of finding the inverse of an invertible matrix of any dimension will be seen in the next Section, but for now, we can easily do it for invertible 2×2 matrices:

Theorem: Suppose that: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then: A is ***invertible*** ***if and only if*** $ad - bc \neq 0$, in which case: $A^{-1} = \dfrac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Proof: First, let us prove that A is invertible ***if and only if*** $ad - bc \neq 0$.

We know that if A is invertible, then the operator $T : \mathbb{R}^2 \to \mathbb{R}^2$ with $[T] = A$ must be invertible. In other words, T is both one-to-one and onto. But this means that $nullity(T) = 0$ and $rank(T) = 2$. But this last condition means that $colspace(A) = \mathbb{R}^2$. Thus the Span of the two columns $\langle a, c\rangle$ and $\langle b, d\rangle$ must be all of $\mathbb{R}^2$. But we know from Chapter 1 that this is possible ***if and only if*** $\langle a, c\rangle$ and $\langle b, d\rangle$ are ***not parallel***, and consequently $ad - bc \neq 0$ from Exercise 21 in Section 1.1.

For the converse, we just reverse this argument. If $ad - bc \neq 0$, then $\langle a, c\rangle$ and $\langle b, d\rangle$ are not parallel and thus their Span must be all of $\mathbb{R}^2$. Thus $rank(T) = 2$, which means that T is onto. But by the Dimension Theorem, $nullity(T) = 0$. Thus, T is both one-to-one and onto, and thus T is invertible, hence A is invertible.

Now, suppose we already know that $ad - bc \neq 0$. Thus, this number has a ***reciprocal***, and we can assemble the matrix:

$$B = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

as prescribed in the Theorem. But we can check that:

$$AB = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} ad-bc & bd-bd \\ -ac+ac & -bc+ad \end{bmatrix} = \boldsymbol{I_2}, \text{ and}$$

$$BA = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \frac{1}{ad-bc}\begin{bmatrix} ad-bc & bd-bd \\ -ac+ac & -bc+ad \end{bmatrix} = \boldsymbol{I_2}.$$

By the Uniqueness of the Inverse, we must have $B = A^{-1}$. ■

Example: Suppose $A = \begin{bmatrix} 3 & 7 \\ 2 & 6 \end{bmatrix}$.

Then $ad - bc = 3 \cdot 6 - 2 \cdot 7 = 4$, and thus A is invertible, and:

$$A^{-1} = \frac{1}{4}\begin{bmatrix} 6 & -7 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 3/2 & -7/4 \\ -1/2 & 3/4 \end{bmatrix}.$$

We can verify that:

$$AA^{-1} = \begin{bmatrix} 3 & 7 \\ 2 & 6 \end{bmatrix}\begin{bmatrix} 3/2 & -7/4 \\ -1/2 & 3/4 \end{bmatrix} = \begin{bmatrix} 9/2-7/2 & -21/4+21/4 \\ 6/2-6/2 & -7/2+9/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

and similarly, $A^{-1}A = \boldsymbol{I_2}$. □

Example: Consider the matrix: $A = \begin{bmatrix} 3 & -7 \\ 12 & -28 \end{bmatrix}$.

This time, $ad - bc = 3(-28) - (-7)12 = 0$, and thus A is ***not*** invertible. Notice that the rows of A are scalar multiples of each other, and the columns of A are also scalar multiples of each other, and thus both the row and column spaces of A are just ***lines*** through the origin in $\mathbb{R}^2$, and not all of $\mathbb{R}^2$. □

The Matrix of T^{-1}

Let us take a step back and see what we have so far: we began by defining an ***invertible operator*** $T : \mathbb{R}^n \to \mathbb{R}^n$ to be an operator that is both ***one-to-one*** and ***onto***. Our first Theorem said that T is invertible ***if and only if*** we can find another operator $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, the ***inverse operator***, such that $T^{-1} \circ T = T \circ T^{-1} = \boldsymbol{I}_{\mathbb{R}^n}$. This inverse is ***unique***, in that there is exactly one operator that satisfies these two equations when ***composed*** with T. Next, we defined an $n \times n$ ***matrix*** A to be ***invertible*** if its associated ***operator*** $T : \mathbb{R}^n \to \mathbb{R}^n$ is ***invertible*** as defined above, where $T(\vec{v}) = A\vec{v}$. But we also showed in the next Theorem that A is invertible ***if and only if*** we can find another $n \times n$ matrix B such that $AB = \boldsymbol{I_n} = BA$, and this inverse matrix $B = A^{-1}$ is likewise ***unique***, in that it is the only matrix that satisfies these two equations when ***multiplied*** by A.

It should therefore seem only logical that the matrix of T^{-1} is A^{-1}. But of course, this is true:

> ***Theorem:*** A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible ***if and only if*** $A = [T]$ is an invertible $n \times n$ matrix. If this is the case, then: $[T^{-1}] = A^{-1} = [T]^{-1}$.

Proof: The first sentence is the definition of an invertible matrix:

An $n \times n$ matrix A is ***invertible*** if the linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ corresponding to $A = [T]$ is an ***invertible operator***. In other words, the operator defined by $T(\vec{v}) = A\vec{v}$, for all $\vec{v} \in \mathbb{R}^n$, is an invertible operator on $\mathbb{R}^n$. Thus, we only have to prove that if A and T are both invertible, then:

$$[T^{-1}] = A^{-1} = [T]^{-1}.$$

We know that $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ exists, such that $T \circ T^{-1}$ and $T^{-1} \circ T$ are both the identity operators on $\mathbb{R}^n$. But then:

$$\boldsymbol{I}_{\mathbb{R}^n}(\vec{v}) = \boldsymbol{I}_n\vec{v} = \vec{v} = (T^{-1} \circ T)(\vec{v}) = T^{-1}([T]\vec{v}) = [T^{-1}][T]\vec{v},$$

for all vectors $\vec{v} \in \mathbb{R}^n$. By the Uniqueness of the Standard Matrix:

$$[\boldsymbol{I}_{\mathbb{R}^n}] = [T^{-1}][T], \text{ and so } \boldsymbol{I}_n = [T^{-1}][T].$$

Similarly, we can show that $[T][T^{-1}] = \boldsymbol{I}_n$. These equations tell us that $[T] = A$ is an invertible matrix, and by the Uniqueness of the Inverse, $[T^{-1}] = A^{-1} = [T]^{-1}$. ■

Example: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $T(\langle x, y\rangle) = \langle 3x + 7y, 2x + 6y\rangle$. Its matrix is:

$$[T] = \begin{bmatrix} 3 & 7 \\ 2 & 6 \end{bmatrix},$$

which is an invertible matrix as we saw in a previous Example, and thus T is an invertible operator. We found the inverse of this matrix to be:

$$[T]^{-1} = \begin{bmatrix} 3/2 & -7/4 \\ -1/2 & 3/4 \end{bmatrix} = [T^{-1}].$$

We can thus give an explicit formula for T^{-1}:

$$T^{-1}(\langle x, y\rangle) = \langle 3x/2 - 7y/4, -x/2 + 3y/4\rangle.$$

From this, we can verify that:

$$\begin{aligned}(T^{-1} \circ T)(\langle x, y\rangle) &= T^{-1}(\langle 3x + 7y, 2x + 6y\rangle) \\ &= \langle 3/2(3x + 7y) - 7/4(2x + 6y), -1/2(3x + 7y) + 3/4(2x + 6y)\rangle \\ &= \langle 9x/2 + 21y/2 - 7x/2 - 21y/2, -3x/2 - 7y/2 + 3x/2 + 9y/2\rangle = \langle x, y\rangle,\end{aligned}$$

and thus $T^{-1} \circ T$ is indeed the identity operator on $\mathbb{R}^2$. Similarly, $T \circ T^{-1}$ is also the identity operator on $\mathbb{R}^2$. □

Bonus Example: The Matrix of a ***Reflection*** Across a Plane in $\mathbb{R}^3$:

We do not know yet how to determine if a 3×3 matrix is invertible or not, much less how to find its inverse, if this exists. But let's think about the ***geometry*** of the operator in $\mathbb{R}^3$ that reflects a vector across a plane Π passing through the origin, as we saw in Section 2.2. If we take the reflection of any vector $\vec{v}$ across Π and reflect it ***again*** across Π, then we should get back our original vector $\vec{v}$. Thus we must have:

$$\mathit{refl}_\Pi(\mathit{refl}_\Pi(\vec{v})) = \vec{v} = \boldsymbol{I}_{\mathbb{R}^3}(\vec{v}),$$

or in other words $[\mathit{refl}_\Pi][\mathit{refl}_\Pi] = [\mathit{refl}_\Pi]^2 = \boldsymbol{I}_3$.

This tells us that $[refl_\Pi]$ is an ***invertible*** matrix, and is its ***own inverse***, that is:

$$[refl_\Pi]^{-1} = [refl_\Pi].$$

Let us verify this using $refl_\Pi$ for the plane Π in $\mathbb{R}^3$ with Cartesian equation:

$$3x - 5y + 2z = 0.$$

We found in the last Example of Section 2.2 that:

$$[refl_\Pi] = \begin{bmatrix} \frac{10}{19} & \frac{15}{19} & -\frac{6}{19} \\ \frac{15}{19} & -\frac{6}{19} & \frac{10}{19} \\ -\frac{6}{19} & \frac{10}{19} & \frac{15}{19} \end{bmatrix} = \frac{1}{19}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix}.$$

Thus we can check that:

$$[refl_\Pi][refl_\Pi] = \frac{1}{19}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix} \cdot \frac{1}{19}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix}$$

$$= \frac{1}{361}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix} = \frac{1}{361}\begin{bmatrix} 361 & 0 & 0 \\ 0 & 361 & 0 \\ 0 & 0 & 361 \end{bmatrix} = \boldsymbol{I_3}.$$

The same can be said for a reflection $refl_L$ across a line L. □

2.6 Section Summary

We say that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible ***if and only if*** T is both ***one-to-one*** and ***onto***. We also say that T is ***bijective***, T is a ***bijection*** or T is an ***isomorphism***. This is only possible if $n = m$, that is, T is an ***operator***.

A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible ***if and only if*** we can find another ***unique*** linear operator, $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, the ***inverse operator*** for T, such that if $\vec{v} \in \mathbb{R}^n$ and $T(\vec{v}) = \vec{w}$, then we define $T^{-1}(\vec{w}) = \vec{v}$, and thus $(T^{-1} \circ T)(\vec{v}) = \vec{v}$ and $(T \circ T^{-1})(\vec{w}) = \vec{w}$.

In other words, $T^{-1} \circ T = T \circ T^{-1} = \boldsymbol{I}_{\mathbb{R}^n}$, the identity operator on $\mathbb{R}^n$. Furthermore, if T is invertible, then T^{-1} is also invertible, and $(T^{-1})^{-1} = T$. Thus, T and T^{-1} are ***inverses of each other***.

We define an $n \times n$ matrix A to be invertible if the linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ corresponding to A is an invertible operator. In other words, the operator defined by $T(\vec{v}) = A\vec{v}$, for all $\vec{v} \in \mathbb{R}^n$, is an invertible operator on $\mathbb{R}^n$. Note: by this definition, $[T] = A$.

An $n \times n$ matrix A is ***invertible*** ***if and only if*** we can find another $n \times n$ matrix B such that $AB = \boldsymbol{I_n} = BA$. We call B the ***inverse matrix*** of A, and denote it by A^{-1}. If A is invertible, then the inverse matrix A^{-1} is likewise invertible, and $(A^{-1})^{-1} = A$. In other words, $B^{-1} = A$. Thus, we can say that A and A^{-1} are ***inverses of each other***.

The inverse of an invertible matrix is likewise ***unique***. This means that if B and C are two $n \times n$ matrices such that: $AB = \boldsymbol{I_n} = BA$, and $AC = \boldsymbol{I_n} = CA$, then $B = C$.

Thus, we can speak of ***the*** inverse of A.

Suppose that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Then: A is invertible ***if and only if*** $ad - bc \neq 0$,

in which case: $A^{-1} = \dfrac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible ***if and only if*** $A = [T]$ is an invertible $n \times n$ matrix. If this is the case, then: $[T^{-1}] = A^{-1} = [T]^{-1}$.

2.6 Exercises

For Exercises (1) to (19): Find the inverses of the following matrices, if possible:

1. $\begin{bmatrix} -2 & 0 \\ 0 & 3 \end{bmatrix}$
2. $\begin{bmatrix} 5 & 7 \\ 0 & -4 \end{bmatrix}$
3. $\begin{bmatrix} 0 & 4 \\ 6 & 0 \end{bmatrix}$
4. $\begin{bmatrix} 7 & 9 \\ 3 & 4 \end{bmatrix}$
5. $\begin{bmatrix} 7 & -6 \\ -4 & 3 \end{bmatrix}$
6. $\begin{bmatrix} 2/3 & 0 \\ 0 & -8/3 \end{bmatrix}$
7. $\begin{bmatrix} -5 & -8 \\ 5 & 6 \end{bmatrix}$
8. $\begin{bmatrix} 20 & 36 \\ -15 & -27 \end{bmatrix}$
9. $\begin{bmatrix} 2/3 & 5/6 \\ -7 & 11/2 \end{bmatrix}$
10. $\begin{bmatrix} -4/3 & 11/6 \\ 8 & 9/2 \end{bmatrix}$
11. $\begin{bmatrix} 5/3 & 2/5 \\ -1/6 & 7/4 \end{bmatrix}$
12. $\begin{bmatrix} \sqrt{6} & \sqrt{15} \\ \sqrt{30} & \sqrt{3} \end{bmatrix}$
13. $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$, where $\theta \in \mathbb{R}$. Note: this is the matrix of rot_θ.

    Follow up: explain the geometric significance of its inverse.
14. $\begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}$, where $\theta \in \mathbb{R}$.
15. $\begin{bmatrix} e^{3x} & -e^{4x} \\ 2e^{-2x} & 3e^{-x} \end{bmatrix}$, where $x \in \mathbb{R}$.
16. $\begin{bmatrix} 6^x & 4^x \\ 15^x & -10^x \end{bmatrix}$, where $x \in \mathbb{R}$.
17. $\begin{bmatrix} \cosh(x) & \sinh(x) \\ \sinh(x) & \cosh(x) \end{bmatrix}$, where $\cosh(x) = \dfrac{e^x + e^{-x}}{2}$, $\sinh(x) = \dfrac{e^x - e^{-x}}{2}$, and $x \in \mathbb{R}$.
18. $\begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}$, where $a, b \in \mathbb{R}$ and $a^2 + b^2 = 1$.
19. $\begin{bmatrix} a^2 - b^2 & 2ab \\ 2ab & b^2 - a^2 \end{bmatrix}$, where $a, b \in \mathbb{R}$ and $a^2 + b^2 = 1$.

20. The matrices in the last two Exercises were the standard matrices of the operators $[proj_L]$ and $[refl_L]$, respectively, where L is a line through the origin in $\mathbb{R}^2$ with ***unit*** direction vector $\langle a, b\rangle$. See Exercise 25 in Section 2.2. Give a geometric argument as to why one of these matrices is invertible and the other matrix is not invertible. Explain also the geometric significance of the inverse of the invertible matrix.

For Exercises (21) to (24): For the given linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$: (a) find $[T]$; (b) find $[T^{-1}]$, if T is invertible, and if so: (c) explicitly give a formula for $T^{-1}(\langle x, y\rangle)$, and finally, (d) check that $T^{-1}(T(\langle x, y\rangle)) = \langle x, y\rangle$.

21. $T(\langle x, y\rangle) = \langle 3x - 7y, -4x + 9y\rangle$

22. $T(\langle x, y\rangle) = \langle 2x - 5y, 6x - 15y\rangle$

23. $T(\langle x, y\rangle) = \langle 3x + 5y, 5x + 9y\rangle$

24. $T(\langle x, y\rangle) = \left\langle \frac{2}{3}x + \frac{5}{3}y, \frac{4}{3}x - \frac{1}{3}y \right\rangle$

25. Prove that an operator T is invertible ***if and only if*** either T is one-to-one ***or*** T is onto.

26. Let's play:

a. Compute: $\begin{bmatrix} 3 \\ 5 \end{bmatrix}\begin{bmatrix} 2 & -7 \end{bmatrix}$. Is the resulting matrix invertible?

b. Repeat part (a) by making your own 2×1 and 1×2 matrices and multiplying them together. Is the resulting matrix invertible? Try a third pair.

c. Let's think about bigger matrices. Suppose that we know that a 9×9 matrix A factors as a product: $A = BC$, where B is a 9×5 matrix and C is a 5×9 matrix. Prove that A is ***not*** invertible. Hint: show that the linear transformation corresponding to A ***cannot*** be one-to-one. Which matrix is responsible for this, B or C?

d. Prove in general that if an $n \times n$ matrix A factors as $A = BC$, where B is an $n \times m$ matrix and C is an $m \times n$ matrix, where $n > m$, then A is ***not*** invertible.

e. Compute: $\begin{bmatrix} 5 & -2 & 1 \\ 0 & 3 & 7 \end{bmatrix}\begin{bmatrix} 4 & -3 \\ -8 & 9 \\ -5 & 6 \end{bmatrix}$. Is the resulting matrix invertible?

f. Compute: $\begin{bmatrix} 5 & -2 & 1 \\ -15 & 6 & -3 \end{bmatrix}\begin{bmatrix} 4 & 16 \\ -8 & -32 \\ -5 & -20 \end{bmatrix}$. Is the resulting matrix invertible?

g. Based on the last two parts, can we determine immediately that A is invertible or not invertible if $A = BC$, where B is an $n \times m$ matrix and C is an $m \times n$ matrix, and $n < m$?

27. Suppose that Π is the plane in $\mathbb{R}^3$ with Cartesian equation $4x + 2y - 3z = 0$.

a. Verify that the matrix of $refl_\Pi$, that you found in Exercise 13 of Section 2.2, also satisfies: $[refl_\Pi][refl_\Pi] = \boldsymbol{I}_3$.

b. Consider now the operator $proj_\Pi$, the projection onto Π. Show that $ker(proj_\Pi) = L$, the normal line $Span(\{\langle 4, 2, -3\rangle\})$, that is the orthogonal complement of Π. Does this mean that $[proj_\Pi]$ is invertible or not invertible? Why?

2.7 Finding the Inverse of a Matrix

Our goal in this Section is to be able to construct the inverse of an invertible square matrix that is 3×3 or bigger, when it is possible to do so. In so doing, we will also be able to find the matrix of the inverse of an invertible operator T.

We begin by going back to the elementary matrices that we saw in Section 2.1:

Multiplicative Properties of Elementary Matrices

Elementary matrices essentially ***encode*** the elementary row operation that was used to produce the matrix from $\boldsymbol{I_n}$:

> ***Theorem:*** If E is an ***elementary*** $n \times n$ matrix and A is any $n \times m$ matrix, then the matrix product EA can be computed by performing the ***same elementary row operation*** on A that was used to produce E from $\boldsymbol{I_n}$.

The proof will be left as an Exercise. Notice that A and EA are both $n \times m$ matrices.

Example: Suppose that $A = \begin{bmatrix} 5 & 7 & -2 & 3 \\ 4 & 1 & 8 & -5 \\ 2 & -3 & 9 & 6 \end{bmatrix}$,

and E_1, E_2 and E_3 are the elementary matrices:

$$E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \text{ and } E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix},$$

that we saw in the Example in Section 2.1. Then:

$$E_1A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 5 & 7 & -2 & 3 \\ 4 & 1 & 8 & -5 \\ 2 & -3 & 9 & 6 \end{bmatrix} = \begin{bmatrix} 5 & 7 & -2 & 3 \\ 12 & 3 & 24 & -15 \\ 2 & -3 & 9 & 6 \end{bmatrix},$$

$$E_2A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 5 & 7 & -2 & 3 \\ 4 & 1 & 8 & -5 \\ 2 & -3 & 9 & 6 \end{bmatrix} = \begin{bmatrix} 2 & -3 & 9 & 6 \\ 4 & 1 & 8 & -5 \\ 5 & 7 & -2 & 3 \end{bmatrix}, \text{ and}$$

$$E_3A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix}\begin{bmatrix} 5 & 7 & -2 & 3 \\ 4 & 1 & 8 & -5 \\ 2 & -3 & 9 & 6 \end{bmatrix} = \begin{bmatrix} 5 & 7 & -2 & 3 \\ 4 & 1 & 8 & -5 \\ 22 & 2 & 49 & -19 \end{bmatrix}.$$

We can see that E_1A is the same as A, except the second row is 3 times that of A, E_2A is the same as A, with the 1st and 3rd rows exchanged, and we get E_3A by adding 5 times row 2 of A to row 3 of A (with the first two rows ***staying the same***).□

Now let us see the primary reason why elementary matrices are important in studying invertible matrices:

> ***Theorem:*** Elementary matrices are ***invertible***, and the inverse of an elementary matrix is another elementary matrix of the ***same type***.

Again, the proof is left as an Exercise, but the following examples will serve as hints.

Examples: We can easily check by direct multiplication that the corresponding inverses of our three sample elementary matrices are as indicated below:

$$\text{For } E_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad E_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/3 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

$$\text{For } E_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}, \quad E_2^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

$$\text{For } E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 5 & 1 \end{bmatrix}, \quad E_3^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 1 \end{bmatrix}. \square$$

A Preliminary Test for Invertibility

In the previous Section, we saw that the 2×2 matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is invertible ***if and only if*** $ad - bc \neq 0$. We would like to have a similar test to know when a larger square matrix is invertible. Our preliminary answer is given below, although we will incorporate it later into the algorithm to find the inverse of our matrix:

> ***Theorem:*** Let A be an $n \times n$ matrix. Then A is ***invertible*** ***if and only if*** the rref of A is $\boldsymbol{I_n}$.

Proof: ($\Rightarrow$) Suppose that A is invertible. Then the operator $T : \mathbb{R}^n \to \mathbb{R}^n$ corresponding to A must be an invertible operator. This means that T is both one-to-one and onto. However, in Chapter 1, we saw that the rref of a square matrix is either $\boldsymbol{I_n}$, or it contains at least one free variable and one row of zeroes. Since T is one-to-one and onto, the second case is impossible. Thus, the rref of A must be $\boldsymbol{I_n}$.

($\Leftarrow$) Now, suppose the rref of A is $\boldsymbol{I_n}$.

This implies that $nullspace(A) = \left\{ \vec{\mathbf{0}}_n \right\}$ and $colspace(A) = \mathbb{R}^{\boldsymbol{n}}$. Thus the operator $T : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{n}}$ corresponding to A is both one-to-one and onto, and so A is invertible.■

Now we can accomplish the main objective of this section:

A Method to Find A^{-1}

The Gauss-Jordan Algorithm will help us find the inverse of an invertible square matrix:

> ***Theorem:*** Let A be an $n \times n$ matrix. If we construct the $n \times 2n$ augmented matrix:
>
> $$\left[\, A \,|\, \boldsymbol{I_n} \,\right],$$
>
> then A is ***invertible*** ***if and only if*** the rref of this augmented matrix contains $\boldsymbol{I_n}$ in the first n columns. If this is the case, then A^{-1} will be found in the last n columns.
>
> In other words, the rref of $\left[\, A \,|\, \boldsymbol{I_n} \,\right]$ is $\left[\, \boldsymbol{I_n} \,|\, A^{-1} \,\right]$.

Proof: We already saw from the previous Theorem that A is invertible ***if and only if*** the rref of A is $\boldsymbol{I_n}$. Thus, suppose we form the augmented matrix $\left[\, A \,|\, \boldsymbol{I_n} \,\right]$ as prescribed above, and find the rref of this augmented matrix. Since the rref of A itself is $\boldsymbol{I_n}$, the first half of the Gauss-Jordan algorithm ***stops*** when we get the final leading 1 at row n, column n of this augmented matrix. Thus, we do not have to worry about finding a leading 1 to the ***right*** of column n. The second half of the Gauss-Jordan algorithm completes the process of producing $\boldsymbol{I_n}$ in the first n columns. Once this is done, the rref of $\left[\, A \,|\, \boldsymbol{I_n} \,\right]$ will have the form: $\left[\, \boldsymbol{I_n} \,|\, \vec{x}_1 \,|\, \vec{x}_2 \,|\, \ldots \,|\, \vec{x}_n \,\right]$.

We will now show that: $B = \left[\, \vec{x}_1 \,|\, \vec{x}_2 \,|\, \ldots \,|\, \vec{x}_n \,\right] = A^{-1}$.

Recall that our original augmented matrix is: $\left[\, A \,|\, \boldsymbol{I_n} \,\right] = \left[\, A \,|\, \vec{e}_1 \,|\, \vec{e}_2 \,|\, \ldots \,|\, \vec{e}_n \,\right]$.

The key to completing the proof is to remember that the Gauss-Jordan algorithm also allows us to solve ***multiple*** systems at the same time, as long as we are dealing with the same coefficient matrix A. The rref above tells us that if we began only with $\left[\, A \,|\, \vec{e}_1 \,\right]$, the final rref will be $\left[\, \boldsymbol{I_n} \,|\, \vec{x}_1 \,\right]$, and thus:

$$A\vec{x}_1 = \vec{e}_1.$$

Similarly, if we began only with $\left[\, A \,|\, \vec{e}_2 \,\right]$, the final rref will be $\left[\, \boldsymbol{I_n} \,|\, \vec{x}_2 \,\right]$, and so on. By finding the rref of $\left[\, A \,|\, \boldsymbol{I_n} \,\right]$ in one fell swoop, we solve n systems of equations, with right sides $\vec{e}_1$ through $\vec{e}_n$. Thus, we also get:

$$A\vec{x}_2 = \vec{e}_2,\ A\vec{x}_3 = \vec{e}_3,\ \ldots,\ A\vec{x}_n = \vec{e}_n.$$

Putting it all together, we get:

$$AB = A\left[\, \vec{x}_1 \,|\, \vec{x}_2 \,|\, \ldots \,|\, \vec{x}_n \,\right] = \left[\, A\vec{x}_1 \,|\, A\vec{x}_2 \,|\, \ldots \,|\, A\vec{x}_n \,\right] = \left[\, \vec{e}_1 \,|\, \vec{e}_2 \,|\, \ldots \,|\, \vec{e}_n \,\right] = \boldsymbol{I_n}.$$

We will see in Section 2.8 that if $AB = \boldsymbol{I_n}$, then $BA = \boldsymbol{I_n}$ as well, and thus $B = A^{-1}$.■

Example: Suppose that $A = \begin{bmatrix} 3 & 2 & 5 \\ 2 & 1 & 4 \\ 4 & 2 & 7 \end{bmatrix}$.

Let us see if we can find A^{-1}. For reasons we will see later, we will actually list out all the elementary row operations that we will use to perform the reduction:

$$\left[\begin{array}{ccc|ccc} 3 & 2 & 5 & 1 & 0 & 0 \\ 2 & 1 & 4 & 0 & 1 & 0 \\ 4 & 2 & 7 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_1 \to R_1 - R_2} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & -1 & 0 \\ 2 & 1 & 4 & 0 & 1 & 0 \\ 4 & 2 & 7 & 0 & 0 & 1 \end{array}\right]$$

$$\xrightarrow{R_2 \to R_2 - 2R_1} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & -1 & 0 \\ 0 & -1 & 2 & -2 & 3 & 0 \\ 4 & 2 & 7 & 0 & 0 & 1 \end{array}\right] \xrightarrow{R_3 \to R_3 - 4R_1} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & -1 & 0 \\ 0 & -1 & 2 & -2 & 3 & 0 \\ 0 & -2 & 3 & -4 & 4 & 1 \end{array}\right]$$

$$\xrightarrow{R_2 \to -R_2} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & -2 & 2 & -3 & 0 \\ 0 & -2 & 3 & -4 & 4 & 1 \end{array}\right] \xrightarrow{R_3 \to R_3 + 2R_2} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & -2 & 2 & -3 & 0 \\ 0 & 0 & -1 & 0 & -2 & 1 \end{array}\right]$$

$$\xrightarrow{R_3 \to -R_3} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & -2 & 2 & -3 & 0 \\ 0 & 0 & 1 & 0 & 2 & -1 \end{array}\right] \xrightarrow{R_2 \to R_2 + 2R_3} \left[\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & 2 & 1 & -2 \\ 0 & 0 & 1 & 0 & 2 & -1 \end{array}\right]$$

$$\xrightarrow{R_1 \to R_1 - R_3} \left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & -3 & 1 \\ 0 & 1 & 0 & 2 & 1 & -2 \\ 0 & 0 & 1 & 0 & 2 & -1 \end{array}\right] \xrightarrow{R_1 \to R_1 - R_2} \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & -4 & 3 \\ 0 & 1 & 0 & 2 & 1 & -2 \\ 0 & 0 & 1 & 0 & 2 & -1 \end{array}\right].$$

Thus A is invertible, and $A^{-1} = \begin{bmatrix} -1 & -4 & 3 \\ 2 & 1 & -2 \\ 0 & 2 & -1 \end{bmatrix}$. □

Factoring Invertible Matrices

In the same way that complex molecules can be broken down into constituent atoms, invertible matrices can be ***factored*** into elementary matrices:

> ***Theorem:*** An $n \times n$ matrix A is ***invertible*** ***if and only if*** it can be expressed as a ***product*** of ***elementary*** matrices. If this is the case, then more precisely, we can factor A as:
>
> $$A = E_1^{-1} E_2^{-1} \cdots E_{k-1}^{-1} E_k^{-1},$$
>
> where $E_1, E_2, \ldots, E_k$ are the elementary matrices corresponding to a sequence of elementary row operations in the Gauss-Jordan Algorithm that transforms A into $\boldsymbol{I_n}$.

Note: The factorization of A into elementary matrices is ***not unique***, since a different sequence of elementary row operations will result in a different sequence of elementary matrices.

Proof of the Theorem: ($\Rightarrow$) If A is invertible, then the rref of A is $\boldsymbol{I_n}$. This means, by the Gauss-Jordan Algorithm, that we can find a finite sequence of elementary row operations that will transform A into $\boldsymbol{I_n}$. But every elementary row operation corresponds to a left multiplication with an elementary matrix. Thus we can find a finite sequence of elementary matrices $E_1, E_2, \ldots, E_k$ such that:

$$E_k E_{k-1} \cdots E_2 E_1 A = \boldsymbol{I_n}.$$

But since each elementary matrix is invertible, we have:

$$E_k^{-1}(E_k E_{k-1} \cdots E_2 E_1 A) = E_k^{-1}\boldsymbol{I_n}, \text{ or simplifying:}$$
$$E_{k-1} \cdots E_2 E_1 A = E_k^{-1}.$$

Notice that we multiplied both sides of the equation on the ***left side*** by E_k^{-1}, because E_k is the ***leftmost factor*** of the matrix product. Continuing thus, we get:

$$E_{k-1}^{-1}(E_{k-1}E_{k-2} \cdots E_2 E_1 A) = E_{k-1}^{-1}E_k^{-1}, \quad \text{or}$$
$$E_{k-2} \cdots E_2 E_1 A = E_{k-1}^{-1}E_k^{-1}, \quad \text{and keep going until}$$
$$A = E_1^{-1}E_2^{-1} \cdots E_{k-1}^{-1}E_k^{-1}.$$

Since the inverse of every elementary matrix is also an elementary matrix, the product on the right is also made up of elementary matrices.

($\Leftarrow$) Conversely, if $A = G_1 G_2 \cdots G_k$, where every G_i is an (invertible) elementary matrix, then $G_k^{-1}G_{k-1}^{-1} \cdots G_2^{-1}G_1^{-1}A = \boldsymbol{I_n}$. This equation tells us that there is a sequence of elementary row operations that will transform A into $\boldsymbol{I_n}$, and thus the rref of A is $\boldsymbol{I_n}$. Hence, A is invertible.■

We can see from this proof that a different sequence of elementary row operations that reduces A to $\boldsymbol{I_n}$ will indeed result in a different factorization, so the factorization into elementary matrices is hardly unique.

Example: We saw that $A = \begin{bmatrix} 3 & 2 & 5 \\ 2 & 1 & 4 \\ 4 & 2 & 7 \end{bmatrix}$ is invertible.

In order to express A as a product of elementary matrices, we need to list down the sequence of elementary row operations that produced A^{-1}, find the inverse matrix corresponding to each of these operations, and form the product of these matrices, going left to right. For easy reference, we list down, in chronological order, the sequence of 9 row operations that we performed:

$$R_1 \to R_1 - R_2, \qquad R_2 \to R_2 - 2R_1, \qquad R_3 \to R_3 - 4R_1,$$
$$R_2 \to -R_2, \qquad R_3 \to R_3 + 2R_2, \qquad R_3 \to -R_3,$$
$$R_2 \to R_2 + 2R_3, \qquad R_1 \to R_1 - R_3 \qquad R_1 \to R_1 - R_2.$$

From the proof of the Theorem, we must ***invert*** each corresponding elementary matrix and multiply them in the ***same order***. Thus, the first three operations will turn into:

$$R_1 \to R_1 + R_2, \qquad R_2 \to R_2 + 2R_1, \qquad R_3 \to R_3 + 4R_1.$$

Since the product has 9 factors, we write our factorization on two lines:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\cdot$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Since matrix multiplication is associative, we can actually perform the tedious task of checking that this is correct by multiplying ***consecutive pairs*** together, collapsing the product (leaving the last matrix on the right alone) to 5 matrices, and proceeding in this way multiplying consecutive pairs:

$$\begin{bmatrix} 3 & 1 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 4 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -2 & -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -1 & 0 \\ 2 & -1 & 0 \\ 4 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -2 \\ 0 & -2 & 3 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & -1 & 5 \\ 2 & -1 & 4 \\ 4 & -2 & 7 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 & 5 \\ 2 & 1 & 4 \\ 4 & 2 & 7 \end{bmatrix} = A. \square$$

Solving Invertible Square Equations

We already know how to solve a system of equations using the Gauss-Jordan algorithm. However, if we have a ***square*** system, and the coefficient matrix is invertible, and we know its ***inverse***, then we have another way to solve the system:

Theorem: If A is an invertible $n \times n$ matrix, then the system:

$$A\vec{x} = \vec{b}$$

has ***exactly one*** solution for any $n \times 1$ matrix $\vec{b}$, namely:

$$\vec{x} = A^{-1}\vec{b}.$$

More generally, if C is any $n \times m$ matrix, then the matrix equation:

$$AB = C$$

has exactly one solution for the $n \times m$ matrix B, namely:

$$B = A^{-1}C.$$

Proof: If $A\vec{x} = \vec{b}$, where A is invertible, then $A^{-1}(A\vec{x}) = A^{-1}\vec{b}$. But by the Associative Property, $A^{-1}(A\vec{x}) = (A^{-1}A)\vec{x} = \boldsymbol{I_n}\vec{x} = \vec{x}$, thus $\vec{x} = A^{-1}\vec{b}$. The generalization is proven in the same way.■

Example: We saw that if $A = \begin{bmatrix} 3 & 2 & 5 \\ 2 & 1 & 4 \\ 4 & 2 & 7 \end{bmatrix}$, then $A^{-1} = \begin{bmatrix} -1 & -4 & 3 \\ 2 & 1 & -2 \\ 0 & 2 & -1 \end{bmatrix}$.

The system $A\vec{x} = \begin{bmatrix} 5 \\ -8 \\ 3 \end{bmatrix}$ therefore has the unique solution:

$$\vec{x} = A^{-1}\begin{bmatrix} 5 \\ -8 \\ 3 \end{bmatrix} = \begin{bmatrix} -1 & -4 & 3 \\ 2 & 1 & -2 \\ 0 & 2 & -1 \end{bmatrix}\begin{bmatrix} 5 \\ -8 \\ 3 \end{bmatrix} = \begin{bmatrix} 36 \\ -4 \\ -19 \end{bmatrix}.$$

Similarly, if we want to solve the matrix equation:

$$\begin{bmatrix} 3 & 2 & 5 \\ 2 & 1 & 4 \\ 4 & 2 & 7 \end{bmatrix}\begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix} = \begin{bmatrix} -5 & 4 \\ 3 & -6 \\ 7 & 3 \end{bmatrix}, \text{ then:}$$

$$\begin{bmatrix} u & v \\ w & x \\ y & z \end{bmatrix} = A^{-1}\begin{bmatrix} -5 & 4 \\ 3 & -6 \\ 7 & 3 \end{bmatrix} = \begin{bmatrix} -1 & -4 & 3 \\ 2 & 1 & -2 \\ 0 & 2 & -1 \end{bmatrix}\begin{bmatrix} -5 & 4 \\ 3 & -6 \\ 7 & 3 \end{bmatrix} = \begin{bmatrix} 14 & 29 \\ -21 & -4 \\ -1 & -15 \end{bmatrix}. \square$$

2.7 Section Summary

If E is an elementary $n \times n$ matrix and A is any $n \times m$ matrix, then the matrix product EA can be computed by performing the same elementary row operation on A that was used to produce E from $\boldsymbol{I_n}$.

Elementary matrices are ***invertible***, and the inverse of an elementary matrix is another elementary matrix of the same type.

Let A be an $n \times n$ matrix. Then: A is ***invertible*** ***if and only if*** the rref of A is $\boldsymbol{I_n}$. Furthermore, if we construct the $n \times 2n$ augmented matrix $\left[\, A \mid \boldsymbol{I_n} \,\right]$, then A is ***invertible*** ***if and only if*** the rref of this augmented matrix contains $\boldsymbol{I_n}$ in the first n columns, in which case A^{-1} will be found in the last n columns. In other words, the rref of $\left[\, A \mid \boldsymbol{I_n} \,\right]$ is $\left[\, \boldsymbol{I_n} \mid A^{-1} \,\right]$.

An $n \times n$ matrix A is ***invertible*** ***if and only if*** it can be expressed as a product of elementary matrices. In this case, $A = E_1^{-1}E_2^{-1}\ldots E_{k-1}^{-1}E_k^{-1}$, where $E_1, E_2, \ldots, E_k$ are the elementary matrices corresponding to the elementary row operations that transformed A into $\boldsymbol{I_n}$, in the same order.

If A is an ***invertible*** $n \times n$ matrix, then the system $A\vec{x} = \vec{b}$ has ***exactly one*** solution for any $n \times 1$ matrix $\vec{b}$, namely $\vec{x} = A^{-1}\vec{b}$. More generally, if C is any $n \times m$ matrix, then the matrix equation $AB = C$ has exactly one solution for B, namely $B = A^{-1}C$.

2.7 Exercises

For Exercises (1) to (12): (a) Use the Gauss-Jordan Algorithm to find the inverse, if possible, of the following matrices; (b) List explicitly each elementary transformation that you perform in the process; (c) Express each invertible matrix as the product of elementary matrices.

1. $\begin{bmatrix} 4 & 7 \\ 3 & 6 \end{bmatrix}$

2. $\begin{bmatrix} 1/2 & 3/4 \\ -5/2 & 7/4 \end{bmatrix}$

3. $\begin{bmatrix} 2 & 4 & -1 \\ -1 & 3 & -2 \\ 3 & 0 & 1 \end{bmatrix}$

4. $\begin{bmatrix} 3 & 1 & 7 \\ -2 & 4 & 5 \\ 4 & 6 & 19 \end{bmatrix}$

5. $\begin{bmatrix} 3 & 6 & -1 \\ 0 & -4 & 8 \\ 0 & 0 & 2 \end{bmatrix}$

6. $\begin{bmatrix} 1/2 & 0 & 0 \\ -1/4 & 2/3 & 0 \\ 3/2 & 1/3 & -3 \end{bmatrix}$

7. $\begin{bmatrix} 3 & 0 & 2 \\ 2 & 4 & -1 \\ -5 & -2 & 3 \end{bmatrix}$

8. $\begin{bmatrix} -2 & 1 & -3 \\ 5 & -4 & 1 \\ 3 & 0 & 2 \end{bmatrix}$

9. $\begin{bmatrix} -1/2 & 1/4 & -3/4 \\ 5/2 & -3/4 & 1/4 \\ 3/2 & 0 & 1/2 \end{bmatrix}$

10. $\begin{bmatrix} -1 & 3 & 5 & 1 \\ 0 & 2 & -4 & 7 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

11. $\begin{bmatrix} 3 & 0 & 0 & 0 \\ 4 & -5 & 0 & 0 \\ 7 & 2 & 0 & 0 \\ 8 & 6 & 9 & 4 \end{bmatrix}$

12. $\begin{bmatrix} 3 & -3 & 0 & 1 \\ 0 & 1 & -1 & -2 \\ -2 & -1 & 1 & -2 \\ 3 & 0 & -2 & 1 \end{bmatrix}$

13. Check explicitly that your factorization for Exercise 3 is correct.

For Exercises (14) to (18): Use your answers for Exercises 7 through 12 to solve the following matrix equations using the inverse of the coefficient matrix:

14. $\begin{bmatrix} 3 & 0 & 2 \\ 2 & 4 & -1 \\ -5 & -2 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 7 \\ -2 \\ -4 \end{bmatrix}$

15. $\begin{bmatrix} -2 & 1 & -3 \\ 5 & -4 & 1 \\ 3 & 0 & 2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ -8 \end{bmatrix}$

16. $\begin{bmatrix} -1/2 & 1/4 & -3/4 \\ 5/2 & -3/4 & 1/4 \\ 3/2 & 0 & 1/2 \end{bmatrix}\begin{bmatrix} u & x \\ v & y \\ w & z \end{bmatrix} = \begin{bmatrix} -3 & 0 \\ 5 & 4 \\ 2 & -7 \end{bmatrix}$

17. $$\begin{bmatrix} -1 & 3 & 5 & 1 \\ 0 & 2 & -4 & 7 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 3 \end{bmatrix}\begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 5 \\ 0 \\ -3 \\ -1 \end{bmatrix}$$

18. $$\begin{bmatrix} 3 & -3 & 0 & 1 \\ 0 & 1 & -1 & -2 \\ -2 & -1 & 1 & -2 \\ 3 & 0 & -2 & 1 \end{bmatrix}\begin{bmatrix} s & w \\ t & x \\ u & y \\ v & z \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ -1 & 7 \\ 5 & 2 \\ 6 & -8 \end{bmatrix}$$

19. Suppose that A is a $3 \times n$ matrix. ***Write a sentence*** describing how to compute the following matrix products:

a) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 1 \end{bmatrix} A$ b) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2/5 \end{bmatrix} A$ c) $\begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} A$

d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 7 & 1 \end{bmatrix} A$ e) $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} A$ f) $\begin{bmatrix} 1 & 0 & -4 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} A$

20. Suppose that A is a $4 \times n$ matrix. ***Write a sentence*** describing how to compute the following matrix products:

a) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} A$ b) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} A$ c) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} A$

d) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix} A$ e) $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 5 & 0 & 1 \end{bmatrix} A$ f) $\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} A$

21. Prove that if E is an elementary $n \times n$ matrix and A is any $n \times m$ matrix, then the product EA can be computed by performing the same elementary row operation on A that was used to produce E from $\boldsymbol{I_n}$. You will need to consider all three types of elementary row operations.

22. Prove that elementary matrices are ***invertible***, and the inverse of an elementary matrix is another elementary matrix of the same type. Again, you will need to consider all three types of elementary row operations.

23. Prove that in general, if A is an ***invertible*** $n \times n$ matrix, and C is a given $n \times m$ matrix, then there is exactly one $n \times m$ matrix B such that $AB = C$, namely $B = A^{-1}C$.

24. Prove that if a square matrix A is ***invertible***, then A^k is invertible for all positive integers k. Express the inverse of A^k in terms of A^{-1}.

25. ***Permutation Matrices:*** An $n \times n$ matrix is called a ***permutation matrix*** if every entry is 0 except for a ***single 1*** that appears on ***each row*** and ***each column***. We saw an Example of a permutation matrix in Exercises 10 and 28 of Section 2.1.

    a. Give an example of a 5×5 permutation matrix where 1 does ***not*** appear on the main diagonal.

    b. Prove that every product of Type 2 elementary matrices is a permutation matrix. Hint: prove by induction. Begin by explaining why every Type 2 elementary matrix is a permutation matrix.

    c. Prove that every permutation matrix is invertible.

    d. Prove that every permutation matrix can be expressed as the product of Type 2 elementary matrices. Hint: one way to do it is by Induction.

26. Analogous to elementary row operations, an ***elementary column operation*** is any one of the following actions on a matrix A (where C_i represents column i of A):

Type:	Notation:
1. Multiply column i by a nonzero scalar k	$C_i \rightarrow kC_i$
2. Exchange column i and column j	$C_i \leftrightarrow C_j$
3. Add k times column j to column i	$C_i \rightarrow C_i + kC_j$

Similarly, an ***elementary column matrix*** is a matrix obtained from $\boldsymbol{I_n}$ using a single elementary column operation.

    a. Prove that multiplying a matrix A on the ***right*** by a compatible elementary column matrix has the same effect as performing the corresponding elementary column operation on A. See Exercise 21.

    b. Show that every elementary matrix (as defined in the beginning of this section) is also an elementary column matrix, and vice versa, and in fact they are of the same respective type.

27. Suppose that A is an $m \times 4$ matrix. ***Write a sentence*** describing how to compute the following matrix products, using the ideas in the previous Exercise (compare your answers to those from Exercise 20):

a) $A\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -3 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ b) $A\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$ c) $A\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 3/2 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

d) $A\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 9 \end{bmatrix}$ e) $A\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 5 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ f) $A\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$

2.8 Conditions for Invertibility

It should be no surprise that the property of a matrix A being invertible is very special, and we will see in this Section that we can test for the invertibility of a matrix A (and consequently, a linear operator T) in a variety of ways. We begin by summarizing several conditions that are equivalent to the invertibility of a linear transformation or matrix:

Theorem — The Really Big Theorem on Invertibility:
The following conditions are equivalent for a linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$, with standard matrix $[T] = A$.

1. T is an invertible operator.
2. A is an invertible matrix.
3. The rref of A is $\boldsymbol{I_n}$.
4. A is the product of elementary matrices.
5. T is one-to-one.
6. $ker(T) = nullspace(A) = \left\{\vec{\mathbf{0}}_n\right\}$.
7. $nullity(T) = nullity(A) = 0$.
8. T is onto.
9. $range(T) = \mathbb{R}^n$.
10. $rank(T) = n$.
11. $colspace(A) = \mathbb{R}^n$.
12. The columns of A form a basis for $\mathbb{R}^n$.
13. The columns of A are linearly independent.
14. The columns of A Span $\mathbb{R}^n$.
15. $rowspace(A) = \mathbb{R}^n$.
16. The rows of A form a basis for $\mathbb{R}^n$.
17. The rows of A are linearly independent.
18. The rows of A Span $\mathbb{R}^n$.
19. The homogeneous equation $A\vec{x} = \vec{\mathbf{0}}_n$ has only the trivial solution.
20. ***For every*** $n \times 1$ matrix $\vec{b}$, the system $A\vec{x} = \vec{b}$ is ***consistent***.
21. ***For every*** $n \times 1$ matrix $\vec{b}$, the system $A\vec{x} = \vec{b}$ has ***exactly one solution***.
22. ***There exists*** an $n \times 1$ matrix $\vec{b}$, such that the system $A\vec{x} = \vec{b}$ has ***exactly one solution***.

Proof: We already know that conditions (1) and (2) are equivalent to each other, and that conditions (2), (3) and (4) are equivalent from the previous Section.

Similarly, our definitions and Theorems from Section 2.5 say that conditions (5) through (7) are equivalent, and that conditions (8) through (11) are equivalent. ***The Dimension Theorem*** says that:

$$rank(T) + nullity(T) = n.$$

Thus $rank(T) = n$ ***if and only if*** $nullity(T) = 0$, so conditions (7) and (10) are equivalent. Thus, conditions (5) through (11) are equivalent.

Let us next show that condition (1) is equivalent to condition (5). If T is invertible, then automatically T is one-to-one. Conversely, suppose that T is one-to-one. By the previous paragraph, T is also onto, and thus T is invertible, so conditions (1) and (5) are equivalent. Thus, conditions (1) through (11) are equivalent.

Since there are exactly n rows and n columns, conditions (11) through (14) are equivalent, and conditions (15) through (18) are equivalent, using the ***Two-for-One Theorem*** in Section 1.9, and the fact that a subspace W of $\mathbb{R}^n$ has dimension n ***if and only if*** $W = \mathbb{R}^n$. But since $rank(T)$ is the common dimension of $rowspace(A)$ and $colspace(A)$, we see that conditions (11) and (15) are equivalent. Thus, conditions (1) through (18) are equivalent.

Notice that the last four conditions have to do with the ***solvability*** of a system of equations. However, condition (19) is equivalent to condition (6). We will leave the equivalence of the last three conditions with conditions (1) to (19) as Exercises. We must warn, though, that there are ***subtle differences*** among them:

Condition (20) says that no matter which $n \times 1$ matrix $\vec{b}$ we choose, we can find ***at least one*** solution $\vec{x}$ to the matrix equation $A\vec{x} = \vec{b}$.

Condition (21) says that no matter which $n \times 1$ matrix $\vec{b}$ we choose, we can find ***exactly one*** solution $\vec{x}$ to the matrix equation $A\vec{x} = \vec{b}$.

Condition (22) says that we can find ***one*** $n \times 1$ matrix $\vec{b}$, that yields ***exactly one*** solution $\vec{x}$ to the matrix equation $A\vec{x} = \vec{b}$. ■

One Sided Inverses

To determine if a square matrix A is invertible, we must find another square matrix B of the same dimensions such that:

$$AB = \boldsymbol{I_n} \;\; \textit{\textbf{and}} \;\; BA = \boldsymbol{I_n}.$$

If B only satisfies the first equation, we could naturally call B a "right" inverse for A. Similarly we could call B a "left" inverse for A if it only satisfies the second equation (in the same way we talk about left limits or right limits in Calculus). However, it turns out we don't need to worry about this at all:

Theorem — Left and Right Inverses:
An $n \times n$ matrix A is ***invertible*** ***if and only if*** we can find an $n \times n$ matrix B such that $AB = \boldsymbol{I_n}$ ***or*** $BA = \boldsymbol{I_n}$. Thus, a "right" inverse is also a "left" inverse, and vice versa.

Proof: We will show that if $BA = \boldsymbol{I_n}$, then $AB = \boldsymbol{I_n}$ also. The proof of the opposite case is essentially the same and will be left as an Exercise.
The idea is to think of the operators corresponding to A and B and view the product as the composition. Let us denote $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}^n$, where $[T_1] = A$ and $[T_2] = B$. Thus:

$$BA = [T_2] \cdot [T_1] = [T_2 \circ T_1],$$

and so $BA = \boldsymbol{I_n}$ represents $T_2 \circ T_1$. Thus, $T_2 \circ T_1 = \boldsymbol{I}_{\mathbb{R}^n}$, so $T_2 \circ T_1$ is both one-to-one and onto. But we saw in Exercise 23 of Section 2.5 that if $T_2 \circ T_1$ is one-to-one, then T_1 is also one-to-one. But since T_1 is an operator, by the Really Big Theorem, T_1 is invertible, and thus $[T_1] = A$ is invertible. Since

A^{-1} exists, we can solve:

$$BA = \boldsymbol{I_n} \Rightarrow (BA)A^{-1} = \boldsymbol{I_n}A^{-1} \Rightarrow B = A^{-1}.$$

Thus: $AB = AA^{-1} = \boldsymbol{I_n}$. ■

The Inverse of a Composition and Matrix Product

It is not surprising that we get an invertible operator when we compose two invertible operators:

> ***Theorem:*** If $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}^n$ are both ***invertible*** operators, then $T_2 \circ T_1$ is also ***invertible***, and furthermore:
>
> $$[T_2 \circ T_1]^{-1} = [T_1]^{-1}[T_2]^{-1}.$$
>
> Analogously, if A and B are ***invertible*** $n \times n$ matrices, then AB is also ***invertible***, and furthermore:
>
> $$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof: If T_1 and T_2 are both invertible operators, then we know that their matrices $[T_1]$ and $[T_2]$ are both invertible matrices. But we also know that $[T_2 \circ T_1] = [T_2][T_1]$, and therefore to show that $[T_2 \circ T_1]$ is invertible, we must show that we can find some matrix, that when multiplied to $[T_2][T_1]$, yields the identity matrix. Since we already know that $[T_1]^{-1}$ and $[T_2]^{-1}$ both exist, their product, ***in this order***, is the perfect candidate for the inverse. All we have to do is show that the product of $[T_2][T_1]$ and $[T_1]^{-1}[T_2]^{-1}$ yields the identity matrix:

$$\begin{aligned}&([T_2][T_1]) \cdot \left([T_1]^{-1}[T_2]^{-1}\right)\\ &= [T_2]\left([T_1] \cdot [T_1]^{-1}\right)[T_2]^{-1} \quad \text{(by the Associative Property)}\\ &= [T_2]\boldsymbol{I_n}[T_2]^{-1} = [T_2][T_2]^{-1} = \boldsymbol{I_n}.\end{aligned}$$

Similarly, $\left([T_1]^{-1}[T_2]^{-1}\right) \cdot ([T_2][T_1]) = \boldsymbol{I_n}$. In exactly the same way, we can show that $(AB)(B^{-1}A^{-1}) = \boldsymbol{I_n} = (B^{-1}A^{-1})(AB)$. ■

Example: Consider the matrices:

$$A = \begin{bmatrix} 3 & 8 \\ 2 & 5 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} -8 & 9 \\ 7 & -8 \end{bmatrix}.$$

We can check that both of these matrices are invertible, and their inverses are:

$$A^{-1} = \begin{bmatrix} -5 & 8 \\ 2 & -3 \end{bmatrix}, \quad \text{and} \quad B^{-1} = \begin{bmatrix} -8 & -9 \\ -7 & -8 \end{bmatrix}.$$

Let us examine AB:

$$AB = \begin{bmatrix} 3 & 8 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} -8 & 9 \\ 7 & -8 \end{bmatrix} = \begin{bmatrix} 32 & -37 \\ 19 & -22 \end{bmatrix}.$$

Again, this product is invertible, and:

$$(AB)^{-1} = \frac{1}{32(-22)-19(-37)}\begin{bmatrix} -22 & 37 \\ -19 & 32 \end{bmatrix} = \begin{bmatrix} 22 & -37 \\ 19 & -32 \end{bmatrix}.$$

We verify that we need to take the product $B^{-1}A^{-1}$ to get the same matrix:

$$B^{-1}A^{-1} = \begin{bmatrix} -8 & -9 \\ -7 & -8 \end{bmatrix}\begin{bmatrix} -5 & 8 \\ 2 & -3 \end{bmatrix} = \begin{bmatrix} 22 & -37 \\ 19 & -32 \end{bmatrix}.$$

Just for fun, let us look at the ***wrong*** order:

$$A^{-1}B^{-1} = \begin{bmatrix} -5 & 8 \\ 2 & -3 \end{bmatrix}\begin{bmatrix} -8 & -9 \\ -7 & -8 \end{bmatrix} = \begin{bmatrix} -16 & -19 \\ 5 & 6 \end{bmatrix}.$$

We note that although this answer is wrong, it is the inverse of:

$$BA = \begin{bmatrix} -6 & -19 \\ 5 & 16 \end{bmatrix}. \square$$

It turns out that the ***converse*** of the previous Theorem is also true, and you will prove it in the Exercises:

> ***Theorem:*** If $T_1,\ T_2\ :\mathbb{R}^n \to \mathbb{R}^n$ are operators, and $T_2 \circ T_1$ is ***invertible***, then ***both*** T_2 and T_1 are also ***invertible***. Analogously if A and B are two $n \times n$ matrices and the product AB is ***invertible***, then ***both*** A and B are ***invertible***.

2.8 Section Summary

The Really Big Theorem on Invertibility on p. 248 gives us 22 conditions that are equivalent to an operator T and its standard matrix A being ***invertible***. A few more will be added in the future.

An $n \times n$ matrix A is ***invertible*** ***if and only if*** we can find an $n \times n$ matrix B such that:

$$AB = \boldsymbol{I_n} \ \ \textbf{\textit{or}} \ \ BA = \boldsymbol{I_n}.$$

Thus, a "right" inverse is also a "left" inverse, and vice versa.

If $T_1, T_2\ :\mathbb{R}^n \to \mathbb{R}^n$ are both ***invertible*** operators, then the composition $T_2 \circ T_1$ is also ***invertible***, and furthermore:

$$[T_2 \circ T_1]^{-1} = [T_1]^{-1}[T_2]^{-1}.$$

Analogously, if A and B are ***invertible*** $n \times n$ matrices, then AB is also ***invertible***, and furthermore:

$$(AB)^{-1} = B^{-1}A^{-1}.$$

Conversely, if $T_2 \circ T_1$ is ***invertible***, then ***both*** T_2 and T_1 are also ***invertible***. Analogously if A and B are two $n \times n$ matrices and the product AB is ***invertible***, then both A and B are ***invertible***.

2.8 Exercises

1. Let $A = \begin{bmatrix} 3 & 5 \\ 4 & 6 \end{bmatrix}$ and $B = \begin{bmatrix} 7 & -4 \\ 6 & -3 \end{bmatrix}$.

   a. Show that A and B are both invertible by computing their inverses.

   b. Compute the product $B^{-1}A^{-1}$.

   c. Find AB.

   d. Show directly that AB is invertible by computing its inverse.

   e. Verify that $(AB)^{-1} = B^{-1}A^{-1}$.

For Exercises (2) to (4): Prove that an $n \times n$ matrix A is ***invertible*** ***if and only if*** it satisfies any of the following conditions:

2. ***For every*** $n \times 1$ matrix $\vec{b}$, the system $A\vec{x} = \vec{b}$ is ***consistent***.

3. ***For every*** $n \times 1$ matrix $\vec{b}$, the system $A\vec{x} = \vec{b}$ has ***exactly one solution***.

4. ***There exists*** an $n \times 1$ matrix $\vec{b}$, such that the system $A\vec{x} = \vec{b}$ has ***exactly one solution***.

5. Prove that if A is an ***invertible*** $n \times n$ matrix, then the system $\vec{y}A = \vec{d}$ has exactly one solution for any $1 \times n$ matrix $\vec{d}$, namely $\vec{y} = \vec{d}A^{-1}$.

6. Prove that if $AB = \boldsymbol{I_n}$, then $BA = \boldsymbol{I_n}$ as well. Hint: study and mimic the proof of the 1st half of the Theorem.

7. Prove that if A and B are $n \times n$ matrices and AB is invertible, then ***both*** A and B are invertible.

   Warning: You ***cannot*** use the formula $(AB)^{-1} = B^{-1}A^{-1}$, because we are trying to prove in the first place that A and B are invertible, so the inverse matrices on the right side of this equation are ***not known to exist***.

   Instead, here's a partial hint: Let us give AB a name, say, X. Then, we can find a matrix X^{-1} with the property that . . . . Stare at what you have and use the Associative Property of Matrix Multiplication. Remember that a left inverse is also a right inverse, and vice-versa, so you only need to show that one equation is true for A, and another one is true for B. In the process, you should be able to provide a formula for A^{-1} and for B^{-1}.

8. Use the ideas in our Theorem on Left and Right Inverses to prove that if $T_2 \circ T_1$ is ***invertible***, then ***both*** T_1 and T_2 are invertible. Explain why this also shows that if A and B are $n \times n$ matrices and AB is invertible, then A and B are also invertible.

9. Suppose that A is a 5×5 matrix, and B is the matrix obtained from A by exchanging rows 1 and 3, and exchanging rows 2 and 5. Describe in a sentence how B^{-1} is related to A^{-1}. Hint: think of this problem in terms of multiplying A by elementary matrices.

10. ***Direct Sums and Matrices in Block Diagonal Form:*** Suppose that A_1 is an $n_1 \times n_1$ matrix, and A_2 is an $n_2 \times n_2$ matrix. We define the ***direct sum*** of these matrices, denoted: $A = A_1 \oplus A_2$, as the matrix of size $(n_1 + n_2) \times (n_1 + n_2)$ where A_1 appears in columns 1 to n_1 and rows 1 to n_1, and A_2 appears in columns $n_1 + 1$ to $n_1 + n_2$ and rows $n_1 + 1$ to $n_1 + n_2$, and all the other entries of A are ***zeroes***.

For example, if $A_1 = \begin{bmatrix} 3 & -7 \\ -2 & 4 \end{bmatrix}$ and $A_2 = \begin{bmatrix} 5 & -2 & 1 \\ -4 & 0 & 7 \\ 3 & -9 & -8 \end{bmatrix}$, then:

$$A = A_1 \oplus A_2 = \begin{bmatrix} 3 & -7 & 0 & 0 & 0 \\ -2 & 4 & 0 & 0 & 0 \\ 0 & 0 & 5 & -2 & 1 \\ 0 & 0 & -4 & 0 & 7 \\ 0 & 0 & 3 & -9 & -8 \end{bmatrix}.$$

Notice that the diagonal entries of A_1 and A_2 also lie on the diagonal of A.

We also use the notation:

$$A = A_1 \oplus A_2 = \begin{bmatrix} A_1 & \mathbf{0} \\ \mathbf{0} & A_2 \end{bmatrix}$$

where $\mathbf{0}$ represents a zero matrix of an appropriate size.

a. Warm-up: Write down the matrices: $B = A_2 \oplus A_1$ and $C = A_1 \oplus A_1$, where A_1 and A_2 are the matrices in the example above.

b. Explain ***in general*** why $A_1 \oplus A_2$ and $A_2 \oplus A_1$ are ***not equal***. In other words, the operation of taking direct sums is ***not commutative***.

c. Suppose that $A_3 = \begin{bmatrix} -4 & 5 \\ 7 & -3 \end{bmatrix}$. Write down $(A_1 \oplus A_2) \oplus A_3$, $A_2 \oplus A_3$, and $A_1 \oplus (A_2 \oplus A_3)$, where A_1 and A_2 are as above.

Notice that $(A_1 \oplus A_2) \oplus A_3$ and $A_1 \oplus (A_2 \oplus A_3)$ are equal. It is not hard to see that in general, the operation of taking direct sums is ***associative:*** for any three square matrices A_1, A_2 and A_3, not necessarily of the same size:

$$(A_1 \oplus A_2) \oplus A_3 = A_1 \oplus (A_2 \oplus A_3) = \begin{bmatrix} A_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & A_3 \end{bmatrix} = A_1 \oplus A_2 \oplus A_3.$$

Thus, we can generalize the process to direct sums of k square matrices: if $A_1, A_2, \ldots, A_k$ have sizes $n_1 \times n_1$, $n_2 \times n_2$, $\ldots$, $n_k \times n_k$, then the direct sum:

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_k,$$

is a square matrix of size $m \times m$, where $m = n_1 + n_2 + \cdots + n_k$, each A_i appears in the given order along the diagonal of A, and all the other entries of A are zeroes.

Conversely, we say that A is in ***block diagonal form***, and the A_i are called the ***blocks*** forming A, if we can find a sequence of two or more matrices $A_1, A_2, \ldots, A_k$ whose direct sum is A. We can also write A showing its blocks in the notation:

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_k = \begin{bmatrix} A_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & A_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & A_k \end{bmatrix},$$

where the zero matrices $\mathbf{0}$ have the appropriate sizes.

d. Suppose that $B_1 = \begin{bmatrix} 8 & -2 & -1 \\ 4 & 6 & -7 \\ -3 & 5 & 9 \end{bmatrix}$, $B_2 = [5]$, and $B_3 = \begin{bmatrix} 0 & 9 \\ -2 & -5 \end{bmatrix}$.

Find $B = B_1 \oplus B_2 \oplus B_3$. What is the size of B?

e. Suppose that $A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ -2 & 4 & -8 & 0 \\ 0 & 0 & 3 & -3 \\ 0 & -1 & 0 & 5 \end{bmatrix}$ and $B = \begin{bmatrix} 3 & -7 & 0 & 0 \\ -2 & 4 & 0 & 0 \\ 0 & 0 & 8 & -1 \\ 0 & 0 & 0 & 5 \end{bmatrix}$.

Which of these matrices, if any, is in block diagonal form? If so, find the blocks, but if not, explain why not.

f. Suppose that $A_1, A_2, \ldots, A_k$ and $B_1, B_2, \ldots, B_k$ are square matrices such that A_i and B_i have the same size for all i, and A and B are the corresponding direct sums. Show that we can also express $A + B$ and AB as the direct sums:

$$A + B = (A_1 + B_1) \oplus (A_2 + B_2) \oplus \cdots \oplus (A_k + B_k), \quad \text{and}$$
$$AB = (A_1 B_1) \oplus (A_2 B_2) \oplus \cdots \oplus (A_k B_k).$$

Hint: Use Induction on the number of blocks k.

g. Prove that $A = A_1 \oplus A_2 \oplus \cdots \oplus A_k$ is invertible ***if and only if*** each A_i is invertible. If so, state and prove a formula for A^{-1}. Note: one direction is easy, but the other direction is ***not*** obvious. Recall that the Gauss-Jordan Algorithm can be used to find the inverse. Use this in your proof.

11. ***Elementary Number Theory:*** In this Chapter, we encountered invertible matrices such as:

$$A = \begin{bmatrix} 3 & 8 \\ 2 & 5 \end{bmatrix}, \text{ where } A^{-1} = \frac{1}{5(3) - 8(2)} \begin{bmatrix} 5 & -8 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} -5 & 8 \\ 2 & -3 \end{bmatrix}.$$

Notice that A contains only integer entries, and since $ad - bc = -1$, A^{-1} also contains integer entries. The purpose of this Exercise is to create more matrices with this property.

The Chinese Remainder Theorem states that if a and b are ***relatively prime*** integers (that is, the only common divisors are 1 and -1), then we can find two integers x and y such that:

$$ax - by = 1.$$

For example, 7 and 12 are relatively prime, and: $7(-5) - 12(-3) = 1$.

The interested student can search the Internet for the ***Euclidean Algorithm***, which lets us systematically find a solution for x and y. When a and b are small, though, we can easily find a solution by trial and error, as we did above.

a. Use this example to create a 2×2 matrix A whose first row contains 7 and 12 so that A is invertible and A^{-1} also contains four integer entries. Compute A^{-1} to check your answer.

b. Use The Chinese Remainder Theorem to show that if a and b are relatively prime, then we can find a 2×2 matrix A whose first row contains a and b and second row contains integers, such that A is invertible and A^{-1} also contains four integer entries.

c. Find an invertible 2×2 matrix A whose first row contains 5 and 8 such that A^{-1} contains integer entries. Compute A^{-1} to check that your answer is correct.

d. Find an invertible 2×2 invertible matrix whose second column contains -7 and 16 and whose inverse only contains integer entries.

2.9 Diagonal, Triangular, and Symmetric Matrices

In this Section, we will explore linear transformations whose matrices have special forms, and consequently, have special properties and actions. The proofs of all of the Theorems in this Section are straightforward, so they will all be left as Exercises.

Diagonal Matrices

We will start with the easiest, and arguably the most ***elegant*** kind of square matrix. Recall that the ***main diagonal*** of an $n \times n$ matrix A are the entries $a_{i,i}$, where $i = 1..n$:

Definition: An $n \times n$ matrix $D = [d_{i,j}]$ is called ***diagonal*** if all the entries that are not on the main diagonal are 0, that is, $d_{i,j} = 0$ if $i \neq j$. In other words, D has the form:

$$D = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}.$$

For the sake of saving space, we write in shorthand:

$$D = Diag(d_1, d_2, \ldots, d_n).$$

Note that the definition only requires that the entries that are ***not*** on the main diagonal must be 0. The entries on the main diagonal ***could*** still be 0. Thus, for example, the zero square matrices are diagonal matrices. It is very easy to compute the action of a linear transformation whose matrix is diagonal:

Theorem: Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation, and:

$$[T] = D = Diag(d_1, d_2, \ldots, d_n)$$

is an $n \times n$ ***diagonal*** matrix. If $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ is any vector from $\mathbb{R}^n$, then:

$$T(\vec{v}) = \langle d_1 v_1, d_2 v_2, \ldots, d_n v_n \rangle.$$

In particular, the action of T on the basic unit vectors is given by:

$$T(\vec{e}_k) = d_k \vec{e}_k \text{ for all } k = 1..n.$$

Example: Suppose $T : \mathbb{R}^3 \to \mathbb{R}^3$ has a diagonal matrix:

$$[T] = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Then:

$$T\left(\begin{bmatrix} 2 \\ 6 \\ -5 \end{bmatrix}\right) = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} 2 \\ 6 \\ -5 \end{bmatrix} = \begin{bmatrix} 3 \cdot 2 \\ -7 \cdot 6 \\ 4 \cdot -5 \end{bmatrix} = \begin{bmatrix} 6 \\ -42 \\ -20 \end{bmatrix},$$

so indeed $T(\langle 2, 6, -5 \rangle) = \langle 3 \cdot 2, -7 \cdot 6, 4 \cdot -5 \rangle = \langle 6, -42, -20 \rangle$. We can also see that:

$$T(\vec{e}_2) = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ -7 \\ 0 \end{bmatrix} = -7\vec{e}_2,$$

as expected.

The product of a diagonal matrix with an arbitrary matrix is also easily computed:

Theorem: Suppose $D = Diag(d_1, d_2, \ldots, d_n)$ is an $n \times n$ ***diagonal*** matrix, A is any $n \times m$ matrix and B is any $m \times n$ matrix. If we write A and B, respectively, in terms of their rows and columns as follows:

$$A = \begin{bmatrix} \vec{r}_1 \\ \vec{r}_2 \\ \vdots \\ \vec{r}_n \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_n \end{bmatrix}, \quad \text{then:}$$

$$DA = \begin{bmatrix} d_1\vec{r}_1 \\ d_2\vec{r}_2 \\ \vdots \\ d_n\vec{r}_n \end{bmatrix}, \quad \text{and} \quad BD = \begin{bmatrix} d_1\vec{c}_1 & d_2\vec{c}_2 & \cdots & d_n\vec{c}_n \end{bmatrix}.$$

In other words, we can obtain DA by multiplying each row of A by the corresponding diagonal entry of D, and we can obtain BD by multiplying each column of B by the corresponding diagonal entry of D.

Example: Suppose:

$$D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & 4 \end{bmatrix},$$

$$A = \begin{bmatrix} 6 & 2 & 9 & -4 \\ 3 & 5 & -2 & 1 \\ -8 & -3 & -1 & 6 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 4 & -8 & 7 \\ 3 & 2 & -6 \end{bmatrix}.$$

Then:

$$DA = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & 4 \end{bmatrix} \begin{bmatrix} 6 & 2 & 9 & -4 \\ 3 & 5 & -2 & 1 \\ -8 & -3 & -1 & 6 \end{bmatrix} = \begin{bmatrix} 3 \bullet 6 & 3 \bullet 2 & 3 \bullet 9 & 3 \bullet -4 \\ -7 \bullet 3 & -7 \bullet 5 & -7 \bullet -2 & -7 \bullet 1 \\ 4 \bullet -8 & 4 \bullet -3 & 4 \bullet -1 & 4 \bullet 6 \end{bmatrix}$$

$$= \begin{bmatrix} 18 & 6 & 27 & -12 \\ -21 & -35 & 14 & -7 \\ -32 & -12 & -4 & 24 \end{bmatrix}, \quad \text{and similarly:}$$

$$BD = \begin{bmatrix} 4 & -8 & 7 \\ 3 & 2 & -6 \end{bmatrix} \begin{bmatrix} 3 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 4 \bullet 3 & -8 \bullet -7 & 7 \bullet 4 \\ 3 \bullet 3 & 2 \bullet -7 & -6 \bullet 4 \end{bmatrix}$$

$$= \begin{bmatrix} 12 & 56 & 28 \\ 9 & -14 & -24 \end{bmatrix}. \square$$

Recall that we say that a subspace of $\mathbb{R}^n$ is ***closed*** under vector addition and scalar multiplication. The following Theorem says that the set of diagonal matrices of the same size also enjoy similar closure properties.

Theorem — Closure Properties for Diagonal Matrices:
If A and B are $n \times n$ ***diagonal*** matrices and c is any scalar, then $A + B$, $A - B$, cA and AB are also $n \times n$ ***diagonal*** matrices. In particular, the positive ***powers*** of a diagonal matrix are also diagonal, and if $D = Diag(d_1, d_2, \ldots, d_n)$, then:

$$D^k = Diag(d_1^k, d_2^k, \ldots, d_n^k)$$

for all positive integers k.

Lastly, the next Theorem says how to easily determine if a diagonal matrix is invertible, and if so, how to find its inverse:

Theorem — Invertibility of Diagonal Matrices:
A diagonal matrix $D = Diag(d_1, d_2, \ldots, d_n)$ is ***invertible*** ***if and only if*** $d_i \neq 0$ for all $i = 1..n$. In this case:

$$D^{-1} = Diag(d_1^{-1}, d_2^{-1}, \ldots, d_n^{-1}).$$

Example: Let: $D = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -7 & 0 \\ 0 & 0 & \frac{2}{5} \end{bmatrix}$.

D is invertible because all the entries on the main diagonal are non-zero, and:

$$D^{-1} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{7} & 0 \\ 0 & 0 & \frac{5}{2} \end{bmatrix}.$$

Similarly, we can compute the 6^{th} power of D, as:

$$D^6 = \begin{bmatrix} 3^6 & 0 & 0 \\ 0 & (-7)^6 & 0 \\ 0 & 0 & \left(\frac{2}{5}\right)^6 \end{bmatrix} = \begin{bmatrix} 729 & 0 & 0 \\ 0 & 117,649 & 0 \\ 0 & 0 & \frac{64}{15,625} \end{bmatrix}.$$

Since D is invertible, we can also compute any ***negative power***, for instance:

$$D^{-3} = (D^{-1})^3 = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & -\frac{1}{7} & 0 \\ 0 & 0 & \frac{5}{2} \end{bmatrix}^3 = \begin{bmatrix} \frac{1}{27} & 0 & 0 \\ 0 & -\frac{1}{343} & 0 \\ 0 & 0 & \frac{125}{8} \end{bmatrix}. \square$$

Triangular Matrices

The next twin families of matrices that we will explore are second-best to the diagonal matrices in terms of simplicity and elegance:

Definition: An $n \times n$ matrix $U = [u_{i,j}]$ is called ***upper triangular*** if all the entries ***below*** the main diagonal are 0, that is, $u_{i,j} = 0$ if $i > j$. Similarly, an $n \times n$ matrix $L = [l_{i,j}]$ is called ***lower triangular*** if all the entries ***above*** the main diagonal are 0, that is, $l_{i,j} = 0$ if $i < j$.
Thus, U and L have the form:

$$U = \begin{bmatrix} u_{1,1} & u_{1,2} & \cdots & u_{1,n} \\ 0 & u_{2,2} & \cdots & u_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{n,n} \end{bmatrix} \text{ and } L = \begin{bmatrix} l_{1,1} & 0 & \cdots & 0 \\ l_{2,1} & l_{2,2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n,1} & l_{n,2} & \cdots & l_{n,n} \end{bmatrix},$$

respectively. We also say the U is ***strictly upper triangular*** if its diagonal entries are also zeroes, that is, $u_{i,j} = 0$ if $i \geq j$. Similarly L is ***strictly lower triangular*** when $l_{i,j} = 0$ if $i \leq j$.

To remember these definitions, notice that the ***interesting*** (that is, the non-zero) entries in an ***upper triangular*** matrix are in the ***upper part*** of the matrix — the rest are zeroes.

Example: The following matrices, L and U, are respectively lower and upper triangular:

$$L = \begin{bmatrix} 5 & 0 & 0 & 0 \\ -3 & -9 & 0 & 0 \\ 2 & 8 & 0 & 0 \\ 6 & 1 & -7 & -2 \end{bmatrix}, \quad U = \begin{bmatrix} -4 & -5 & 1 \\ 0 & 7 & -3 \\ 0 & 0 & 6 \end{bmatrix}. \square$$

Again, closure properties also hold for each kind of triangular matrix:

Theorem — Closure Properties for Triangular Matrices:

If A and B are $n \times n$ ***upper triangular*** matrices and c is any scalar, then $A + B$, $A - B$, cA and AB are also $n \times n$ ***upper triangular*** matrices. In particular, the positive ***powers*** of an upper triangular matrix are also upper triangular.
An analogous statement is true for ***lower triangular*** matrices.

Finally, like diagonal matrices, we can also tell immediately if a triangular matrix is invertible by looking at the main diagonal, and if so, its inverse will have the same form:

Theorem — Invertibility of Triangular Matrices:

An upper triangular matrix is ***invertible*** ***if and only*** if all of its entries on the main diagonal are non-zero. If so, its inverse is again an upper triangular matrix.
Analogous statements are also true for lower triangular matrices.

Example: The matrix:

$$A = \begin{bmatrix} -4 & -5 & 1 \\ 0 & 7 & -3 \\ 0 & 0 & 6 \end{bmatrix}$$

from our previous example is invertible because the diagonal entries are non-zero. We can compute its inverse by the Gauss-Jordan method, as usual, and we get:

$$A^{-1} = \begin{bmatrix} -\frac{1}{4} & -\frac{5}{28} & -\frac{1}{21} \\ 0 & \frac{1}{7} & \frac{1}{14} \\ 0 & 0 & \frac{1}{6} \end{bmatrix}.$$

Notice that the diagonal entries of A^{-1} are the reciprocals of the corresponding diagonal entries of A. On the other hand, the lower triangular matrix:

$$B = \begin{bmatrix} 6 & 0 & 0 \\ -2 & 0 & 0 \\ -5 & 4 & -2 \end{bmatrix}$$

is ***not*** invertible because the second diagonal entry of B is 0. $\square$

The Transpose of a Matrix

Recall that in Section 1.8, we defined the transpose of a matrix A as the matrix obtained from A by writing row 1 of A as column 1 of $A^\top$, and so on. Now that we are accustomed to matrix notation, let us rewrite that definition using the individual entries of A:

> ***Definition:*** Let A be an $m \times n$ matrix. The ***transpose*** of A, denoted $A^\top$, is the $n \times m$ matrix whose entries are given by:
>
> $$(A^\top)_{i,j} = a_{j,i}.$$
>
> In particular, the transpose of a row matrix is a column matrix, and the transpose of a column matrix is a row matrix.

Example: If $A = \begin{bmatrix} 3 & 7 & -4 \\ 0 & 9 & 2 \end{bmatrix}$, then $A^\top$ is a 3×2 matrix, and:

$$A^\top = \begin{bmatrix} 3 & 0 \\ 7 & 9 \\ -4 & 2 \end{bmatrix}. \square$$

As before, we see that the transpose operation has the effect of turning the i^{th} row of A into the i^{th} column of $A^\top$, with the entries going top to bottom instead of left to right. Similarly, the j^{th} column of A become the j^{th} row of $A^\top$. The transpose operation has many interesting properties:

> ***Theorem — Properties of the Transpose Operation:***
> Suppose that A and B are $m \times k$ matrices, C is a $k \times n$ matrix, and r is any scalar. Then:
>
> $$\begin{aligned} (A^\top)^\top &= A, \\ (A+B)^\top &= A^\top + B^\top, \\ (rA)^\top &= rA^\top, \quad \text{and} \\ (BC)^\top &= C^\top B^\top. \end{aligned}$$
>
> (Notice that B and C ***switch places*** in the last formula.) Furthermore, if A is a square matrix, then A is ***invertible*** ***if and only if*** $A^\top$ is ***invertible***, in which case: $(A^\top)^{-1} = (A^{-1})^\top$.

Now we are ready to describe our final family of special matrices.

Symmetric Matrices

> ***Definition:*** Let A be an $n \times n$ matrix. We say that A is ***symmetric*** if:
>
> $$A^\top = A.$$

Notice that a matrix which is not square ***cannot*** be symmetric.

Example: The matrix $A = \begin{bmatrix} 3 & 8 & -2 & 5 \\ 8 & 0 & 6 & -1 \\ -2 & 6 & 4 & 3 \\ 5 & -1 & 3 & -7 \end{bmatrix}$ is symmetric.□

Example: In Section 2.2, we found the matrix of the reflection operation across the plane $\Pi : 3x - 5y + 2z = 0$ to be:

$$[refl_\Pi] = \frac{1}{19}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix}.$$

We mentioned in that Section that this matrix is symmetric. In general, the matrix for any reflection operator, whether across a line or a plane, is symmetric, as we can see from the Exercises in Section 2.2. We will generalize the concept of a projection onto a subspace W of $\mathbb{R}^n$ in Chapter 7, and we will prove that their matrices are always symmetric.□

It should not be surprising that the set of symmetric matrices of the same size enjoy closure properties.

Theorem — Closure Properties of Symmetric Matrices:
Suppose A and B are symmetric $n \times n$ matrices and c is any scalar. Then: $A + B$, $A - B$ and cA are also symmetric. If A is invertible, then A^{-1} is also symmetric.

Notice that the closure properties above do ***not*** say that the ***product*** of two symmetric matrices is again symmetric. This is because an additional condition is both necessary and sufficient:

Theorem: Suppose A and B are ***symmetric*** $n \times n$ matrices. Then: AB is also ***symmetric*** ***if and only if*** A and B ***commute*** with each other, that is, $AB = BA$.

We will also see at the end of Chapter 7 (as later proven in Chapter 8) that symmetric matrices have a certain magical "diagonalizability" property.

2.9 Section Summary

An $n \times n$ matrix $D = [d_{i,j}]$ is called ***diagonal*** if all the entries that are not on the main diagonal are 0, that is, $d_{i,j} = 0$ if $i \neq j$. We will write D in shorthand as: $D = Diag(d_1, d_2, \ldots, d_n)$, where $d_1, d_2, \ldots, d_n$ are the diagonal entries of D.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation and $[T] = D$. If $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ is any vector from $\mathbb{R}^n$, then $T(\vec{v}) = \langle d_1 v_1, d_2 v_2, \ldots, d_n v_n \rangle$. In particular, the action of T on the basic unit vectors is given by: $T(\vec{e}_k) = d_k \vec{e}_k$ for all $k = 1..n$.

We can obtain DA by multiplying each row of A by the corresponding diagonal entry of D, and we can obtain BD by multiplying each column of B by the corresponding diagonal entry of D.

If A and B are $n \times n$ diagonal matrices and c is any scalar, then $A+B$, $A-B$, cA and AB are also $n \times n$ diagonal matrices. In particular, the positive powers of a diagonal matrix are also diagonal, and if $D = Diag(d_1, d_2, \ldots, d_n)$, then $D^k = Diag(d_1^k, d_2^k, \ldots, d_n^k)$ for all positive integers k.

A diagonal matrix D is ***invertible*** ***if and only if*** all the entries on the main diagonal are non-zero, and in this case $D^{-1} = Diag(d_1^{-1}, d_2^{-1}, \ldots, d_n^{-1})$.

An $n \times n$ matrix U is called ***upper triangular*** if all the entries ***below*** the main diagonal are 0, that is, $u_{i,j} = 0$ if $i > j$. Similarly, an $n \times n$ matrix L is called ***lower triangular*** if all the entries ***above*** the main diagonal are 0, that is, $l_{i,j} = 0$ if $i < j$.

If A and B are $n \times n$ ***upper triangular*** matrices and c is any scalar, then $A+B$, $A-B$, cA and AB are also $n \times n$ ***upper triangular*** matrices. In particular, the positive powers of an upper triangular matrix are also upper triangular. Similar closure properties hold for the set of lower triangular matrices.

An upper triangular matrix is ***invertible*** ***if and only if*** all of its entries on the main ***diagonal*** are ***non-zero***. If so, its ***inverse*** is again an upper triangular matrix. Analogous statements can be made for the set of ***lower triangular*** matrices.

Let A be an $m \times n$ matrix. The ***transpose*** of A, denoted $A^\top$, is the $n \times m$ matrix whose entries are given by $(A^\top)_{i,j} = a_{j,i}$. In particular, the transpose of a row matrix is a column matrix, and the transpose of a column matrix is a row matrix.

Suppose that A and B are $m \times k$ matrices, C is a $k \times n$ matrix, and r is any scalar. Then: $(A^\top)^\top = A$, $(A+B)^\top = A^\top + B^\top$, $(kA)^\top = kA^\top$, and $(BC)^\top = C^\top B^\top$.

If A is a square matrix, then A is ***invertible*** ***if and only if*** $A^\top$ is ***invertible***, in which case: $(A^\top)^{-1} = (A^{-1})^\top$.

Let A be an $n \times n$ matrix. We say that A is ***symmetric*** if $A^\top = A$. Suppose A and B are ***symmetric*** $n \times n$ matrices and c is any scalar. Then: $A+B$, $A-B$ and cA are also symmetric. If A is ***invertible***, then A^{-1} is also ***symmetric***. The product AB is ***symmetric*** ***if and only if*** $AB = BA$.

2.9 Exercises

1. State ***all*** the words in the following list that correctly describe each matrix: diagonal, upper triangular, lower triangular, symmetric, all of the above, and none of the above:

a. $\begin{bmatrix} 0 & 0 & 0 \\ 3 & 4 & 0 \\ -5 & 2 & 7 \end{bmatrix}$ b. $\begin{bmatrix} 6 & 0 & 0 \\ 0 & -9 & 0 \\ 0 & 0 & 3 \end{bmatrix}$ c. $\begin{bmatrix} 2 & 5 & -8 \\ 5 & -3 & 2 \\ -8 & 2 & 3 \end{bmatrix}$

d. The identity matrices $\boldsymbol{I_n}$.

e. The square zero matrices $\boldsymbol{0_{n \times n}}$.

2. An $n \times n$ matrix A is defined by the formula:

$$a_{i,j} = i + j, \quad \text{for all } i, j = 1..n.$$

a. Assemble the matrix A when $n = 4$.

b. Which of the words in Exercise 1 above describe these matrices? Prove your answer in general.

3. Let $D = \begin{bmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -7 \end{bmatrix}$, $A = \begin{bmatrix} 4 & 7 & -3 & 2 & 0 \\ 9 & -2 & 1 & 4 & 6 \\ 5 & 3 & 2 & -9 & -1 \end{bmatrix}$ and $B = \begin{bmatrix} 9 & -5 & 2 \\ 1 & 4 & -3 \end{bmatrix}$.

a. Compute DA.

b. Compute BD.

c. Prove ***in general*** that if D is an $n \times n$ diagonal matrix, and A is any $n \times m$ matrix, then we can compute DA by multiplying each row of A by the corresponding diagonal entry of D.

d. Prove ***in general*** that if D is an $n \times n$ diagonal matrix, and B is any $m \times n$ matrix, then we can compute BD by multiplying each column of B by the corresponding diagonal entry of D.

4. Let A and B be $n \times n$ diagonal matrices and c is any scalar. Prove that the following are also diagonal:

a. $A + B$ b. $A - B$ c. cA d. AB

5. Let $D = Diag(d_1, d_2, \ldots, d_n)$ be a diagonal matrix.

a. Prove by induction that $D^k = Diag(d_1^k, d_2^k, \ldots, d_n^k)$, for all positive integers k.

b. Prove that D is invertible if and only if all of the d_i are non-zero. Hint: Use the fact that any matrix A is invertible if and only if the reduced row echelon form of A is the identity matrix.

c. Prove that if $D = Diag(d_1, d_2, \ldots, d_n)$ is invertible, then:

$$D^{-1} = Diag(d_1^{-1}, d_2^{-1}, \ldots, d_n^{-1}).$$

6. Prove that all diagonal matrices are symmetric.

7. Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation with $[T] = Diag(d_1, d_2, \ldots, d_n)$, a diagonal matrix. Show that $T(\vec{e}_k) = d_k \vec{e}_k$ for $k = 1..n$.

8. Let A and B be $n \times n$ upper triangular matrices and c any scalar. Prove that the following are also $n \times n$ upper triangular matrices:

a. $A + B$ b. $A - B$ c. cA d. AB

9. Let $A = \begin{bmatrix} 3 & -5 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & -7 \end{bmatrix}$ and $B = \begin{bmatrix} 8 & 4 & 1 \\ 0 & -3 & 6 \\ 0 & 0 & -4 \end{bmatrix}$.

a. Compute the matrix product AB step-by-step.

b. Use your observations in (a) to prove ***in general*** that the product of two upper triangular matrices is again upper triangular. Hint: use the dot product formula for the matrix product. We remind you that you only need to show that $(AB)_{i,j} = 0$ if $i > j$. Look at the locations of the zeroes.

10. Prove that the transpose of an upper triangular matrix is a lower triangular matrix, and vice versa.

11. Prove that an upper triangular matrix is invertible ***if and only if*** none of the entries in the main diagonal is 0. Use the same hint as in Exercise 5 (b). State and prove a similar statement for lower triangular matrices (Warning: the proof for upper triangular matrices is not exactly the same idea for lower triangular matrices).

12. Let $T : \mathbb{R}^3 \to \mathbb{R}^3$, with $[T] = A = \begin{bmatrix} 3 & -5 & 4 \\ 0 & 2 & 1 \\ 0 & 0 & -7 \end{bmatrix}$.

a. Compute $T(\vec{e}_1)$, $T(\vec{e}_2)$, and $T(\vec{e}_3)$. Express your answers in terms of $\vec{e}_1$, $\vec{e}_2$, and $\vec{e}_3$.

b. Use your computations in (a) to find three vectors $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$, such that $T(\vec{v}_1) = \vec{e}_1$, $T(\vec{v}_2) = \vec{e}_2$, and $T(\vec{v}_3) = \vec{e}_3$. Hint: fully exploit the linearity properties of T.

c. Use your answers in (b) to write A^{-1}.

13. Prove that if an upper triangular matrix is invertible, then its inverse is also upper triangular. Hints: Let T be the linear transformation corresponding to this matrix. Explicitly construct T^{-1} by defining it on the standard basis using The Principle of Mathematical Induction. Generalize the ideas from the previous Exercise. Start by showing that you can define $T^{-1}(\vec{e}_1)$. Assume that you can define $T^{-1}(\vec{e}_1)$ through $T^{-1}(\vec{e}_k)$. Finally, show that you can define $T^{-1}(\vec{e}_{k+1})$.

14. Let A and B be $m \times k$ matrices, let C be a $k \times n$ matrix, and let $r \in \mathbb{R}$. Prove that:

a. $(A^\top)^\top = A$ b. $(A + B)^\top = A^\top + B^\top$ c. $(rA)^\top = r(A^\top)$

d. $(BC)^\top = C^\top B^\top$ (Hint: Use the dot product formula for the matrix product)

15. Prove that if A is an invertible $n \times n$ matrix, then $A^\top$ is also invertible, and $(A^\top)^{-1} = (A^{-1})^\top$.

Hint: all you need to show is that $A^\top(A^{-1})^\top = \boldsymbol{I}_n$.

Part (d) from the previous Exercise will be useful.

16. Let A and B be symmetric $n \times n$ matrices and let c be any scalar. Prove that $A^\top$, $A + B$, $A - B$ and cA are also symmetric.

17. Prove that if A is invertible and symmetric, then A^{-1} is also symmetric.

18. Suppose A and B are symmetric $n \times n$ matrices. Prove that AB is also symmetric ***if and only if*** A and B ***commute*** with each other, that is, $AB = BA$.

Hint: Exercise 14 (d) will again be very useful.

19. Prove the converse of Exercise 5 from Section 2.8: If the equation $\vec{y}A = \vec{d}$ is solvable (for $\vec{y}$) for all $1 \times n$ matrices $\vec{d}$, then A is invertible. Hint: use the transpose operation and Exercise 14 (d).

20. Suppose that A is a strictly upper triangular $n \times n$ matrix. Prove that $A^n = \mathbf{0}_{n \times n}$, the zero $n \times n$ matrix. Hint: compute the powers of a strictly upper triangular 4×4 matrix and observe what happens to each power, and why this happens.

For Exercises (21) to (32): (a) identify whether the matrix is diagonal, upper triangular (but not diagonal), lower triangular (but not diagonal), or symmetric (but not diagonal); (b) show that each matrix is invertible by finding its inverse, and (c) verify that the inverse of the matrix is of the same type.

21. $\begin{bmatrix} 5 & 3 \\ 3 & 4 \end{bmatrix}$

22. $\begin{bmatrix} -4/7 & -2/3 \\ 0 & 3/5 \end{bmatrix}$

23. $\begin{bmatrix} 3 & 0 \\ -7 & 8 \end{bmatrix}$

24. $\begin{bmatrix} -9 & 0 \\ 0 & 3/4 \end{bmatrix}$

25. $\begin{bmatrix} 5 & 0 & 0 \\ -4 & 7 & 0 \\ 3 & -8 & 4/3 \end{bmatrix}$

26. $\begin{bmatrix} 5 & 3 & -2 \\ 3 & -4 & 0 \\ -2 & 0 & 1 \end{bmatrix}$

27. $\begin{bmatrix} 3/5 & 0 & 0 \\ 0 & -2/3 & 0 \\ 0 & 0 & -7/9 \end{bmatrix}$

28. $\begin{bmatrix} -3 & 6 & 2 \\ 0 & -4 & -8 \\ 0 & 0 & 7 \end{bmatrix}$

29. $\begin{bmatrix} 5 & 0 & 0 & 0 \\ -2 & 9 & 0 & 0 \\ 7 & 6 & -1 & 0 \\ 0 & -4 & 8 & 3 \end{bmatrix}$

30. $\begin{bmatrix} 3 & -2 & 1 & 4 \\ -2 & 0 & 5 & -1 \\ 1 & 5 & 2 & 3 \\ 4 & -1 & 3 & 7 \end{bmatrix}$

31. $\begin{bmatrix} 2 & -5 & 9 & -1 \\ 0 & 4 & 3 & 0 \\ 0 & 0 & -7 & 8 \\ 0 & 0 & 0 & 6 \end{bmatrix}$

32. $\begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & -3/2 & 0 & 0 \\ 0 & 0 & -7 & 0 \\ 0 & 0 & 0 & 8/5 \end{bmatrix}$

33. Let $A = \begin{bmatrix} -2 & 10 & -20 \\ 10 & 13 & 10 \\ -20 & 10 & -2 \end{bmatrix}$, $B = \begin{bmatrix} 10 & 4 & 19 \\ 4 & 43 & 4 \\ 19 & 4 & 10 \end{bmatrix}$, and $C = \begin{bmatrix} 1 & 0 & -9 \\ 0 & -3 & 4 \\ -9 & 4 & 2 \end{bmatrix}$.

Notice that all three matrices are obviously symmetric.

a. Compute AB and BA and verify that they are equal. Look at the resulting matrix, and check that it is also symmetric.

b. On the other hand, compute AC and CA. Is either matrix symmetric? Are these two products equal to each other?

34. ***Matrices in Block Diagonal Form:*** Suppose that A_1, A_2, …, A_k are all square matrices, not necessarily of the same size, with $k \geq 2$. We defined the direct sum of these matrices:

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_k$$

in Exercises 10 of Section 2.8. Prove the following statements about certain kinds of special direct sums:

a. A is diagonal ***if and only if*** every A_i is also diagonal.

b. A is upper triangular ***if and only if*** every A_i is also upper triangular.

c. A is lower triangular ***if and only if*** every A_i is also lower triangular.

d. $A^\top = A_1^\top \oplus A_2^\top \oplus \cdots \oplus A_k^\top$.

e. A is symmetric ***if and only if*** every A_i is also symmetric.

A Summary of Chapter 2

A ***linear transformation*** $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function that satisfies, for all $\vec{u}, \vec{v} \in \mathbb{R}^n$ and for all $k \in \mathbb{R}$:

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}) \text{ and } T(k\vec{u}) = kT(\vec{u}).$$

In the special case when $T : \mathbb{R}^n \to \mathbb{R}^n$, we call T a ***linear operator***.

A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if we can find an $m \times n$ matrix A so that $T(\vec{x}) = A\vec{x}$ for all $\vec{x} \in \mathbb{R}^n$. We call A the ***standard matrix*** of T, and:

$$\left[\,T\,\right] = A = \left[\; T(\vec{e}_1) \mid T(\vec{e}_2) \mid \cdots \mid T(\vec{e}_n) \;\right].$$

An $n \times n$ matrix E is called an ***elementary matrix*** if it is obtained by performing a ***single*** elementary row operation on the identity matrix I_n.

In $\mathbb{R}^2$, linear operators have geometric effects, such as ***dilations***, ***contractions***, ***shear*** operators, ***rotations*** by an angle θ, ***projections*** onto a line L through the origin, and ***reflections*** across a line L through the origin. In $\mathbb{R}^3$, linear operators can geometrically represent a ***projection*** onto or ***reflection*** across a plane Π through the origin, or a line L through the origin.

If $T_1, T_2 : \mathbb{R}^n \to \mathbb{R}^m$ are linear transformations, and $k \in \mathbb{R}$, then we $T_1 + T_2$, $T_1 - T_2$, and kT_1 as linear transformations, also from $\mathbb{R}^n$ to $\mathbb{R}^m$, with actions given by: $(T_1 + T_2)(\vec{v}) = T_1(\vec{v}) + T_2(\vec{v})$, $(T_1 - T_2)(\vec{v}) = T_1(\vec{v}) - T_2(\vec{v})$, and $(kT_1)(\vec{v}) = kT_1(\vec{v})$.

Analogously, if A and B are both $m \times n$ matrices, and $k \in \mathbb{R}$, then we can define $A + B$, $A - B$, and kA as $m \times n$ matrices with entries given by:

$$(A+B)_{i,j} = (A)_{i,j} + (B)_{i,j}, (A-B)_{i,j} = (A)_{i,j} - (B)_{i,j}, \text{ and } (kA)_{i,j} = k(A)_{i,j}.$$

If $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ are linear transformations, the ***composition*** $T_2 \circ T_1 : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, with action given by $(T_2 \circ T_1)(\vec{v}) = T_2(T_1(\vec{v}))$ for all $\vec{v} \in \mathbb{R}^n$.

If A is an $m \times k$ matrix, and $B = \left[\; \vec{c}_1\, \vec{c}_2 \ldots \vec{c}_n \;\right]$ is a $k \times n$, then we can construct the $m \times n$ ***matrix product*** $AB = \left[\; A\vec{c}_1\, A\vec{c}_2 \ldots A\vec{c}_n \;\right]$.

Under compatible conditions, matrix arithmetic enjoys the following properties: $A + B = B + A$, $A + (B + C) = (A + B) + C$, $(r + s)A = rA + sA$, $r(A + B) = rA + rB$, $r(sA) = (rs)A = s(rA)$, $(A + B)C = AC + BC$, $A(C + D) = AC + AD$, $r(BC) = (rB)C = B(rC)$, and $A(BC) = (AB)C$. However, ***matrix multiplication***, in general, ***is not commutative***.

If $T_1 : \mathbb{R}^n \to \mathbb{R}^k$ and $T_2 : \mathbb{R}^k \to \mathbb{R}^m$ are linear transformations, then $[T_2 \circ T_1] = [T_2][T_1]$.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and $B = \left\{ \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \right\}$ is a basis for $\mathbb{R}^n$, and if $\vec{v} \in \mathbb{R}^n$ is written (uniquely) as $\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$, then:

$$T(\vec{v}) = c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n).$$

The ***kernel*** of T is: $ker(T) = \left\{ \vec{v} \in \mathbb{R}^n \mid T(\vec{v}) = \vec{0}_m \right\} = nullspace([T]) \trianglelefteq \mathbb{R}^n$.

The ***range*** of T is: $range(T) = \left\{ \vec{w} \in \mathbb{R}^m | \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in \mathbb{R}^n \right\} = colspace([T]) \trianglelefteq \mathbb{R}^m$.

We call $dim(ker(T))$ the ***nullity*** of T, and $dim(range(T))$ the ***rank*** of T.

We say that T is ***full rank*** if $rank(T) = \min(m, n)$.

The Dimension Theorem for Linear Transformations states that:

$$rank(T) + nullity(T) = n = dim\big(\text{domain of } T\big).$$

A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***one-to-one*** if $T(\vec{v}_1) \neq T(\vec{v}_2)$ whenever $\vec{v}_1 \neq \vec{v}_2$.

T is ***one-to-one*** or ***injective*** ***if and only if*** $ker(T) = \left\{ \vec{0}_n \right\}$.

T is ***onto*** or ***surjective*** ***if and only if*** $range(T) = \mathbb{R}^m$, or equivalently, $rank(T) = m$.

T ***cannot*** be one-to-one if $n > m$, and T ***cannot*** be onto if $n < m$.

Depending on m and n, the property of an $m \times n$ matrix being of ***full rank*** is equivalent to the property of being one-to-one, onto, or both.

T is ***invertible*** or ***bijective*** ***if and only if*** T is both ***one-to-one*** and ***onto***.

If $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible, then $n = m$. In other words, T must be an ***operator***.

A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is invertible ***if and only if*** we can find another linear operator, denoted by $T^{-1} : \mathbb{R}^n \to \mathbb{R}^n$, the ***inverse*** of T, such that T^{-1} has the properties: $T \circ T^{-1} = T^{-1} \circ T = I_{\mathbb{R}^n}$. Analogously, an $n \times n$ matrix A is ***invertible*** if the corresponding operator T is invertible. This is true ***if and only if*** we can find another $n \times n$ matrix B such that $AB = \boldsymbol{I_n} = BA$. We call B the ***inverse*** of A, and write $B = A^{-1}$. The inverse of A is ***unique***.

A linear operator $T : \mathbb{R}^n \to \mathbb{R}^n$ is ***invertible*** ***if and only if*** $A = [T]$ is an invertible $n \times n$ matrix. If this is the case, then $[T^{-1}] = A^{-1} = [T]^{-1}$.

An $n \times n$ matrix A is ***invertible*** ***if and only if*** the rref of $[A \,|\, \boldsymbol{I_n}]$ contains $\boldsymbol{I_n}$ in the first n columns, in which case, A^{-1} will be found in the last n columns of the rref.

An $n \times n$ matrix A is ***invertible*** ***if and only if*** A can be expressed as a ***product*** of ***elementary*** matrices.

If A is an ***invertible*** $n \times n$ matrix, then the system $A\vec{x} = \vec{b}$ has ***exactly one*** solution for any $n \times 1$ matrix $\vec{b}$, namely $\vec{x} = A^{-1}\vec{b}$. More generally, if C is any $n \times m$ matrix, then the matrix equation $AB = C$ has exactly one solution for B, namely $B = A^{-1}C$.

The Really Big Theorem on Invertibility gives us 22 conditions that are equivalent to an operator T and its matrix A being ***invertible***.

If $T_1,\ T_2 : \mathbb{R}^n \to \mathbb{R}^n$ are both ***invertible*** operators, then $T_2 \circ T_1$ is also invertible, and furthermore: $[T_2 \circ T_1]^{-1} = [T_1]^{-1} \bullet [T_2]^{-1}$. Analogously, if A and B are invertible $n \times n$ matrices, then AB is also invertible, and furthermore: $(AB)^{-1} = B^{-1}A^{-1}$. Conversely, if $T_1,\ T_2 : \mathbb{R}^n \to \mathbb{R}^n$ are operators and $T_2 \circ T_1$ is invertible, then both T_1 and T_2 are also invertible.

An $n \times n$ matrix $D = [d_{i,j}]$ is called ***diagonal*** if all the entries that are not on the main diagonal are 0, that is, $d_{i,j} = 0$ if $i \neq j$. We write D in shorthand as $D = Diag(d_1, d_2, \ldots, d_n)$, where $d_1,\ d_2,\ \ldots,\ d_n$ are the diagonal entries of D.

An $n \times n$ matrix U is called ***upper triangular*** if all the entries ***below*** the main diagonal are 0, that is, $(U)_{i,j} = 0$ if $i > j$. Similarly, an $n \times n$ matrix L is called ***lower triangular*** if all the entries ***above*** the main diagonal are 0, that is, $(L)_{i,j} = 0$ if $i < j$. If A and B are $n \times n$ upper triangular matrices and c is any scalar, then $A + B$, $A - B$, cA and AB are also $n \times n$ upper triangular matrices. Similar closure properties hold for the set of lower triangular matrices.

Let A be an $m \times n$ matrix. The ***transpose*** of A, denoted $A^\top$, is the $n \times m$ matrix whose entries are given by $(A^\top)_{i,j} = a_{j,i}$. In particular, the transpose of a row matrix is a column matrix, and the transpose of a column matrix is a row matrix.

Suppose that A and B are $m \times k$ matrices, C is a $k \times n$ matrix, and r is any scalar. Then: $(A^\top)^\top = A$, $(A + B)^\top = A^\top + B^\top$, $(kA)^\top = kA^\top$, and $(BC)^\top = C^\top B^\top$. Furthermore, if A is a square matrix, then A is ***invertible*** ***if and only if*** $A^\top$ is ***invertible***, in which case: $(A^\top)^{-1} = (A^{-1})^\top$.

Let A be an $n \times n$ matrix. We say that A is ***symmetric*** if $A^\top = A$. Suppose A and B are symmetric $n \times n$ matrices and c is any scalar. Then: $A^\top$, $A + B$, $A - B$ and cA are also symmetric. If A is invertible, then A^{-1} is also symmetric. The product AB is symmetric ***if and only if*** $AB = BA$.

Chapter 3

From the Real to the Abstract:

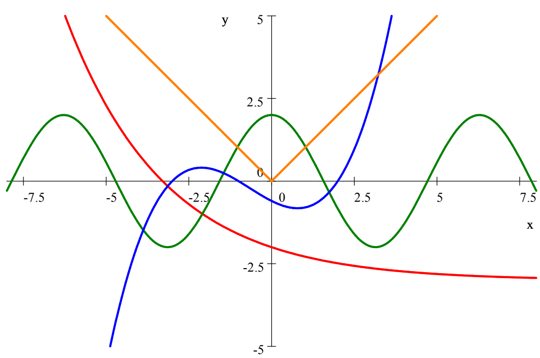
General Vector Spaces

We began our study of Linear Algebra by constructing ***Euclidean Spaces*** and studying ***linear combinations*** and the ***Span*** of a set of vectors. From this, we defined ***subspaces*** of a Euclidean space and saw how each subspace can be described as the Span of a ***basis*** consisting of a finite set of vectors. Aside from Spanning the subspace, a basis is also required to be ***linearly independent***. We saw that any two bases for the same subspace must contain exactly the same number of elements, called the ***dimension*** of the subspace.

In Chapter 2, we saw how a ***linear transformation*** $T : \mathbb{R}^n \to \mathbb{R}^m$ maps one Euclidean Space into another, and that such a transformation can be described using its $m \times n$ ***standard matrix*** $[T]$. We saw how the ***nullspace*** of these matrices corresponds to the ***kernel*** of the linear transformation it represents, and similarly that the ***columnspace*** corresponds to the ***range***. We were able to determine if such a linear transformation is ***one-to-one***, ***onto***, both or neither, and how to find the matrix of the ***inverse*** transformation when T is invertible.

Now we begin the process of generalizing all these concepts.

We will see that many objects that we are already familiar with from Algebra and Calculus enjoy the same properties as Euclidean spaces, and as such, we will refer to them as ***general*** or ***abstract vector spaces***. Analogously, we will define the concepts of linear combinations, Span, linear independence, subspaces, basis and dimension, as we did in Chapter 1. One of the most useful examples of an abstract vector space would be the space of all continuous functions on an interval I, denoted $C(I)$. In this case, a vector is a continuous function, and so we can visualize them graphically:



Four Vectors From $C(-\infty, \infty)$

We will generalize the concept of a ***linear transformation*** that will map one vector space to another vector space. We will see that many (but not all) operations from Algebra as well as Calculus are linear transformations. We will also construct ***coordinates*** with respect to a basis for a finite-dimensional vector space, and use these coordinates in order to construct ***matrices*** for linear transformations and study their attributes and properties, as we did in Chapter 2.

3.1 Axioms for a Vector Space

Although they are the most important example, Euclidean spaces are not the only kind of vector spaces. We will now generalize the central subject of Linear Algebra:

Definition — The Ten Axioms of an Abstract Vector Space:

A ***vector space*** $(V, \oplus, \odot)$ is a ***non-empty*** set V, together with two operations: $\oplus$ (***vector addition***), and $\odot$ (***scalar multiplication***), such that: for all $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$ and all r, $s \in \mathbb{R}$, $(V, \oplus, \odot)$ satisfies:

1. The Closure Property of Vector Addition:

$$\vec{u} \oplus \vec{v} \in V.$$

2. The Closure Property of Scalar Multiplication:

$$r \odot \vec{u} \in V.$$

3. The Commutative Property of Vector Addition:

$$\vec{u} \oplus \vec{v} = \vec{v} \oplus \vec{u}.$$

4. The Associative Property of Vector Addition:

$$(\vec{u} \oplus \vec{v}) \oplus \vec{w} = \vec{u} \oplus (\vec{v} \oplus \vec{w}).$$

5. The Existence of a Zero Vector:

$$\text{There exists } \vec{\mathbf{0}}_V \in V, \text{ such that: } \vec{\mathbf{0}}_V \oplus \vec{v} = \vec{v} = \vec{v} \oplus \vec{\mathbf{0}}_V.$$

6. The Existence of Additive Inverses:

$$\text{There exists } -\vec{v} \in V, \text{ such that: } \vec{v} \oplus (-\vec{v}) = \vec{\mathbf{0}}_V = (-\vec{v}) \oplus \vec{v}.$$

7. The Distributive Property of Ordinary Addition over Scalar Multiplication:

$$(r+s) \odot \vec{v} = (r \odot \vec{v}) \oplus (s \odot \vec{v}).$$

8. The Distributive Property of Vector Addition over Scalar Multiplication:

$$r \odot (\vec{u} \oplus \vec{v}) = (r \odot \vec{u}) \oplus (r \odot \vec{v}).$$

9. The Associative Property of Scalar Multiplication:

$$r \odot (s \odot \vec{v}) = (rs) \odot \vec{v} = s \odot (r \odot \vec{v}).$$

10. The Unitary Property of Scalar Multiplication:

$$1 \odot \vec{v} = \vec{v}.$$

Notice that we need ***three objects*** to define a vector space: (1) a non-empty set of ***vectors*** V, (2) a rule for ***vector addition*** $\oplus$ that tells us how to add two vectors to get another vector, and (3) a rule for ***scalar multiplication*** $\odot$ that tells us how to multiply a real number with a vector to get another vector. Also, note that the addition on the left side of Axiom 7 is the ordinary addition of real numbers, and similarly the middle multiplication in Axiom 9 is ordinary multiplication of real numbers. We also abbreviate the phrase "additive inverse" of a vector as its ***negative***.

Clearly, all Euclidean spaces $\mathbb{R}^n$ satisfy these Ten Axioms, as seen in our Theorem in Section 1.1.

Although we put a "circle" around the vector addition and scalar multiplication to distinguish them from ordinary addition and multiplication of real numbers, with practice and further experience, we will be shedding these circles and just write $\vec{u} + \vec{v}$ and $k \cdot \vec{v}$ when the context indicates which operation we are using.

Let us look at some examples of Vector Spaces that you are already familiar with from your study of Algebra and Calculus:

Polynomial Spaces

Consider the set of all ***polynomials*** in one variable with real coefficients and degree at most n. This set is denoted by:

$$\mathbb{P}^n = \{\, p(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n \mid a_0, a_1, a_2, \ldots, a_n \in \mathbb{R} \,\}.$$

We will define the vector addition of two polynomials in the usual way we add polynomials in ordinary algebra, and similarly with scalar multiplication. Thus, for example, if $p(x) = 3 - 5x + 7x^2$ and $q(x) = 4 - 3x^2 \in \mathbb{P}^2$, then:

$$p(x) \oplus q(x) = (3 - 5x + 7x^2) + (4 - 3x^2) = 7 - 5x + 4x^2, \quad \text{and}$$

$$3 \odot p(x) = 3(3 - 5x + 7x^2) = 9 - 15x + 21x^2.$$

It is easy to see that these operations are closed, and that addition is both commutative and associative. The zero vector for this space is:

$$\vec{\mathbf{0}}_{\mathbb{P}^n} = z(x) = 0 + 0x + \cdots + 0x^n$$

i.e., the ***zero polynomial***. Consequently, the negative of a polynomial is:

$$-p(x) = -a_0 - a_1 x - a_2 x^2 - \cdots - a_n x^n$$

with the desired property in Axiom 6. The rest of the axioms are also obviously satisfied.

Functions Spaces

If I is a non-empty interval on the real number line, we denote the set of all ***functions*** that are defined on I as:

$$F(I) = \left\{ f(x) \,|\, f(a) \text{ is defined for all } a \in I \right\}.$$

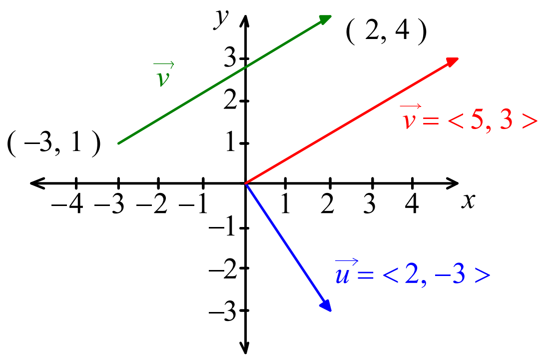
For example, $f(x) = \sqrt{x+2}$ and $g(x) = \ln(x) \in F((0,+\infty))$, but $h(x) = \dfrac{1}{x-1}$ is not, because it is undefined at $a = 1$. Notice also that ***all*** polynomials are members of $F(I)$ for all intervals I, since polynomials are defined for all real numbers. We will define vector addition and scalar multiplication in the natural way, as sums of two functions and the product of a real constant with a function as we do in algebra:

$$(f \oplus g)(x) = f(x) + g(x), \quad \text{and}$$
$$(k \odot f)(x) = k \cdot f(x).$$

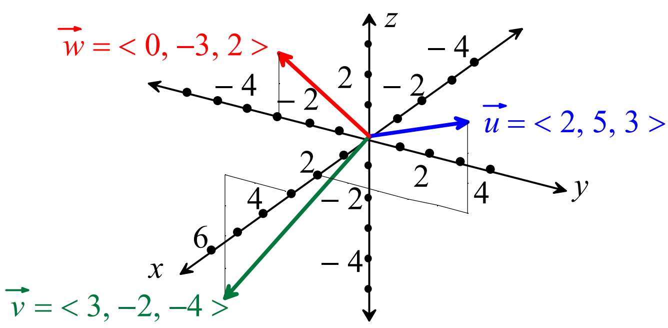
The zero vector is the function $z(x)$ that outputs the value 0 for all $a \in I$. We call this the ***zero function***, but note that its graph is exactly the same as the graph of the zero polynomial, and so we use the same symbol $z(x)$ for both. The negative of a function is the function that outputs as its value of $-f(a)$ when we input $x = a$. Again, all the Axioms for a vector space are easily verified.

How Can We Visualize Vectors?

In Chapter 1, we saw that vectors in $\mathbb{R}^2$ can be visualized or represented as arrows on the Cartesian plane, whereas vectors in $\mathbb{R}^3$ can be visualized as arrows in Cartesian space. Let us bring back the diagrams from Chapter 1 that demonstrate these representations:



Two Vectors, $\vec{u}$ and $\vec{v}$, in $\mathbb{R}^2$



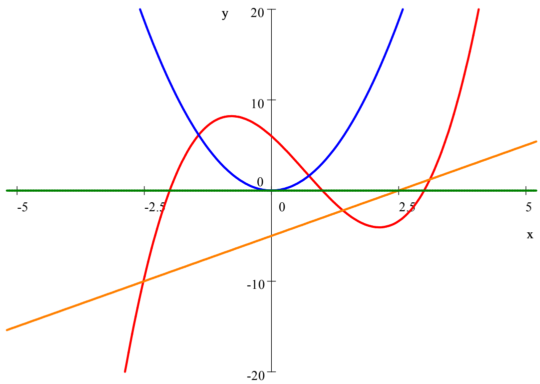
Three Vectors, $\vec{u}$, $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^3$

Unfortunately, it is not possible to visualize vectors in $\mathbb{R}^4$ or in any other Euclidean space of higher dimension, at least not in a simple way like our arrows above. The same can be said about vectors from more general vector spaces. It is sometimes possible to visualize them, while sometimes it is not.

The vectors from $\mathbb{P}^n$ can certainly be visualized because we know how to graph polynomials, or at least some simple ones. For example, in $\mathbb{P}^3$, let us consider the polynomials:

$$p(x) = (x+2)(x-1)(x-3),$$
$$q(x) = 3x^2, \text{ and}$$
$$r(x) = 2x-5.$$

We show these three vectors below, along with the zero polynomial $z(x)$:

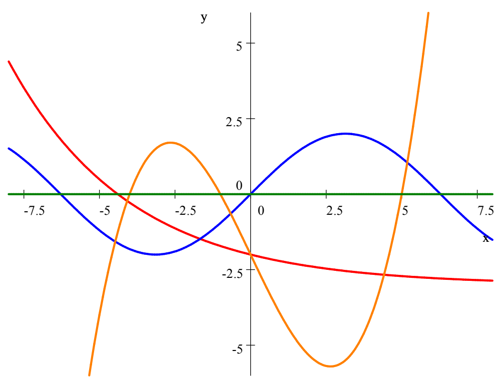


Four Vectors from $\mathbb{P}^3$

Similarly, we can show the functions:

$$f(x) = e^{-x/4} - 3, \quad g(x) = 2\sin\left(\frac{1}{2}x\right), \text{ and}$$
$$h(x) = \frac{1}{10}(x-5)(x+4)(x+1),$$

which are vectors from $F((-\infty,\infty))$, in the graphs below, along with the zero function $z(x)$:



Four Vectors from $F((-\infty,\infty))$

Notice that $h(x)$ is also a vector from $\mathbb{P}^3$. In the same way, $z(x)$ is also the ***zero function*** in all function spaces $F(I)$. This is a major advantage of abstract vector spaces, especially function spaces: there is no restriction to the number of spaces that a particular function can belong to. This does not happen to vectors from Euclidean spaces: the vector $\langle 3,-7\rangle$ is from $\mathbb{R}^2$, but it is ***not*** a vector from $\mathbb{R}^3$. On the other hand, $h(x)$ above is a member of $\mathbb{P}^3$ and $\mathbb{P}^4$ and any $\mathbb{P}^n$ with $n \geq 3$. In general, all polynomials $p(x)$ are members of all function spaces, such as $F((-\infty,\infty))$, $F((-\infty,3])$, where we restrict the domain of $p(x)$, and in fact any $F(I)$ where I is any interval of any type, since $p(x)$ is defined for all real numbers. To simplify our notation, we will use the symbol $z(x)$ to denote the zero function in ***all*** these function spaces, regardless of the domain.

Matrix Spaces

Our old friends, the set of all matrices of the same size, also form their own vector spaces. We denote them by the symbol:

$$Mat(m,n) = \{ A \mid A \text{ is an } m \times n \text{ matrix} \}.$$

We add two matrices in the same way as in Section 2.2, and similarly multiply a matrix by a scalar in the usual way. We can easily see that the zero matrix $\mathbf{0}_{m\times n}$ is the zero vector of this space, and we know how to construct the negative of a matrix as before. We can verify that all the Axioms for a vector space are satisfied, as we saw them proved in the properties in Section 2.4. As far as visualizing these vectors is concerned, this would again not be easy for matrices with more than three entries. In general, all we can do is stare at the entries, as illustrated below:

$$A = \begin{bmatrix} 3 & -5 & 2 \\ -1 & 4 & 7 \end{bmatrix}$$

A Single Vector A from $Mat(2,3)$

The Smallest Example

The definition of a vector space requires that V is not empty, so it has ***at least one*** vector. Is it possible for V to have ***exactly one*** member? Axiom 5 requires that this member be the zero vector $\vec{\mathbf{0}}_V$. Now, vector addition and scalar multiplication have to be closed, so we are forced to define:

$$\vec{\mathbf{0}}_V \oplus \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V, \text{ and } r \odot \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V \quad \text{for all } r \in \mathbb{R},$$

both of which make perfect sense from our experience with the zero vectors of $\mathbb{R}^n$. We must also impose that $-\vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V$, which again makes perfect sense. We can now easily check that the rest of the Axioms are also satisfied. Thus, we have a vector space $V = \{ \vec{\mathbf{0}}_V \}$ with exactly one member.

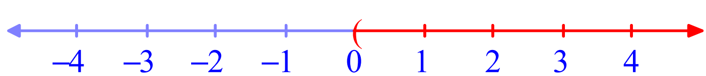
On the other hand, you will see in the Exercises that if you have a vector space with ***at least two*** vectors, then you will have an ***infinite number*** of vectors.

We're Not in Kansas Anymore

Let us take a look at a vector space that will challenge your idea of the word "natural." Consider the set:

$$\mathbb{R}^+ = \{ \vec{x} \mid x \in \mathbb{R}, \text{ and} x > 0 \},$$

the set of all ***positive*** real numbers. We can thus visualize this entire space as an ***interval***:



The Vector Space $\mathbb{R}^+$ (opening right)

To avoid confusion with scalars, we will put an ***arrow*** on top of our vectors, as is our usual notation. We will define ***vector addition*** by: $\vec{x} \oplus \vec{y} = \overrightarrow{xy}$ (ordinary multiplication).

We will define ***scalar multiplication*** by: $r \odot \vec{x} = \overrightarrow{x^r} = \overrightarrow{e^{r\ln(x)}}$ (ordinary exponentiation).

We note that the second formula is often derived in a Calculus course, and is needed in case r is an ***irrational*** number. Thus, for example:

$$\vec{3} \oplus \vec{5} = \overrightarrow{3 \cdot 5} = \overrightarrow{15},$$

$$3 \odot \vec{5} = \overrightarrow{5^3} = \overrightarrow{125}, \quad \text{and}$$

$$-\frac{1}{3} \odot \vec{8} = \overrightarrow{8^{-1/3}} = \overrightarrow{\frac{1}{2}}.$$

Clearly, these are strange, and you might say, unnatural ways to define "vector ***addition***" and "scalar ***multiplication***." However, notice that both operations yield ***positive*** numbers, so they are certainly ***closed***. Also, "addition" is both commutative and associative, because ordinary multiplication satisfies both properties. Now we have to scratch our heads a little and think about the ***zero vector***. Notice that the ***number*** 0 is not a member of $\mathbb{R}^+$. However, the correct question to ask is this: Can we find a positive number, let us call it $\vec{z}$, so that:

$$\vec{z} \oplus \vec{y} = \overrightarrow{zy} = \vec{y}$$

for ***all*** positive numbers $\vec{y}$? This number is 1! Thus: $\vec{\mathbf{0}}_{\mathbb{R}^+} = \vec{1}$.

How about the ***negative*** of a vector? Again, the negative real numbers are not in $\mathbb{R}^+$. However, the correct question to ask is this: If $\vec{x}$ is a positive number, can find another positive number $\vec{y}$ such that:

$$\vec{x} \oplus \vec{y} = \overrightarrow{xy} = \vec{1} = \vec{\mathbf{0}}_{\mathbb{R}^+}?$$

Solving this equation, we get $y = 1/x$, which is again positive. Thus: $-\vec{x} = \overrightarrow{1/x}$.

Axiom 6 is therefore satisfied. For example: $-\vec{5} = \overrightarrow{\frac{1}{5}}$.

This might look like a completely ridiculous equation, but remember that these are no longer numbers but ***vectors***. This says that the ***additive inverse*** of the ***vector*** $\vec{5}$ is the ***vector*** $\overrightarrow{\frac{1}{5}}$. We can check that:

$$\vec{5} \oplus \overrightarrow{\frac{1}{5}} = \overrightarrow{5\left(\frac{1}{5}\right)} = \vec{1} = \vec{\mathbf{0}}_{\mathbb{R}^+}.$$

Let us check the last four Axioms:

$$(r+s) \odot \vec{x} = \overrightarrow{x^{r+s}} = \overrightarrow{x^r x^s} = \overrightarrow{x^r} \oplus \overrightarrow{x^s} = (r \odot \vec{x}) \oplus (s \odot \vec{x})$$

$$r \odot (\vec{x} \oplus \vec{y}) = r \odot \left(\overrightarrow{xy}\right) = \overrightarrow{(xy)^r} = \overrightarrow{x^r y^r} = \overrightarrow{x^r} \oplus \overrightarrow{y^r} = (r \odot \vec{x}) \oplus (r \odot \vec{y})$$

$$(rs) \odot \vec{x} = \overrightarrow{x^{rs}} = \overrightarrow{(x^s)^r} = r \odot \overrightarrow{x^s} = r \odot (s \odot \vec{x}), \quad \text{and}$$

$$1 \odot \vec{x} = \overrightarrow{x^1} = \vec{x}.$$

Thus $\mathbb{R}^+$ is a vector space under multiplication and exponentiation. This example shows that sometimes you have to be ***open minded*** to see things as they are, both in life as well as in mathematics.

Additional Properties of Vector Spaces

The Ten Axioms for a vector space allow us to prove other general properties that are true for ***any*** vector space. These properties might give one a sense of déjà vu, but we remind you that we proved them only for the ***Euclidean Spaces***. Let us begin with:

> ***Theorem — The Uniqueness of the Zero Vector:***
> The ***zero vector*** $\vec{\mathbf{0}}_V$ of any vector space $(V, \oplus, \odot)$ is ***unique***. This means that if $\vec{z} \in V$ is another vector that satisfies: $\vec{z} \oplus \vec{v} = \vec{v}$ for ***all*** $\vec{v} \in V$, then we must have: $\vec{z} = \vec{\mathbf{0}}_V$.

Proof: We are allowed to use only the Ten Axioms for a vector space and we must completely forget about Euclidean spaces as we prove this Theorem.

Suppose that $\vec{z} \in V$ is a vector with the magical property given above. This property also holds if we substitute $\vec{v} = \vec{\mathbf{0}}_V$ in both sides of the equation. Thus, we get: $\vec{z} \oplus \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V$.

But since $\vec{\mathbf{0}}_V$ is the zero vector, by Axiom 5, we also have: $\vec{z} \oplus \vec{\mathbf{0}}_V = \vec{z}$.

Thus we must have: $\vec{z} = \vec{z} \oplus \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V$. Notice that we only used Axiom 5 in this proof.■

Similarly, we will leave the proof of the following as an Exercise:

> ***Theorem — The Uniqueness of Additive Inverses:***
> The ***additive inverse*** $-\vec{v}$ of any vector $\vec{v} \in V$ in a vector space $(V, \oplus, \odot)$ is ***unique***. This means that if $\vec{n} \in V$ is another vector that satisfies: $\vec{v} \oplus \vec{n} = \vec{\mathbf{0}}_V$, then we must have: $\vec{n} = -\vec{v}$.
> As a further consequence: $-\vec{v} = -1 \odot \vec{v}$.

We have the usual bonus properties of the zero vector and multiplication by the scalar 0:

> ***Theorem — The Multiplicative Properties of Zeroes:***
> Let $(V, \oplus, \odot)$ be a vector space, with zero vector $\vec{\mathbf{0}}_V$. Then we have the following properties:
> 1. ***The Multiplicative Property of the Scalar Zero:***
> $$0 \odot \vec{v} = \vec{\mathbf{0}}_V \text{ for all } \vec{v} \in V.$$
> 2. ***The Multiplicative Property of the Zero Vector:***
> $$r \odot \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V \text{ for all } r \in \mathbb{R}.$$
> 3. ***The Zero-Factors Theorem:*** For all $\vec{v} \in V$ and $r \in \mathbb{R}$:
> $$r \odot \vec{v} = \vec{\mathbf{0}}_V \textit{ if and only if } \text{ either } r = 0 \textit{ or } \vec{v} = \vec{\mathbf{0}}_V.$$

Notice that this Zero-Factors Theorem generalizes that in Chapters Zero and 1. Also in Chapter 1, we were able to define the notion of ***parallel*** vectors in $\mathbb{R}^n$ by generalizing the picture in $\mathbb{R}^2$ and $\mathbb{R}^3$: two vectors are parallel to each other if one of them is a scalar multiple of the other. Consequently, the zero vector $\vec{\mathbf{0}}_n$ will be parallel to all vectors. This further motivates us to generalize this concept to any vector space:

Definition — Axiom for Parallel Vectors:
Let $(V, \oplus, \odot)$ be a vector space, and let $\vec{u}, \vec{v} \in V$. We say that $\vec{u}$ and $\vec{v}$ are ***parallel to each other*** if there exists either $a \in \mathbb{R}$ ***or*** $b \in \mathbb{R}$ such that:

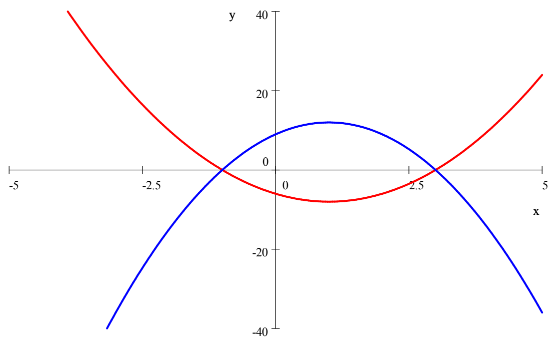
$$\vec{u} = a \odot \vec{v} \quad \textbf{\textit{or}} \quad \vec{v} = b \odot \vec{u}.$$

Consequently, this means that $\vec{\mathbf{0}}_V$ is parallel to ***all*** vectors $\vec{v} \in V$, since $\vec{\mathbf{0}}_V = 0 \odot \vec{v}$.

Example: Consider the polynomials $p(x) = 2x^2 - 4x - 6$ and $q(x) = -3x^2 + 6x + 9$ from $\mathbb{P}^2$. Note that we can factor them completely as:

$$p(x) = 2(x+1)(x-3), \text{ and } q(x) = -3(x+1)(x-3).$$

Thus, $q(x) = -\frac{3}{2}p(x)$. We see the graphs of these quadratics below:



Parallel Vectors $p(x) = 2x^2 - 4x - 6$, and $q(x) = -3x^2 + 6x + 9$

We can see from these pictures that the meaning of "parallel vectors" in $\mathbb{P}^2$ (or in other function spaces, for that matter), is not the same as it is in basic algebra. These two functions are ***not*** vertical (or horizontal) shifts of each other. We do see, though, that one is a ***dilation*** or ***contraction*** of the other (along with a ***reflection*** across the x-axis). In particular, these quadratic polynomials have exactly the ***same roots***, and so they must be scalar multiples of each other.□

Things Don't Always Work Out

It is easy to get carried away and start believing that we can define vector addition and scalar multiplication in almost any way possible and still satisfy the Ten Axioms of a vector space. Of course, this is not true, otherwise vector spaces will not be special.

Example: Suppose $V = Mat(2,3)$, with vector addition defined as matrix addition, as before. However, we will define scalar multiplication by:

$$r \odot A = r \odot \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} = \begin{bmatrix} ra_{1,1} & ra_{1,2} & ra_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix},$$

for all $r \in \mathbb{R}$, that is, we will only multiply the entries in the ***first row*** by r. Axioms 1 through 6 are easily seen to be satisfied, since aside from Axiom 2 (which is clearly satisfied), these 6 Axioms only have to do with matrix addition. Now, the last four Axioms are a bit trickier. Axiom 10 looks easy to check: $1 \odot A = A$, is clearly satisfied. Now let's check Axiom 7: $(r+s) \odot A = (r \odot A) \oplus (s \odot A)$ for all $r, s \in \mathbb{R}$ and all $A \in Mat(2,3)$.

The left side is:

$$(r+s) \odot A = (r+s) \odot \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} = \begin{bmatrix} (r+s)a_{1,1} & (r+s)a_{1,2} & (r+s)a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}.$$

The right side is:

$$(r \odot A) \oplus (s \odot A) = \begin{bmatrix} ra_{1,1} & ra_{1,2} & ra_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix} + \begin{bmatrix} sa_{1,1} & sa_{1,2} & sa_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{bmatrix}$$

$$= \begin{bmatrix} (r+s)a_{1,1} & (r+s)a_{1,2} & (r+s)a_{1,3} \\ 2a_{2,1} & 2a_{2,2} & 2a_{2,3} \end{bmatrix}.$$

Thus, Axiom 7 is ***not*** satisfied by all matrices. Therefore, this is ***not*** a vector space. □

We also want to note that if $(V, \oplus, \odot)$ does not contain a zero vector $\vec{0}_V$, then automatically, there is no such thing as the additive inverse $-\vec{v}$ of a vector, because the equation $\vec{v} + (-\vec{v}) = \vec{\mathbf{0}}_V$ does not make any sense.

Example: Suppose we let $V = \mathbb{R}^2$, but with addition defined by:

$$\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle 2x_1 + 2x_2, y_1 + y_2 \rangle,$$

and scalar multiplication defined as usual. Let us see if this space contains a zero vector. We ***cannot*** automatically assume that the only possible choice for the zero vector is $\vec{\mathbf{0}}_2 = \langle 0, 0 \rangle$, since our addition is now ***different***. Thus, let us suppose that: $\vec{\mathbf{0}}_V = \langle a, b \rangle$, for some $a, b \in \mathbb{R}$.
If $\langle x, y \rangle$ is any vector in $\mathbb{R}^2$, we must satisfy the equations:

$$\langle x, y \rangle \oplus \langle a, b \rangle = \langle x, y \rangle = \langle a, b \rangle \oplus \langle x, y \rangle.$$

But then: $\langle x, y \rangle \oplus \langle a, b \rangle = \langle 2x + 2a, 2y + 2b \rangle = \langle a, b \rangle \oplus \langle x, y \rangle$.
Thus, we need to satisfy the two equations:

$$2x + 2a = x, \text{ and } 2y + 2b = y, \text{ so we get: } a = -x/2, \text{ and } b = -y/2.$$

This means that if we pick two ***different*** vectors $\langle x_1, y_1 \rangle$ and $\langle x_2, y_2 \rangle$, then we will need two different solutions to $\langle a, b \rangle$. This means that there is no ***single*** vector $\langle a, b \rangle$ that can satisfy the equation $\langle x, y \rangle \oplus \langle a, b \rangle = \langle x, y \rangle$ for ***all*** $\langle x, y \rangle$. We saw in the previous sub-section that the zero vector, if it exists, has to be ***unique***, and thus V does not contain a zero vector. Consequently, we cannot speak of an additive inverse $-\vec{v}$ for a vector $\vec{v}$, either. □

Note: Sometimes, we try to construct new vectors spaces by changing only the vector addition or the scalar multiplication of a known vector space $(V, \oplus, \odot)$. If we only change the vector addition, then Axioms 2, 9 and 10 are still valid, since these only involve scalar multiplication. Thus, for the previous Example, we do not need to check these three Axioms. If we only change scalar multiplication, then Axioms 1, 3, 4, 5 and 6 are still valid, since these only involve vector addition. Axioms 7 and 8 will always have to be re-checked if either addition or scalar multiplication is changed.

3.1 Section Summary

We say that $(V, \oplus, \odot)$ is a ***vector space*** if V is a non-empty set, and the operations $\oplus$ and $\odot$ satisfy the Ten Axioms for a vector space: for all $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$ and all r, $s \in \mathbb{R}$, $(V, \oplus, \odot)$ satisfies:

1. $\vec{u} \oplus \vec{v} \in V$; 2. $r \odot \vec{u} \in V$; 3. $\vec{u} \oplus \vec{v} = \vec{v} \oplus \vec{u}$; 4. $(\vec{u} \oplus \vec{v}) \oplus \vec{w} = \vec{u} \oplus (\vec{v} \oplus \vec{w})$;
5. There exists $\vec{\mathbf{0}}_V \in V$, such that: $\vec{\mathbf{0}}_V \oplus \vec{v} = \vec{v} = \vec{v} \oplus \vec{\mathbf{0}}_V$;
6. There exists $-\vec{v} \in V$ such that: $\vec{v} \oplus (-\vec{v}) = \vec{\mathbf{0}}_V = (-\vec{v}) \oplus \vec{v}$;
7. $(r+s) \odot \vec{v} = (r \odot \vec{v}) \oplus (s \odot \vec{v})$; 8. $r \odot (\vec{u} \oplus \vec{v}) = (r \odot \vec{u}) \oplus (r \odot \vec{v})$;
9. $r \odot (s \odot \vec{v}) = s \odot (r \odot \vec{v}) = (rs) \odot \vec{v}$; 10. $1 \odot \vec{v} = \vec{v}$.

Some examples of abstract vector spaces include:

- ***polynomials spaces***, $\mathbb{P}^n$, consisting of polynomials of degree at most n;
- ***function spaces***, $F(I)$, consisting of functions defined on an interval I;
- the set of all $m \times n$ ***matrices*** $Mat(m, n)$;

all under their naturally defined vector addition and scalar multiplication.

The ***zero vector*** $\vec{\mathbf{0}}_V$ of any vector space $(V, \oplus, \odot)$ is ***unique***. This means that if $\vec{z} \in V$ is another vector that satisfies $\vec{z} \oplus \vec{v} = \vec{v}$ for ***all*** $\vec{v} \in V$ then we must have: $\vec{z} = \vec{\mathbf{0}}_V$.

The ***additive inverse*** $-\vec{v}$ of any vector $\vec{v} \in V$ in a vector space $(V, \oplus, \odot)$ is ***unique***. This means that if $\vec{n} \in V$ is another vector that satisfies $\vec{v} \oplus \vec{n} = \vec{\mathbf{0}}_V$, then we must have: $\vec{n} = -\vec{v}$.
Furthermore, $-\vec{v} = -1 \odot \vec{v}$.

Let $(V, \oplus, \odot)$ be a vector space, with zero vector $\vec{\mathbf{0}}_V$, let $\vec{v} \in V$, and let $r \in \mathbb{R}$. Then:

1. $0 \odot \vec{v} = \vec{\mathbf{0}}_V$. 2. $r \odot \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V$. 3. $r \odot \vec{v} = \vec{\mathbf{0}}_V$ ***if and only if*** either $r = 0$ ***or*** $\vec{v} = \vec{\mathbf{0}}_V$.

Let $(V, \oplus, \odot)$ be a vector space, and let $\vec{u}$, $\vec{v} \in V$. We say that $\vec{u}$ and $\vec{v}$ are ***parallel to each other*** if there exists either $a \in \mathbb{R}$ ***or*** $b \in \mathbb{R}$ such that: $\vec{u} = a \odot \vec{v}$ ***or*** $\vec{v} = b \odot \vec{u}$.

Consequently, this means that $\vec{\mathbf{0}}_V$ is parallel to ***all*** vectors $\vec{v} \in V$, since $\vec{\mathbf{0}}_V = 0 \odot \vec{v}$.

3.1 Exercises

1. Consider the set of all 2×2 ***diagonal*** matrices: $D_2 = \left\{ \begin{bmatrix} d_1 & 0 \\ 0 & d_2 \end{bmatrix} \middle| \, d_1, d_2 \in \mathbb{R} \right\}$,

   under ordinary matrix addition and scalar multiplication.

   a. Prove that D_2 is a vector space under these two operations.

   b. Consider the set of all $n \times n$ diagonal matrices:

   $$D_n = \left\{ \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \middle| \, d_1, d_2, \ldots, d_n \in \mathbb{R} \right\},$$

   under ordinary matrix addition and scalar multiplication. Generalize your proof and notation in (a) to show that D_n is a vector space under these two operations for any n.

2. Repeat Exercise 1 to show that the set of all 2×2 ***upper triangular*** matrices:

$$U_2 = \left\{ \begin{bmatrix} u_{11} & u_{12} \\ 0 & u_{22} \end{bmatrix} \middle| \, u_{11}, u_{12}, u_{22} \in \mathbb{R} \right\},$$

is a vector space under ordinary matrix addition and scalar multiplication. Generalize your proof and notation to show that the set of all $n \times n$ upper triangular matrices:

$$U_n = \left\{ \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \middle| \, u_{ij} \in \mathbb{R}, i = 1 \ldots n, j = i \ldots n \right\},$$

is a vector space under ordinary matrix addition and scalar multiplication.

3. Consider the set of all 2×2 ***symmetric*** matrices:

$$Sym(2) = \left\{ \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix} \middle| \, a_{11}, a_{12}, a_{22} \in \mathbb{R} \right\},$$

again under ordinary matrix addition and scalar multiplication. (Notice that the entries off the diagonal are equal.)

a. Show that $Sym(2)$ is a vector space under these two operations.

b. Review the properties of the transpose operation and symmetric $n \times n$ matrices at the end of Section 2.9.

c. Use these properties in order to prove that the set of all $n \times n$ symmetric matrices:

$$Sym(n) = \{ A \in Mat(n,n) \, | \, A = A^{\top} \}$$

is a vector space under ordinary matrix addition and scalar multiplication. Denote the $n \times n$ symmetric matrices by A, B, C etc. instead of explicitly writing the entries of each matrix like you did in Exercise 1 and 2.

4. We saw that the set: $\mathbb{R}^+ = \left\{ \vec{x} \, | \, x \in \mathbb{R}, \text{and} \, x > 0 \right\}$ is a vector space under the operations:

$$\vec{x} \oplus \vec{y} = \overrightarrow{xy} \quad \text{and} \quad r \odot \vec{x} = \overrightarrow{x^r} = \overrightarrow{e^{r\ln(x)}}.$$

Let us construct a new vector space: $\mathbb{R}_2^+ = \left\{ \langle x_1, x_2 \rangle \, | \, x_1, x_2 \in \mathbb{R}, \text{and} \, x_1 > 0, x_2 > 0 \right\}$ under the operations: $\langle x_1, x_2 \rangle \oplus \langle y_1, y_2 \rangle = \langle x_1 y_1, x_2 y_2 \rangle$, and $r \odot \langle x_1, x_2 \rangle = \langle x_1^r, x_2^r \rangle$.
Show that $\mathbb{R}_2^+$ is a vector space under these two operations. In the course of checking the Ten Axioms, you may use the fact that the analogous Axioms have been proven to be true for $\mathbb{R}^+$.

For Exercises (5) to (7): Explain why the following are ***not*** vector spaces:

5. The set of ***non-negative*** real numbers under ordinary addition and scalar multiplication.

6. The set of ***integers*** under ordinary addition and scalar multiplication.

7. The set of all $n \times n$ ***invertible*** matrices, under the usual matrix addition and scalar multiplication.

For Exercises (8) to (19): For each of these problems, the vector space $V = \mathbb{R}^2$ is modified, by changing the vector addition and/or scalar multiplication. Decide whether or not V is still a vector space. If it is not a vector space, specify ***all*** the Axioms that are violated. Review the Note on p. 278 to save some time if only one of the operations is changed. To make sure that you understand the indicated vector addition and/or scalar multiplication, do the warm-up computations first and check your answer before proceeding to check the Ten Axioms.

8. The usual vector addition, but with scalar multiplication changed to: $k \odot \langle x, y \rangle = \langle kx, y \rangle$.
   Warm-up: compute $-3 \odot \langle 5, -2 \rangle$.

9. The usual vector addition, but with scalar multiplication defined by: $k \odot \langle x, y \rangle = \langle -kx, -ky \rangle$.
   Warm-up: compute $-3 \odot \langle 5, -2 \rangle$.

10. Change the vector addition to: $\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle x_1 + x_2 - 2, y_1 + y_2 + 1 \rangle$, but keep the usual scalar multiplication. Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$.

    Hint/warning: To determine if $\mathbb{R}^2$ has a zero vector under this addition, suppose $\vec{0}_V = \langle a, b \rangle$, ***not necessarily*** $\langle 0, 0 \rangle$. Solve for a and b in the equation that $\vec{0}_V$ is supposed to satisfy in Axiom 5. Similarly, what should be the equation that $-\vec{v} = \langle x', y' \rangle$ should satisfy in Axiom 6?

11. Change the vector addition to: $\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle x_1 + x_2, y_1 - y_2 \rangle$, but keep the usual scalar multiplication. Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$.

12. Change the vector addition to: $\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle -x_1 - x_2, -y_1 - y_2 \rangle$, but keep the usual scalar multiplication. Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$.

13. Change the vector addition to: $\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle x_1 + y_2, x_2 + y_1 \rangle$, but keep the usual scalar multiplication. Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$.

14. Change both vector addition and scalar multiplication to:
    $$\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle x_1 + x_2, 2y_1 + 2y_2 \rangle; \quad k \odot \langle x, y \rangle = \langle kx, 2ky \rangle.$$
    Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$ and $-3 \odot \langle 5, -2 \rangle$.

15. Change both vector addition and scalar multiplication to:
    $$\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle x_1 + x_2, 2y_1 + 2y_2 \rangle; \quad k \odot \langle x, y \rangle = \langle 2kx, ky \rangle.$$
    Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$ and $-3 \odot \langle 5, -2 \rangle$.
    Compare this with the previous Exercise.

16. Change both vector addition and scalar multiplication to:
    $$\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle y_1 + y_2, x_1 + x_2 \rangle; \quad k \odot \langle x, y \rangle = \langle ky, kx \rangle.$$
    Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$ and $-3 \odot \langle 5, -2 \rangle$.

17. Change both vector addition and scalar multiplication to:
    $$\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle -x_1 - x_2, y_1 + y_2 \rangle; \quad k \odot \langle x, y \rangle = \langle kx, -ky \rangle.$$
    Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$ and $-3 \odot \langle 5, -2 \rangle$.

18. Change both vector addition and scalar multiplication to:
    $$\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \langle x_1 + x_2, 0 \rangle, \text{ and } k \odot \langle x, y \rangle = \langle kx, 0 \rangle.$$
    Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$ and $-3 \odot \langle 5, -2 \rangle$.

19. Change both vector addition and scalar multiplication to:

$$\langle x_1, y_1 \rangle \oplus \langle x_2, y_2 \rangle = \left\langle x_1 + x_2 - 2,\ y_1 + y_2 + 3 \right\rangle, \text{ and } k \odot \left\langle x,\ y \right\rangle = \left\langle kx + 2,\ ky - 3 \right\rangle.$$

Warm-up: compute $\langle 7, -3 \rangle \oplus \langle 2, 6 \rangle$ and $-3 \odot \langle 5, -2 \rangle$.

20. Let V be the set of all ***rational functions***, that is, functions of the form: $r(x) = \dfrac{p(x)}{q(x)}$,

where $p(x)$ and $q(x)$ are ordinary polynomials, with no common factor, under the usual addition of functions and scalar multiplication by a real constant:

$$\frac{p_1(x)}{q_1(x)} \oplus \frac{p_2(x)}{q_2(x)} = \frac{p_1(x) \cdot q_2(x) + p_2(x) \cdot q_1(x)}{q_1(x) q_2(x)}, \text{ and } k \odot \frac{p(x)}{q(x)} = \frac{k \cdot p(x)}{q(x)}.$$

The final answer may be reduced by cancelling out common factors in the numerator and denominator, as usual. Decide whether or not V is a vector space.

Warm-up: Compute $\dfrac{3}{x+3} \oplus \dfrac{2x+24}{x^2-9}$ and $-3 \odot \dfrac{2x-5}{x+3}$.

Further Properties of Vector Spaces: For Exercises (21) to (29): Prove the following properties. You are only allowed to use the Ten Axioms for a Vector Space. You are ***not*** allowed to rely on the coordinates that we have for vectors in $\mathbb{R}^n$. You may only assume that $(V, \oplus, \odot)$ is any abstract vector space, with zero vector $\vec{\mathbf{0}}_V$, $r \in \mathbb{R}$ and $\vec{v} \in V$, and all Ten Axioms are satisfied by $(V, \oplus, \odot)$.

21. Prove ***The Multiplicative Property of the Scalar Zero:*** $0 \odot \vec{v} = \vec{\mathbf{0}}_V$.

Hint: use the fact that $0 + 0 = 0$.

22. Which of the Ten Axioms directly implies that $\vec{\mathbf{0}}_V \oplus \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V$?

What does this equation say about $-\vec{\mathbf{0}}_V$?

23. Use the previous Exercise to prove ***The Multiplicative Property of the Zero Vector:***

$$r \odot \vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V.$$

24. Prove that if $r \odot \vec{v} = \vec{\mathbf{0}}_V$, then ***either*** $r = 0$ or $\vec{v} = \vec{\mathbf{0}}_V$.

Hint: Review the similar property from Chapter 1.

25. Use the previous Exercises to give a ***complete*** proof of ***The Zero Factors Theorem:***

$$r \odot \vec{v} = \vec{\mathbf{0}}_V \textit{ if and only if } \text{either } r = 0 \text{ or } \vec{v} = \vec{\mathbf{0}}_V.$$

26. ***The Uniqueness of Additive Inverses:*** Prove that the additive inverse $-\vec{v}$ of any vector $\vec{v} \in V$ in a vector space $(V, \oplus, \odot)$ is ***unique***. This means that if $\vec{n} \in V$ is another vector that satisfies:

$$\vec{v} \oplus \vec{n} = \vec{\mathbf{0}}_V,$$

then we must have: $\vec{n} = -\vec{v}$.

27. Use some of the Exercises above to prove that for any vector $\vec{v} \in V$ in a vector space $(V, \oplus, \odot)$:

$$-\vec{v} = (-1) \odot \vec{v}.$$

28. In the Axiom for Parallel Vectors, prove that if $\vec{u}$ and $\vec{v}$ are ***non-zero*** vectors from a vector space V that are ***parallel*** to each other, then the a and the b in the definition must both exist and be ***non-zero scalars***, and furthermore, $a = 1/b$.

29. Prove that if V has ***at least two distinct vectors***, then it has an ***infinite*** number of vectors. Hint: since there are two vectors, one of them, say $\vec{v}$, must be ***non-zero***. Use some of the Exercises above (especially the Zero Factors Theorem) to show that if $x \neq y$ are real numbers, then $x \odot \vec{v} \neq y \odot \vec{v}$. Explain why this creates an infinite family of vectors.

30. ***The Direct Sum of Vector Spaces:*** Suppose that $(V, \oplus_V, \odot_V)$ and $(W, \oplus_W, \odot_W)$ are two vector spaces. Note that we distinguish that addition of each space with the corresponding subscript, and similarly with the scalar multiplication of each space. Let us define the ***direct sum*** of V and W as:

$$(V, W) = \left\{ (\vec{v}, \vec{w}) \,|\, \vec{v} \in V \text{ and } \vec{w} \in W \right\}.$$

In other words, (V, W) consists of all ordered pairs of vectors, where the first vector is from V, and the second vector is from W. We will define addition and scalar multiplication in (V, W) by:

$$(\vec{v}_1, \vec{w}_1) \oplus (\vec{v}_2, \vec{w}_2) = (\vec{v}_1 \oplus_V \vec{v}_2, \vec{w}_1 \oplus_W \vec{w}_2) \quad \text{and} \quad k \odot (\vec{v}, \vec{w}) = (k \odot_V \vec{v}, k \odot_W \vec{w}).$$

Prove that (V, W) under the vector addition and scalar multiplication as defined above is a vector space. Think carefully about the zero vector and additive inverse in (V, W).

By induction, if U_1 through U_m are vector spaces with their corresponding vector additions and scalar multiplications, we can construct the vector space:

$$(U_1, U_2, \ldots, U_m) = \left\{ (\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m) \,|\, \vec{u}_i \in U_i \text{ for } i = 1 \ldots m \right\}$$

under the corresponding vector addition and scalar multiplication.

In particular, if n is a positive integer, we can define:

$$V^n = (V, V, \ldots, V),$$

in the same way that we construct the Euclidean space $\mathbb{R}^n$.

31. ***Vector Spaces of Linear Transformations:*** Consider the set of all linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$, denoted:

$$\boldsymbol{Hom}(\mathbb{R}^n, \mathbb{R}^m) = \left\{ T : \mathbb{R}^n \to \mathbb{R}^m \,|\, T \text{ is a linear transformation} \right\}.$$

This space denotes the set of all vector space ***homomorphisms*** from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Define vector addition as:

$$T_1 \oplus T_2 : \mathbb{R}^n \to \mathbb{R}^m \quad \text{is the linear transformation with action:}$$

$$(T_1 \oplus T_2)(\vec{v}) = T_1(\vec{v}) + T_2(\vec{v}),$$

where the addition on the right is the addition in $\mathbb{R}^m$ (i.e. this is the definition of $T_1 + T_2$ from Chapter 2). Similarly, define scalar multiplication as:

$$k \odot T : \mathbb{R}^n \to \mathbb{R}^m \quad \text{is the linear transformation with action:}$$

$$(k \odot T)(\vec{v}) = k \cdot T(\vec{v}),$$

where the scalar multiplication on the right is the scalar multiplication in $\mathbb{R}^m$ (again, this is the definition of $k \cdot T$ from Chapter 2). Prove that $\boldsymbol{Hom}(\mathbb{R}^n, \mathbb{R}^m)$ forms a vector space under these two operations.

3.2 Linearity Properties for Finite Sets of Vectors

Let us now proceed with generalizing our fundamental constructions from Chapter 1 to abstract vector spaces. We begin with the two most important constructions:

Linear Combinations and Spans of Finite Sets

As with ordinary vectors from n-space, we can generalize the concept of a linear combination and the Span of a set of vectors from any vector space. We begin by considering only ***finite*** sets of vectors:

Definition: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors from a vector space $(V, \oplus, \odot)$, and let $c_1, c_2, \ldots, c_n \in \mathbb{R}$. Then, a ***linear combination*** of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ with ***coefficients*** $c_1, c_2, \ldots, c_n$ is an expression of the form:

$$(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n).$$

Similarly, the ***Span*** of the set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is the set of ***all possible linear combinations*** of these vectors:

$$\begin{aligned} Span(S) &= Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}) \\ &= \{(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) \,|\, c_1, c_2, \ldots, c_n \in \mathbb{R}\}. \end{aligned}$$

If it is clear that we understand the scalar multiplication and addition in V, we simply write:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n$$

for the linear combination, and similarly use this expression when defining Spans.

Unfortunately, in general, the concept of a Span of even a small set of vectors is an abstract concept, and often cannot be visualized. Let us see some easily grasped examples.

Example: The vector space $\mathbb{P}^n$ consists of all polynomials of degree at most n, and therefore every member of $\mathbb{P}^n$ can be written as:

$$\begin{aligned} &c_0 + c_1x + c_2x^2 + \cdots + c_nx^n \\ &= c_0 \bullet 1 + c_1 \bullet x + c_2 \bullet x^2 + \cdots + c_n \bullet x^n. \end{aligned}$$

Thus, every polynomial of degree at most n is a linear combination of the monomials $1, x, x^2, \ldots, x^n$. Therefore, we can write:

$$\mathbb{P}^n = Span(\{1, x, x^2, \ldots, x^n\}). \square$$

We can extrapolate from this Example that many vector spaces can be described as the Span of a set of vectors. We will be seeing more of this in the next two Sections.

Example: Consider $\mathbb{P}^5 = \{p(x) = c_0 + c_1x + c_2x^2 + \cdots + c_5x^5 \,|\, c_0, c_1, c_2, \ldots, c_5 \in \mathbb{R}\}$.

We can form:

$$Span(\{1, x^2, x^4\}) = \{p(x) = c_0 + c_2x^2 + c_4x^4 \,|\, c_0, c_2, c_4 \in \mathbb{R}\},$$

which we should recognize as the set of all ***even*** polynomials of $\mathbb{P}^5$, that is, those polynomials $p(x)$ of $\mathbb{P}^5$ that satisfy $p(-x) = p(x)$. Similarly, we can construct:

$$Span(\{x, x^3, x^5\}) = \{p(x) = c_1x + c_3x^3 + c_5x^5 \,|\, c_1, c_3, c_5 \in \mathbb{R}\},$$

as the set of all ***odd*** polynomials of $\mathbb{P}^5$, that is, those polynomials $p(x)$ that satisfy $p(-x) = -p(x)$. $\square$

Membership in A Span

In Chapter 1, we determined whether or not a particular vector $\vec{b}$ is a member of $Span(S)$ for some set of vectors S from some Euclidean space $\mathbb{R}^n$. We can certainly ask a similar question if S is a set of vectors from some abstract vector space V. Unfortunately, testing for membership in $Span(S)$ can be a tricky question. Let us begin by looking at Spans from polynomial spaces $\mathbb{P}^n$. We will first need the following old and important Theorem:

Theorem — The Fundamental Theorem of Algebra:
Every non-constant ***polynomial*** $p(x)$ (that is, of degree $n \geq 1$), with complex (or possibly real) coefficients, has exactly n complex ***roots***, counting multiplicities.

This should be familiar from Precalculus, where we see, for example, that a cubic polynomial has three roots, counting multiplicities, and can thus be factored into three linear factors. For example, if:

$$p(x) = 2x^3 - 13x^2 - 22x + 105$$

we have the factorization:

$$p(x) = (2x-5)(x+3)(x-7)$$

(which one can check to be correct by expanding). Thus, the roots are $x = 5/2$, -3 and 7. Of course, some cubics have nasty irrational roots, imaginary roots, or repeated roots.

The Fundamental Theorem of Algebra has the following consequence:

Theorem — Equality of Polynomials: Suppose that:

$$p(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n \text{ and } q(x) = d_0 + d_1x + d_2x^2 + \cdots + d_nx^n.$$

Then, ***as functions***, $p(x) = q(x)$ ***if and only if*** $c_0 = d_0,\ c_1 = d_1,\ \ldots,\ c_n = d_n$.
Consequently:

$$p(x) = c_0 + c_1x + c_2x^2 + \cdots + c_nx^n = z(x),$$

the zero polynomial, ***if and only if*** $c_0 = 0,\ c_1 = 0,\ \ldots,\ c_n = 0$.

Note: We say that $p(x) = q(x)$ ***as functions*** if the values of the two functions agree for ***all*** real numbers $a \in \mathbb{R}$, that is:

$$p(a) = q(a) \text{ for all } a \in \mathbb{R}.$$

The Theorem above seems completely obvious, but this is far from the truth.

Proof: ($\Leftarrow$) Only the converse is obvious. If $p(x)$ and $q(x)$ have exactly the same coefficients as polynomials, then if we input the number a into either polynomial, we get the same value, that is, $p(a) = q(a)$ for all $a \in \mathbb{R}$.

($\Rightarrow$) Suppose $p(x)$ and $q(x)$ are as written in the statement of the Theorem, and $p(a) = q(a)$ for all real numbers a. We have to show that the coefficients of $p(x)$ and $q(x)$ are exactly the same. Subtracting the two equations, we get:

$$\begin{aligned} p(x) - q(x) &= (c_0 + c_1x + c_2x^2 + \cdots + c_nx^n) - (d_0 + d_1x + d_2x^2 + \cdots + d_nx^n) \\ &= (c_0 - d_0) + (c_1 - d_1)x + (c_2 - d_2)x^2 + \cdots + (c_n - d_n)x^n \end{aligned}$$

Suppose we call $r(x)$ the polynomial $r(x) = p(x) - q(x)$. Since $p(a) = q(a)$ for all $a \in \mathbb{R}$, $r(a) = p(a) - q(a) = 0$ for ***all*** $a \in \mathbb{R}$. Thus $r(x)$ has an ***infinite*** number of roots. Let us now show that all the coefficients c_i must be the same as d_i, starting with the highest degree. Suppose $c_n \neq d_n$. Then $r(x)$ has degree n, and by the Fundamental Theorem of Algebra, $r(x)$ has exactly n roots. This contradicts the fact that $r(x)$ has an infinite number of roots. Thus $c_n = d_n$. Now that we know that $c_n - d_n = 0$, we can apply the same reasoning to c_{n-1} and d_{n-1}: if these two coefficients were different, $r(x)$ will have degree $n-1$ and thus exactly $n-1$ roots, leading to the same contradiction. Continuing thus, we can see that $c_1 = d_1$, and finally, $c_0 = d_0$. ■

Now, we are ready to determine if a polynomial is a member of the Span of a set of polynomials. We will be able to exploit the Gauss-Jordan Algorithm, as before, for this type of problem.

Example: Let $S = \{3 - x + 2x^3, 5 + 3x - 4x^2, 2 + x - x^3\} \subseteq \mathbb{P}^3$. Consider:

$$p(x) = 16 - 3x + 8x^2 + x^3 \in \mathbb{P}^3.$$

Let us try to decide whether or not $p(x)$ is a member of $Span(S)$. To do this, we have to find coefficients c_1, c_2 and c_3, if possible, such that:

$$c_1(3 - x + 2x^3) + c_2(5 + 3x - 4x^2) + c_3(2 + x - x^3) = 16 - 3x + 8x^2 + x^3.$$

Expanding the left side and collecting like terms, we get:

$$\begin{aligned} &3c_1 - c_1x + 2c_1x^3 + 5c_2 + 3c_2x - 4c_2x^2 + 2c_3 + c_3x - c_3x^3 \\ &= 3c_1 + 5c_2 + 2c_3 - c_1x + 3c_2x + c_3x - 4c_2x^2 + 2c_1x^3 - c_3x^3 \\ &= 3c_1 + 5c_2 + 2c_3 + (-c_1 + 3c_2 + c_3)x - 4c_2x^2 + (2c_1 - c_3)x^3. \end{aligned}$$

According to the Equality of Polynomials Theorem, we have to satisfy the system of (linear!) equations:

$$\begin{array}{rcrcrcr} 3c_1 & + & 5c_2 & + & 2c_3 & = & 16 \\ -c_1 & + & 3c_2 & + & c_3 & = & -3 \\ & & -4c_2 & & & = & 8 \\ 2c_1 & & & - & c_3 & = & 1 \end{array}$$

As in Chapter 1, we can form the corresponding augmented matrix:

$$\begin{bmatrix} 3 & 5 & 2 & | & 16 \\ -1 & 3 & 1 & | & -3 \\ 0 & -4 & 0 & | & 8 \\ 2 & 0 & -1 & | & 1 \end{bmatrix} \text{ with rref: } \begin{bmatrix} 1 & 0 & 0 & | & 4 \\ 0 & 1 & 0 & | & -2 \\ 0 & 0 & 1 & | & 7 \\ 0 & 0 & 0 & | & 0 \end{bmatrix}.$$

Thus, there is exactly one solution:

$$c_1 = 4, \; c_2 = -2, \; \text{and} \; c_3 = 7.$$

We can indeed check that:

$$4(3 - x + 2x^3) - 2(5 + 3x - 4x^2) + 7(2 + x - x^3) = 16 - 3x + 8x^2 + x^3 = p(x).$$

Thus, $p(x)$ is a member of $Span(S)$. □

Linear Independence of a Finite Set of Vectors

The concept of the linear dependence or independence of a finite set of vectors is virtually identical to the concept from Chapter 1, with only the notation needing modification:

> ***Definition:*** Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors from a vector space $(V, \oplus, \odot)$. We say that S is ***linearly independent*** if the only solution to the equation:
>
> $$(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) = \vec{\mathbf{0}}_V$$
>
> is the ***trivial solution*** $c_1 = 0, c_2 = 0, \ldots, c_n = 0$. As before, we will refer to this equation as a ***dependence test equation*** and sometimes just say "independent" to mean linearly independent. The opposite of being linearly independent is being ***linearly dependent***, which means there is a ***non-trivial solution*** to the dependence test equation, that is, where at least one c_i is ***non-zero***.
>
> In particular, if $S = \{f_1(x), f_2(x), \ldots, f_n(x)\}$ is a subset of $F(I)$ for some interval I, then S is linearly independent if the only solution to the dependence test equation:
>
> $$c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) = z(x), \quad \text{for } \textbf{\textit{all}} \;\; x \in I,$$
>
> is the trivial solution $c_1 = 0, c_2 = 0, \ldots, c_n = 0$. We note that it may be possible to satisfy the equation above for ***some*** $x \in I$ using non-zero coefficients c_i, but the equation needs to be satisfied for ***all*** $x \in I$.

Since we are no longer dealing with Euclidean n-spaces, we do not have the luxury of using the coordinates of a vector or a linear combination in order to solve for the coefficients c_1 through c_n. However, as we shall see, there are often ideas from Algebra, Trigonometry or Calculus that will enable us to show that only the trivial solution exists, especially in the case of sets of functions.

Example: Suppose that $S = \{\sin(x), \cos(x), \sin(2x)\}$, a set of three vectors from $F((-\infty, \infty))$. Although we have the famous ***double angle*** formula:

$$\sin(2x) = 2\sin(x)\cos(x),$$

the right side is ***not*** a linear combination of $\sin(x)$ and $\cos(x)$, and therefore this formula is ***irrelevant***.

Let us consider the dependence test equation:

$$c_1 \sin(x) + c_2 \cos(x) + c_3 \sin(2x) = z(x).$$

Recall that $z(x)$, the function that is identically zero for ***any*** value of x, is the zero vector $\vec{\mathbf{0}}_V$ for $V = F((-\infty, \infty))$. Let us see if there is a non-trivial solution to this dependence test equation. Since this equation must be true for ***all*** values of x, our strategy to solve this equation is to substitute some ***convenient*** values for x. For example, suppose $x = 0$. Since $\sin(0) = 0$, and $\cos(0) = 1$, we get: $0 + c_2 + 0 = 0$.

Thus, $c_2 = 0$ is the ***only*** possible solution to c_2. Now we are down to solving:

$$c_1 \sin(x) + c_3 \sin(2x) = z(x).$$

This time, let us substitute $x = \pi/2$. Since $\sin(\pi/2) = 1$ and $\sin(\pi) = 0$, we get: $c_1 + 0 = 0$.

Thus, $c_1 = 0$, and we are left with $c_3 \sin(2x) = z(x)$. Since $\sin(2x)$ is not the zero function, we have $c_3 = 0$ by the Zero-Factors Theorem. Thus, $c_1 = c_2 = c_3 = 0$, so S is linearly independent.□

The last step in our proof leads us to generalize a similar statement from Chapter 1:

Theorem: Let $(V, \oplus, \odot)$ be a vector space, and $\vec{v} \in V$. Then: $S = \{\vec{v}\}$ is ***linearly independent*** ***if and only if*** $\vec{v} \neq \vec{\mathbf{0}}_V$.

Similarly, we have the following generalization regarding sets of two vectors:

Theorem: Let $(V, \oplus, \odot)$ be a vector space, and $\vec{v}_1,\ \vec{v}_2 \in V$. Then: $S = \{\vec{v}_1, \vec{v}_2\}$ is ***linearly independent*** ***if and only if*** $\vec{v}_1$ and $\vec{v}_2$ are ***not parallel*** to each other.

We will leave the proof of both Theorems as Exercises. As before, once we have three or more vectors, it becomes a tricky problem to determine whether or not a set of vectors is linearly independent. We can generalize the last Theorem, though, in the following, whose proof we also leave as an Exercise:

Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors from a vector space $(V, \oplus, \odot)$. Then: S is ***linearly dependent*** ***if and only if*** at least one vector in S (which, without loss of generality, we can set to be $\vec{v}_1$) is a linear combination of the other vectors of S, that is:

$$\vec{v}_1 = (c_2 \odot \vec{v}_2) \oplus (c_3 \odot \vec{v}_3) \oplus \cdots \oplus (c_n \odot \vec{v}_n),$$

for some scalars $c_2,\ c_3,\ \ldots,\ c_n \in \mathbb{R}$.

The technique of substituting particular values of x from the common domain I of the functions in S, is very useful in determining whether or not S is linearly independent. Let us look at another set containing trigonometric functions and see a different strategy:

Example: Let us continue with the same space $V = F((-\infty, \infty))$, and consider:

$$S = \{\cos(2x),\ \sin^2(x),\ \cos^2(x)\}.$$

We know the Double Angle Formula: $\cos(2x) = \cos^2(x) - \sin^2(x)$, and thus:

$$1 \cdot \cos(2x) + 1 \cdot \sin^2(x) - 1 \cdot \cos^2(x) = z(x)$$

is a true equation with non-zero coefficients. In fact, the Double Angle Formula itself tells us that $\cos(2x)$ is a linear combination of $\sin^2(x)$ and $\cos^2(x)$, and thus S is linearly ***dependent*** by the Theorem above. □

Let us turn now to an Example involving polynomials. Once again, the Fundamental Theorem of Algebra will be useful:

Example: Let $S = \{7x^4 + 2x^3 - 5x^2 + 3x - 2, 2x^3 - x^2 + 3x + 5, 2x - 7, 3\}$. A linear combination of the vectors of S looks like:

$$\begin{aligned} & c_1(7x^4 + 9x^3 - 5x^2 + 3x - 2) + c_2(2x^3 - x^2 + 3x + 5) + c_3(2x - 7) + k_4(3) \\ = {} & 7c_1x^4 + 9c_1x^3 - 5c_1x^2 + 3c_1x - 2c_1 + 2c_2x^3 - c_2x^2 + 3c_2x + 5c_2 + 2c_3x - 7c_3 + 3c_4 \\ = {} & 7c_1x^4 + (9c_1 + 2c_2)x^3 + (-5c_1 - c_2)x^2 + (2c_3 + 3c_1 + 3c_2)x - 7c_3 + 3c_4 - 2c_1 + 5c_2 \end{aligned}$$

Like our earlier Example, this polynomial is $z(x)$ ***if and only if*** all of the coefficients of the individual monomials are 0. From the last line above where we collected common terms, we must have:

$$\begin{array}{rcrcrcrcl} 7c_1 & & & & & & & = & 0 \\ 9c_1 & + & 2c_2 & & & & & = & 0 \\ -5c_1 & - & c_2 & & & & & = & 0 \\ 3c_1 & + & 3c_2 & + & 2c_3 & & & = & 0 \\ -2c_1 & + & 5c_2 & - & 7c_3 & + & 3c_4 & = & 0 \end{array}$$

From the first equation, we immediately see that $c_1 = 0$. But since this is the case, we also get $c_2 = 0$ using the second and third equations. But now that c_1 and c_2 are both 0, we next get $c_3 = 0$ from the fourth equation, and finally $c_4 = 0$ from the fifth equation. Thus, all the coefficients must be 0, and S is ***linearly independent***.

Now that we know that we can only have the trivial linear combination in order to produce $z(x)$ let us see if we can show this without expanding the linear combination:

$$c_1(7x^4 + 9x^3 - 5x^2 + 3x - 2) + c_2(2x^3 - x^2 + 3x + 5) + c_3(2x - 7) + c_4(3).$$

The key observation here is that the polynomials are in the order of ***decreasing degrees***. The only polynomial that contributes x^4 is the first polynomial, and thus to make the coefficient of x^4 zero, we must have $c_1 = 0$. Now we are left with a shorter linear combination:

$$c_2(2x^3 - x^2 + 3x + 5) + c_3(2x - 7) + c_4(3)$$

By the same reasoning as before, we must have $c_2 = 0$, because the first polynomial above is the only one that contributes x^3. Proceeding thus, we also get $c_3 = 0$, and finally $c_4 = 0$. □

Clearly, this reasoning works as long as all the polynomials are of different degrees:

Theorem: Suppose $S = \{p_1(x), p_2(x), \ldots, p_k(x)\}$ is a set of polynomials from $\mathbb{P}^n$ with ***distinct*** degrees. Then S is ***linearly independent***. In particular, the set $\{1, x, x^2, \ldots, x^n\}$ is linearly independent.

Now let us see an Example where Calculus can be useful:

Example: Let us go back to the space $V = F((-\infty, \infty))$, this time with $S = \{e^x, e^{3x}, e^{7x}\}$. Let us see whether or not this set is linearly independent. We set up the dependence test equation:

$$c_1 e^x + c_2 e^{3x} + c_3 e^{7x} = z(x)$$

and determine if this has a non-trivial solution. Let us play a little trick: ***divide*** both sides of the equation above by e^x, which is allowable because e^x is strictly ***positive*** for any real number x. We get:

$$c_1 + c_2 e^{2x} + c_3 e^{6x} = z(x)/e^x = z(x).$$

Now, we know that e^{2x} and e^{6x} both approach 0 as x approaches $-\infty$, that is:

$$\lim_{x \to -\infty} e^{2x} = 0 = \lim_{x \to -\infty} e^{6x}.$$

From this, we get:

$$\lim_{x \to -\infty} (c_1 + c_2 e^{2x} + c_3 e^{6x}) = \lim_{x \to -\infty} z(x), \text{ and thus:}$$

$$c_1 + 0 + 0 = 0.$$

So we get $c_1 = 0$. Similarly, by dividing both sides of the new equation:

$$c_2 e^{2x} + c_3 e^{6x} = z(x)$$

by e^{2x} and taking the same limit, we get $c_2 = 0$. Finally, since e^{6x} is obviously not the zero vector, we must have $c_3 = 0$. Thus this set S is linearly independent.□

It is easy to see that we can repeat this argument for any set of functions of the form e^{kx}, so we leave the proof of the following as an Exercise:

> ***Theorem:*** Suppose $S = \{e^{k_1 x}, e^{k_2 x}, \ldots, e^{k_n x}\}$, where $k_1 < k_2 < \cdots < k_n$ are n ***distinct*** real numbers. Then S is ***linearly independent***.

Example: We saw a set of exponential functions in the previous Example. Let us now take a look at a set of logarithmic functions. Suppose:

$$S = \{log_2(x), log_5(x), log_7(x)\}$$

It is not obvious whether or not these functions are linearly independent. However, we do know the ***change of base formula:***

$$log_a(x) = \frac{log_b(x)}{log_b(a)}$$

for all bases $a,\ b > 0,\ a,\ b \neq 1$. Since this is valid for any of the three bases above, let us pick $b = 2$ and convert the other two functions. Thus:

$$log_5(x) = \frac{log_2(x)}{log_2(5)} \quad \text{and} \quad log_7(x) = \frac{log_2(x)}{log_2(7)}$$

Thus, both $log_5(x)$ and $log_7(x)$ are scalar multiples of $log_2(x)$, and we can therefore conclude that S is ***linearly dependent***.□

So far, we have seen only examples where the functions have a natural domain which is an interval such as $(-\infty, \infty)$ or $(0, \infty)$. However, we have to be careful when we are dealing with ***piecewise defined*** functions, because the dependence test equation must be satisfied on the entire (common) domain of all the functions involved, and not just one interval.

Example: Consider the two piecewise defined functions:

$$f(x) = \begin{cases} \frac{1}{4}x + 1 & \text{if } x \in [0,2) \\ 2 - x & \text{if } x \in [2,3] \end{cases}, \quad \text{and} \quad g(x) = \begin{cases} \frac{1}{2}x + 2 & \text{if } x \in [0,2) \\ x - 2 & \text{if } x \in [2,3] \end{cases}.$$

Note that $f(x), g(x) \in F([0,3])$. We easily notice that on the interval $[0,2)$:

$$g(x) = \frac{1}{2}x + 2 = 2\left(\frac{1}{4}x + 1\right) = 2f(x),$$

so on $[0,2)$, $g(x)$ is parallel to $f(x)$. However, on $[2,3]$:

$$g(x) = x - 2 = -(2 - x) = -f(x),$$

so on $[2,3]$, $g(x)$ is also parallel to $f(x)$, but with a ***different*** proportionality constant. Since we cannot find a ***single*** constant k such that:

$$g(x) = k \cdot f(x) \ \text{ for } \boldsymbol{all}\ x \in [0,3],$$

we conclude that $S = \{f(x), g(x)\}$ is an ***independent*** set from $F([0,3])$. However, S is a ***dependent*** set in $F([0,2))$ and in $F([2,3])$.

In Section 1.6, we showed the various connections and relationships between the concepts of the Span of a set of vectors and whether or not a set of vectors is linearly independent. Obviously, the Theorems we proved in that Section generalize to abstract vector spaces, so we will find similar ones in the Exercises, such as ***The Equality of Spans Theorem***, ***The Elimination Theorem***, ***The Extension Theorem***, and ***The Dependent/Independent Sets from Spanning Sets Theorem***. We will see a general version of ***The Minimizing Theorem*** in Section 3.6.

3.2 Section Summary

Let $(V, \oplus, \odot)$ be a vector space, and $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in V$. A ***linear combination*** of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ with ***coefficients*** $c_1, c_2, \ldots, c_n$ has the form:

$$(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n).$$

Similarly, the ***Span*** of the set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is the set of ***all possible linear combinations*** of these vectors:

$$\begin{aligned} Span(S) &= Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}) \\ &= \{(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) \mid c_1, c_2, \ldots, c_n \in \mathbb{R}\}. \end{aligned}$$

We say that S is ***linearly independent*** if any linear combination of vectors from S results in the zero vector if and only if all the coefficients of these vectors are 0. This means that the only solution to the ***dependence test equation***: $(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) = \vec{\mathbf{0}}_V$, is the ***trivial solution:*** $c_1 = 0, c_2 = 0, \ldots, c_n = 0$.

If we have a ***non-trivial solution***, we say that S is ***linearly dependent***, and an equation above with at least one ***non-zero*** coefficient is called a ***dependence equation*** for S.

$S = \{\vec{v}\}$ is ***linearly independent*** ***if and only if*** $\vec{v} \neq \vec{\mathbf{0}}_V$.

$S = \{\vec{v}_1, \vec{v}_2\}$ is ***linearly independent*** ***if and only if*** $\vec{v}_1$ and $\vec{v}_2$ are ***not parallel*** to each other.

Suppose $S = \{p_1(x), p_2(x), \ldots, p_n(x)\}$ is a set of polynomials of ***different degrees***. Then S is ***linearly independent***.

Similarly, if $S = \{e^{k_1 x}, e^{k_2 x}, \ldots, e^{k_n x}\}$, where $k_1, k_2, \ldots, k_n$ are ***distinct*** real numbers, then S is ***linearly independent***.

We may try to determine if S is linearly dependent or independent by substituting particular values for x in the dependence equation for S and solving for the coefficients.

3.2 Exercises

For Exercises (1) to (6): Decide whether the given vector is a member of $Span(S)$ for each given set S. If so, express it as a linear combination from S:

1. $S = \{6 + 3x - 4x^2, 5 - 2x + 7x^2\} \subset \mathbb{P}^2$; $p(x) = -7 + 19x - 47x^2$.
2. $S = \{2 - 4x + 5x^3, 7 + 3x^2 - 2x^3\} \subset \mathbb{P}^3$; $p(x) = 105 - 28x + 39x^2 + 9x^3$.
3. $S = \left\{\dfrac{1}{x}, \dfrac{1}{x^2}, \dfrac{1}{x^3}\right\} \subset F((0,\infty))$; $f(x) = \dfrac{2x^2 - 7x - 10}{x^3}$.
4. $S = \{2 - 3x + 4x^2, 5 + 7x - 2x^3, 4 + 6x - 5x^2 - 2x^3\} \subset \mathbb{P}^3$; $p(x) = 22 + 11x - 39x^2 - 13x^3$.
5. $S = \left\{\dfrac{1}{x+1}, \dfrac{1}{x-2}\right\} \subset F((3,\infty))$; $f(x) = \dfrac{4x + 25}{(x+1)(x-2)}$
6. $S = \left\{\dfrac{1}{x-1}, \dfrac{1}{x+3}\right\} \subset F((2,\infty))$; $f(x) = \dfrac{5x^2 + 8x + 3}{(x-1)(x+3)}$
7. Show that: $\left\{\dfrac{1}{x-1}, \dfrac{1}{x+1}, \dfrac{x}{x^2-1}\right\} \subset F((-1,1))$ is linearly dependent.

   Note that these three functions are all defined on $(-1,1)$.
8. Show that $\{\sin(x), \cos(x), \tan(x)\} \subset F((-\pi/2, \pi/2))$ is linearly independent.
9. Show that $\{x^{1/2}, x^{1/3}, x^{1/4}\} \subset F([0,\infty))$ is linearly independent.

   For Exercises (10) to (34): Decide if the set of functions is independent or dependent, and prove your answer:
10. $\{x^2, x^2 - 1, x^2 + 3\} \subset F((-\infty,\infty))$.
11. $\{x^2, (x-1)^2, (x+3)^2\} \subset F((-\infty,\infty))$.
12. $\{\cos^2(x), \sin^2(x), 1\} \subset F((-\infty,\infty))$.
13. $\{\cos^2(x), \sin^2(x), \sin(2x)\} \subset F((-\infty,\infty))$.
14. $\{\cos^2(x), \sin^2(x), \cos(2x)\} \subset F((-\infty,\infty))$.
15. $\{\cos^{-1}(x), \sin^{-1}(x), 1\} \subset F([-1,1])$.
16. $\{\tan^2(x), \sec^2(x), 1\} \subset F((-\pi/2, \pi/2))$.
17. $\{\cot^2(x), \csc^2(x), 3\} \subset F((0,\pi))$.

18. $\{\cos^{-1}(x),\ \sin^{-1}(x),\ \tan^{-1}(x)\} \subset F([-1,1])$.

19. $\left\{\dfrac{1}{x},\ \dfrac{1}{x+1},\ \dfrac{1}{x-2}\right\} \subset F((2,\infty))$.

20. $\left\{\dfrac{1}{x},\ \dfrac{1}{x^2},\ \dfrac{1}{x^3}\right\} \subset F((0,\infty))$.

21. $\{e^x, xe^x, x^2e^x, x^3e^x\} \subset F((-\infty,\infty))$.

22. $\left\{\dfrac{1}{x+5},\ \dfrac{1}{x+4},\ \dfrac{x+2}{(x+5)(x+4)}\right\} \subset F((-4,\infty))$.

23. $\{e^x, e^{x-1}, e^{x-2}, e^{x-3}\} \subset F((-\infty,\infty))$.

24. $\left\{\sin(x),\ \sin\left(x-\frac{\pi}{6}\right),\ \sin\left(x-\frac{\pi}{3}\right)\right\} \subset F((-\infty,\infty))$.

25. $\left\{\sqrt{x-1},\ \sqrt{x-2},\ \sqrt{x-3}\right\} \subset F((3,\infty))$.

26. $\{\log_3(x),\ \log_5(x),\ \log_8(x)\} \subset F((0,\infty))$.

27. $\left\{\dfrac{1}{x^2},\ \dfrac{1}{(x+1)^2},\ \dfrac{1}{(x-2)^2}\right\} \subset F((2,\infty))$.

28. $\{x-x^2, x^2-x^3, x-x^3\} \subset F((-\infty,\infty))$.

29. $\{1+x,\ 1+x^2,\ 1+x^3, x+x^2, x+x^3, x^2+x^3\} \subset F((-\infty,\infty))$.

30. $\left\{\dfrac{1}{x+2},\ \dfrac{1}{x+1},\ \dfrac{1}{x-1},\ \dfrac{6x^2+19x+23}{(x+2)(x+1)(x-1)}\right\} \subset F((1,\infty))$.

31. $\left\{\dfrac{1}{x-1},\ \dfrac{1}{x+1},\ \dfrac{1}{x+2},\ \dfrac{9x+5}{(x+2)(x+1)(x-1)}\right\} \subset F((1,\infty))$.

32. $\left\{\dfrac{1}{x+1},\ \dfrac{1}{(x+1)^2},\ \dfrac{1}{x-2},\ \dfrac{9x^2-7x-4}{(x+1)^2(x-2)}\right\} \subset F((2,\infty))$.

33. $\left\{\dfrac{1}{x+3},\ \dfrac{1}{(x+3)^2},\ \dfrac{1}{x-1},\ \dfrac{1}{(x-1)^2},\ \dfrac{6x^3+10x^2-14x-82}{(x+3)^2(x-1)^2}\right\} \subset F((1,\infty))$.

34. $\left\{\dfrac{1}{x+2},\ \dfrac{1}{x+1},\ \dfrac{1}{x-1},\ \dfrac{5x+2}{(x-2)(x+1)(x-1)}\right\} \subset F((2,\infty))$.

35. Use the idea in one of the Examples to prove that the set:
$$S = \{e^{k_1x}, e^{k_2x}, \ldots, e^{k_nx}\} \subset F((-\infty,\infty))$$
is linearly independent, for any set of ***distinct*** real numbers $k_1 < k_2 < k_3 < \cdots < k_n$.

36. Show that the set of functions $\{3^x, 5^x, 8^x\} \subset F((-\infty,\infty))$ is linearly independent. Hint: modify slightly the idea behind the previous Exercise.

37. Generalize the previous Exercise: Show that if $b_1, b_2, \ldots, b_n$ are n ***distinct*** positive numbers, none of them equal to 1, (and you may assume that they are in ascending order), then:
$$\{b_1^x, b_2^x, \ldots, b_n^x\}$$
is a linearly independent set of functions.

38. Prove that the set $S = \left\{\dfrac{1}{x-a},\ \dfrac{1}{x-b},\ \dfrac{cx+d}{(x-a)(x-b)}\right\} \subset F((b,\infty))$ is ***dependent*** for any real numbers a, b, c, d, where $a < b$.

39. ***The Domain Matters:*** Consider the functions:

$$f(x) = x^2 \text{ and } g(x) = x|x|.$$

Both functions are defined over all $\mathbb{R}$, and so we can also consider them as functions defined on a smaller interval.

a. Show that $\{f(x), g(x)\}$, restricted to $[0,\infty)$, is a dependent set.
b. Show that $\{f(x), g(x)\}$, restricted to $(-\infty,0]$, is a dependent set.
c. Show that $\{f(x), g(x)\}$, defined over all $\mathbb{R}$, is an independent set.
d. Is $\{f(x), g(x)\}$, restricted to $(-3,2)$, a dependent or independent set?

40. Prove that any set S from a vector space V, that contains $\vec{\mathbf{0}}_V$, is ***linearly dependent***.

41. Let $(V,\oplus,\odot)$ be a vector space, and $\vec{v} \in V$. Prove that the set $S = \{\vec{v}\}$ is ***linearly independent if and only if*** $\vec{v} \neq \vec{\mathbf{0}}_V$.

42. Let $(V,\oplus,\odot)$ be a vector space, and $\vec{v}_1,\ \vec{v}_2 \in V$. Prove that the set $S = \{\vec{v}_1, \vec{v}_2\}$ is ***linearly independent if and only if*** $\vec{v}_1$ and $\vec{v}_2$ are ***not parallel*** to each other.

43. Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a set of vectors from a vector space $(V,\oplus,\odot)$. Prove that S is ***linearly dependent if and only if*** at least one vector (which, without loss of generality, we can set to be $\vec{v}_1$) is a ***linear combination*** of $\vec{v}_2,\ \vec{v}_3,\ \ldots,\ \vec{v}_n$, that is:

$$\vec{v}_1 = (c_2 \odot \vec{v}_2) \oplus (c_3 \odot \vec{v}_3) \oplus \cdots \oplus (c_n \odot \vec{v}_n),$$

for some scalars $c_2,\ c_3,\ \ldots,\ c_n \in \mathbb{R}$.

44. Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***linearly independent*** set of vectors from a vector space $(V,\oplus,\odot)$. Prove that any ***subset*** of S is also ***linearly independent***.

For Exercises (45) to (48): Review the corresponding Theorem from Section 1.6 and rewrite their proofs in the language and notation of a general vector space.

45. ***The Equality of Spans Theorem:*** Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ and $S^{/} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be two sets of vectors from some vector space V. Prove that $Span(S) = Span(S^{/})$ ***if and only if*** every $\vec{v}_i$ can be written as a ***linear combination*** of the $\vec{w}_1$ through $\vec{w}_m$, ***and*** every $\vec{w}_j$ can also be written as a ***linear combination*** of the $\vec{v}_1$ through $\vec{v}_n$.

46. ***The Elimination Theorem:*** Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***linearly dependent*** set of vectors from a vector space $(V,\oplus,\odot)$, with dependence equation:

$$(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) = \vec{\mathbf{0}}_V,$$

where we assume without loss of generality that $c_n \neq 0$. Prove that:

$$Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}) = Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-1}\}),$$

that is, we can ***eliminate*** one vector $\vec{v}_n$ from S, and the smaller set still has the ***same Span*** as the original set S.

47. ***The Extension Theorem:*** Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***linearly independent*** set of vectors from a vector space $(V,\oplus,\odot)$, and let $\vec{w} \in V$ be a vector which is ***not*** in $Span(S)$. Prove that the ***extended*** set $S^{/} = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{w}\}$ is still ***linearly independent***.

48. ***The Dependent/Independent Sets from Spanning Sets Theorem:*** Let $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be any set of vectors from a vector space V, and we form $Span(S)$. Suppose now we randomly choose a set of m vectors from $Span(S)$ to form a new set: $L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$.
Prove that ***if*** $m > n$, ***then*** L is automatically linearly ***dependent***.
Consequently, ***if*** L is ***independent***, ***then*** $m \leq n$.

3.3 Linearity Properties for Infinite Sets of Vectors

Our next goal is to define the concepts of linear combinations, Spans and linear independence for ***infinite*** sets of vectors. For example, we will study the infinite set of ***monomials:***

$$S = \{1, x, x^2, x^3, \ldots, x^n, \ldots\}.$$

These are nuanced concepts, and so we must first introduce key concepts from ***Set Theory***, which is the subject at the very core of Mathematics. We will only present the concepts that are relevant to Linear Algebra. A more thorough treatment is usually found in an advanced course such as ***Logic*** or ***Analysis***.

A Primer on Infinite Sets

Definition: A non-empty set X is ***finite*** if the number of elements in the set is finite, that is, a positive integer n. In other words, we can choose to ***list*** the elements of X in some particular ***order***, say:

$$X = \{x_1, x_2, \ldots, x_n\},$$

where the list eventually ***terminates***. In this case, we call n the ***cardinality*** of our set, and we use the notation: $|X| = n$, pronounced as "the cardinality of X is n."
We agree that the ***empty set*** has cardinality 0, and we also consider it to be a finite set.
A set that is ***not*** finite is called an ***infinite set***.

Examples: Some of the best Examples of infinite sets are those sets of numbers that we saw in the early part of Chapter Zero. We started with the most important infinite set: the ***natural numbers:***

$$\mathbb{N} = \{0, 1, 2, 3, 4, \ldots\}.$$

The next infinite set of numbers that we saw was the set of all ***integers***, where we enlarge the set $\mathbb{N}$ to include the ***negative integers:***

$$\mathbb{Z} = \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

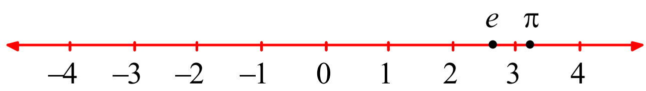
The letter $\mathbb{Z}$ stands for ***Zahlen***, which is German for "number." These sets of numbers are easy to grasp using the ***decimal or base 10 system*** to tell us what an integer represents. We know what the number 537 means, and we know that 537 is not the same as 735, and neither is 537 the same as -537. The digits $\{0,1,2,3,4,5,6,7,8,9\}$, the symbol $(-)$ for negatives, and a knowledge of ***place value*** (one's digit, ten's digit, hundred's digit, and so on) are enough to define the decimal system.

By dividing pairs of integers, we further enlarge $\mathbb{Z}$ to get the set of fractions or ***rational numbers:***

$$\mathbb{Q} = \left\{ \frac{a}{b} \mid a \text{ and } b \text{ are integers, with } b \neq 0 \right\},$$

where $\mathbb{Q}$ stands for ***quotients***. We agree that b is the lowest denominator possible.

Finally, we make a big jump and end up with the set of real numbers $\mathbb{R}$, which contains not just the rational numbers, but also irrational numbers such as e, π, $\sqrt{2}$, $\sqrt{3}$, $\sqrt[3]{2}$, and so on. We can again use the decimal system, this time with the use of a ***decimal point***. A real number that has either a repeating or terminating decimal representation, such as 0.783 or $5.21212121\ldots$ is a rational number, but a real number with a non-repeating, non-terminating decimal representation is an ***irrational number***. As before, we can visualize $\mathbb{R}$ using points on the real number line:



The Real Number Line With Some Members of $\mathbb{R}$

Notice that we have a ***nesting*** of our four sets of numbers:

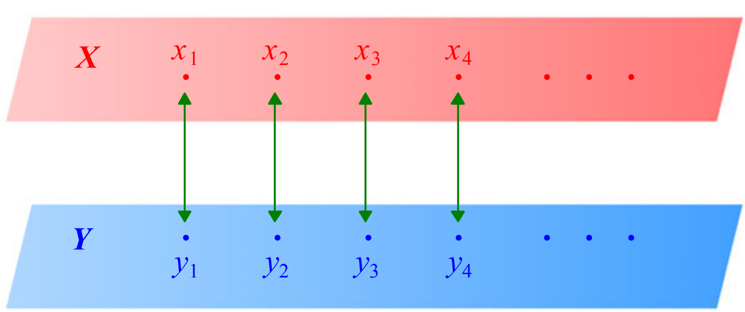
$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

These are ***strict*** nestings, in the sense that there are members of $\mathbb{Z}$ that are ***not*** members of $\mathbb{N}$, there are members of $\mathbb{Q}$ that are ***not*** members of $\mathbb{Z}$, and there are members of $\mathbb{R}$ that are ***not*** members of $\mathbb{Q}$, namely, the ***irrational numbers***, denoted $\mathbb{R} - \mathbb{Q}$.□

In the same way that finite sets come in different cardinalities, infinite sets likewise come in different cardinalities. Thus, our next step is to define the concept of the ***cardinality*** of an infinite set X, again denoted $|X|$. The cardinality of X is basically a measure of the ***size*** of a set. We cannot say anymore that the cardinality of an infinite set is "the number of elements" in the set, because ***infinity is not a number***. To avoid some technical issues, we will use the following Theorem as a definition that will let us decide when two sets have the same cardinality:

Definition/Theorem — The Schroeder-Bernstein Theorem:
Suppose that X and Y are two sets (they can be finite or infinite).
Then: $|X| = |Y|$, that is, X and Y have ***equal cardinality***, ***if and only if*** there exists a function $f : X \to Y$ which is both ***one-to-one*** and ***onto***.

This Theorem tells us that for two sets X and Y to have the same cardinality, the members of X must be in a ***one-to-one correspondence*** with the members of Y: for every $x \in X$, there corresponds exactly one $y \in Y$, and for every $y \in Y$, there corresponds exactly one $x \in X$. Thus, we have a ***pairing*** (x, y), where every $x \in X$ appears ***exactly once***, and every $y \in Y$ appears ***exactly once***.



Two Infinite Sets of Equal Cardinality

Since the generic symbol ∞ is no longer a good symbol for the cardinalities of infinite sets, we will need a new symbol. We use the Hebrew letter $\aleph$ (***Aleph***), together with a subscript, to denote the cardinalities of infinite sets. The smallest infinite cardinality is that of $\mathbb{N}$, which will be denoted $\aleph_0$ (pronounced "Aleph nought" or "Aleph zero"), that is:

$$|\mathbb{N}| = \aleph_0.$$

Thus, any set for which we can find a one-to-one correspondence with $\mathbb{N}$ will also have cardinality $\aleph_0$. Since the members of $\mathbb{N}$ are $0,\ 1,\ 2,\ 3,\ ...$ and so on, to prove that a set X has cardinality $\aleph_0$, we must be able to construct a function $f : \mathbb{N} \to X$ which is both one-to-one and onto. In other words, we must be able to list the members of X in a ***sequence:***

$$X = \{x_0, x_1, x_2, x_3, \ldots\}, \text{ where } x_i = f(i),$$

and every element $x \in X$ appears ***exactly once*** on this list in order for f to be both one-to-one and onto. For this reason, we say that any set with cardinality $\aleph_0$ is ***countable***. It turns out that two other sets in our Examples above are also countable, but one of them is not.

Examples: As we saw above, we can list the set of integers $\mathbb{Z}$ as:

$$\mathbb{Z} = \{\ldots -3, -2, -1, 0, 1, 2, 3, \ldots\}.$$

In this form, though, the set goes to infinity both to the right ***and*** the left. However, we can list the members of $\mathbb{Z}$ in the following order instead:

$$0, 1, -1, 2, -2, 3, -3, \ldots, n, -n, \ldots$$

Notice that this list goes off to infinity only to the ***right***, so we have succeeded in writing the members of $\mathbb{Z}$ in a ***sequence***, and it is clear that every integer appears on this list exactly once. Thus:

$$|\mathbb{Z}| = \aleph_0 = |\mathbb{N}|.$$

We emphasize that this equality is ***not*** saying that $\mathbb{N}$ and $\mathbb{Z}$ are the same sets. This equation says that the sets $\mathbb{N}$ and $\mathbb{Z}$ have ***equal cardinality***. Again, this means that the members of $\mathbb{N}$ are in a ***one-to-one correspondence*** with the members of $\mathbb{Z}$.

It turns out that the cardinality of the set of rational numbers $\mathbb{Q}$ is also $\aleph_0$, but this time it is not so obvious how to write the rational numbers as a sequence. We will see this in the Exercises.

We will now show that the set of real numbers $\mathbb{R}$ is ***not*** countable, in other words, it is impossible to list ***all*** the real numbers in a sequence. To accomplish this, we will represent any real number in the usual decimal notation, just as we saw above. We must be careful, though, because a repeating 9 could be rounded up at the digit immediately to the left of the first repeated 9. For example:

$$0.782\overline{9} = 0.78299999\ldots = 0.783.$$

We will therefore agree to write this number as 0.783, which makes more sense. Now, suppose, for the sake of argument that we have a sequence of real numbers, and the first five numbers in our sequence are:

$$5.\boxed{7}38257\ldots,\ 3.0\boxed{4}2963\ldots,\ -8.21\boxed{5}732\ldots,\ 15.732\boxed{9}46\ldots,\ -9.3285\boxed{2}4\ldots.$$

We must show that this sequence is ***missing*** at least one number x. We will construct this missing number x digit by digit: to determine the *nth* digit of x, look at the *nth* digit to the right of the decimal point of the *nth* real number (which we have boxed in the numbers above) and ***add one*** to this digit (make this 0 if the *nth* digit is 9). Thus, according to our sequence above, the first few digits of x are:

$$x = 0.85603\ldots$$

By construction, x is ***not*** the first term, nor the second, nor the third, and so on, because the *nth* digit of x doesn't ***match*** the corresponding *nth* digit of the *nth* term! Thus, our sequence is missing this real number x. This reasoning can be used to construct a missing number from ***any*** sequence of real numbers, in general, and so it is impossible to list all the members of $\mathbb{R}$ in a sequence.$\square$

This last Example is extremely important. It tells us that the set of natural numbers ***cannot*** be put in a one-to-one correspondence with the set of real numbers. Thus these sets have ***different cardinalities*** even though both sets are infinite. Since $\mathbb{N}$ is a subset of $\mathbb{R}$, it is reasonable to say that the cardinality of $\mathbb{R}$ is ***strictly bigger*** than $\aleph_0$. We can use this Example to motivate the following:

Definitions — Comparing Cardinalities:

Suppose that X and Y are two sets (they can be finite or infinite). Then we say that $|X| < |Y|$, that is, the cardinality of X is ***strictly smaller*** than the cardinality of Y, if there exists a function $f: X \to Y$ which is ***one-to-one***, but there is ***no*** such function which is both one-to-one and ***onto***. In this case, we can also write: $|Y| < |X|$ and say that the cardinality of Y is ***strictly bigger*** than the cardinality of X.

We can also say that $|X| \leq |Y|$, that is, the cardinality of X is ***at most*** the cardinality of Y, if there exists a function $f: X \to Y$ which is ***one-to-one***. Such a function may or may not be ***onto***. In this case, we can also write: $|Y| \geq |X|$ and say that the cardinality of Y is ***at least*** the cardinality of X.

We denote the cardinality of $\mathbb{R}$ by $\aleph_1$, pronounced "Aleph one," that is:

$$|\mathbb{R}| = \aleph_1 > \aleph_0 = |\mathbb{N}|.$$

Any infinite set such as $\mathbb{R}$ whose cardinality is strictly bigger than $\aleph_0$ is called ***uncountable***. This means that we ***cannot*** list all of the elements of an uncountable set in a ***sequence*** like we have above for the members of $\mathbb{N}$ or $\mathbb{Z}$. This distinction will become important in the next sub-section. We also note that it is possible to construct sets which have cardinality that are strictly larger than $\aleph_1$. For example, by constructing the set of all possible subsets of $\mathbb{R}$, which is denoted by $\wp(\mathbb{R})$ and pronounced the ***power set*** of $\mathbb{R}$, we produce a set of cardinality $\aleph_2$. This process can be continued indefinitely, producing set of strictly larger and larger cardinalities. Thus begins an infinite chain:

$$|\mathbb{N}| = \aleph_0 < |\mathbb{R}| = \aleph_1 < |\wp(\mathbb{R})| = \aleph_2 < \cdots$$

We summarize below our discussion of infinite sets of numbers, as well as other results that are found in the Exercises:

Theorem — Countable and Uncountable Sets of Numbers:

The set of ***natural numbers***, ***integers***, and ***rational numbers*** are all ***countable:***

$$|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| = \aleph_0.$$

However, the set of ***real numbers***, the set of ***irrational numbers***, and all ***intervals*** of the real number line that contain at least two points are all ***uncountable*** and have cardinality $\aleph_1$:

$$|\mathbb{R}| = |\mathbb{R} - \mathbb{Q}| = |(a,b)| = |[a,b]| = |[a,b)| = |(a,b]| = \aleph_1,$$

where $a < b \in \mathbb{R}$. More generally, these infinite intervals also have cardinality $\aleph_1$:

$$|(-\infty,b)| = |(-\infty,b]| = |(a,\infty)| = |[a,\infty)| = \aleph_1.$$

Describing Infinite Sets of Vectors

We are now in a position to establish notation and guidelines in order to describe an infinite set of vectors S from a vector space $(V, \oplus, \odot)$. We will restrict our attention to sets that have cardinality that are either $\aleph_0$ or $\aleph_1$. The most convenient way to do this is by using what is called an ***indexing set***, denoted I, which is typically a non-empty set which is an easily described subset of $\mathbb{R}$ (not necessarily an interval, as the symbol I might lead you to think). The simplest examples will have indexing sets $I = \mathbb{N}$ or $I = \mathbb{R}$ itself. The vectors of S will be in a one-to-one correspondence with the elements of I through the use of set-builder notation, which we will write in general as:

$$S = \{\vec{v}_i \,|\, i \in I\} \subset (V, \oplus, \odot), \text{ where } I \subset \mathbb{R} \text{ is some non-empty indexing set.}$$

To avoid ambiguity, we will insist that $\vec{v}_i \neq \vec{v}_j$ if i and j are distinct indices in I. In other words, distinct indices correspond to distinct vectors, and vice versa. Thus, if I is a countable indexing set, then S is also a countable set of vectors (and similarly for uncountable index sets).

Note that if $I = \{1, 2, \ldots, n\}$, we get our old sets of vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$.

Example: Let us consider the infinite set of vectors, first in ***roster*** form:

$$S = \{1, x, x^2, x^3, \ldots, x^n, \ldots\} \subset F(\mathbb{R}),$$

that is, the set of ***all monomials*** in x. We can view S as a set of vectors from the vector space $F(\mathbb{R})$ of functions defined on $\mathbb{R}$. Since we can list the members of S, this set is ***countable***. Since the powers of x are natural numbers, we can also write S in set-builder notation as:

$$S = \{x^n \,|\, n \in \mathbb{N}\}.$$

In this case, we describe S using the ***indexing set*** $\mathbb{N}$. In the same way, we can construct the infinite sets of ***even monomials*** and ***odd monomials:***

$$E = \{x^{2n} \,|\, n \in \mathbb{N}\} = \{1, x^2, x^4, \ldots, x^{2n}, \ldots\}, \text{ and}$$
$$O = \{x^{2n+1} \,|\, n \in \mathbb{N}\} = \{x, x^3, x^5, \ldots, x^{2n+1}, \ldots\}.$$

Notice that for both sets, we used the same indexing set $\mathbb{N}$. Since the monomials within each set are in one-to-one correspondence with $\mathbb{N}$, we can say that all three sets S, E and O are ***countably infinite*** sets of vectors.□

As seen above, if S is ***countable***, we can describe it using the ***roster form***, where we list the members of S in a sequence with a clear pattern, or in ***set-builder notation***, using a countable set such as $\mathbb{N}$ as our indexing set. However, if S is ***uncountable***, we can no longer list the members of S in a sequence, so we have no choice but to use ***set-builder notation*** using a non-empty indexing set I.

Example: We will consider several more subsets from $F(\mathbb{R})$. Let us start with:

$$S_1 = \{e^{kx} \,|\, k \in \mathbb{Z}\} = \{\ldots, e^{-3x}, e^{-2x}, e^{-x}, 1, e^x, e^{2x}, e^{3x}, \ldots\}.$$

The indexing set for S_1 is $\mathbb{Z}$. Since $\mathbb{Z}$ is countable, and the functions in S_1 are ***distinct*** (no two of them have the same graph), S_1 is also ***countable***. Similarly, the set:

$$S_2 = \{e^{kx} \,|\, k \in \mathbb{Q}\}$$

has $\mathbb{Q}$ as an indexing set. S_2 is also ***countable***, since $\mathbb{Q}$ is countable, as we shall see in the Exercises. However, since the way we list $\mathbb{Q}$ is not very convenient, it is certainly better to describe S_2 using set-builder notation instead of roster form. Since every integer is also a rational number, $S_1 \subset S_2$. However, the functions $e^{2x/3}$ and $e^{-3x/5}$ are functions in S_2 which are not in S_1. Now, consider the set:

$$S_3 = \{e^{kx} \mid k \in \mathbb{R}\}.$$

Since the indexing set is $\mathbb{R}$, S_3 is also ***uncountable***, and we cannot list the vectors of S_3 in roster form. Note that $S_1 \subset S_2 \subset S_3$, but $e^{\sqrt{2}x}$ and $e^{\pi x}$ are functions in S_3 which are not in S_1 or S_2. □

Linearity Concepts for Infinite Sets of Vectors

Fortunately, in order to define and understand linearity concepts for an infinite set of vectors S, we will test only ***finite*** subsets of S. Suppose we are given the infinite set of vectors:

$$S = \{\vec{v}_i \mid i \in I\} \subset (V, \oplus, \odot), \text{ where } I \subset \mathbb{R} \text{ is some non-empty indexing set.}$$

A ***finite subset*** of S can be listed explicitly, and written in roster form:

$$\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_n}\},$$

where $i_1, i_2, \ldots, i_n$ are numbers from I, which are called ***indices*** (the plural of ***index***), with $i_1 < i_2 < \cdots < i_n$. This notation is particularly important if I is ***uncountable***. This notation is called a ***double subscript notation***, because the subscript of $\vec{v}$ (which is i) also contain a subscript (which is 1 through n).

We are now ready to generalize the concepts of linear combinations, Span and linear independence to include infinite sets:

Definition: Let $(V, \oplus, \odot)$ be a vector space. Suppose that $S = \{\vec{v}_i \mid i \in I\} \subset (V, \oplus, \odot)$, where $I \subset \mathbb{R}$ is some non-empty indexing set. A ***linear combination*** of vectors from S can be constructed in the following way:

(a) Choose a ***finite*** subset of vectors: $\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_n}\}$ from S, where $i_1 < i_2 < \cdots < i_n$ are from I.

(b) Choose a corresponding finite list of ***coefficients*** $c_1, c_2, \ldots, c_n \in \mathbb{R}$, as before.

(c) Form the vector expression: $c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_n \odot \vec{v}_{i_n}$.

Similarly, the ***Span of S***, denoted $Span(S)$ as before, is the set of ***all possible linear combinations*** of vectors from ***all finite subsets*** of S.

Based on the description above, we can construct $Span(S)$ as follows:

1. Form ***all finite subsets*** of S: $\{\vec{v}_{i_1}\}$, $\{\vec{v}_{i_1}, \vec{v}_{i_2}\}$, $\{\vec{v}_{i_1}, \vec{v}_{i_2}, \vec{v}_{i_3}\}$, . . . and so on.

   In other words, form ***all*** subsets consisting of exactly one vector, exactly two vectors, exactly three vectors, and so on.

2. For each of these subsets, form ***all possible linear combinations:***

   $c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_n \odot \vec{v}_{i_n}$ of these finite sets.

3. Collect all of these linear combinations in one enormous set which will be $Span(S)$.

Note that there is no such thing as the linear combination of an ***infinite*** number of vectors. We can only form the linear combination of a ***finite*** number of vectors. We will also write $c_1\vec{v}_{i_1} + c_2\vec{v}_{i_2} + \cdots + c_n\vec{v}_{i_n}$ for simplicity when the context is clear. Fortunately, if a set of vectors S is ***countable***, there is a simpler way to think of linear combinations and Spans:

Theorem: Suppose that $S = \{\vec{v}_i \,|\, i \in \mathbb{N}\} = \{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n, \ldots\}$ is a ***countable*** set of vectors from a vector space $(V, \oplus, \odot)$. Then, a linear combination of the vectors in S is an expression of the form:

$$c_0 \odot \vec{v}_0 \oplus c_1 \odot \vec{v}_1 \oplus c_2 \odot \vec{v}_2 \oplus \cdots \oplus c_k \odot \vec{v}_k,$$

for some $k \in \mathbb{N}$, and coefficients $c_0, c_1, \ldots, c_k \in \mathbb{R}$.
Similarly, $Span(S)$ is the set of all linear combinations from S of the form given above.

Proof: Since the set S is countable, a finite subset of n vectors from S has the form:

$$\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_n}\},$$

where we can assume that $i_1 < i_2 < \cdots < i_n$, and these subscripts are all ***natural numbers***. Now, according to the Definition above, a linear combination of this finite set has the form:

$$r_1 \odot \vec{v}_{i_1} \oplus r_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus r_n \odot \vec{v}_{i_n},$$

for some scalars r_1 through r_n. Since i_n is the largest subscript, suppose $i_n = k$. By making a coefficient 0 if need be, the linear combination we have above can be rewritten to include ***all*** of the vectors from $\vec{v}_0$ to $\vec{v}_k$. For example, if we formed the finite set:

$$\{\vec{v}_2, \vec{v}_5, \vec{v}_7\},$$

then the linear combination $r_1 \odot \vec{v}_2 \oplus r_2 \odot \vec{v}_5 \oplus r_3 \odot \vec{v}_7$ can be written as:

$$0 \cdot \vec{v}_0 + 0 \cdot \vec{v}_1 + r_1\vec{v}_2 + 0 \cdot \vec{v}_3 + 0 \cdot \vec{v}_4 + r_2\vec{v}_5 + 0 \cdot \vec{v}_6 + r_3\vec{v}_7.$$

This is a linear combination from the finite subset $\{\vec{v}_0, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_7\}$.
Thus, in general, any linear combination of a finite subset of S can be written in the form:

$$c_0\vec{v}_0 + c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k,$$

for some natural number k, and coefficients $c_0, c_1, c_2, \ldots, c_k$. ■

This Theorem gives us a significant advantage when we construct the Span of a countable set of vectors S, and later on as well, when we test if S is linearly independent or not. For example, we do not need to form the finite subset $\{\vec{v}_1, \vec{v}_3, \vec{v}_6, \vec{v}_8\}$ of S and form all of its linear combinations, because the linear combinations of this set appear in the set of all linear combinations of the set $\{\vec{v}_0, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_8\}$, by setting some of the coefficients to zero.

Example: Let us consider the infinite set of monomials:

$$S = \{x^n \,|\, n \in \mathbb{N}\} = \{1, x, x^2, x^3, \ldots, x^n, \ldots\} \subset F(\mathbb{R}),$$

that we saw in an earlier Example. The set S has a monomial with degree n for ***every*** positive integer n, as well as the constant monomial 1. Thus, S is ***not*** a subset of ***any*** of the polynomial spaces $\mathbb{P}^n$, since the polynomials of $\mathbb{P}^n$ have degree at most n, and there will always be an infinite number of monomials in S whose degree is bigger than any particular n.

According to the Theorem above, a linear combination of the members of S has the form:

$$c_0 \cdot 1 + c_1 \cdot x + c_2 \cdot x^2 + c_3 \cdot x^3 + \cdots + c_n \cdot x^n$$
$$= c_0 + c_1 x + c_2 x^2 + c_3 x^3 + \cdots + c_n x^n.$$

But this is a ***polynomial*** of degree n. Thus, we can conclude that $Span(S)$ consists of all possible polynomials, of any degree. Since the sum of two polynomials and a scalar multiple of a polynomial are again polynomials, this set forms a vector space, which we will denote as $\mathbb{P}$:

$$\mathbb{P} = Span(\{1, x, x^2, x^3, \ldots, x^n, \ldots\}) = Span(\{x^n \,|\, n \in \mathbb{N}\}).\square$$

The previous Example involved a countable set of vectors. If the set S is ***uncountable***, though, we will have to follow the original formula to find the Span of S. In some cases, though, it is still possible to have a grasp as to what the Span of S is.

Example: Consider the uncountable set: $S_3 = \{e^{kx} \,|\, k \in \mathbb{R}\} \subset F(\mathbb{R})$.

To form a finite subset of n vectors, we pick n real numbers: $k_1 < k_2 < \cdots < k_n$, and form the set:

$$\{e^{k_1 x}, e^{k_2 x}, \ldots, e^{k_n x}\}.$$

A linear combination of this finite set therefore has the form:

$$c_1 e^{k_1 x} + c_2 e^{k_2 x} + \cdots + c_n e^{k_n x},$$

for some scalars $c_1, c_2, \ldots, c_n$.

The Span of S consists of all functions of this form, that is, for ***all*** possible choices of real numbers $k_1 < k_2 < \cdots < k_n$, and ***all*** possible coefficients $c_1, c_2, \ldots, c_n$, and ***all*** positive integers n. Obviously this enormous set is certainly uncountable, since $Span(S)$ must contain S, but it is still possible to visualize what each member of $Span(S)$ looks like, at least in terms of a formula.$\square$

Now, let us generalize the concept of linear independence to an infinite set of vectors:

Definition: Suppose that $S = \{\vec{v}_i \,|\, i \in I\}$, where $I \subset \mathbb{R}$ is some non-empty indexing set, is an ***infinite*** set of vectors from a vector space $(V, \oplus, \odot)$. We will say that S is ***linearly independent*** if every ***finite*** subset of S is linearly independent. In other words, the only solution to the ***dependence test equation:***

$$c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_k \odot \vec{v}_{i_k} = \vec{\mathbf{0}}_V$$

is the ***trivial solution*** $c_0 = 0,\ c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0,$ for all indices $i_1 < i_2 < \cdots < i_k$ from I.

Once again, if S is countable, then it becomes easier to test if S is independent or not:

Theorem: Suppose that $S = \{\vec{v}_i \,|\, i \in \mathbb{N}\} = \{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n, \ldots\}$ is a ***countable*** set of vectors from a vector space $(V, \oplus, \odot)$. Then, S is linearly independent ***if and only if*** the only solution to the ***dependence test equation:***

$$c_0 \odot \vec{v}_0 \oplus c_1 \odot \vec{v}_1 \oplus c_2 \odot \vec{v}_2 \oplus \cdots \oplus c_k \odot \vec{v}_k = \vec{\mathbf{0}}_V$$

is the ***trivial solution*** $c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0,$ for all $k \in \mathbb{N}$. In other words, every ***finite*** subset $\{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_k\}$ is linearly independent, for ***every*** $k \in \mathbb{N}$.

The proof is based on exactly the same idea as the proof of the previous Theorem, and will be left as an Exercise. It tells us that we do not have to test if the set $\{\vec{v}_3, \vec{v}_6, \vec{v}_8\}$ is independent, because this set will be tested in the process of testing the bigger (more complete) set:

$$\{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_8\}.$$

Example: Let us return to the set of all monomials $S = \{1, x, x^2, x^3, \ldots, x^n, \ldots\} \subset F(\mathbb{R})$. According to the Theorem above, in order to test if S is linearly independent, we have to show that every finite subset of the form:

$$S^{/} = \{1, x, x^2, x^3, \ldots, x^k\}$$

is linearly independent. But $S^{/}$ consists of polynomials (in fact, monomials) of ***distinct degrees***. By our Theorem in Section 3.2, $S^{/}$ is linearly independent. Since every finite subset $S^{/}$ is linearly independent, S is also linearly independent by our Theorem.□

Example: Let us decide if the infinite uncountable set $S_3 = \{e^{kx} \mid k \in \mathbb{R}\}$ is linearly dependent or independent. We saw that every finite subset of S_3 has the form:

$$\{e^{k_1 x}, e^{k_2 x}, \ldots, e^{k_n x}\},$$

where $k_1 < k_2 < \cdots < k_n$ and n is a positive integer. But we saw in Exercise 35 of Section 3.2 that these finite sets are all linearly independent. Thus, by our definition, S_3 is also linearly independent.□

3.3 Section Summary

A non-empty set X is ***finite*** if the number of elements in the set is finite, that is, a positive integer n. In other words, we can choose to list the elements of X in some particular order, say $X = \{x_1, x_2, \ldots, x_n\}$, where the list eventually terminates. We call n the ***cardinality*** of our set, and we use the notation: $|X| = n$. We agree that the ***empty set*** has cardinality 0, and we also consider it to be a finite set. A set that is ***not*** finite is called ***infinite***. The sets of numbers: $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ and $\mathbb{R}$ are all infinite sets.

The Schroeder-Bernstein Theorem: Suppose that X and Y are two sets (they can be finite or infinite). Then, $|X| = |Y|$, that is, X and Y have the same cardinality, ***if and only if*** there exists a function $f : X \to Y$ that is both ***one-to-one*** and ***onto***.

The sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$ all have the same cardinality, and we call this cardinality $\aleph_0$. We refer to these as ***countable*** sets. We write: $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}| = \aleph_0$.

The set $\mathbb{R}$ has cardinality that is strictly bigger than $\aleph_0$, and we call this cardinality $\aleph_1$. Any set with cardinality strictly bigger than $\aleph_0$ is called ***uncountable***. We write: $|\mathbb{R}| = \aleph_1$.

Let $(V, \oplus, \odot)$ be a vector space. Suppose that $S = \{\vec{v}_i \mid i \in I\}$ is an infinite set of vectors from V, where $I \subset \mathbb{R}$ is a non-empty ***indexing set***. A ***linear combination*** of vectors from S can be constructed in the following way:

(a) Choose a finite subset of vectors: $\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_n}\}$ from S, where $i_1 < i_2 < \cdots < i_n$ are from I.

(b) Choose a finite list of scalars $c_1, c_2, \ldots, c_n \in \mathbb{R}$, as before.

(c) Form the vector expression: $c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_n \odot \vec{v}_{i_n}$.

Similarly, the ***Span of S***, denoted ***Span(S)*** as before, is the set of ***all possible linear combinations*** of vectors from ***all finite subsets*** of S.

We say that S is ***linearly independent*** if every ***finite*** subset of S is linearly independent. In other words, the only solution to the dependence test equation:

$$c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_k \odot \vec{v}_{i_k} = \vec{\mathbf{0}}_V$$

is the trivial solution $c_0 = 0,\ c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0,$ for all indices $i_1 < i_2 < \cdots < i_k$ from I.

Suppose that $S = \{\vec{v}_i \mid i \in \mathbb{N}\} = \{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n, \ldots\}$ is a ***countable*** set of vectors from a vector space $(V, \oplus, \odot)$. Then, a ***linear combination*** of the vectors in S is an expression of the form:

$$c_0 \odot \vec{v}_0 \oplus c_1 \odot \vec{v}_1 \oplus c_2 \odot \vec{v}_2 \oplus \cdots \oplus c_k \odot \vec{v}_k,$$

for some $k \in \mathbb{N}$ and coefficients $c_0,\ c_1,\ \ldots,\ c_k \in \mathbb{R}$. Similarly, ***Span(S)*** is the set of all linear combinations from S of the form given above. Furthermore, S is ***linearly independent*** ***if and only if*** the only solution to the dependence test equation:

$$c_0 \odot \vec{v}_0 \oplus c_1 \odot \vec{v}_1 \oplus c_2 \odot \vec{v}_2 \oplus \cdots \oplus c_k \odot \vec{v}_k = \vec{\mathbf{0}}_V$$

is the trivial solution $c_1 = 0,\ c_2 = 0,\ \ldots,\ c_k = 0,$ for all $n \in \mathbb{N}$. In other words, every finite subset $\{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_k\}$ is linearly independent, for ***every*** $k \in \mathbb{N}$.

3.3 Exercises

1. Show that the set of ***odd integers*** $O = \{\ldots -7, -5, -3, -1, 1, 3, 5, 7, \ldots\}$ is countable by constructing an explicit bijection from $\mathbb{N}$ to O.
2. Show that the set of ***even integers*** $E = \{\ldots -6, -4, -2, 0, 2, 4, 6, \ldots\}$ is countable by constructing an explicit bijection from $\mathbb{N}$ to E.
3. More generally, let m be a fixed positive integer. Show that the set of all multiples of m:
$$M = \{\ldots -4m, -3m, -2m, -m, 0, m, 2m, 3m, 4m, \ldots\}$$
is countable by constructing an explicit bijection from $\mathbb{N}$ to M.
4. Consider the set of ***even monomials:*** $E(x) = \{1, x^2, x^4, \ldots, x^{2n}, \ldots\}$
   a. Rewrite S using $\mathbb{N}$ as an indexing set in set-builder notation.
   b. Describe in words $Span(E(x))$. Hint: what can we say about the symmetry of the polynomials in $Span(E(x))$?
   c. Show that $E(x)$ is linearly independent.
5. Repeat the previous Exercise for the set of odd monomials: $O(x) = \{x, x^3, x^5, \ldots, x^{2n+1}, \ldots\}$.
6. Consider the countable set of basic ***rational*** functions:
$$S = \left\{\frac{1}{x}, \frac{1}{x^2}, \frac{1}{x^3}, \frac{1}{x^4}, \ldots, \frac{1}{x^n}, \ldots\right\} \subset F((0,\infty)).$$
   a. Rewrite S using $\mathbb{N}$ as an indexing set in set-builder notation.
   b. Write down explicitly what a linear combination of vectors from S looks like. Hint: $\mathbb{N}$ is countable.
   c. Prove that S is linearly independent. Hint: Combine all the terms using a least common denominator, and use the Fundamental Theorem of Algebra.
7. Consider the uncountable set of ***exponential*** functions:
$$S = \{b^x \mid b > 0\} \subset F((-\infty,\infty)).$$
   a. What is the indexing set of S?

b. Write down explicitly what a linear combination of vectors from S looks like.

c. Prove that S is linearly independent.

d. In Calculus, we usually exclude $b = 1$. Why can we allow $b = 1$ in this context?

8. Consider the countable set of ***radical*** functions:

$$S = \{x^{1/2}, x^{1/3}, x^{1/4}, \ldots, x^{1/n}, \ldots\} \subset F([0,\infty)).$$

a. Rewrite S using $\mathbb{N}$ as an indexing set in set-builder notation.

b. Write down explicitly what a linear combination of vectors from S looks like.

c. Prove that S is linearly independent. Hint: mimic the idea we used for exponential functions.

For Exercises (9) to (20): Decide whether or not the set of functions is linearly dependent or independent. Assume that any finite subset is defined on a common interval.

9. $S = \left\{ \frac{1}{x-k} \mid k \in \mathbb{R} \right\}$

10. $S = \{ \log_b(x) \mid b > 0, b \neq 1\}$ Note: this time, we ***cannot*** allow $b = 1$, as compared to Exercise 7. Why?

11. $S = \{(x-a)^n \mid n \in \mathbb{N}\}$, where $a \in \mathbb{R}$ is a ***fixed*** real number.

12. $S = \{(x-a)^n \mid a \in \mathbb{R}\}$, where n is a ***fixed*** positive integer.
Why is this set different from that in the previous Exercise?

13. $S = \{ \ln(x+k) \mid k \in \mathbb{R}\}$

14. $S = \{e^{3x}, xe^{3x}, x^2e^{3x}, x^3e^{3x}, \ldots, x^ne^{3x}, \ldots\}$

15. $S = \{e^x \sin(x), e^{2x}\sin(x), e^{3x}\sin(x), \ldots, e^{nx}\sin(x), \ldots\}$

16. $S = \{5^{-x}, x5^{-x}, x^2 5^{-x}, x^3 5^{-x}, \ldots, x^n 5^{-x}, \ldots\}$

17. $S = \{ \sin(2x), x \cdot \sin(2x), x^2 \cdot \sin(2x), x^3 \cdot \sin(2x), \ldots, x^n \cdot \sin(2x), \ldots\}$

18. $S = \{x^{1/2}, x^{3/2}, x^{5/2}, \ldots, x^{n/2}, \ldots\}$

19. $S = \{1+x, 1+x^2, x+x^2, x+x^3, x^2+x^3, x^2+x^4, \ldots, x^n+x^{n+1}, x^n+x^{n+2}, \ldots\}$

20. $S = \left\{\ldots, \frac{1}{x^3}, \frac{1}{x^2}, \frac{1}{x}, 1, x, x^2, x^3, \ldots\right\}$.
Hint: what would a general linear combination from S look like?

21. Suppose that $S = \{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n, \ldots\}$ is a ***countable*** set of vectors from a vector space V. Prove that S is linearly independent ***if and only if*** every subset of the form:

$$\{\vec{v}_0, \vec{v}_1, \vec{v}_2, \vec{v}_3, \ldots, \vec{v}_n\}$$

is linearly independent, for every $n \in \mathbb{N}$.

22. In this Section, we showed that $\mathbb{N}$ and $\mathbb{Z}$ have the same cardinality by listing the members of $\mathbb{Z}$ as:

$$0, 1, -1, 2, -2, 3, -3, \ldots, n, -n, \ldots$$

Use this to explicitly construct a function $f : \mathbb{N} \to \mathbb{Z}$ that is both one-to-one and onto. Hint: an easy way to do it would be to use a piecewise definition.

23. Suppose that X is a ***subset*** of Y. Prove that $|X| \leq |Y|$. Hint: state the definition of this symbol and create an easy function f that satisfies the definition.

24. Suppose that X and Y are both countable sets, and assume for the sake of simplicity that $X \cap Y = \varnothing$, that is, they have no element in common. Prove that $X \cup Y$ is also countable. Hint: list X and Y in a countable way and show how to list the elements of $X \cup Y$ also in a countable way.

25. Show that the set of irrational numbers $\mathbb{R} - \mathbb{Q}$ is also uncountable. Hint: Use Proof by Contradiction and Exercise 24.

26. In the Exercises of Section 3.2, we saw that the Elimination Theorem and the Extension Theorem are still valid in an abstract vector space if S is a finite set of vectors. Review the Proofs that you wrote in that Section, and convince yourself that the Proofs are still valid if S is an infinite set. On the other hand, explain why the Proof of the Dependence vs. Spanning Sets Theorem does not extend to an infinite set S (where $L \subset Span(S)$ may also be infinite).

27. ***The Countability of the Rational Numbers:***

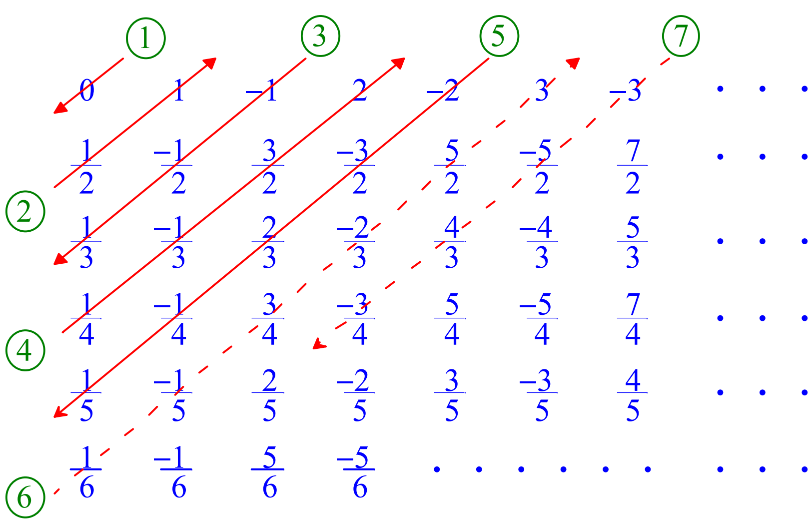
The purpose of this Exercise is to show that $\mathbb{Q}$ is countable, that is: $|\mathbb{Q}| = \aleph_0$.

Recall that every rational number a/b can be written as a fraction where b is as small as possible, i.e., in ***lowest terms***. We will create an ***infinite table*** that will contain the members of $\mathbb{Q}$ in the order of increasing denominators. On the 1st row we see the members of $\mathbb{Z}$ (rational numbers with denominator 1), listed as a sequence as we saw in the Examples. On the 2nd row are the rational numbers with denominator 2, then on the 3rd row those with denominator 3, and so on.

$$\begin{array}{cccccccc}
0 & 1 & -1 & 2 & -2 & 3 & -3 & \cdots \\
\frac{1}{2} & \frac{-1}{2} & \frac{3}{2} & \frac{-3}{2} & \frac{5}{2} & \frac{-5}{2} & \frac{7}{2} & \cdots \\
\frac{1}{3} & \frac{-1}{3} & \frac{2}{3} & \frac{-2}{3} & \frac{4}{3} & \frac{-4}{3} & \frac{5}{3} & \cdots \\
\frac{1}{4} & \frac{-1}{4} & \frac{3}{4} & \frac{-3}{4} & \frac{5}{4} & \frac{-5}{4} & \frac{7}{4} & \cdots \\
\frac{1}{5} & \frac{-1}{5} & \frac{2}{5} & \frac{-2}{5} & \frac{3}{5} & \frac{-3}{5} & \frac{4}{5} & \cdots \\
\frac{1}{6} & \frac{-1}{6} & \frac{5}{6} & \frac{-5}{6} & \cdots & \cdots & \cdots & \cdots
\end{array}$$

a. List the first 7 rational numbers on each of the next 3 rows of this table.

Now, in order to prove that $\mathbb{Q}$ is countable, we have to list ***all*** the rational numbers in ***one sequence***, without repetition. The idea is to traverse this table in a diagonal or zigzag manner, following the arrows numbered 1, 2, 3 and so on, as shown in the figure below:



Thus, the first 6 rational numbers in this sequence are $0, \frac{1}{2}, 1, -1, -\frac{1}{2}$ and $\frac{1}{3}$.

b. List the next 25 rational numbers in this sequence according to this diagonal traversal. Be sure to include in your traversal appropriate members from the next three rows from (a).

c. Suppose the rational number a/b is found in row i, column j in the table above. This rational number will be found on the kth arrow as described in the diagonal traversal above. Find a formula for k in terms of i and j. For example, the number $-2/3$ in row 3 and column 4 is on the 6th arrow.

d. Explain why every rational number will be found exactly once in this sequence.

e. The Big Picture: Summarize the steps above into a proof that $|\mathbb{Q}| = \aleph_0$.

The idea above is attributed to ***Georg Cantor*** (1845, Russia - 1918, Germany), professor at the University of Halle, and the Father of ***Set Theory***.
It is known as ***The Cantor Diagonalization Argument***.

28. ***The Uncountability of Sub-intervals of the Real Numbers:***

We know from Algebra that there are different kinds of intervals of real numbers: finite open intervals of the form (a,b), finite closed intervals of the form $[a,b]$, finite half-open/half-closed intervals of the form $(a,b]$ or $[a,b)$, infinite open intervals of the form (a,∞) and $(-\infty,b)$, and infinite closed intervals of the form $[a,\infty)$, and $(-\infty,b]$.

The purpose of this Exercise is to show that ***all*** of these eight interval types have cardinality $\aleph_1$, and thus they are all ***uncountable***.

a. Suppose that $[a,b]$ is any closed interval of $\mathbb{R}$. Find a ***linear*** function:

$$f : [0,1] \to [a,b],$$

that is, of the form $f(x) = mx + k$, with a ***positive slope*** m, which is both one-to-one and onto $[a,b]$. Hint: what should the graph look like? Explain why this proves that all finite closed intervals $[a,b]$ have the same cardinality.

b. Show that the same linear function f that you found in the previous part is also a function:

$$f : (0,1] \to (a,b],$$
$$f : [0,1) \to [a,b), \text{ and}$$
$$f : (0,1) \to (a,b),$$

and ***each*** is one-to-one and onto the indicated range when restricted to the corresponding domain. Hint: this just means that the graph you drew in (a) will have a hole or two. Explain why this proves that all finite intervals of the form $(a,b]$ have the same cardinality, and similarly for the intervals of the other two forms.

Parts (a) and (b) show that in order to prove that a finite interval of any of the four forms seen above has cardinality $\aleph_1$, we have to prove that the four intervals $[0,1]$, $(0,1]$, $[0,1)$ and $(0,1)$ — or ***any*** particular example of each of these four forms — all have cardinality $\aleph_1$.

c. Sketch the graph of ***the restricted tangent function*** from Trigonometry:

$$\tan(x) : \left(-\frac{\pi}{2}, \frac{\pi}{2}\right) \to \mathbb{R},$$

and explain why it is both one-to-one and onto. Explain why this proves that

$$\left|\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\right| = |\mathbb{R}| = \aleph_1.$$

d. Use parts (b) and (c) to prove that for all finite open intervals (a,b):

$$|(a,b)| = |\mathbb{R}| = \aleph_1.$$

e. Next, let us consider infinite open intervals of the form (a,∞), starting with $a = 0$. Sketch the graph of the ***natural logarithmic function***:

$$\ln(x) : (0,\infty) \to \mathbb{R},$$

and explain why this function is one-to-one and onto. Explain why this proves that:

$$|(0,\infty)| = |\mathbb{R}| = \aleph_1.$$

f. Let $a \in \mathbb{R}$. Find a (very simple) ***linear*** function $f : (a,\infty) \to (0,\infty)$, that is one-to-one and onto. Explain why this proves that for any $a \in \mathbb{R}$:

$$|(a,\infty)| = |\mathbb{R}| = \aleph_1.$$

g. Show that the same function f that you found in the previous part is also a function:

$$f : [a,\infty) \to [0,\infty),$$

that is again one-to-one and onto. Explain why this proves that for any $a \in \mathbb{R}$:

$$|[a,\infty)| = |[0,\infty)|.$$

h. Let $b \in \mathbb{R}$. Find a ***linear*** function $f : (-\infty,b) \to (0,\infty)$, that is one-to-one and onto, but with a ***negative*** slope. Explain why this proves that for any $b \in \mathbb{R}$:

$$|(-\infty,b)| = |\mathbb{R}| = \aleph_1.$$

i. Show that the same function f that you found in the previous part is also a function:

$$f : (-\infty,b] \to [0,\infty),$$

that is again one-to-one and onto. Explain why this proves that for any $b \in \mathbb{R}$:

$$|(-\infty,b]| = |[0,\infty)|.$$

j. Show that the function:

$$f(x) = \frac{x}{1-x},$$

restricted to the domain $[0,1)$ is a one-to-one function. Show that its range is $[0,\infty)$. In other words:

$$f : [0,1) \to [0,\infty)$$

is both one-to-one and onto. Explain why this shows that $|[0,1)| = |[0,\infty)|$.

k. Find a ***linear*** function:

$$f : (0,1] \to [0,1),$$

which is both one-to-one and onto, but with a ***negative*** slope. Explain why this proves that:

$$|(0,1]| = |[0,1)|.$$

Explain why this shows that all finite half-open/half-closed intervals have the same cardinality.

l. Generalize the idea from the previous part to create a linear function:

$$f : (a,b] \to [a,b),$$

that is both one-to-one and onto. We will call this function a ***flip***.

Now comes the hardest part. We will construct a function:

$$f : [0,1] \to [0,1),$$

that is both one-to-one and onto. The idea is to break up $[0,1]$ into ***halves***, then ***quarters***, then ***eighths***, and so on, and ***flip*** the new interval on the ***right***:

$$\begin{aligned}
&[0,1] \\
&= \left[0,\tfrac{1}{2}\right] \cup \left(\tfrac{1}{2},1\right]; \text{ flip } \left(\tfrac{1}{2},1\right]: \\
&\to \left[0,\tfrac{1}{2}\right] \cup \left[\tfrac{1}{2},1\right); \text{ subdivide } \left[0,\tfrac{1}{2}\right] \text{ into two:} \\
&= \left[0,\tfrac{1}{4}\right] \cup \left(\tfrac{1}{4},\tfrac{1}{2}\right] \cup \left[\tfrac{1}{2},1\right); \text{ flip } \left(\tfrac{1}{4},\tfrac{1}{2}\right]: \\
&\to \left[0,\tfrac{1}{4}\right] \cup \left[\tfrac{1}{4},\tfrac{1}{2}\right) \cup \left[\tfrac{1}{2},1\right); \text{ subdivide } \left[0,\tfrac{1}{4}\right] \text{ into two:} \\
&= \left[0,\tfrac{1}{8}\right] \cup \left(\tfrac{1}{8},\tfrac{1}{4}\right] \cup \left[\tfrac{1}{4},\tfrac{1}{2}\right) \cup \left[\tfrac{1}{2},1\right); \text{ flip } \left(\tfrac{1}{8},\tfrac{1}{4}\right]: \\
&\to \left[0,\tfrac{1}{8}\right] \cup \left[\tfrac{1}{8},\tfrac{1}{4}\right) \cup \left[\tfrac{1}{4},\tfrac{1}{2}\right) \cup \left[\tfrac{1}{2},1\right); \text{ subdivide } \left[0,\tfrac{1}{8}\right] \text{ into two:} \\
&= \left[0,\tfrac{1}{16}\right] \cup \left(\tfrac{1}{16},\tfrac{1}{8}\right] \cup \left[\tfrac{1}{8},\tfrac{1}{4}\right) \cup \left[\tfrac{1}{4},\tfrac{1}{2}\right) \cup \left[\tfrac{1}{2},1\right); \text{ flip } \left(\tfrac{1}{16},\tfrac{1}{8}\right]: \ldots
\end{aligned}$$

Notice that $\frac{1}{2^n}$ appears ***twice*** after the *nth* flip, but this is a ***temporary*** concern, because the leftmost interval will be divided into two once again in the next iteration and the next subinterval flipped, and so $\frac{1}{2^n}$ will appear ***only once*** after the ***next*** step.

m. Write the algorithm above as a piecewise function with an infinite number of pieces in the domain. Explain why $f(0) = 0$ and why this function is one-to-one and onto, with range $[0,1)$. Hint: the previous part should be very useful. The formula should be in this form:

$$f(x) = \begin{cases} 0 & \text{if } x = 0, \\ \cdots\cdots & \text{if } x \in \left(\frac{1}{2},1\right], \\ \cdots\cdots & \text{if } x \in \left(\frac{1}{4},\frac{1}{2}\right], \\ \cdots\cdots & \quad\vdots \\ \cdots\cdots & \text{if } x \in \left(\frac{1}{2^{n+1}},\frac{1}{2^n}\right], \\ \cdots\cdots & \quad\vdots \end{cases}$$

Note that we have no choice but to have $f(0) = 0$. Explain why this part shows that $|[0,1]| = |[0,1)|$. Note that these are ***almost*** the same interval, except the second interval is open at $x = 1$. In other words, they are only different by ***exactly one point***.

n. Sketch the graph of $f(x)$ from part (m).

o. Apply the same ideas found in parts (l) and (m) to define a function:

$$f : (0,1) \to (0,1],$$

which is one-to-one and onto. Hint: begin with:

$$(0,1) = \left(0,\tfrac{1}{2}\right) \cup \left[\tfrac{1}{2},1\right).$$

p. Summarize all the parts above to show that all of the eight interval types of real numbers as listed at the beginning of this Exercise have cardinality $\aleph_1$.

3.4 Subspaces, Basis and Dimension

We continue generalizing the ideas that we saw in Chapter 1, specifically, the concepts of a ***subspace*** of an abstract vector space, a ***basis*** for a vector space, and the ***dimension*** of a vector space.

Subspaces

We will define a subspace W of a vector space V in exactly the same way as we did in Chapter 1:

> ***Definition:*** A ***non-empty*** subset W of a vector space $(V, \oplus, \odot)$ is called a ***subspace*** of V if W is ***closed*** under vector addition and scalar multiplication.
> In other words, for all $\vec{w}_1$ and $\vec{w}_2 \in W$, and $k \in \mathbb{R}$: $\vec{w}_1 \oplus \vec{w}_2 \in W$, and $k \odot \vec{w}_1 \in W$.
> As before, we write $W \trianglelefteq V$, and we refer to V as the ***ambient space*** of W.

Notice that the closure properties are the first two properties for a vector space. Let us see if W satisfies the other eight. Since the members of W are also members of V, then W ***inherits*** the six arithmetic properties of V: the commutative and associative properties of addition,the two distributive properties, the associative property of scalar multiplication, and the unitary property. This leaves only the existence of a zero vector in W and additive inverses in W.

The requirement that W be non-empty means that there is at least one vector $\vec{w} \in W$. However, since W is closed under $\odot$, this means that $0 \odot \vec{w} = \vec{\mathbf{0}}_V \in W$ as well. Likewise, $-1 \odot \vec{w} = -\vec{w} \in W$ as well. Thus, we have proven the following:

> ***Theorem:*** Let W be a non-empty subset of $(V, \oplus, \odot)$.
> Then: W is a ***subspace*** of V ***if and only if*** $(W, \oplus, \odot)$ is ***itself*** a ***vector space***.

This Theorem also tells us that if we want to show that W is a subspace, then we should be able to show that $\vec{\mathbf{0}}_V \in W$. In other words, W is not just a non-empty set, but in particular, $\vec{\mathbf{0}}_V$ is one of its members. The ***contrapositive*** of this Theorem also says that if $\vec{\mathbf{0}}_V$ is ***not*** in W, then it is impossible for W to be a subspace of V.

Example: Let $V = F((-\infty, \infty))$. Consider the set:

$$W = \{f(x) \in V \,|\, f(3) = 0\}.$$

The zero function $z(x)$ satisfies the condition that $z(3) = 0$, and thus $z(x) \in W$. Thus, W is not empty, and it contains the zero vector of V. Now, suppose $f(x)$ and $g(x)$ are members of W. Thus:

$$f(3) = 0 \quad \text{and} \quad g(3) = 0, \text{ and so:}$$

$$f(3) + g(3) = 0 + 0 = 0.$$

Thus, $f(x) + g(x) \in W$ as well, and so W is closed under vector addition. Now, let $k \in \mathbb{R}$. Then:

$$(k \cdot f)(3) = k \cdot f(3) = k \cdot 0 = 0,$$

and thus $(k \cdot f)(x) \in W$ as well. Thus W is also closed under scalar multiplication, and so W is a subspace of V. On the other hand, consider the set:

$$U = \{f(x) \in V \mid f(3) = 2\}.$$

Note that U is not an empty set either, because $f(x) = x - 1$ satisfies $f(3) = 2$, and this function is certainly a member of V. However, $z(x)$ does not satisfy the condition $z(3) = 2$, and so U is ***not*** a subspace of V. □

As before, the Span of a set of vectors is the easiest example of a subspace:

Theorem: Suppose $S = \{\vec{v}_i \mid i \in I\} \subset (V, \oplus, \odot)$, where $I \subset \mathbb{R}$ is some non-empty indexing set, and let $W = Span(S)$.
Then: $(W, \oplus, \odot)$ is a ***subspace*** of $(V, \oplus, \odot)$.

Proof: The idea behind the Proof of the analogous Theorem in Chapter 1 still works, but the difference lies in our new notation. In Chapter 1, we were only dealing with a finite set of vectors: $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. This time, S could be an ***infinite*** set, so the vectors chosen to create a linear combination for $\vec{v} \in Span(S)$ may be completely ***different*** vectors from the vectors chosen to create another linear combination for a second vector $\vec{w} \in Span(S)$. To avoid using cumbersome notation, suppose that $S_1 \subset S$ is the finite set of vectors used to create $\vec{v}$, and $S_2 \subset S$ is the finite set of vectors used to create $\vec{w}$. Let us combine these two into a single finite set:

$$S_1 \cup S_2 = S_3,$$

and we can list the members of the finite set S_3 in the usual notation:

$$S_3 = \{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}.$$

Since S_1 and S_2 are both subsets of S_3, we can express both $\vec{v}$ and $\vec{w}$ as linear combinations of S_3, by placing a coefficient of ***zero*** beside a vector in S_3 which is not needed in order to produce either $\vec{v}$ or $\vec{w}$, if there are any. In other words, we can write:

$$\begin{aligned}\vec{v} &= c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_k \odot \vec{v}_{i_k}, \text{ and}\\ \vec{w} &= d_1 \odot \vec{v}_{i_1} \oplus d_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus d_k \odot \vec{v}_{i_k},\end{aligned}$$

where we allow some coefficients to be zero, if necessary. Now, we can write the sum as:

$$\vec{v} + \vec{w} = (c_1 + d_1) \odot \vec{v}_{i_1} \oplus (c_2 + d_2) \odot \vec{v}_{i_2} \oplus \cdots \oplus (c_k + d_k) \odot \vec{v}_{i_k},$$

which is again another member of $Span(S)$. Thus, $Span(S)$ is closed under vector addition. Similarly:

$$\begin{aligned}r \odot \vec{v} &= r \odot (c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_k \odot \vec{v}_{i_k})\\ &= (rc_1) \odot \vec{v}_{i_1} \oplus (rc_2) \odot \vec{v}_{i_2} \oplus \cdots \oplus (rc_k) \odot \vec{v}_{i_k},\end{aligned}$$

by the distributive and associative properties. The result is therefore in $Span(S)$ as well. Thus, $Span(S)$ is a subspace of $(V, \oplus, \odot)$. ■

Non-Example: In our most unusual example in Section 3.1, we defined $\mathbb{R}^+$ as a vector space under ordinary multiplication and exponentiation. The vector space $\mathbb{R} = \mathbb{R}^1$ is also a vector space, but with respect to ordinary addition and multiplication. Thus, even though $\mathbb{R}^+$ is a ***subset*** of $\mathbb{R}$, it is ***not a subspace*** of $\mathbb{R}$, for the simple reason that the operations in $\mathbb{R}^+$ are different from the operations in $\mathbb{R}$. □

Now let us consider some important families of subspaces and their relationships among each other.

Example: We saw in Section 3.3 that the polynomial spaces $\mathbb{P}^n$ can be written as:

$$\mathbb{P}^n = Span(\{1, x, x^2, \ldots, x^n\}).$$

By the same token:

$$\mathbb{P}^{n+1} = Span(\{1, x, x^2, \ldots, x^n, x^{n+1}\}),$$

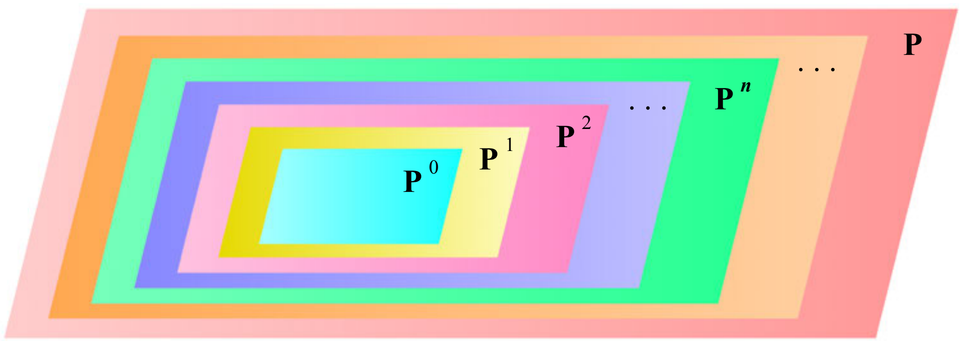
and thus, the polynomial spaces $\mathbb{P}^n$ form an ***ascending sequence*** of ***nested subspaces:***

$$\mathbb{P}^0 \trianglelefteq \mathbb{P}^1 \trianglelefteq \mathbb{P}^2 \trianglelefteq \mathbb{P}^3 \cdots\cdots \trianglelefteq \mathbb{P}^n \trianglelefteq \mathbb{P}^{n+1} \trianglelefteq \cdots\cdots \trianglelefteq \mathbb{P}.$$

As indicated at the end of the chain above, all of the $\mathbb{P}^n$ are subspaces of $\mathbb{P}$, the vector space of ***all polynomials***, of all degrees, which we defined as:

$$\mathbb{P} = Span(\{1, x, x^2, \ldots, x^n, \ldots\})$$

We can thus visualize the polynomial spaces nested as follows:



The Nesting of Polynomial Spaces

Continuing with this example, $\mathbb{P}$ is a subspace of $F((-\infty, \infty))$, the space of all functions defined over all real numbers. This is equivalent to saying that the set of all polynomials is closed under addition and scalar multiplication, that is, the sum of any two polynomials is again a polynomial, and the scalar multiple of any polynomial is again a polynomial.$\square$

Subspaces of Function Spaces

Let us consider functions with special properties. We saw in Section 3.1 the vector space $F(I)$ consisting of all functions defined on an interval I. We will refer to any subspace of $F(I)$, for any interval I, as a ***function space***.

In Calculus, we encountered ***continuous*** functions on I. It was a basic result that the sum of two continuous functions and a scalar (constant) multiple of a continuous function are again continuous. Thus the set $C(I)$ of continuous functions on I is a subspace of $F(I)$, and we write:

$$C(I) = \left\{ f(x) \in F(I) \,|\, f(x) \text{ is } \textbf{\textit{continuous}} \text{ on } I \right\} \trianglelefteq F(I).$$

Now, we liked a function even more when it is ***differentiable***, that is, its derivative exists at every point on the interval I (if I has a closed end-point, say $I = [a, b)$, we have to consider only the right-differentiability of the function at $x = a$). Again, we saw in Calculus that the sum of two differentiable functions and a scalar multiple of a differentiable function are again differentiable. However, we will further require that the derivative of these functions also be continuous (in the Exercises, we show an example of a differentiable function whose derivative is discontinuous.). We call these functions $C^1(I)$, to denote that their ***first derivative is continuous:***

$$C^1(I) = \{ f(x) \in F(I) \mid f(x) \text{ is } \textbf{\textit{differentiable}}, \text{ and } f'(x) \text{ is } \textbf{\textit{continuous}} \text{ on } I \} \trianglelefteq F(I).$$

But recall that a function that is differentiable on an open interval I is itself also continuous on I, and therefore the space $C^1(I)$ is a subspace of $C(I)$, so we can write: $C^1(I) \trianglelefteq C(I)$.

Continuing with this logic, a function that is twice-differentiable on I with a continuous second derivative also possesses a continuous first derivative, and so we call the set of these functions:

$$C^2(I) = \{ f(x) \in F(I) \mid f(x) \text{ is } \textbf{\textit{twice-differentiable}}, \text{ and } f''(x) \text{ is } \textbf{\textit{continuous}} \text{ on } I \} \trianglelefteq C^1(I).$$

At this point, let us write the subspace inclusions (or nestings) that we have so far:

$$C^2(I) \trianglelefteq C^1(I) \trianglelefteq C(I) \trianglelefteq F(I).$$

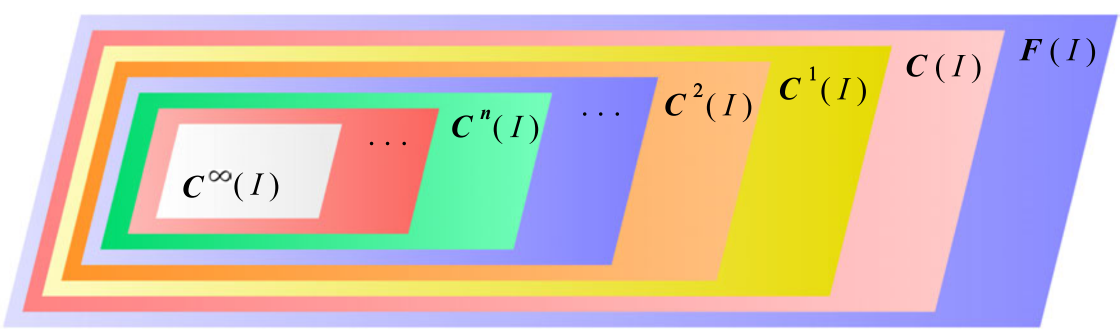
By Induction, we can define the subspace $C^n(I)$ consisting of all functions that are differentiable n times, and whose n^{th} derivative is also continuous. (As mentioned above, this automatically guarantees that ***all*** derivatives, up to the n^{th} derivative, are also continuous.)

However, since we now seem to be going right to left, and as mathematicians we normally like ***reading*** from left to right, we will ***reverse*** the subspace symbol and write the ***descending sequence*** of nested subspaces, where the space on the left ***contains*** the space on its right:

$$F(I) \trianglerighteq C(I) \trianglerighteq C^1(I) \trianglerighteq C^2(I) \trianglerighteq \cdots C^n(I) \trianglerighteq C^{n+1}(I) \trianglerighteq \cdots\cdots$$

Finally, we also have the subspace $C^\infty(I)$ of functions that have derivatives of ***all*** possible orders. Since we require that each derivative is likewise differentiable, we can conclude that all of these derivatives are also continuous. We sometimes call this subspace the set of ***real analytic*** or ***smooth functions***. Our friends the polynomials, the sine and cosine functions, and the exponential functions of any base, are all members of this somewhat exclusive club.

Notice also that $C^\infty(I)$ is a subspace of all the $C^n(I)$, but since the list above does not terminate, we will exclude $C^\infty(I)$ from the "end" of the list. However, we can visualize these nested subspaces below, with $C^\infty(I)$ living at the very core:



The Nesting of Spaces of Continuous and n-Differentiable Functions

Basis for a Vector Space

We can define the concept of a basis for a vector space just as we did for $\mathbb{R}^n$:

Definition — Basis for a Vector Space:
A set of vectors B from a vector space $(V, \oplus, \odot)$ is a ***basis*** for V if it is ***linearly independent*** and ***Spans*** V.
We will agree that the zero vector space $V = \left\{ \vec{\mathbf{0}}_V \right\}$ does not have a basis, since any set containing $\vec{\mathbf{0}}_V$ is automatically dependent.

Example: We saw in Section 3.2 that the finite set of monomials:

$$B_n = \{1, x, x^2, \ldots, x^n\}$$

is a linearly independent subset of $\mathbb{P}^n$. But we also saw that $\mathbb{P}^n = Span(B_n)$, and so B_n is a basis for $\mathbb{P}^n$, for every positive integer n. Similarly, if we consider the countably infinite set:

$$B = \{1, x, x^2, \ldots, x^n, \ldots\},$$

then B is also linearly independent, and $Span(B) = \mathbb{P}$, the vector space of all polynomials. Thus, B is a basis for $\mathbb{P}$.□

The Spanning property of a basis B tells us that every $\vec{v} \in V$ can be expressed as a linear combination of the vectors in B in ***at least one*** way. It turns out that if we combine this property with linear independence, then this expression becomes ***unique***, that is, every $\vec{v} \in V$ can be expressed as a linear combination of the vectors in B in ***exactly one*** way. The following Theorem tells us that the converse is also true:

Theorem — Uniqueness of Representation:

Suppose that $S = \{\vec{v}_i \mid i \in I\}$, for some non-empty indexing set I, is a set of vectors from some vector space $(V, \oplus, \odot)$.

Then: S is a basis for $(V, \oplus, \odot)$ ***if and only if*** every vector $\vec{v} \in V$ can be represented ***uniquely*** as a linear combination of a ***finite*** subset of vectors $\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}$ from S:

$$\vec{v} = c_1\vec{v}_{i_1} + c_2\vec{v}_{i_2} + \cdots + c_k\vec{v}_{i_k}.$$

Proof: We will borrow notation from the proof of our Theorem that $Span(S)$ is a subspace of V.

($\Rightarrow$) We are given that S is a basis for $(V, \oplus, \odot)$. Thus, S Spans V and S is linearly independent. Therefore, every vector $\vec{v} \in V$ can be expressed as a linear combination from S in at least one way. In other words, there is a finite subset $S_1 \subset S$ such that $\vec{v}$ is a linear combination of the vectors in S_1. Suppose that there exists another finite subset $S_2 \subset S$ such that $\vec{v}$ is also a linear combination of the vectors in S_2. As before, let $S_3 = S_1 \cup S_2 = \{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}$. Thus, the two linear combinations that we have for $\vec{v}$ can both be written in terms of the vectors in S_3, that is:

$$\vec{v} = c_1\vec{v}_{i_1} + c_2\vec{v}_{i_2} + \cdots + c_k\vec{v}_{i_k}, \text{ and}$$
$$\vec{v} = d_1\vec{v}_{i_1} + d_2\vec{v}_{i_2} + \cdots + d_k\vec{v}_{i_k}.$$

But now, subtracting the corresponding sides of the two equations, we get:

$$\vec{\mathbf{0}}_V = (c_1 - d_1)\vec{v}_{i_1} + (c_2 - d_2)\vec{v}_{i_2} + \cdots + (c_k - d_k)\vec{v}_{i_k}.$$

Since S is linearly independent and S_3 is a finite subset of S, S_3 is also linearly independent, and so all the coefficients above have to be zero, that is:

$$c_1 = d_1, c_2 = d_2, \ldots, c_k = d_k.$$

($\Leftarrow$) Now, suppose that S has the uniqueness of representation property, that is, every $\vec{v} \in V$ can be represented ***uniquely*** as a linear combination from S:

$$\vec{v} = c_1\vec{v}_{i_1} + c_2\vec{v}_{i_2} + \cdots + c_k\vec{v}_{i_k}.$$

This property directly shows that S Spans V. To prove linear independence, let us construct the dependence test equation:

$$c_1\vec{v}_{i_1} + c_2\vec{v}_{i_2} + \cdots + c_k\vec{v}_{i_k} = \vec{\mathbf{0}}_V.$$

for any finite subset $\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}\}$ of S. But the Multiplicative Property of Zero also tells us that:

$$0\vec{v}_{i_1} + 0\vec{v}_{i_2} + \cdots + 0\vec{v}_{i_k} = \vec{\mathbf{0}}_V.$$

By the Uniqueness of Representation Property, we must have:

$$c_1 = 0, c_2 = 0, \ldots, c_k = 0.$$

Thus, S is linearly independent.■

Our next task, as before, is to show that any non-zero vector space has at least one basis B. Let us try to construct a basis for V by using the same algorithm that we saw in Chapter 1. The key idea in the proof is the Extension Theorem, that again, thanks to the 10 Axioms and the additional constructions and properties of vector spaces, is still true in general, as you proved in Exercise 45 of Section 3.2:

Theorem — The Extension Theorem:

Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***finite***, linearly ***independent*** set of vectors from some vector space $(V, \oplus, \odot)$, and suppose $\vec{v}_{n+1}$ is ***not*** a member of $Span(S)$.

Then, the ***extended*** set: $S' = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{v}_{n+1}\}$ is ***still linearly independent***.

Let us now attempt to mimic the proof from Chapter 1 of the existence of a basis of a non-zero subspace V:

We initialize B to contain any nonzero vector, say $B = \{\vec{v}_1\}$. If $V = Span(B)$, then we are finished.

If not, then enlarge B by including a vector that is not in $Span(B)$, say a vector we will call $\vec{v}_2$. We now have $B = \{\vec{v}_1, \vec{v}_2\}$, and B is linearly independent by the Extension Theorem. If $V = Span(B)$, then we are finished.

If not, then enlarge B by including another vector that is not in $Span(B)$, say a vector we will call $\vec{v}_3$, and repeat the process of checking if $V = Span(B)$, where now $B = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$. Again, since $\vec{v}_3$ is not in the Span of $\{\vec{v}_1, \vec{v}_2\}$, the extended set $B = \{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ is linearly independent.

The process sounds simple enough, but the problem is that we are ***not guaranteed*** that this process will ***terminate***, much less terminate in a ***finite*** number of steps. In Chapter 1, we were able to use the fact that we are dealing with a subspace of $\mathbb{R}^n$, and therefore if we ever had n vectors in B, we would have to stop because including one more vector would have resulted in a ***dependent*** set.

In order to show that this algorithm will in fact produce a basis for us in an abstract vector space, we need an extremely powerful Axiom from Set Theory that is called ***The Axiom of Choice***, or equivalently, ***Zorn's Lemma*** or a process called ***Transfinite Induction***. We will not elaborate on this Axiom (the interested reader is invited to do an internet search), which is usually a subject for a more advanced treatment of Linear Algebra. Suffice it to say that this algorithm ***will*** result in a basis B for our vector space V, and although we did not complete the proof, we will formally state the Theorem as follows:

Theorem — Existence of a Basis: Every non-zero vector space $(V, \oplus, \odot)$ has a ***basis*** B.

The ***cardinality*** of this basis B will be the next focus of our attention.

The Dimension of a Vector Space

The fact that our subspaces no longer live in some Euclidean space $\mathbb{R}^n$ leads to some interesting complications. We will proceed with the following distinction:

Definition: A non-zero vector space $(V, \oplus, \odot)$ is called ***finite dimensional*** if we can find a ***finite*** set B which is a basis for V. We call such a set a ***finite basis*** for V.

Otherwise, we say that V is ***infinite dimensional***.

We will agree that the zero vector space $V = \left\{\vec{\mathbf{0}}_V\right\}$ has dimension 0, and is also finite-dimensional.

Our next goal is to show that any two bases for a ***finite-dimensional*** space have the ***same*** number of vectors. The key to this in Chapter 1 was The Dependent Sets from Spanning Sets Theorem, and its contrapositive, the Independent Sets from Spanning Sets Theorem. This says that if a set of vectors from the Span of a set of n vectors has ***more*** than n vectors, then it is automatically ***dependent***. The proof that we saw there carries over to the Span of a set of vectors from an abstract vector space, thanks to the Ten Axioms. Thus, we generalize (as seen in the Exercises of Section 3.2):

Theorem — The Dependent/Independent Sets from Spanning Sets Theorem:
Suppose we have a set of n vectors: $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\} \subset (V, \oplus, \odot)$,
and we form $W = Span(S)$.
Suppose now we randomly choose a set of m vectors from W to form a new set:

$$L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}.$$

Then, we can conclude that: ***if*** $m > n$, ***then*** L is linearly ***dependent***.

Consequently, the ***contrapositive*** states that: ***if*** L is linearly ***independent***, ***then*** $m \leq n$.

Again, this is exactly the Theorem that will allow us to prove:

Definition/Theorem — The Dimension of a Vector Space:

Any two ***bases*** for a ***finite-dimensional*** vector space $(V, \oplus, \odot)$ have exactly the same number of elements. We call this common number the ***dimension*** of V and is denoted as $dim(V)$. If $dim(V) = k$, we also say that V is a ***k-dimensional*** vector space.

The Proof is exactly the same as the analogous Theorem in Section 1.6. This Theorem is actually true as well even if V is infinite dimensional, though in this case "number of elements" must now be replaced by the more general term ***cardinality***. Thus, V may have dimension $\aleph_0$ or $\aleph_1$ and so on. Like our Theorem on the Existence of a Basis, this Theorem will require Transfinite Induction in its proof when V is infinite-dimensional.

Example: We saw earlier that $B_n = \{1, x, x^2, x^3, \ldots, x^n\}$ is a basis for $\mathbb{P}^n$. Since there are $n+1$ monomials in this set:

$$\dim(\mathbb{P}^n) = n+1.$$

Similarly, we saw that the countable set $B = \{1, x, x^2, \ldots, x^n, \ldots\}$ is a basis for $\mathbb{P}$, and thus:

$$\dim(\mathbb{P}) = \aleph_0.$$

In Section 3.4, we also saw the linearly independent, uncountable set:

$$S_3 = \{e^{kx} \mid k \in \mathbb{R}\} \subset F(\mathbb{R}), \text{ with } |S_3| = \aleph_1.$$

Since the Span of any set of vectors is a subspace, we can construct:

$$W = Span(S_3) \trianglelefteq F(\mathbb{R}),$$

and S_3 is a basis for W. Thus:

$$\dim(W) = \aleph_1. \square$$

We can also generalize the Theorem that says that any subspace of $\mathbb{R}^n$ is at most n-dimensional:

Theorem: Let $(W, \oplus, \odot)$ be a subspace of a ***finite-dimensional*** vector space $(V, \oplus, \odot)$. If $dim(V) = n$, then $dim(W) \leq n$, that is, $dim(W) \leq dim(V)$. Furthermore, $dim(W) = n = dim(V)$ ***if and only if*** $W = V$.

Again, the proof carries over exactly as we saw it in Section 1.7, thanks to the Extension Theorem.

Example: Consider $\mathbb{P}^2$, the vector space of polynomials with degree at most 2. Let:

$$W = \{p(x) \in \mathbb{P}^2 \mid p(3) = 2p(-1)\}.$$

First let us show that W is indeed a subspace of $\mathbb{P}^2$. First, W is ***not empty***, because the zero polynomial $z(x)$ satisfies the condition $z(3) = 0 = 2z(-1)$. We must show next that W is ***closed*** under addition and scalar multiplication. Suppose $p(x)$ and $q(x)$ are members of W and c is a scalar. Then:

$$(p+q)(3) = p(3) + q(3) = 2p(-1) + 2q(-1) = 2(p+q)(-1), \quad \text{and}$$

$$(c \cdot p)(3) = c \cdot p(3) = c \cdot 2p(-1) = 2(c \cdot p)(-1),$$

and thus W is indeed closed under both operations. Now, let us find a basis and the dimension of W. We begin by forming a typical linear combination from the ambient space $\mathbb{P}^2$ and determine what restrictions are imposed on the coefficients. Using our basis $\{1, x, x^2\}$, a typical member of $\mathbb{P}^2$ can be written as a linear combination:

$$p(x) = c_0 + c_1 x + c_2 x^2.$$

The ***defining condition*** says that $p(3) = 2p(-1)$, so we must have:

$$p(3) = c_0 + 3c_1 + 9c_2 = 2(c_0 - c_1 + c_2) = 2p(-1).$$

We can rewrite this equation as: $c_0 - 5c_1 - 7c_2 = 0$.

This is a single homogeneous equation. We can find solutions for it by making c_1 and c_2 our free variables (this is because we wrote our linear combination in ascending degree). From this, we have:

$$c_0 = 5c_1 + 7c_2,$$

and thus we have: $p(x) = 5c_1 + 7c_2 + c_1 x + c_2 x^2$.

Let us collect the terms with ***common coefficients:***

$$p(x) = c_1(5 + x) + c_2(7 + x^2),$$

and thus we see that any member of W must be a ***linear combination*** of the two polynomials $5 + x$ and $7 + x^2$ (both of which satisfy the required condition: $5 + 3 = 2(5 - 1)$, and likewise for $7 + x^2$). In other words:

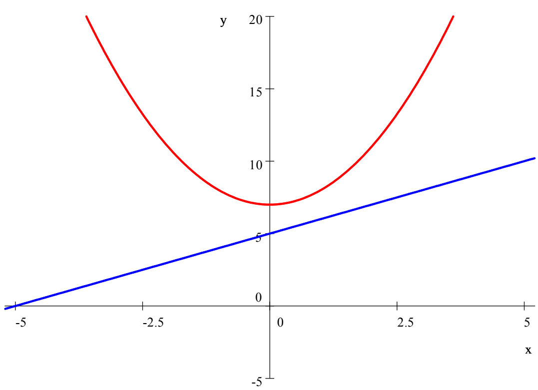
$$W = Span(\{5 + x, 7 + x^2\}).$$

Since these are polynomials of different degrees, they are ***linearly independent*** by our Theorem in Section 3.2. Thus, we have found a ***basis*** for W:

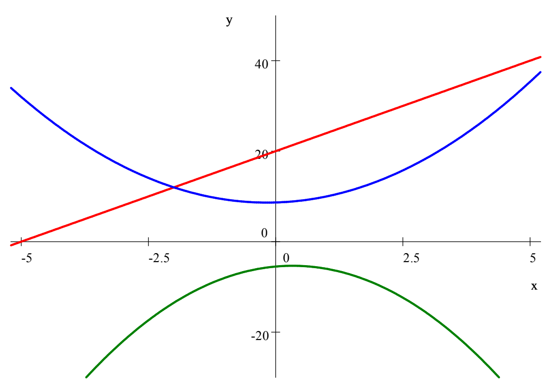
$$B = \{5 + x, 7 + x^2\},$$

and from this, $dim(W) = 2$. As expected, it is smaller than $dim(\mathbb{P}^2) = 3$. We graph below the two members of B on the left. We also graph on the right three examples of ***linear combinations*** from B:

$$p_1(x) = 4(5 + x) + 0(7 + x^2) = 20 + 4x,$$
$$p_2(x) = (5 + x) - \frac{3}{2}(7 + x^2) = -\frac{11}{2} + x - \frac{3}{2}x^2, \text{ and}$$
$$p_3(x) = \frac{1}{3}(5 + x) + (7 + x^2) = \frac{26}{3} + \frac{1}{3}x + x^2.$$



Two Basis Vectors for W



Other Members of W

We can see both algebraically and graphically that $p(3) = 2p(-1)$ for all these polynomials.□

Let us see what we can do if we have a slightly bigger ambient space and we play with some Calculus:

Example: Let $V = \mathbb{P}^3$, and consider:

$$W = \left\{ p(x) \in \mathbb{P}^3 \mid p(-2) = 0 \textbf{\textit{ and }} p'(3) = 0 \right\},$$

under the same addition and scalar multiplication, of course. First let us show that W is indeed a subspace of V. Again, the zero polynomial $z(x)$ satisfies these two conditions, so W is ***not empty***. We must show next that W is ***closed*** under vector addition and scalar multiplication. If $p(x)$ and $q(x)$ are two vectors in W, and k is any scalar, then:

$$(p+q)(-2) = p(-2) + q(-2) = 0 + 0 = 0, \text{ and}$$
$$(p+q)'(3) = p'(3) + q'(3) = 0 + 0 = 0.$$

Thus, W is closed under addition, since both defining properties are satisfied by $p + q$. Similarly:

$$(kp)(-2) = k \cdot p(-2) = k \cdot 0 = 0, \quad \text{and}$$
$$(kp)'(3) = k \cdot p'(3) = k \cdot 0 = 0,$$

and thus W is closed under scalar multiplication. Notice that we used two properties of the derivative:

$$(p+q)'(x) = p'(x) + q'(x), \text{ and}$$
$$(k \cdot p)'(x) = k \cdot p'(x).$$

Now, let us further use our knowledge of Calculus in order to find a basis for W. We know that $\{1, x, x^2, x^3\}$ is a basis for $\mathbb{P}^3$, so first we write:

$$p(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3,$$

as a typical generic member of $\mathbb{P}^3$. In order to be a member of W, it must satisfy the two defining properties. Since the second property involves a derivative, we first compute:

$$p'(x) = c_1 + 2c_2 x + 3c_3 x^2.$$

Now, the two properties say:

$$p(-2) = 0 = c_0 - 2c_1 + 4c_2 - 8c_3, \quad \text{and}$$
$$p'(3) = 0 = c_1 + 6c_2 + 27c_3.$$

But this is a ***homogeneous system*** of two equations in four variables. We can solve this using our techniques from Chapter 1. We assemble the matrix:

$$\begin{bmatrix} 1 & -2 & 4 & -8 \\ 0 & 1 & 6 & 27 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 16 & 46 \\ 0 & 1 & 6 & 27 \end{bmatrix}.$$

The ***nullspace*** of this matrix gives us the solutions to the four coefficients. Since the leading variables are c_0 and c_1, and the free variables are c_2 and c_3, we can sight-read the basis for the nullspace:

$$\langle c_0, c_1, c_2, c_3 \rangle = c_2\langle -16, -6, 1, 0 \rangle + c_3\langle -46, -27, 0, 1 \rangle = \langle -16c_2 - 46c_3,\ -6c_2 - 27c_3, c_2, c_3 \rangle.$$

Thus, $p(x)$ must have the form:

$$\begin{aligned} p(x) &= (-16c_2 - 46c_3) + (-6c_2 - 27c_3)x + c_2 x^2 + c_3 x^3 \\ &= -16c_2 - 6c_2 x + c_2 x^2 - 46c_3 - 27c_3 x + c_3 x^3 \\ &= c_2(-16 - 6x + x^2) + c_3(-46 - 27x + x^3). \end{aligned}$$

These last two polynomials have different degrees, so they are ***independent***, and since every member of W must be a linear combination of these two polynomials, we conclude that W has basis:

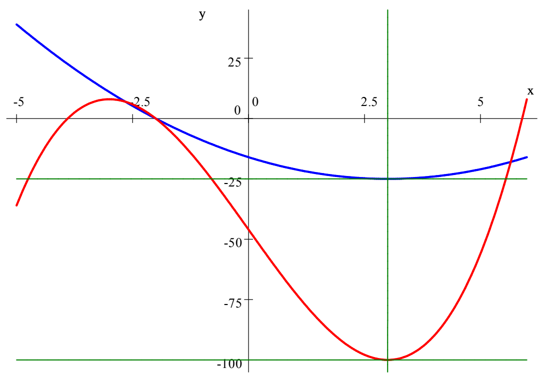
$$B = \{-16 - 6x + x^2,\ -46 - 27x + x^3\}.$$

We also conclude that W is 2-dimensional. Notice also that the basis $\{\langle -16, -6, 1, 0\rangle, \langle -46, -27, 0, 1\rangle\}$ for the nullspace closely corresponds to the coefficients in the two basis vectors in B. Thus, by ***decoding*** the basis for the nullspace, we can find a basis for W.

We can check that both polynomials in B satisfy the defining properties of W:

$$\begin{aligned} p_1(x) &= -16 - 6x + x^2; & p_2(x) &= -46 - 27x + x^3; \\ p_1(-2) &= -16 + 12 + 4 = 0; & p_2(-2) &= -46 + 54 - 8 = 0; \\ p_1'(x) &= -6 + 2x; & p_2'(x) &= -27 + 3x^2, \text{ and thus:} \\ p_1'(3) &= -6 + 6 = 0; & p_2'(3) &= -27 + 27 = 0. \end{aligned}$$

We also graph the two polynomials below, and indeed we see that both polynomials have -2 as an x-intercept, and the tangent line at $x = 3$ is horizontal for both graphs.



The Polynomials $p_1(x) = -16 - 6x + x^2$ and $p_2(x) = -46 - 27x + x^3$ □

The last portion of the previous Theorem which states:

$$dim(W) = dim(V) \ \textbf{\textit{if and only if}} \ W = V,$$

is ***false*** in the case when W is a subspace of an ***infinite dimensional*** vector space V. This is one of the reasons why infinite dimensional vector spaces are best left in the appropriate field of study, which is called ***Functional Analysis***.

Example: Consider the vector space of ***all polynomials:***

$$\mathbb{P} = Span(\{1, x, x^2, x^3, \ldots, x^n, \ldots\}).$$

Since the set of monomials is ***countable*** and linearly independent, $dim(\mathbb{P}) = \aleph_0$. Now, consider the subspace:

$$\mathbb{P}_e = Span(E), \text{ where } E = \{1, x^2, x^4, x^6, \ldots, x^{2n}, \ldots\},$$

consisting of all ***even polynomials***, that is, the polynomials $p(x)$ that satisfy the equation $p(x) = p(-x)$. As seen in the Exercises in the previous Section, E is likewise countable and linearly independent, and so $dim(\mathbb{P}_e) = \aleph_0$ as well. But obviously $\mathbb{P}_e$ is not all of $\mathbb{P}$, and so we see that our Theorem on subspaces of finite dimensional spaces is ***false*** if the ambient space is infinite dimensional. □

Using dim(W) to Find Other Bases for W

In Section 1.9, we saw that if we knew the dimension of W, it becomes easier to verify if a subset B of W is a basis for W. The Theorem that we saw there generalizes to all finite-dimensional subspaces:

> ***Theorem — The Two-for-the-Price-of-One or Two-for-One Theorem:***
> Let $(W, \oplus, \odot)$ be a ***finite-dimensional subspace*** of a vector space $(V, \oplus, \odot)$, with $dim(W) = n$, and suppose that $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ is any subset of vectors from W. Then: B is a ***basis*** for W ***if and only if*** either B is ***linearly independent*** ***or*** B ***Spans*** W. In other words, it is necessary and sufficient to check B for ***only one condition*** (and this would more easily be the condition of linear ***independence***) without checking the other if B already contains the ***correct number of vectors***.

3.4 Section Summary

A non-empty subset W of a vector space $(V, \oplus, \odot)$ is called a ***subspace*** of V if W is ***closed*** under $\oplus$ and $\odot$. In other words, for all $\vec{w}_1$ and $\vec{w}_2 \in W$, and $k \in \mathbb{R}$: $\vec{w}_1 \oplus \vec{w}_2 \in W$, and $k \odot \vec{w}_1 \in W$. As before, we write $W \trianglelefteq V$, and we refer to V as the ***ambient space*** of W.

Let $W \subset (V, \oplus, \odot)$. Then: W is a ***subspace*** of V ***if and only if*** $(W, \oplus, \odot)$ is itself a ***vector space***.

Suppose $S \subset (V, \oplus, \odot)$, where S is non-empty. Then $(Span(S), \oplus, \odot)$ is a ***subspace*** of $(V, \oplus, \odot)$.

The polynomial spaces: $\mathbb{P}^0 \trianglelefteq \mathbb{P}^1 \trianglelefteq \mathbb{P}^2 \trianglelefteq \mathbb{P}^3 \cdots\cdots \trianglelefteq \mathbb{P}^n \trianglelefteq \mathbb{P}^{n+1} \trianglelefteq \cdots\cdots \trianglelefteq \mathbb{P}$, form an ***increasing*** sequence of nested subspaces.

The continuous functions, differentiable functions with continuous derivatives, and so on, form a ***decreasing*** sequence of nested subspaces: $C^0(I) \trianglerighteq C^1(I) \trianglerighteq C^2(I) \trianglerighteq \ldots C^n(I) \trianglerighteq C^{n+1}(I) \trianglerighteq \ldots$

All these subspaces contain $C^\infty(I)$, the space of all functions with continuous derivatives of ***all*** orders. Polynomials are members of this space, along with $\sin(x)$, $\cos(x)$ and e^x, among many others.

A set of vectors B from $(V, \oplus, \odot)$ is a ***basis*** for V if it is ***linearly independent*** and ***Spans*** V.

Uniqueness of Representation: A set of vectors $S = \{\vec{v}_i \,|\, i \in I\}$ is a basis for $(V, \oplus, \odot)$ ***if and only if*** every vector $\vec{v} \in V$ can be represented ***uniquely*** as a linear combination of a ***finite*** set of members $\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_k}$ from S: $\vec{v} = c_1\vec{v}_{i_1} + c_2\vec{v}_{i_2} + \cdots + c_k\vec{v}_{i_k}$.

Every non-zero vector space $(V, \oplus, \odot)$ has a basis B. A vector space $(V, \oplus, \odot)$ is called ***finite dimensional*** if we can find a ***finite basis*** B for V, otherwise we say that V is ***infinite dimensional***.

The Dependent/Independent Sets from Spanning Sets Theorem: Suppose we have a set of n vectors: $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\} \subset (V, \oplus, \odot)$, and we form $W = Span(S)$. Suppose now we randomly choose a set of m vectors from W to form a new set: $L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$. Then, we can conclude that: ***if*** $m > n$, ***then*** L is linearly ***dependent***. Consequently, the ***contrapositive*** states that: ***if*** L is linearly ***independent***, ***then*** $m \leq n$.

Let $(V, \oplus, \odot)$ be a ***finite-dimensional*** vector space. Any two ***bases*** for $(V, \oplus, \odot)$ have the ***same*** number of elements. We call this common number the ***dimension*** of V. Let $(W, \oplus, \odot)$ be a subspace of $(V, \oplus, \odot)$. If $dim(V) = n$, then $dim(W) \leq n$. Furthermore, $dim(W) = n$ ***if and only if*** $W = V$.

The "Two-for-One" Theorem: Let $(W, \oplus, \odot)$ be a ***finite-dimensional subspace*** of a vector space $(V, \oplus, \odot)$, with $dim(W) = n$, and suppose that $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ is any subset of vectors from W. Then: B is a basis for W ***if and only if*** either B is ***linearly independent*** ***or*** B ***Spans*** W.

3.4 Exercises

1. Show that the set of all ***diagonal*** $n \times n$ matrices is a subspace of the set of ***upper triangular*** $n \times n$ matrices (under the usual matrix addition and scalar multiplication). Is it also true if we replace "upper" with "lower?"
2. Is the set of all diagonal $n \times n$ matrices also a subspace of the set of all **symmetric** $n \times n$ matrices?
3. Show that the set D of all vectors in $\mathbb{R}^n$ that are of the form: $\langle r, r, \ldots, r\rangle$ for some $r \in \mathbb{R}$ forms a subspace of $\mathbb{R}^n$. Find a basis for D and its dimension.
4. Does the set E of all vectors in $\mathbb{R}^n$ that are of the form: $\langle r, 2r, 3r, \ldots, nr\rangle$ for some $r \in \mathbb{R}$ form a subspace of $\mathbb{R}^n$? Why or why not? If so, find a basis for E and its dimension.
5. Is the set F of all vectors in $\mathbb{R}^3$ that are of the form: $\left\langle r, r^2, r^3\right\rangle$, for some $r \in \mathbb{R}$, a subspace of $\mathbb{R}^3$? Why or why not? If so, find a basis for F and its dimension.

For Exercises (6) to (12): These Exercises all involve a possible subspace for $\mathbb{P}^2$ or $\mathbb{P}^3$. In order to get the same basis as in the Answers (for those sets that are actually subspaces), let $p(x) = a + bx + cx^2$ be the generic member for $\mathbb{P}^2$ (add dx^3 for Exercises involving $\mathbb{P}^3$). Make the leftmost variables the leading variables as we did in the Examples.

6. Show that the subset W of $\mathbb{P}^3$ defined by: $W = \left\{ p(x) \in \mathbb{P}^3 \,|\, p(-2) = p(1) \text{ and } p'(2) = 0 \right\}$ is a subspace of $\mathbb{P}^3$. Find a basis for W and its dimension.
7. Show that the subset W of $\mathbb{P}^3$ defined by:
$$W = \left\{ p(x) \in \mathbb{P}^3 \,|\, p(-2) = p'(3) \text{ and } p(3) = -2p'(-1) \right\}$$
is a subspace of $\mathbb{P}^3$. Find a basis for W and its dimension.
8. Show that the subset W of $\mathbb{P}^2$ defined by: $W = \{ p(x) \in \mathbb{P}^2 \,|\, 2p(1) = 3p'(-1) \}$ is a subspace of $\mathbb{P}^2$. Find a basis for W and its dimension.
9. Show that the subset W of $\mathbb{P}^2$ defined by: $W = \left\{ p(x) \in \mathbb{P}^2 \,\middle|\, \int_{-1}^{2} p(x)\, dx = 0 \right\}$ is a subspace of $\mathbb{P}^2$. Find a basis for W and its dimension.
10. Show that the subset W of $\mathbb{P}^3$ defined by: $W = \left\{ p(x) \in \mathbb{P}^3 \,\middle|\, \int_{1}^{3} p(x)\, dx = 0 \right\}$ is a subspace of $\mathbb{P}^3$. Find a basis for W and its dimension.
11. Show that: $W = \{ p(x) \in \mathbb{P}^2 \,|\, p(3) = -2\}$ is ***not*** a subspace of $\mathbb{P}^2$.
12. Show that the subset W of $\mathbb{P}^3$ defined by:
$$W = \left\{ p(x) \in \mathbb{P}^3 \,|\, p(-1) + p(2) = 2p'(3), p(1) = p(2) \text{ and } p''(-2) = p'(0) \right\}$$
is a subspace of $\mathbb{P}^3$. Find a basis for W and its dimension.
13. Show that the subset W_1 of $V = Span(\{e^{2x}, e^{3x}, e^{5x}\})$ defined by:
$$W_1 = \left\{ f(x) \in V \,|\, f(0) = 0 \text{ and } f'(0) = 0 \right\}$$
is a subspace of V. Find a basis for W_1 and its dimension.
14. Show that the subset W_2 of $V = Span(\{e^{2x}, e^{3x}, e^{5x}\})$ defined by:
$$W_2 = \{ f(x) \in V \,|\, f(0) = f'(0)\}$$
is a subspace of V. Find a basis for W_2 and its dimension. Why is this problem different from the last one? Without looking at a basis for each space, would it be possible to say that W_1 is a subspace of W_2, or W_2 is a subspace of W_1, or neither?

15. Show that the subset W of $V = Span\left(\left\{\sin(x), \cos(x), \tan(x)\right\}\right)$ defined by:

$$W = \{ f(x) \in V \mid f(0) = f(\pi/4) \}$$

is a subspace of V. Find a basis for W and its dimension.

16. Let F be the set of all ***functions*** $f(x)$ defined on an interval I, ***except*** possibly at a specific point $x = a \in I$.

    a. Show that F is a vector space under the usual addition and scalar multiplication of functions.

    b. Show that: $W = \left\{ f(x) \in F \mid \lim\limits_{x \to a} f(a) = 0 \right\}$ is a subspace of F.

    c. Show that: $U = \left\{ f(x) \in F \mid \lim\limits_{x \to a} f(a) = -2 \right\}$ is ***not*** a subspace of F.

17. Think carefully about this one before you start making any computations: Show that the subset W of $\mathbb{P}^3$ defined by:

$$W = \left\{ p(x) \in \mathbb{P}^3 \mid p(-1) = 0,\ p(1) = 0, \text{ and } p(4) = 0 \right\}$$

is a subspace of $\mathbb{P}^3$. Find a basis for W and its dimension. Again, think smartly!

18. In our Examples, we defined subspaces by listing two (or more) conditions joined by the word "***and***." Consider the set: $W = \left\{ p(x) \in \mathbb{P}^2 \mid p(-2) = 0 \textbf{ or } p(3) = 0 \right\}$.

    a. Show that $z(x)$ is a member of W.

    b. Show that W is closed under scalar multiplication.

    c. Show that W is ***not*** closed under vector addition, by producing two vectors from W whose sum is not in W. Conclude that W is ***not*** a subspace.

For Exercises (19) to (27): Use the ***Two-for-One Theorem*** (and the answer to the corresponding Exercise) to determine if the indicated set of vectors is also a basis for the subspace W in the corresponding Exercise. In other words, check if each member of the basis is actually a vector from the subspace, and check if the set is independent (it is easier to check independence instead of Spanning).

19. Exercise 6: $S = \{3 - 24x - 9x^2 + 5x^3, 7 + 72x + 27x^2 - 15x^3\}$
20. Exercise 7: $S = \{10x^3 + 16x^2 - 99x + 85, 3x^3 + 20x^2 - 43x + 16\}$
21. Exercise 8: $S = \{x^2 + 8x, x^2 - 4\}$
22. Exercise 9: $S = \{x^2 - 2x, 2x - 1\}$
23. Exercise 10: $S = \{x^3 - x - 8, x^3 + 3x^2 - 23, 3x^2 + x - 15\}$
24. Exercise 10: $S = \{x^3 - 5x, 13x^3 - 30x^2, 6x^2 - 13x\}$
25. Exercise 12: $S = \{22 + 10x - x^2 - x^3\}$
26. Exercise 14: $S = \{e^{5x} - 2e^{3x}, e^{5x} - 4e^{2x}\}$
27. Exercise 15: $S = \left\{ \tan(x) + \cos(x) - \sin(x), \sqrt{2}\tan(x) - 2\sin(x) \right\}$
28. Prove that W is a ***subspace*** of a vector space $(V, \oplus, \odot)$ ***if and only if*** for all vectors $\vec{v}, \vec{w} \in W$ and all scalars $r, s \in \mathbb{R}$: $(r \odot \vec{v}) \oplus (s \odot \vec{w}) \in W$.
29. Let W be a ***subspace*** of a vector space $(V, \oplus, \odot)$. Show that $(W, \oplus, \odot)$ is also a ***vector space***, that is, $(W, \oplus, \odot)$ satisfies all 10 Axioms for a vector space in its own right. List all 10 Axioms and decide if the property is inherited from V, or follows from the definition of the word subspace. Warning: carefully prove that W has a zero vector and every vector has a negative.

30. ***Equivalent Conditions for a Basis of a Subspace:***

Suppose that V is a non-zero ***finite-dimensional*** vector space, and $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is any subset of V (we are ***not*** assuming that S is a basis for V, and neither are we assuming that $dim(V) = m$). Prove the following statements:

a. S is a basis for V ***if and only if*** S is a ***maximal*** linearly independent subset of V. This means that if S' is another subset of V, with more vectors than S, then S' must be linearly ***dependent***. Hint: use Proof by Contradiction and the Extension Theorem to show that S must also Span V in the converse direction.

b. S is a basis for V ***if and only if*** S is a ***minimal*** Spanning set of V. This means that if S'' is another subset of V, with fewer vectors than S, then S'' ***cannot*** Span V. Hint: use Proof by Contradiction and the Elimination Theorem to show that S must also be linearly independent in the converse direction. Note that we do not have a Minimizing Theorem to help us.

c. S is a basis for V ***if and only if*** for ***every*** $\vec{v} \in V$, the equation:

$\vec{v} = c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_m\vec{w}_m$ has ***exactly one*** solution in $c_1, c_2, \ldots, c_m$.

This is called the ***Uniqueness of Representation Property*** of a basis.

d. Use (a) to show that $dim(W) = k$ ***if and only if*** there exists a maximal linearly independent subset of W consisting of k vectors.

e. Use (b) to show that $dim(W) = k$ ***if and only if*** there exists a minimal Spanning subset of W consisting of k vectors.

31. The objective of this Exercise is to demonstrate that the ***Two-for-One Theorem*** is sometimes ***false*** if we try to apply it to an ***infinite dimensional*** vector space. Consider $\mathbb{P}$, the vector space of all polynomials. Note that $dim(\mathbb{P}) = \aleph_0$.

a. Show that $S_1 = \{1, x^2, x^4, \ldots, x^{2n}, \ldots\}$ is a linearly independent subset of $\mathbb{P}$, but S_1 does ***not*** Span $\mathbb{P}$. Note that $|S_1| = \aleph_0$, but S_1 is not a basis for $\mathbb{P}$.

b. Let $S_2 = \{1+x, 1+x^2, x+x^2, x+x^3, x^2+x^3, x^2+x^4, \ldots, x^n+x^{n+1}, x^n+x^{n+2}, \ldots\}$.

Show that Spans $\mathbb{P}$, but S_2 is ***not*** linearly independent. Note that $|S_2| = \aleph_0$ as well, but S_2 is not a basis for $\mathbb{P}$ either.

32. Let $S = \{\sin(x+k) \,|\, k \in [0, 2\pi)\}$. Show that S is an infinite set (in other words, if $k_1 \neq k_2$ and both are from $[0, 2\pi)$, then $\sin(x+k_1) \neq \sin(x+k_2)$), but $Span(S)$ is a finite dimensional vector space. What is its dimension?

33. Consider the vector space $Mat(3,2)$, the set of all 3×2 matrices under the usual matrix addition and scalar multiplication. Show that the set:

$$\left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \right\}$$

is a basis for $Mat(3,2)$. Hint: show that every 3×2 can be expressed as a linear combination of these matrices, and they are linearly independent.

34. Use the idea in the previous Exercise to:

a. find a basis for $Mat(m,n)$,

b. prove that the set you constructed is indeed a basis, and

c. prove in general that $Mat(m,n)$ has dimension $m \cdot n$. In particular, this implies that $Mat(n,n)$, the space of $n \times n$ (square) matrices, has dimension n^2.

35. Show that the set:

$$\left\{\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}\right\}$$

is a basis for the vector space of ***diagonal*** 3×3 matrices. Modify the hint in Exercise 32.

36. Let $Diag(n)$ be the vector space of all ***diagonal*** $n \times n$ matrices. Use idea of the previous Exercise to find a general basis for $Diag(n)$ and state its dimension.

37. Show that the set:

$$\left\{\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&1&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&1\\0&0&0\\0&0&0\end{bmatrix},\right.$$
$$\left.\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&1\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix}\right\}$$

is a basis for the vector space of ***upper triangular*** 3×3 matrices.

38. Let $Upper(n)$ be the vector space of all ***upper triangular*** $n \times n$ matrices. Use the idea of the previous Exercise to find a general basis for $Upper(n)$ and state its dimension.

Hint: you will need the formula: $1 + 2 + 3 + \cdots + k = \dfrac{k(k+1)}{2}$

that is usually seen in Precalculus when we study Mathematical Induction.

39. Let $Lower(n)$ be the vector space of all ***lower triangular*** $n \times n$ matrices. Explain why the dimension of $Lower(n)$ should be exactly the same as that of $Upper(n)$.

40. Show that the set:

$$\left\{\begin{bmatrix}1&0&0\\0&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&0\\0&0&1\end{bmatrix},\right.$$
$$\left.\begin{bmatrix}0&1&0\\1&0&0\\0&0&0\end{bmatrix},\begin{bmatrix}0&0&1\\0&0&0\\1&0&0\end{bmatrix},\begin{bmatrix}0&0&0\\0&0&1\\0&1&0\end{bmatrix}\right\}$$

is a basis for the vector space of ***symmetric*** 3×3 matrices. Hint: complete the matrix below depicting the general form of a symmetric 3×3 matrix:

$$\begin{bmatrix}a&b&\cdots\\b&\cdots&\cdots\\\cdots&\cdots&\cdots\end{bmatrix}$$

41. Let $Sym(n)$ be the vector space of all ***symmetric*** $n \times n$ matrices. Use the idea of the previous Exercise to find a general basis for $Sym(n)$, and find its dimension.

42. ***Bisymmetric Matrices:*** We say that an $n \times n$ matrix A is ***bisymmetric*** if the entries of A are symmetric across the main diagonal as well as the opposite diagonal, which is made of the entries $a_{1,n}, a_{2,n-1}, \ldots, a_{n,1}$. Algebraically, this means:

$$a_{i,j} = a_{j,i}, \quad \textbf{\textit{and}} \quad a_{i,j} = a_{n+1-j,n+1-i} \quad \text{for all } i, j = 1 \ldots n.$$

For example, the most general form of a bisymmetric 1×1, 2×2 and 3×3 matrix would be, respectively:

$$[a], \quad \begin{bmatrix} a & b \\ b & a \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} a & b & c \\ b & d & b \\ c & b & a \end{bmatrix}.$$

Let us denote by $Bisym(n)$ the set of all $n \times n$ bisymmetric matrices. Notice that every 1×1 matrix is automatically bisymmetric.

a. Show that $Bisym(2)$ is a subspace of $Sym(2)$.

b. Find a basis for $Bisym(2)$ and state $dim(Bisym(2))$. Hint: decompose the matrix above into two matrices which contain only one distinct letter and the other entries are zeroes.

c. Show that $Bisym(3)$ is a subspace of $Sym(3)$.

d. Find a basis for $Bisym(3)$ and state $dim(Bisym(3))$.

e. Find the general form of all 4×4 bisymmetric matrices and repeat parts (a) and (b). Replace 2 with 4 in the instructions.

f. Find the general form of all 5×5 bisymmetric matrices and repeat parts (a) and (b). Replace 2 with 5 in the instructions.

g. Use your answer in (e) to show that if you erase the 1st and 4th rows and 1st and 4th columns of a 4×4 bisymmetric matrix, what remains is a 2×2 bisymmetric matrix.

h. Use your answer in (f) to show that if you erase the 1st and 5th rows and 1st and 5th columns of a 5×5 bisymmetric matrix, what remains is a 3×3 bisymmetric matrix.

i. Now, let us begin to generalize: show that $Bisym(n)$ is a subspace of $Sym(n)$.

j. Show how to construct a basis for $Bisym(n)$ consisting of two kinds of matrices: (1) matrices where the only non-zero entries are in rows 1 and n and in columns 1 and n, and (2) matrices where all the entries in rows 1 and n and columns 1 and n are zeroes. Show that the matrices of the 2nd kind are in one-to-one correspondence with a basis for $Bisym(n-2)$. Draw from your observations in parts (e) and (f).

k. Use part (j) and Induction to show that:

$$dim(Bisym(n)) = \begin{cases} k^2 & \text{if } n = 2k-1, \text{ an odd number, and} \\ k^2 + k & \text{if } n = 2k, \text{ an even number.} \end{cases}$$

Note: divide your proof into the case when n is odd and n is even.

43. ***The Centralizer of a Matrix:*** If A and B are $n \times n$ matrices, we know in general that $AB \neq BA$. However, let us ***fix*** an $n \times n$ matrix A. Let us define the set:

$$\textbf{\textit{Centralizer(A)}} = \{B \in Mat(n) \,|\, AB = BA\}.$$

In other words, $Centralizer(A)$ consists of all the matrices B that ***commute*** with A.

a. Warm-up: Let $A = \begin{bmatrix} 3 & -2 \\ 4 & 7 \end{bmatrix}$. Describe all the matrices in $Centralizer(A)$.

Hint: let $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$. Solve a system of equations involving the entries of B.

What kind of a system do you get?

b. Prove in general that for ***any*** $n \times n$ matrix A, $Centralizer(A)$ contains an infinite number of matrices. Who are these matrices?

c. Prove in general that for ***any*** $n \times n$ matrix A, $Centralizer(A)$ is a ***subspace*** of $Mat(n)$. Hint: you only have to use the Definition of a subspace.

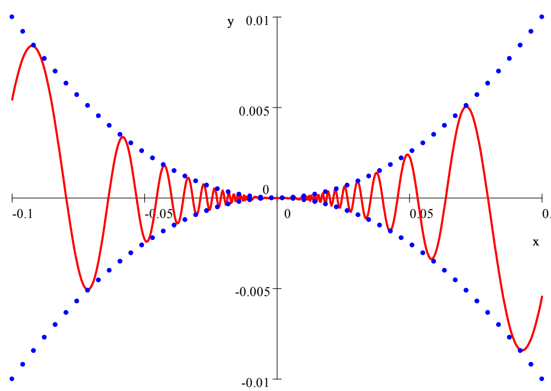
d. Use your answer in (a) to find a basis for the centralizer of the matrix A in that part. What is the dimension of $Centralizer(A)$?

e. Suppose that $A = k\boldsymbol{I}_n$, for some $k \in \mathbb{R}$. What is $Centralizer(A)$?

44. ***A "Pathological" Example:*** We required the members of $C^1(I)$ to be differentiable functions whose derivatives are also continuous. The objective of this Exercise is to investigate a ***differentiable*** function with domain $\mathbb{R}$ whose ***derivative*** is ***not*** continuous.

Consider the function: $f(x) = \begin{cases} x^2 \sin\left(\frac{1}{x}\right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$

with its graph over the interval $[-0.1, 0.1]$ shown below, with its parabolic asymptotes $y = \pm x^2$:



a. Warm-up: Explain why $f(x)$ is continuous at all $x \neq 0$. Then, use the Squeeze Theorem to show that $f(x)$ is also continuous at $x = 0$ (see the graph for a hint). Thus, $f(x)$ is continuous at all $x \in \mathbb{R}$.

b. Show that $f(x)$ is in fact ***differentiable*** for all $x \in \mathbb{R}$, and:

$$f'(x) = \begin{cases} 2x \cdot \sin\left(\frac{1}{x}\right) - \cos\left(\frac{1}{x}\right) & \text{if } x \neq 0, \text{ and} \\ 0 & \text{if } x = 0. \end{cases}$$

Hint: The first formula follows from basic rules of derivatives. The value of $f'(0)$ follows from the ***definition*** of the derivative: $f'(0) = \lim_{h\to 0} \dfrac{f(0+h)-f(0)}{h}$.

The Squeeze Theorem will again be useful.

c. Show that $f'(x)$ is ***not*** continuous at $x = 0$, and in particular: $\lim_{x\to 0} f'(x)$ does not exist.

45. ***Vector Spaces of Infinite Sequences and Series:*** Let us recall some terms from Calculus:

A ***sequence*** $\{ a_n \}_{n=1}^{\infty}$ ***converges*** to a limit $L \in \mathbb{R}$ if $\lim_{n\to\infty} a_n = L$.

This means that for any $\varepsilon > 0$, there exists an integer N such that if $n > N$, then $|a_n - L| < \varepsilon$.

We define: $\{ a_n \}_{n=1}^{\infty} \oplus \{ b_n \}_{n=1}^{\infty} = \{ a_n + b_n \}_{n=1}^{\infty}$, and $k \odot \{ a_n \}_{n=1}^{\infty} = \{ k \cdot a_n \}_{n=1}^{\infty}$.

An ***infinite series*** $\sum_{n=1}^{\infty} a_n$ ***converges*** to a limit $T \in \mathbb{R}$ if the sequence of ***partial sums*** $\{ S_n \}_{n=1}^{\infty}$, where $S_n = a_1 + a_2 + \cdots + a_n$, converges to T. In this case, we say that the sum of the series is T. We know the exact sum for geometric series and telescoping series, among others. If the partial sums do not converge, or have an infinite limit, we say that the series ***diverges***.

We define: $\sum_{n=1}^{\infty} a_n \oplus \sum_{n=1}^{\infty} b_n = \sum_{n=1}^{\infty} (a_n + b_n)$, and $k \odot \sum_{n=1}^{\infty} a_n = \sum_{n=1}^{\infty} (k \cdot a_n)$.

We also say that $\sum_{n=1}^{\infty} a_n$ converges even though we do not know the exact sum T, but we know that T exists, thanks to one of our tests for convergence. These tests include the ***Integral Test***, ***Comparison Tests***, and the ***Alternating Series Test***.

We distinguish between two kinds of convergence:

A series $\sum_{n=1}^{\infty} a_n$ is ***absolutely convergent*** if $\sum_{n=1}^{\infty} |a_n|$ converges. We often use the ***Root*** and ***Ratio Tests*** to determine if a series is absolutely convergent. An absolutely convergent series is also convergent as defined above.

A series $\sum_{n=1}^{\infty} a_n$ is ***conditionally convergent*** if $\sum_{n=1}^{\infty} a_n$ converges but $\sum_{n=1}^{\infty} |a_n|$ diverges.

Exactly ***one*** adjective applies to any infinite series: it is either absolutely convergent, or conditionally convergent, or divergent.

a. Show that the set C of all sequences $\{ a_n \}_{n=1}^{\infty}$ that converge (to ***any*** limit) is a vector space, under the addition and scalar multiplication of sequences defined above.

b. Show that the set C_0 of all sequences $\{ a_n \}_{n=1}^{\infty}$ that converge to 0 is a subspace of C.

c. Show that the set C_3 of all sequences $\{ a_n \}_{n=1}^{\infty}$ that converge to 3 is ***not*** a subspace of C.

d. Show that the set S of infinite series $\sum_{n=1}^{\infty} a_n$ that converge (to ***any*** sum) is a vector space, under the addition and scalar multiplication of infinite series defined above.

e. Show that the set A of ***absolutely convergent*** series is a subspace of S. Hint: use the Triangle Inequality for Real Numbers from Chapter Zero: $|x+y| \le |x| + |y|$.
Which convergence Theorem for positive series would finish the proof?

f. Show that the set D of ***divergent*** series is ***not*** a vector space.

g. Is the set of all ***conditionally convergent*** series a vector space? Why or why not?

3.5 Linear Transformations on General Vector Spaces

Now we begin generalizing the terms, constructions and Theorems from Chapter 2:

Definition: A ***linear transformation*** $T : (V, \oplus_V, \odot_V) \to (W, \oplus_W, \odot_W)$ is a function that assigns a ***unique*** member $\vec{w} \in W$ to every vector $\vec{v} \in V$, such that T satisfies the following conditions for all $\vec{u}, \vec{v} \in V$ and all scalars $k \in \mathbb{R}$:

The Additivity Property: $T(\vec{u} \oplus_V \vec{v}) = T(\vec{u}) \oplus_W T(\vec{v})$, and

The Homogeneity Property: $T(k \odot_V \vec{v}) = k \odot_W T(\vec{v})$.

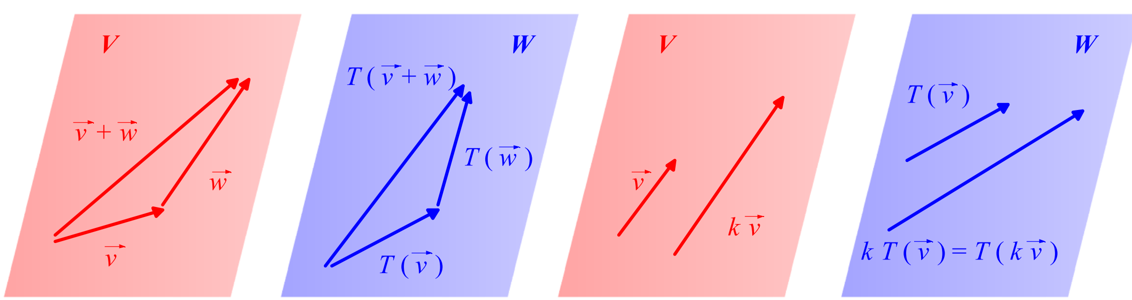
As usual, we write $T(\vec{v}) = \vec{w}$, the ***image*** of $\vec{v}$ under T.

It is of course cumbersome to explicitly specify that the addition and scalar multiplication on the left side of the equations (inside the parentheses) are those of the space V, and those on the right side of the equation are the operations in W, and indeed we will simply write:

$T : V \to W$ is a function, and T satisfies:

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}), \text{ and}$$
$$T(k \cdot \vec{v}) = k \cdot T(\vec{v}),$$

when the addition and scalar multiplication on each side of the equations are clear. We can visualize the two linearity properties with essentially the same diagram as in Section 2.1:



The Additivity Property The Homogeneity Property

As before, we call V the ***domain*** of T and W the ***codomain*** of T. A linear transformation is also called a ***vector space homomorphism***, the last word literally meaning "same form," because T essentially preserves the vector addition and scalar operation in both spaces. When the domain is the same space as the codomain, i.e. $T : V \to V$, we once again call T a ***linear operator*** as we did for Euclidean spaces.

Let us generalize the simplest kinds of linear transformations:

Definitions/Theorem: Let $(V, \oplus_V, \odot_V)$ and $(W, \oplus_W, \odot_W)$ be any two vector spaces. The following three examples are all linear transformations:

The ***zero transformation*** from V to W is the function:

$$Z : V \to W, \text{ where } Z(\vec{v}) = \vec{\mathbf{0}}_W \text{ for all } \vec{v} \in V.$$

The ***identity operator*** of V is the function:

$$\boldsymbol{I}_V : V \to V, \text{ where } \boldsymbol{I}_V(\vec{v}) = \vec{v} \text{ for all } \vec{v} \in V.$$

More generally, for any scalar k, the ***scaling operator of V by k*** is the function:

$$\boldsymbol{S}_k : V \to V, \text{ where } \boldsymbol{S}_k(\vec{v}) = k \odot \vec{v} \text{ for all } \vec{v} \in V.$$

Proving that these are all linear transformations is easy, and will be assigned in the Exercises.

We began our study of linear transformations of Euclidean spaces by constructing the ***standard matrix*** $[T]$ and deriving information about T from the properties of matrix multiplication. However, it is only possible to create a matrix for a linear transformation of abstract vector spaces if the domain is ***finite dimensional***, as we shall see in the next Section. In the meantime, we will try to be as general as possible in re-defining the terms and proving the properties from Chapter 2 in the context of abstract vector spaces, even if these spaces are infinite dimensional.

Evaluation Transformations

Let $V = F(I)$ and $W = \mathbb{R}$. We can construct the ***evaluation transformation*** (sometimes called the ***evaluation homomorphism***):

$$E_a : F(I) \to \mathbb{R}, \text{ where } E_a(f(x)) = f(a),$$

for some ***fixed*** number $a \in I$. For example, if $a = 1$, then $E_1(x^2 - 5x + 2) = -2$, $E_1(\ln(x)) = 0$, $E_1(\sin^{-1}(x)) = \pi/2$.

Let us show that this is a linear transformation. If $f(x), g(x) \in V$ and c is a scalar, then:

$$E_a(f(x) + g(x)) = (f + g)(a) = f(a) + g(a) = E_a(f(x)) + E_a(g(x)).$$

Similarly:

$$E_a(c \cdot f(x)) = c \cdot f(a) = c \cdot E_a(f(x)),$$

so indeed E_a is a linear transformation.

We can extend this definition by choosing several members $a_1, a_2, \ldots, a_n \in I$ and writing $\vec{a} = \langle a_1, a_2, \ldots, a_n \rangle \in \mathbb{R}^n$. We define:

$$E_{\vec{a}} : V \to \mathbb{R}^n, \text{ where } E_{\vec{a}}(f(x)) = \langle f(a_1), f(a_2), \ldots, f(a_n) \rangle.$$

It is easy to verify, using the ideas above, that $E_{\vec{a}}$ is indeed a linear transformation. For example, if $\vec{a} = \langle 0, \pi/6, \pi/2 \rangle$, then:

$$E_{\vec{a}}(\sin(x)) = \langle \sin(0), \sin(\pi/6), \sin(\pi/2) \rangle = \langle 0, 1/2, 1 \rangle.$$

Differentiation and Integration as Linear Transformations

The basic operations of differentiation and integration from Calculus are arguably the most sophisticated examples of linear transformations that we will see in this course, and demonstrate the relevance and relationship of Linear Algebra with the fields of Calculus and Differential Equations.

Example: Let $V = C^1(I)$, the space of all differentiable functions defined on the ***open*** interval I, whose derivatives are also continuous, and let $W = C(I)$, the space of all continuous functions with domain I. Define the ***differentiation*** linear transformation:

$$\boldsymbol{D} : V \to W, \quad \text{where:}$$
$$D(f(x)) = \frac{d}{dx} f(x) = f'(x) \ \text{ for all } f(x) \in V.$$

For example, $D(x^3) = 3x^2$, $D(\sin(3x)) = 3\cos(x)$, and $D(e^{4x}) = 4e^{4x}$.

Let us show that D possesses the two required properties. If $f(x),\ g(x) \in V$ and c is a scalar, then:

$$D(f(x) + g(x)) = \frac{d}{dx}[f(x) + g(x)] = \frac{d}{dx} f(x) + \frac{d}{dx} g(x) = D(f(x)) + D(g(x)),$$

where we used the famous property of the differentiation operation that the derivative of the sum of two differentiable functions is the sum of their derivatives. Similarly, we use the other famous property regarding constant multiples to show:

$$D(c \cdot f(x)) = \frac{d}{dx}[c \cdot f(x)] = c \cdot \frac{d}{dx} f(x) = c \cdot D(f(x)). \square$$

Example: Let $V = C(I)$, the vector space of continuous functions on the ***closed*** interval $I = [a, b]$, and $W = \mathbb{R}$, the vector space of real numbers under ordinary addition and scalar multiplication. Recall from Calculus that any ***continuous*** function on a closed interval is ***integrable*** on this interval, so define the function that finds the ***definite integral:***

$$Def : C(I) \to \mathbb{R}, \quad \text{where:}$$
$$Def(f(x)) = \int_a^b f(x)\, dx.$$

For example, if $I = [0, \pi]$, then:

$$Def(\sin(x)) = \int_0^\pi \sin(x)\, dx = -\cos(x)|_0^\pi = -\cos(\pi) + \cos(0) = 2.$$

Notice that if $f(x) \geq 0$ on the interval I, $Def(f(x))$ is the ***area*** between $f(x)$ and the x-axis above $[a, b]$. Again, we will use some famous properties, this time of the definite integral, to show that Def is in fact a linear transformation. If $f(x)$ and $g(x) \in V$ and c is a scalar, then:

$$Def(f(x) + g(x)) = \int_a^b [f(x) + g(x)]\, dx = \int_a^b f(x)\, dx + \int_a^b g(x)\, dx = Def(f(x)) + Def(g(x)),$$

and similarly:

$$Def(c \cdot f(x)) = \int_a^b c \cdot f(x)\, dx = c \cdot \int_a^b f(x)\, dx = c \cdot Def(f(x)).$$

Thus, Def is indeed a linear transformation. $\square$

Example: Similarly, we can construct the linear transformation:

$$Ind : C(I) \to C^1(I), \quad \text{where:} \quad Ind(f(x)) = \int_a^x f(t)\,dt,$$

and a is the left endpoint of I. The symbol Ind stands for ***indefinite integral***. We know from the Fundamental Theorem of Calculus (traditionally, Part I) that the function on the right, usually denoted $g(x)$, is in fact ***differentiable***, and its derivative is precisely $f(x)$. Since $f(x)$ was assumed to be continuous, we obtain a function that is indeed in $C^1(I)$. The proof of linearity is exactly the same as that for the definite integral.□

Function Spaces Preserved by the Derivative

We saw above that the derivative operation D can be viewed as a linear transformation with domain $C^1(I)$ and codomain $C(I)$, for some interval I. However, suppose we have a function space $W \trianglelefteq C^1(I)$, $W = Span(B)$, such that the derivative of any function $f(x)$ from W is again from W. When this happens, we say that D ***preserves*** W, and D becomes an ***operator***, that is, $D : W \to W$.

By induction, of course, all higher derivatives of functions from W will also be in W, so all higher derivatives D^2, D^3, …, D^n, … likewise preserve W. In particular, D preserves $C^\infty(I)$, since we can take derivatives of all order for any function in $C^\infty(I)$. This will be of particular importance when W is ***finite dimensional*** in Sections 3.6 through 3.8.

Example: We know from Calculus that:

$$\frac{d}{dx}e^x = e^x, \quad \text{so by the } \textbf{\textit{Chain Rule:}} \quad \frac{d}{dx}e^{kx} = ke^{kx}.$$

Let us apply this to compute the following derivative:

$$\frac{d}{dx}(x^2e^{-3x}) = 2xe^{-3x} - 3x^2e^{-3x},$$

which follows from the ***Product Rule***. Notice that the first term involves a ***new*** function, xe^{-3x}. Similarly, we get:

$$\frac{d}{dx}(xe^{-3x}) = e^{-3x} - 3xe^{-3x},$$

and so we get another ***new*** function, e^{-3x}. However, we already know that:

$$\frac{d}{dx}(e^{-3x}) = -3e^{-3x},$$

so we no longer get a function that we didn't see earlier. The computation above tells us that the function space:

$$W = Span(B), \quad \text{where } B = \{x^2e^{-3x}, xe^{-3x}, e^{-3x}\},$$

is ***preserved*** by the derivative. More precisely:

$$\begin{aligned} D(c_1x^2e^{-3x} + c_2xe^{-3x} + c_3e^{-3x}) &= c_1(2xe^{-3x} - 3x^2e^{-3x}) + c_2(e^{-3x} - 3xe^{-3x}) + c_3(-3e^{-3x}) \\ &= -3c_1x^2e^{-3x} + (2c_1 - 3c_2)xe^{-3x} + (c_2 - 3c_3)e^{-3x}, \end{aligned}$$

a linear combination of the three functions in B. Thus, the derivative transformation can be ***restricted*** to W, and becomes an ***operator*** $D : W \to W$. Furthermore, W is the ***smallest*** function space which contains x^2e^{-3x} and is preserved by D. This means that if U is preserved by D and contains x^2e^{-3x}, then $W \trianglelefteq U$.□

Arithmetic Operations on Linear Transformations

Just like linear transformations on the same Euclidean spaces, linear transformations that have the same domains and codomains can be combined using addition, subtraction and scalar multiplication:

Definition/Theorem: Let $T_1, T_2 : (V, \oplus_V, \odot_V) \to (W, \oplus_W, \odot_W)$ be linear transformations, and $k \in \mathbb{R}$. Then, we can define the ***sum***, ***difference*** and ***scalar product*** of these transformations as:

$$(T_1 + T_2) : (V, \oplus_V, \odot_V) \to (W, \oplus_W, \odot_W),$$
$$(T_1 - T_2) : (V, \oplus_V, \odot_V) \to (W, \oplus_W, \odot_W), \text{ and}$$
$$(k \cdot T_1) : (V, \oplus_V, \odot_V) \to (W, \oplus_W, \odot_W),$$

the linear transformations with actions given by:

$$(T_1 + T_2)(\vec{v}) = T_1(\vec{v}) \oplus_W T_2(\vec{v}),$$
$$(T_1 - T_2)(\vec{v}) = T_1(\vec{v}) \ominus_W T_2(\vec{v}), \text{ and}$$
$$(k \cdot T_1)(\vec{v}) = k \odot_W T_1(\vec{v}).$$

Note that the vector addition, subtraction and scalar multiplication on the right side of these equations are those of the ***codomain*** W.

The fact that these functions are indeed linear transformations follow from the linearity properties of T_1 and T_2 and will be left as Exercises.

Example: Let $V = Span(\{\sin(x), \cos(x)\})$, and suppose:

$$T : V \to V, \quad \text{where:} \quad T(f(x)) = 2f'(x) - 5f(x).$$

We know from Calculus that: $\frac{d}{dx}\sin(x) = \cos(x)$ and $\frac{d}{dx}\cos(x) = -\sin(x)$,

and thus the derivative of a linear combination of $\sin(x)$ and $\cos(x)$ is once again a linear combination of these two functions. This gives us another example of a finite-dimensional space which is ***preserved*** by the derivative D. Thus $T(f(x)) \in V$ if $f(x) \in V$. For example:

$$\begin{aligned} T(3\sin(x) + 4\cos(x)) &= 2(3\cos(x) - 4\sin(x)) - 5(3\sin(x) + 4\cos(x)) \\ &= -14\cos(x) - 23\sin(x). \end{aligned}$$

Notice that if D is the differentiation operation and I is the identity operator, then we can write T as the linear combination:

$$T = 2D - 5I. \square$$

More generally, we can construct the ***linear combination*** of a finite list of linear transformations T_1, $T_2, \ldots, T_n$ all of which have domain V and codomain W, with coefficients $c_1, c_2, \ldots, c_n$, via:

$$(c_1T_1 + c_2T_2 + \cdots + c_nT_n)(\vec{v}) = c_1T_1(\vec{v}) + c_2T_2(\vec{v}) + \cdots + c_nT_n(\vec{v}),$$

where again, the scalar multiplications and additions on the right side are those of the codomain W.

The Kernel and Range of a Linear Transformation

We will now generalize two important subspaces associated with a linear transformation that were introduced in Section 2.5:

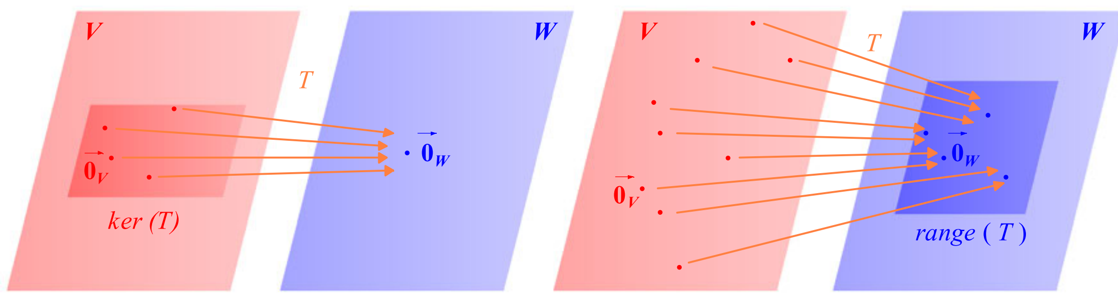
Definition/Theorem: If $T : V \to W$ is a linear transformation, the ***kernel*** of T is the set:

$$ker(T) = \left\{ \vec{v} \in V \mid T(\vec{v}) = \vec{\mathbf{0}}_W \right\}.$$

The set $ker(T)$ is a ***subspace*** of the ***domain*** V.
Similarly, we define the ***range*** of T as the set:

$$range(T) = \left\{ \vec{w} \in W \mid \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in V \right\}.$$

The set $range(T)$ is a ***subspace*** of the ***codomain*** W.



$$ker(T) = \left\{ \vec{v} \in V \mid T(\vec{v}) = \vec{\mathbf{0}}_W \right\}$$ is a subspace of V.

$$range(T) = \left\{ \vec{w} \in W \mid \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in V \right\}$$ is a subspace of W.

Recall that in Section 2.5, we were able to use the fact that for a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$:

$$ker(T) = nullspace([T]) \trianglelefteq \mathbb{R}^n, \text{ and } range(T) = colspace([T]) \trianglelefteq \mathbb{R}^m,$$

which are obviously subspaces of the domain and codomain, respectively. Using only the definitions above, we will be able to prove that $ker(T)$ and $range(T)$ are subspaces of V and W, respectively, without the help of any matrix.

Proof: In the Exercises, you will show that $T\left(\vec{\mathbf{0}}_V\right) = \vec{\mathbf{0}}_W$, as implied in the diagrams, and so by the definitions, $\vec{\mathbf{0}}_V \in ker(T)$ and $\vec{\mathbf{0}}_W \in range(T)$. We need to show that both sets are closed under vector addition and scalar multiplication. Suppose $\vec{v}_1$ and $\vec{v}_2$ are both members of $ker(T)$. Thus:

$$T(\vec{v}_1) = \vec{\mathbf{0}}_W = T(\vec{v}_2), \text{ and so:}$$

$$T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2) = \vec{\mathbf{0}}_W + \vec{\mathbf{0}}_W = \vec{\mathbf{0}}_W.$$

Thus, $\vec{v}_1 + \vec{v}_2$ likewise satisfies the definition of a vector in $ker(T)$. Similarly:

$$T(k\vec{v}_1) = kT(\vec{v}_1) = k \cdot \vec{\mathbf{0}}_W = \vec{\mathbf{0}}_W,$$

and so $k\vec{v}_1$ likewise satisfies the definition of a vector in ker(T). Now for $range(T)$: suppose $\vec{w}_1$ and $\vec{w}_2$ are members of $range(T)$. Then there must exist two vectors $\vec{v}_1$ and $\vec{v}_2$ from V such that:

$$\vec{w}_1 = T(\vec{v}_1) \text{ and } \vec{w}_2 = T(\vec{v}_2).$$

Note: these are not the same $\vec{v}_1$ and $\vec{v}_2$ that we used in the proof for $ker(T)$ above. Now:

$$\vec{w}_1 + \vec{w}_2 = T(\vec{v}_1) + T(\vec{v}_2) = T(\vec{v}_1 + \vec{v}_2).$$

But $\vec{v}_1 + \vec{v}_2 \in V$, and so $\vec{w}_1 + \vec{w}_2$ likewise satisfies the definition of a vector in $range(T)$. Similarly:

$$k\vec{w}_1 = kT(\vec{v}_1) = T(k\vec{v}_1),$$

and since $k\vec{v}_1 \in V$, $k\vec{w}_1$ likewise satisfies the definition of a vector in $range(T)$. ■

Example: Let $D : C^1(I) \to F(I)$ be the derivative transformation. To find $ker(D)$, we must find all the functions $f(x) \in C^1(I)$ such that $f'(x) = z(x)$, the function that is identically 0 for ***all*** $x \in I$.

We know that if $f(x) = k$, a constant, for all $x \in I$, then $f'(x) = z(x)$. Conversely, suppose $f(x)$ is a differentiable function on I and $f'(x) = 0$ for all x in I. If $f(x)$ is ***not*** a constant function, then there must be two points c_1 and c_2 in I where $f(c_1) \neq f(c_2)$. In this case, the slope of the ***secant line*** through $(c_1, f(c_1))$ and $(c_2, f(c_2))$ would be:

$$m = \frac{f(c_2) - f(c_1)}{c_2 - c_1} \neq 0.$$

But ***The Mean Value Theorem*** from Calculus tells us that there must be a $c \in (c_1, c_2)$ where $f'(c) = m$. This violates the given condition that $f'(x) = 0$ for all x in I. Thus, every non-constant function is not in $ker(D)$. This tells us that $f(x) \in ker(D)$ ***if and only if*** $f(x) = k$, a constant. Since k can be written as $k = k \cdot 1$, we can conclude that:

$$ker(D) = Span(\{1\}). \square$$

Example: Let $Def : C([0,1]) \to \mathbb{R}$ be the definite integral transformation:

$$Def(f(x)) = \int_0^1 f(x)\,dx.$$

The range of this transformation is a subspace of $\mathbb{R}$, which is a 1-dimensional vector space. Thus, its only subspaces are either 0 or 1 dimensional, and so they are either only $\{0\}$ or $\mathbb{R}$. But:

$$\int_0^1 x\,dx = \frac{x^2}{2}\Big|_0^1 = \frac{1}{2}.$$

so the range cannot be only $\{0\}$. Thus, $range(Def) = \mathbb{R}$.

We can make this more explicit: if r is any real number, let us find a multiple of the function $f(x) = x$ that has r as its definite integral, by solving the equation:

$$\int_0^1 cx\,dx = \frac{cx^2}{2}\Big|_0^1 = \frac{c}{2} = r.$$

Thus, $c = 2r$. In other words, $Def(2r \cdot x) = r$, so for any $r \in \mathbb{R}$, we can find a function $g(x) = 2r \cdot x \in C([0,1])$, such that $Def(g(x)) = r$.

Once again, we can conclude that $range(Def) = \mathbb{R}$. There are ***other*** continuous functions that have r as their definite integral, but all we need is ***one*** such function. $\square$

Example: Let $V = Span(B)$, where $B = \{\sin(x), \cos(x)\}$. We already know that B is linearly independent, and so $\dim(V) = 2$. We saw in an Example above that V is also preserved by the derivative operator D, that is: $D : V \to V$. More explicitly, if:

$$f(x) = c_1 \sin(x) + c_2 \cos(x) \in V, \text{ then}$$
$$D(f(x)) = c_1 \cos(x) - c_2 \sin(x) \in V \text{ also.}$$

From the equation above, though, we can also see that $D(f(x)) = z(x)$ ***if and only if*** both c_1 and c_2 are zero. Thus, $ker(D) = \{z(x)\}$.

To find $range(D)$, consider an arbitrary vector $g(x)$ in V:

$$g(x) = d_1 \sin(x) + d_2 \cos(x) \in V.$$

We must ask: under what conditions on d_1 and d_2 can we find $f(x) \in V$ such that:

$$D(f(x)) = g(x)?$$

But if we let $f(x) = c_1 \sin(x) + c_2 \cos(x)$ as we did above, then we must solve the equation:

$$c_1 \cos(x) - c_2(\sin(x)) = d_1 \sin(x) + d_2 \cos(x).$$

By the Uniqueness of Representation Theorem, we must have:

$$c_1 = d_2 \text{ and } -c_2 = d_1.$$

This tells us that for any coefficients d_1 and d_2 in $g(x)$, we can find a function $f(x) \in V$ such that $D(f(x)) = g(x)$. Thus, $range(D) = V$.

Notice that D in this Example has domain V and not $C^1(I)$, and so we obtain a different answer for the kernel.□

Example: Let $V = Span(B)$, where $B = \{e^{2x}, e^{-4x}\}$. We know from Section 3.2 that B is linearly independent. It is also easy to see that V is preserved by the derivative operator D, and thus it is also preserved by the 2nd derivative D^2. Furthermore, the function:

$$T : V \to V, \text{ given by } T(f(x)) = f^{//}(x) + 3f^{/}(x) - 10f(x)$$

can also easily be checked to be a linear transformation. Consider the typical member of V:

$$f(x) = c_1 e^{2x} + c_2 e^{-4x}.$$

Its derivatives are:

$$f^{/}(x) = 2c_1 e^{2x} - 4c_2 e^{-4x}, \text{ and}$$
$$f^{//}(x) = 4c_1 e^{2x} + 16c_2 e^{-4x}.$$

Thus:

$$T(f(x)) = 4c_1 e^{2x} + 16c_2 e^{-4x} + 3(2c_1 e^{2x} - 4c_2 e^{-4x}) - 10(c_1 e^{2x} + c_2 e^{-4x})$$
$$= 4c_1 e^{2x} + 16c_2 e^{-4x} + 6c_1 e^{2x} - 12c_2 e^{-4x} - 10c_1 e^{2x} - 10c_2 e^{-4x} = -6c_2 e^{-4x}.$$

We can therefore conclude that $f(x) \in ker(T)$ ***if and only if*** $c_2 = 0$, that is, $f(x) = c_1 e^{2x}$. Similarly, since c_2 can be any real number, $range(T)$ consists of all multiples of e^{-4x}. Thus:

$$ker(T) = Span(\{e^{2x}\}) \text{ and}$$
$$range(T) = Span(\{e^{-4x}\}).\square$$

In the next Section, we will construct a matrix for T when V and W are finite dimensional, and we will use these matrices in Section 3.7 to solve for the kernel and range of T as before — using information obtained from the ***nullspace*** and ***columnspace*** of this matrix.

3.5 Section Summary

A ***linear transformation*** $T : (V, \oplus_V, \odot_V) \to (W, \oplus_W, \odot_W)$ is a function that assigns a ***unique*** member $\vec{w} \in W$ to every vector $\vec{v} \in V$, such that T satisfies for all $\vec{u}$, $\vec{v} \in V$ and all scalars c:

$$T(\vec{u} \oplus_V \vec{v}) = T(\vec{u}) \oplus_W T(\vec{v}), \text{ and } T(c \odot_V \vec{u}) = c \odot_W T(\vec{u}).$$

We call V the ***domain*** of T and W the ***codomain*** of T. When the domain is the same space as the codomain, i.e. $T : V \to V$, we call T a ***linear operator***.

We can construct the evaluation transformation (sometimes called the ***evaluation homomorphism***): $E_a : F(I) \to \mathbb{R}$, where $E_a(f(x)) = f(a)$, for some ***fixed*** number $a \in I$. We can extend this definition by choosing several members $a_1, a_2, \ldots, a_n \in I$ and writing $\vec{a} = \langle a_1, a_2, \ldots, a_n \rangle \in \mathbb{R}^n$. We define: $E_{\vec{a}} : V \to \mathbb{R}^n$, where: $E_{\vec{a}}(f(x)) = \langle f(a_1), f(a_2), \ldots, f(a_n) \rangle$.

The ***differentiation*** operation and both definite and indefinite ***integration*** operations are linear transformations on the appropriate function spaces.

In some instances, the derivative of every function in a function space W is again a function from W. In this case, we say that D ***preserves*** W, and we can view the restriction of the derivative operation D to W to be an operator, $D : W \to W$. In particular, D preserves $C^\infty(I)$.

We can add and subtract two linear transformations and take a scalar multiple of a linear transformation using the arithmetic in the codomain of these linear transformations.

Suppose that $T : V \to W$ is a linear transformation.
The ***kernel*** of T: $ker(T) = \left\{ \vec{v} \in V \mid T(\vec{v}) = \vec{\mathbf{0}}_W \right\}$, is a subspace of the domain V, and the ***range*** of T: $range(T) = \left\{ \vec{w} \in W \mid \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in V \right\}$, is a subspace of the codomain W.

3.5 Exercises

1. Compute the evaluation transformations:

   a. $E_2(5x^2 - 3x + 7)$. b. $E_{\pi/3}(\tan(x))$ c. $E_e(\ln(x))$

2. Let $\vec{a} = \langle \pi, \sin^{-1}(3/5), \pi/6 \rangle$. Compute:

   a. $E_{\vec{a}}(\sin(x))$ b. $E_{\vec{a}}(\cos(2x))$ c. $E_{\vec{a}}(\tan(x))$

3. Let $\vec{a} = \langle -3, 1 \rangle$, and consider the linear transformation: $E_{\vec{a}} : \mathbb{P}^2 \to \mathbb{R}^2$.

   a. Compute $E_{\vec{a}}(-5x^2 + 8x + 3)$.

   b. Find a basis for $ker(E_{\vec{a}})$. Hint: think of a factored form for $p(x)$ if $E_{\vec{a}}(p(x)) = z(x)$.

4. Let $\vec{a} = \langle -5, 3, -2 \rangle$, and consider the linear transformation: $E_{\vec{a}} : \mathbb{P}^3 \to \mathbb{R}^3$.

   a. Compute $E_{\vec{a}}(7x^3 - 4x^2 + 3x - 6)$

   b. Find a basis for $ker(E_{\vec{a}})$. Hint: think of a factored form for $p(x)$ if $E_{\vec{a}}(p(x)) = z(x)$.

5. Let $\vec{a} = \langle -5, 3, -2 \rangle$, and consider the linear transformation: $E_{\vec{a}} : \mathbb{P}^2 \to \mathbb{R}^3$.

   a. Compute $E_{\vec{a}}(4x^2 - 5x - 8)$

   b. Describe $ker(E_{\vec{a}})$. Hint: how many roots can a quadratic polynomial have? Why is this Exercise different from the two previous ones?

6. Let $T : \mathbb{P}^2 \to \mathbb{R}^3$ be given by: $T(p(x)) = \langle p(1), p'(-2), 2p(3) - 5p'(1) \rangle$.
   a. Compute $T(-5x^2 + 8x + 3)$.
   b. Show that T is in fact a linear transformation.
7. Let $T : \mathbb{P}^2 \to \mathbb{R}^4$ be given by: $T(p(x)) = \left\langle p(-2), p'(1), p''(3), \int_0^1 p(x)dx \right\rangle$.
   a. Compute $T(-5x^2 + 8x + 3)$.
   b. Show that T is in fact a linear transformation.
8. Let $T : \mathbb{P}^3 \to \mathbb{P}^1$ be given by: $T(p(x)) = p''(x)$.
   a. Compute $T(2x^3 + 5x^2 - 4x + 3)$.
   b. Explain why, in general, if $p(x) \in \mathbb{P}^3$, then $T(p(x)) \in \mathbb{P}^1$.
   c. Show that T is in fact a linear transformation.
   d. Show that $ker(T) = \mathbb{P}^1 \trianglelefteq \mathbb{P}^3$. Hint: Start with a generic cubic: $p(x) = c_0 + c_1 x + c_2 x^2 + c_3 x^3$, and show that $c_2 = c_3 = 0$ ***if and only if*** $p(x) \in ker(T)$.
   e. Show that $range(T) = \mathbb{P}^1$.
9. Let $T : \mathbb{P}^2 \to \mathbb{P}^4$ be given by: $T(p(x)) = p(x) \cdot x^2$.
   a. Compute $T(3x^2 + 2x - 7)$.
   b. Explain why, in general, if $p(x) \in \mathbb{P}^2$, then $T(p(x)) \in \mathbb{P}^4$.
   c. Show that T is in fact a linear transformation.
   d. Show that $ker(T) = \{z(x)\}$. Hint: Look at the highest degree.
   e. Show that $range(T) = Span(\{x^2, x^3, x^4\})$. In other words, $range(T)$ consists exactly of those polynomials of the form $p(x) = c_2 x^2 + c_3 x^3 + c_4 x^4$, that is, polynomials missing the constant and linear terms.
   f. Prove in general that if $q(x)$ is a ***fixed*** polynomial of degree k, then:
   $$M_{q(x)} : \mathbb{P}^n \to \mathbb{P}^{n+k} \quad \text{given by:} \quad M_{q(x)}(p(x)) = p(x) \cdot q(x)$$
   (i.e., multiplication by $q(x)$) is a linear transformation.
   g. Show that $ker(M_{q(x)}) = \{z(x)\}$. Use the same idea as part (d).
   h. In contrast, show that: $T : \mathbb{P}^2 \to \mathbb{P}^2$ given by: $T(p(x)) = p(x) + x^2$, is ***not*** a linear transformation.
10. Consider $Ind : \mathbb{P}^2 \to \mathbb{P}^3$, given by: $Ind(p(x)) = \int_0^x p(t)dt$.

    We know from the text that Ind is a linear transformation.
    a. Compute $Ind(3x^2 + 2x - 7)$.
    b. Explain why, in general, if $p(x) \in \mathbb{P}^2$, then $Ind(p(x)) \in \mathbb{P}^3$.
    c. Show that $ker(Ind) = \{z(x)\}$. Hint: Start with a generic quadratic, $p(x) = c_0 + c_1 x + c_2 x^2$ and show that if $\int_0^x p(t)dt = z(x)$, then $c_0 = c_1 = c_2 = 0$.
    d. Use the same idea to show that $range(Ind) = Span(\{x, x^2, x^3\})$.

For Exercises (11) to (18): We saw that $W = Span(\{x^2 e^{-3x}, xe^{-3x}, e^{-3x}\})$ and $V = Span(\{\sin(x), \cos(x)\})$ are preserved by the derivative D. For the following Exercises, we are given a function space $W = Span(B)$, where $B = \{f_1(x), f_2(x), \ldots f_n(x)\}$ and a function $f(x) \in W$. (a) Find $D(f(x))$ for the indicated $f(x)$; (b) Construct a generic function $f(x) = c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) \in W$; (c) Show that $D(f(x)) \in W$ for the $f(x)$ you constructed in part (b), and thus D ***preserves*** W; (d) Find a basis for $ker(D)$ or show that $ker(D) = \{z(x)\}$; (e) Find a basis for $range(D)$, or show that $range(D) = W$.

11. $W = Span(B)$, where $B = \{e^{-x}, e^{2x}\}$; $f(x) = 5e^{-x} - 3e^{2x}$.

12. $W = Span(B)$, where $B = \{e^x \sin(x), e^x \cos(x)\}$; $f(x) = 4e^x \sin(x) - 3e^x \cos(x)$.

13. $W = Span(B)$, where $B = (\{e^{-3x} \sin(2x), e^{-3x} \cos(2x)\})$; $f(x) = 5e^{-3x} \sin(2x) - 9e^{-3x} \cos(2x)$.

14. $W = Span(B)$, where $B = (\{xe^{5x}, e^{5x}\})$; $f(x) = -2xe^{5x} + 7e^{5x}$.

15. $W = Span(B)$, where $B = (\{x^2e^{-4x}, xe^{-4x}, e^{-4x}\})$; $f(x) = -5x^2e^{-4x} + 2xe^{-4x} - 7e^{-4x}$.

16. $W = Span(B)$, where $B = (\{x^2 \cdot 5^x, x \cdot 5^x, 5^x\})$; $f(x) = -4x^2 \cdot 5^x + 9x \cdot 5^x - 2 \cdot 5^x$.

17. $W = \mathbb{P}^3 = Span(B)$, where $B = \{1, x, x^2, x^3\}$; $f(x) = 2x^3 - 8x^2 + 3x + 7$.

18. $W = Span(B)$, where $B = \{x\sin(2x), x\cos(2x), \sin(2x), \cos(2x)\}$;
$f(x) = 4x\sin(2x) + 9x\cos(2x) - 5\sin(2x) + 8\cos(2x)$.

19. Let $B = \{\sin(x), \cos(x)\}$ and $W = Span(B)$.
Consider the function $T : W \to W$ given by: $T(f(x)) = f''(x) - 3f'(x) + 2f(x)$.
   a. Compute $T(3\sin(x) + 8\cos(x))$.
   b. Explain why, in general, if $f(x) \in W$, then $T(f(x)) \in W$ also.
   c. Show that T is in fact a linear operator. Hint: since $W \trianglelefteq C^2(\mathbb{R})$, it is sufficient to show that the additivity and homogeneity properties are enjoyed by any two twice-differentiable functions $f(x)$ and $g(x)$.

20. Let $S = \{e^{4x} \sin(3x), e^{4x} \cos(3x)\}$, and let $U = Span(S)$.
Consider the linear operator $T : U \to U$ given by: $T(f(x)) = f''(x) - 3f'(x) + 2f(x)$.
Notice that this is the same formula as that of Exercise 19, but the domain and codomain are now a different function space U.
   a. Compute $T(5e^{4x} \sin(3x) - 9e^{4x} \cos(3x))$.
   b. Explain why U is closed under T, that is, if $f(x) \in U$, then $T(f(x)) \in U$ also.
   c. Explain why your work in Exercise 19 (c) is enough to show that this T is also a linear operator (even though it is acting on a subspace of $C^2(\mathbb{R})$).

21. The purpose of this Exercise is to generalize the previous Exercise. Let $S = \{e^{ax} \sin(bx), e^{ax} \cos(bx)\}$, for some real constants a and b, and let $U = Span(S)$.
   a. Compute $D(c_1e^{ax} \sin(bx) + c_2e^{ax} \cos(bx))$.
   b. Use (a) to explain why, in general, if $f(x) \in U$, then $D(f(x)) \in U$ also, and thus, D is an operator: $D : U \to U$.

22. Let $S = \{e^{-4x}, e^{3x}, e^{5x}\}$, and let $W = Span(S)$.
   a. Compute $D(c_1e^{-4x} + c_2e^{3x} + c_3e^{5x})$.
   b. Use (a) to explain why, in general, if $f(x) \in W$, then $D(f(x)) \in W$ also, and thus, D is an operator: $D : W \to W$.
   c. Use induction to show that all higher derivatives D^n preserve W, for every integer $n \geq 1$.
   d. Consider the linear transformation $T : W \to W$, given as the linear combination:
   $$T = 5D^2 - 8D - 21\boldsymbol{I}_W.$$
   Find $T(c_1e^{-4x} + c_2e^{3x} + c_3e^{5x})$.
   e. Use (d) to find a basis for $ker(T)$.
   f. Use (d) to find a basis for $range(T)$.

23. Suppose that $W = Span(B)$ and W is preserved by the derivative, that is, $D : W \to W$ is an operator. Prove that any linear combination: $c_0 \boldsymbol{I}_W + c_1 D + c_2 D^2 + \cdots c_n D^n$, of the identity operator on W, the derivative D, and all higher derivatives, also preserves W. Hint: use induction.

24. Suppose that $\vec{a} = \langle a_0, a_1, \ldots, a_n \rangle$, where a_0, a_1, $\ldots$, a_n are $n+1$ ***distinct*** real numbers. Consider the evaluation homomorphism:

$$E_{\vec{a}} : \mathbb{P}^n \to \mathbb{R}^{\boldsymbol{n+1}}, \text{ where } E_{\vec{a}}(p(x)) = \langle\, p(a_0), p(a_1), \ldots, p(a_n) \,\rangle.$$

Prove that $ker(E_{\vec{a}}) = \{z(x)\}$, where $z(x)$ is the zero polynomial. Hint: use the Fundamental Theorem of Algebra. Hint: review Exercises 1 to 5.

25. Suppose instead that $\vec{a} = \langle a_1, a_2, \ldots, a_n \rangle$, where a_1, $\ldots$, a_n are n ***distinct*** real numbers. Consider the evaluation homomorphism:

$$E_{\vec{a}} : \mathbb{P}^n \to \mathbb{R}^{\boldsymbol{n}}, \text{ where } E_{\vec{a}}(p(x)) = \langle\, p(a_1), p(a_2), \ldots, p(a_n) \,\rangle.$$

Prove that $ker(E_{\vec{a}})$ is a 1-dimensional subspace of $\mathbb{P}^n$. What polynomial can serve as a basis for the kernel? Hint: you may leave your answer in factored form.

26. Let $T : Mat(2,3) \to Mat(3,2)$ be given by: $T(A) = A^{\top}$, the transpose of A.

    a. Let $A = \begin{bmatrix} 4 & -3 & 5 \\ 0 & 1 & -7 \end{bmatrix}$. Compute $T(A)$.

    b. Explain why, in general, if $A \in Mat(2,3)$ then $T(A) \in Mat(3,2)$.

    c. Show that T is indeed a linear transformation.

    d. Show that $ker(T) = \{\mathbf{0}_{2,3}\}$.

    e. Show that $range(T) = Mat(3,2)$.

27. State and prove analogous statements to (b) through (e) from the previous Exercise if in general $T : Mat(m,n) \to Mat(n,m)$ is the transpose operation: $T(A) = A^{\top}$.

28. ***The Trace of a Square Matrix:*** Define the ***trace*** function: $tr : Mat(n,n) \to \mathbb{R}$, given by:

$$tr(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}.$$

Prove that tr is a linear transformation.

29. Let $T : V \to W$ be any linear transformation. Prove that $T\left(\vec{\mathbf{0}}_V\right) = \vec{\mathbf{0}}_W$ in two different ways:

    a. Complete the property: $\vec{\mathbf{0}}_V \oplus \vec{\mathbf{0}}_V = ???$ and use it in your proof.

    b. Complete the property: $0 \odot \vec{v} = ???$ and use it in your proof.

30. Show that the zero transformation $Z : V \to W$, identity operator $\boldsymbol{I}_V : V \to V$, and the scaling operators $\boldsymbol{S}_k : V \to V$, are in fact linear transformations.

31. Let $T : V \to W$ be a linear transformation. Prove that T is the zero transformation Z ***if and only if*** $ker(T) = V$ ***or*** $range(T) = \left\{\vec{\mathbf{0}}_W\right\}$.

32. Let T_1 and T_2 be linear transformations from V to W. Show that $T_1 + T_2$, $T_1 - T_2$ and $k \cdot T_1$ are also linear transformations from V to W.

33. Flashback to Section 3.1: Prove that the set of all linear transformations $T : V \to W$ is itself a ***vector space*** under the addition and scalar multiplication from the previous Exercise. This vector space is called $Hom(V, W)$. Hint: Which of the 10 Axioms are verified by the previous Exercise?

34. ***From Kansas to Oz:*** Show that $T : (\mathbb{R}, +, \cdot) \to (\mathbb{R}^+, \oplus, \odot)$ given by: $T(x) = e^x$ is a linear transformation. Here, $\mathbb{R}$ is a vector space under ordinary addition and scalar multiplication, but recall that the vector operations in $\mathbb{R}^+$ are: $\vec{x} \oplus \vec{y} = \overrightarrow{xy}$, and $r \odot \vec{x} = \overrightarrow{x^r}$.

3.6 Coordinate Vectors and Matrices for Linear Transformations

We saw in Section 2.1 that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ corresponds to an $m \times n$ matrix, the ***standard matrix*** of T. Our major goal in this Section is to see that if V and W are ***finite dimensional*** vector spaces, then we can likewise ***simulate*** the action of $T : V \to W$ using ***matrix multiplication***, and that there are in fact an infinite number of ways to do this. First, we need to generalize the concept of ***coordinates*** for a vector in $\mathbb{R}^n$:

Definition: Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be an ordered basis for an n-dimensional vector space V. If $\vec{v}$ is any vector in V, we know that $\vec{v}$ can be expressed ***uniquely*** as a linear combination of the vectors of B: $\vec{v} = c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_n\vec{w}_n$.
We call the vector $\langle c_1, c_2, \ldots, c_n \rangle$ the ***coordinate vector*** of $\vec{v}$ with respect to B, written as:

$$\langle \vec{v} \rangle_B = \langle c_1, c_2, \ldots, c_n \rangle.$$

The $n \times 1$ matrix corresponding to $\langle \vec{v} \rangle_B$ is called the ***coordinate matrix*** of $\vec{v}$ with respect to B, written as:

$$[\vec{v}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

Example: Let $B = \{\langle 1, 0, 1 \rangle, \langle -1, 1, 0 \rangle, \langle 0, -1, 1 \rangle\}$. We could leave it as an Exercise to check that this set is indeed a basis for $\mathbb{R}^3$, but we will see below that it will not be necessary. Now, suppose $\vec{v} = \langle -4, 8, -2 \rangle$. To find the coordinates for $\vec{v}$ with respect to B, we have to solve:

$$\langle -4, 8, -2 \rangle = c_1\langle 1, 0, 1 \rangle + c_2\langle -1, 1, 0 \rangle + c_3\langle 0, -1, 1 \rangle.$$

As before, we assemble the vectors of B into columns and solve the augmented system:

$$\left[\begin{array}{ccc|c} 1 & -1 & 0 & -4 \\ 0 & 1 & -1 & 8 \\ 1 & 0 & 1 & -2 \end{array}\right], \quad \text{whose rref is} \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -3 \end{array}\right].$$

Thus, $\langle \vec{v} \rangle_B = \langle 1, 5, -3 \rangle$, and indeed, we can check:

$$1 \bullet \langle 1, 0, 1 \rangle + 5 \bullet \langle -1, 1, 0 \rangle - 3 \bullet \langle 0, -1, 1 \rangle = \langle -4, 8, -2 \rangle.$$

At this point, we also notice that the left side of the rref of our augmented matrix is $\boldsymbol{I}_3$, so as a bonus, we can see that B is indeed a ***basis*** for $\mathbb{R}^3$.

Now let us work ***backwards***. Suppose $\vec{w} \in \mathbb{R}^3$, and $\langle \vec{w} \rangle_B = \langle -5, 2, 7 \rangle$. Then:

$$\vec{w} = -5\langle 1, 0, 1 \rangle + 2\langle -1, 1, 0 \rangle + 7\langle 0, -1, 1 \rangle = \langle -7, -5, 2 \rangle. \square$$

Notice that if we start with a basis B for $\mathbb{R}^n$ and a vector $\vec{v} \in \mathbb{R}^n$, then $\langle \vec{v} \rangle_B$ is another vector from $\mathbb{R}^n$. The process of computing the coordinate vector turns out to be a familiar special kind of operation:

Theorem: For any ordered ***basis*** $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ of an n-dimensional vector space V, the function $T : V \to \mathbb{R}^n$ given by:

$$T(\vec{v}) = \langle \vec{v} \rangle_B$$

is a ***linear transformation***. In particular, if $V = \mathbb{R}^n$ and B is a basis for $\mathbb{R}^n$, then T is in fact ***one-to-one*** and ***onto***, i.e., an ***isomorphism*** of $\mathbb{R}^n$.

Proof: Suppose that $\langle \vec{u} \rangle_B = \langle c_1, c_2, \ldots, c_n \rangle$, and $\langle \vec{v} \rangle_B = \langle d_1, d_2, \ldots, d_n \rangle$. These mean that:

$$\begin{aligned} \vec{u} &= c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_n\vec{w}_n, \quad \text{and} \\ \vec{v} &= d_1\vec{w}_1 + d_2\vec{w}_2 + \cdots + d_n\vec{w}_n. \end{aligned}$$

Thus, we have:

$$\vec{u} + \vec{v} = (c_1 + d_1)\vec{w}_1 + (c_2 + d_2)\vec{w}_2 + \cdots + (c_n + d_n)\vec{w}_n,$$

and therefore:

$$\begin{aligned} \langle \vec{u} + \vec{v} \rangle_B &= \langle c_1 + d_1, c_2 + d_2, \ldots, c_n + d_n \rangle \\ &= \langle c_1, c_2, \ldots, c_n \rangle + \langle d_1, d_2, \ldots, d_n \rangle \\ &= \langle \vec{u} \rangle_B + \langle \vec{v} \rangle_B. \end{aligned}$$

Similarly, we can show that $\langle c \cdot \vec{u} \rangle_B = c \cdot \langle \vec{u} \rangle_B$, which we leave as an Exercise. Thus the process of finding the coordinates of an arbitrary vector with respect to B is indeed a linear transformation.

Now, we have to show that T is one-to-one and onto in the special case when $V = \mathbb{R}^n$. But this follows automatically because every vector $\vec{v}$ is ***uniquely expressible*** as a linear combination of the vectors of B. Thus T is onto because every coordinate vector lets us compute the corresponding vector using the entries as the coefficients for the vectors of B, and T is one-to-one because two different vectors must have different coordinates by the Uniqueness of Representation property. We will see later in this Chapter that for a fixed basis B, the process of finding coordinates with respect to B is also an "isomorphism" from any n-dimensional vector space V to $\mathbb{R}^n$. ■

Coordinates for Abstract Vector Spaces

Let us look at some examples of bases and coordinates for polynomial and function spaces.

Example: Consider $\mathbb{P}^n = \{ p(x) = c_0 + c_1 x + \cdots + c_n x^n \,|\, c_0, c_1, \ldots, c_n \in \mathbb{R} \}$.

We know that the ***monomials*** $B = \{1, x, x^2, \ldots, x^n\}$ form a basis for $\mathbb{P}^n$.

The ***coefficients*** of these monomials are exactly the ***coordinates*** of $p(x) \in \mathbb{P}^n$ with respect to this basis: $\langle p(x) \rangle_B = \langle c_0 + c_1 x + \cdots + c_n x^n \rangle_B = \langle c_0, c_1, c_2, \ldots, c_n \rangle$.
It is thus natural to think of B as a ***standard basis*** for $\mathbb{P}^n$, and also shows why the ***ascending*** order is more ***natural***. Let us look at a specific polynomial space, say $\mathbb{P}^3$, with standard basis $B = \{1, x, x^2, x^3\}$. Thus, for example:

$$\begin{aligned} \langle 3 - 5x + 7x^2 + x^3 \rangle_B &= \langle 3, -5, 7, 1 \rangle, \\ \langle 4x - x^3 \rangle_B &= \langle 0, 4, 0, -1 \rangle, \quad \text{and} \\ \langle 9x^2 - 6x + 8 \rangle_B &= \langle 8, -6, 9, 0 \rangle. \end{aligned}$$

Notice that we have to be careful with "missing" powers, and also respect the ***order*** of the basis. □

Example: Let us stay with the polynomial space $\mathbb{P}^3$. It is a 4-dimensional space, so let us consider:

$$B' = \{x^3 + 4x^2 - 5x + 2,\ x^2 + 7,\ x - 3,\ 2\}.$$

There are 4 polynomials in B', and the degrees of these 4 polynomials are all ***distinct***, and thus from Section 3.2, we know that B' is ***linearly independent***, and therefore is a basis for $\mathbb{P}^3$ by the ***Two-For-One Theorem***. Finding the coordinates of a random polynomial with respect to this basis, though, is not as obvious as it was for our standard basis. For example, let us find:

$$\langle 3 - 5x + 7x^2 + x^3 \rangle_{B'}.$$

We have to solve for coefficients c_1, c_2, c_3 and c_4, such that:

$$3 - 5x + 7x^2 + x^3 = c_1(x^3 + 4x^2 - 5x + 2) + c_2(x^2 + 7) + c_3(x - 3) + c_4(2).$$

We can do this systematically by comparing coefficients, as we did in Section 3.2. Looking at the x^3 terms, we must have $c_1 = 1$. But the x^2 terms tell us that: $7 = 4 + c_2$, and thus $c_2 = 3$. Now, the x terms tell us that: $-5 = -5 + c_3$, so $c_3 = 0$. Finally, the constant terms tell us that: $3 = 2 + 3 \cdot 7 + 2c_4$, so $c_4 = -10$. Thus $\langle 3 - 5x + 7x^2 + x^3 \rangle_{B'} = \langle 1, 3, 0, -10 \rangle$. We can check that:

$$3 - 5x + 7x^2 + x^3 = 1 \cdot (x^3 + 4x^2 - 5x + 2) + 3 \cdot (x^2 + 7) + 0 \cdot (x - 3) - 10(2). \square$$

Coordinate Vectors for W = Span(B)

If $W = Span(B)$, where B is a ***linearly independent*** subset of some vector space V, then we know that B is automatically a **basis** for W. We can thus think of B as a **standard basis** for W.

Example: Consider $B = \{\sin(x), \cos(x)\}$ and $W = Span(B)$. Then:

$$\langle 4\cos(x) + 7\sin(x) \rangle_B = \langle 7, 4 \rangle.$$

The formula for $\sin(\alpha + \beta)$ tells us that $f(x) = \sin(x + \pi/6)$ is a member of W, because:

$$\begin{aligned} \sin(x + \pi/6) &= \sin(x)\cos(\pi/6) + \cos(x)\sin(\pi/6) \\ &= \frac{\sqrt{3}}{2}\sin(x) + \frac{1}{2}\cos(x), \end{aligned}$$

and thus $\langle \sin(x + \pi/6) \rangle_B = \left\langle \sqrt{3}/2, 1/2 \right\rangle$. $\square$

Constructing A Matrix For T

Now we come to the main goal of this section. We know that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ has an $m \times n$ ***standard matrix:***

$$[T] = [T(\vec{e}_1) \,|\, T(\vec{e}_2) \,|\, \cdots \,|\, T(\vec{e}_n)].$$

We compute $T(\langle x_1, x_2, \ldots, x_n \rangle)$ using a matrix product:

$$T(\langle x_1, x_2, \ldots, x_n \rangle) = x_1 T(\vec{e}_1) + x_2 T(\vec{e}_2) + \cdots + x_n T(\vec{e}_n) = [T(\vec{e}_1) \,|\, T(\vec{e}_2) \,|\, \cdots \,|\, T(\vec{e}_n)] \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Let us use this to motivate the construction of a matrix for a linear transformation of abstract vector spaces $T : V \to W$. We must assume that V and W are finite dimensional vector spaces, say with $dim(V) = n$ and $dim(W) = m$. Our first step is to choose a basis for V, say $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, and a basis for W, say $B^{/} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$.

We know that any vector $\vec{v} \in V$ can be written ***uniquely*** as a linear combination:

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n.$$

By the two ***linearity properties***, we have:

$$\begin{aligned} T(\vec{v}) &= T(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n) \\ &= c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n). \end{aligned}$$

Now, we saw that the process of finding the coordinates of a vector with respect to a basis is a ***linear transformation***. Since $T(\vec{v}) \in W$, we will now proceed to find the coordinates of $T(\vec{v})$ with respect to the basis $B^{/}$ of W:

$$\begin{aligned} \langle T(\vec{v})\rangle_{B^{/}} &= \langle c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_nT(\vec{v}_n)\rangle_{B^{/}} \\ &= c_1\langle T(\vec{v}_1)\rangle_{B^{/}} + c_2\langle T(\vec{v}_2)\rangle_{B^{/}} + \cdots + c_n\langle T(\vec{v}_n)\rangle_{B^{/}}. \end{aligned}$$

But each of the n coordinate vectors $\langle T(\vec{v}_i)\rangle_{B^{/}}$ in this sum is a vector in $\mathbb{R}^m$, and thus can be written as an $m \times 1$ coordinate matrix $[T(\vec{v}_i)]_{B^{/}}$. We can thus assemble these n coordinate matrices into a single $m \times n$ matrix which we will call $[T]_{B,B^{/}}$, and defined by:

$$[T]_{B,B^{/}} = [[T(\vec{v}_1)]_{B^{/}} \mid [T(\vec{v}_2)]_{B^{/}} \mid \cdots \mid [T(\vec{v}_n)]_{B^{/}}].$$

Again, notice that the dimension of $[T]_{B,B^{/}}$ is:

$$m \times n = dim\big(\textbf{codomain of } T\big) \times dim\big(\textbf{domain of } T\big).$$

If we change $\langle T(\vec{v})\rangle_{B^{/}}$ to the coordinate matrix $[T(\vec{v})]_{B^{/}}$, our last equation above says:

$$[T(\vec{v})]_{B^{/}} = [[T(\vec{v}_1)]_{B^{/}} \mid [T(\vec{v}_2)]_{B^{/}} \mid \cdots \mid [T(\vec{v}_n)]_{B^{/}}] \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = [T]_{B,B^{/}}[\vec{v}]_B.$$

Let us summarize this construction:

Definition/Theorem: Let $T : V \to W$ be a ***linear transformation***, where $dim(V) = n$ and $dim(W) = m$. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a ***basis*** for V, and let $B^{/} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be a ***basis*** for W. The $m \times n$ matrix $[T]_{B,B^{/}}$, given by:

$$[T]_{B,B^{/}} = [[T(\vec{v}_1)]_{B^{/}} \mid [T(\vec{v}_2)]_{B^{/}} \mid \cdots \mid [T(\vec{v}_n)]_{B^{/}}],$$

is called the ***matrix*** of T ***relative*** to B and $B^{/}$.

For any $\vec{v} \in V$, we can compute $T(\vec{v})$ via:

$$[T(\vec{v})]_{B^{/}} = [T]_{B,B^{/}}[\vec{v}]_B.$$

If $T : V \to V$ is an ***operator*** and we use the same basis B for the domain and codomain (that is, $B = B^{/}$), we write $[T]_B$ instead of $[T]_{B,B}$.

We can think of computing $T(\vec{v})$ via $[T(\vec{v})]_{B'} = [T]_{B,B'}[\vec{v}]_B$ as a three-step process:

1. ***ENCODE:*** Given $\vec{v} \in V$, find coefficients $c_1, c_2, \ldots, c_n$, such that:
$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n,$$
and assemble the coordinate matrix:
$$[\vec{v}]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} \in \mathbb{R}^{\boldsymbol{n}}.$$

2. ***MULTIPLY:*** Compute the product:
$$[T]_{B,B'}[\vec{v}]_B = [T]_{B,B'}\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = [T(\vec{v})]_{B'} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{bmatrix} \in \mathbb{R}^{\boldsymbol{m}}.$$

3. ***DECODE:*** Use the coefficients $d_1, d_2, \ldots, d_m$ of $[T(\vec{v})]_{B'}$ and the basis B' to explicitly find:
$$T(\vec{v}) = d_1\vec{w}_1 + d_2\vec{w}_2 + \cdots + d_m\vec{w}_m \in W.$$

Example: Let $T : \mathbb{P}^3 \to \mathbb{P}^2$ be the function given by:
$$T(p(x)) = 5p'(x) - p''(x)(2x+7) + p(-3) \cdot x^2.$$
Since each derivative drops the degree by 1, the first term will have degree at most 2, the second derivative has degree at most 1, so when we multiply it by $2x+7$, the product will have a degree of at most 2. The last term has degree either 0 or 2. Thus, T indeed sends a polynomial from $\mathbb{P}^3$ to $\mathbb{P}^2$. For example:
$$\begin{aligned} T(2+3x-5x^2+x^3) &= 5(3-10x+3x^2) - (-10+6x)(2x+7) + (2-9-45-27)x^2 \\ &= -76x^2 - 72x + 85 \in \mathbb{P}^2. \end{aligned}$$
Let us first prove that T is indeed a linear transformation:
$$\begin{aligned} &T(p(x)+q(x)) \\ &= 5(p+q)'(x) - (p+q)''(x)(2x+7) + (p+q)(-3) \cdot x^2 \\ &= 5[p'(x)+q'(x)] + (p''(x)+q''(x))(2x+7) + [p(-3)+q(-3)] \cdot x^2 \\ &= 5p'(x) + 5q'(x) + p''(x)(2x+7) + q''(x)(2x+7) + p(-3) \cdot x^2 + q(-3) \cdot x^2 \\ &= 5p'(x) + p''(x)(2x+7) + p(-3) \cdot x^2 + 5q'(x) + q''(x)(2x+7) + q(-3) \cdot x^2 \\ &= T(p(x)) + T(q(x)). \end{aligned}$$
Notice that the additivity property essentially follows from the ***distributive*** and ***commutative*** properties. Similarly, we can show that $T(c \cdot p(x)) = c \cdot T(p(x))$.

Now, let us use the standard bases:
$$B = \{1, x, x^2, x^3\} \quad \text{and} \quad B' = \{1, x, x^2\}$$
for $\mathbb{P}^3$ and $\mathbb{P}^2$, and find $[T]_{B,B'}$. We apply T to each vector of B, in the given order:

$$\begin{aligned} T(p(x)) &= 5p'(x) - p''(x)(2x+7) + p(-3)\cdot x^2, \text{ and so:} \\ T(1) &= 0 - 0 + 1\cdot x^2 = x^2; \\ T(x) &= 5\cdot 1 - 0 + (-3)\cdot x^2 = -3x^2 + 5; \\ T(x^2) &= 5\cdot 2x - 2(2x+7) + 9\cdot x^2 = 9x^2 + 6x - 14; \\ T(x^3) &= 5\cdot 3x^2 - 6x(2x+7) + (-27)\cdot x^2 = -24x^2 - 42x. \end{aligned}$$

Thus:

$$\begin{aligned} \langle T(1)\rangle_{B'} &= \langle 0, 0, 1\rangle, \\ \langle T(x)\rangle_{B'} &= \langle 5, 0, -3\rangle, \\ \langle T(x^2)\rangle_{B'} &= \langle -14, 6, 9\rangle, \text{ and} \\ \langle T(x^3)\rangle_{B'} &= \langle 0, -42, -24\rangle. \end{aligned}$$

We assemble these 4 coordinate vectors into the ***columns*** of a 3×4 matrix:

$$[T]_{B,B'} = \begin{bmatrix} 0 & 5 & -14 & 0 \\ 0 & 0 & 6 & -42 \\ 1 & -3 & 9 & -24 \end{bmatrix}.$$

Let us recompute $T(2+3x-5x^2+x^3)$ using $[T]_{B,B'}$:

1. ***ENCODE:*** Find $\langle 2+3x-5x^2+x^3\rangle_B$:
$$\langle 2+3x-5x^2+x^3\rangle_B = \langle 2, 3, -5, 1\rangle.$$
2. ***MULTIPLY:*** Compute the product $[T]_{B,B'}[\vec{v}]_B$:
$$[T(\vec{v})]_{B'} = [T]_{B,B'}[\vec{v}]_B = \begin{bmatrix} 0 & 5 & -14 & 0 \\ 0 & 0 & 6 & -42 \\ 1 & -3 & 9 & -24 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \\ -5 \\ 1 \end{bmatrix} = \begin{bmatrix} 85 \\ -72 \\ -76 \end{bmatrix}.$$
3. ***DECODE:*** Use the coefficients in $[T(\vec{v})]_{B'}$ and B' to find $T(\vec{v})$:
$$\begin{aligned} T(\vec{v}) &= 85\cdot 1 - 72\cdot x - 76\cdot x^2 \\ &= 85 - 72x - 76x^2, \end{aligned}$$

which confirms our previous computation.□

Example: Let us suppose that we are given a linear transformation $T : \mathbb{P}^2 \to \mathbb{P}^1$, with matrix:

$$[T]_{B,B'} = \begin{bmatrix} 5 & -3 & 2 \\ 4 & 1 & -7 \end{bmatrix},$$

where $B = \{x^2-2, x+3, 1\}$ and $B' = \{x+1, x-1\}$. The polynomials in B have distinct degrees, and there are three polynomials in B from $\mathbb{P}^2$, which just happens to be 3-dimensional, and so B is a basis for $\mathbb{P}^2$. On the other hand, B' contains two polynomials from $\mathbb{P}^1$ which are not parallel to each other, and so B' is likewise a basis for $\mathbb{P}^1$. Now, suppose we want to find $T(p(x))$, where:

$$p(x) = 7x^2 + 4x - 8.$$

We follow the three steps, as usual:

1. ***ENCODE:*** Find $\langle p(x)\rangle_B$. Since there is only one basis vector of degree 2 and one of degree 1, it is easy to figure out the required coefficients. With a little effort, we find:
$$7x^2 + 4x - 8 = 7(x^2 - 2) + 4(x + 3) - 6(1), \quad \text{and so}$$
$$\langle 7x^2 + 4x - 8\rangle_B = \langle 7, 4, -6\rangle.$$
2. ***MULTIPLY:*** Compute the product $[T]_{B,B'}[p(x)]_B$:
$$\begin{bmatrix} 5 & -3 & 2 \\ 4 & 1 & -7 \end{bmatrix}\begin{bmatrix} 7 \\ 4 \\ -6 \end{bmatrix} = \begin{bmatrix} 11 \\ 74 \end{bmatrix}.$$
3. ***DECODE:*** We find $T(p(x))$ using the coefficients 11 and 74 on the ordered basis B':
$$T(p(x)) = 11(x + 1) + 74(x - 1) = 85x - 63. \square$$

Unfortunately, as far as this Example is concerned, there is absolutely no way that we can ***check*** that this answer is correct, since we have no idea about the inner workings of T. The example also begs the question: why on earth would anyone use the two strange and inconvenient bases that we saw above? In a little bit, though, we will revisit how to create the projection matrix onto a plane Π from Chapter 2. Since we already have one way to do it, we can certainly verify our answer. We will be using a basis which on face value looks complicated, but is chosen because the ***action*** of the projection operator follows easily.

Function Spaces Preserved by the Derivative

We have seen that in general, the derivative transformation:
$$D : C^1(I) \to C(I),$$
has domain $C^1(I)$ and codomain $C(I)$, for some interval I. Both spaces are unfortunately infinite dimensional, and so it is impossible to create a matrix to represent this operation. However, we have seen ***finite dimensional*** subspaces $W \trianglelefteq C^1(I)$, $W = Span(B)$ for some finite set of differentiable functions B, which are preserved by D. In this case, D becomes an ***operator:***
$$D : W \to W,$$
and so we can assemble $[D]_B$.

Example: We saw in Section 3.5 that $W = Span(B)$, where $B = \{x^2e^{-3x}, xe^{-3x}, e^{-3x}\}$, is preserved by D. The derivatives of the three functions in B are:
$$D(x^2e^{-3x}) = -3x^2e^{-3x} + 2xe^{-3x},$$
$$D(xe^{-3x}) = -3xe^{-3x} + e^{-3x}, \text{ and}$$
$$D(e^{-3x}) = -3e^{-3x}.$$
Assembling the coefficients in the columns of a 3×3 matrix, in the order given by B (both ***left-to-right*** and ***top-to-bottom***), we get:

$$[D]_B = \begin{bmatrix} -3 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 1 & -3 \end{bmatrix}.$$

Thus, if we want to compute the derivative of $f(x) = 5x^2e^{-3x} + 9xe^{-3x} - 4e^{-3x}$, we perform the ***matrix product:***

$$\begin{bmatrix} -3 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 1 & -3 \end{bmatrix} \begin{bmatrix} 5 \\ 9 \\ -4 \end{bmatrix} = \begin{bmatrix} -15 \\ -17 \\ 21 \end{bmatrix}.$$

Decoding the result, the derivative is: $f'(x) = -15x^2e^{-3x} - 17xe^{-3x} + 21e^{-3x}$. □

Revisiting Projections and Reflections

Let us see how to apply the ideas in this Section to find another way to create the standard matrix of the projection and reflection operators that we saw in Section 2.2.

Example: Suppose that Π is the plane with equation: $5x + 2y - 6z = 0$.

We will find the standard matrix of $proj_\Pi$ by first finding the matrix of $proj_\Pi$ with respect to a basis B which is ***not*** the standard basis. Instead, we will pick any two vectors on Π as well as the normal vector $\vec{n}$ which we can easily extract from the equation above. To keep it simple, we will pick vectors from Π where one coordinate is zero. For example, we can choose $\langle 2, -5, 0 \rangle$ and $\langle 0, 3, 1 \rangle$. Together with the obvious normal vector for Π, we get the following basis B for $\mathbb{R}^3$:

$$B = \{ \langle 2, -5, 0 \rangle, \langle 0, 3, 1 \rangle, \langle 5, 2, -6 \rangle \}.$$

It is indeed easy to check that the first two vectors are on Π. Now, since $\vec{n} = \langle 5, 2, -6 \rangle$ is obviously not on Π, we know from Chapter 1 that these three vectors indeed form a basis for $\mathbb{R}^3$.

There is a good reason why we chose this basis. If a vector is already on Π, then its projection onto Π is ***itself***. If a vector is ***orthogonal*** to Π, then its projection is the ***zero vector***. Thus:

$$\begin{aligned} proj_\Pi(\langle 2, -5, 0 \rangle) &= \langle 2, -5, 0 \rangle, \\ proj_\Pi(\langle 0, 3, 1 \rangle) &= \langle 0, 3, 1 \rangle, \text{ and} \\ proj_\Pi(\langle 5, 2, -6 \rangle) &= \langle 0, 0, 0 \rangle. \end{aligned}$$

To form $[proj_\Pi]_B$, we need to find the coordinates of the three ***image*** vectors with respect to B. Since $\langle 2, -5, 0 \rangle$ and $\langle 0, 3, 1 \rangle$ are the first two members of B, we obtain:

$$\langle \langle 2, -5, 0 \rangle \rangle_B = \langle 1, 0, 0 \rangle, \ \langle \langle 0, 3, 1 \rangle \rangle_B = \langle 0, 1, 0 \rangle, \text{ and } \langle \langle 0, 0, 0 \rangle \rangle_B = \langle 0, 0, 0 \rangle.$$

Thus:

$$[proj_\Pi]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} = Diag(1, 1, 0).$$

This looks almost too easy, but unfortunately, this is not the ***standard*** matrix $[proj_{\Pi}]$. To find the standard matrix, we will need to do a bit more work. We need to find $proj_{\Pi}(\vec{e}_1)$, $proj_{\Pi}(\vec{e}_2)$, and $proj_{\Pi}(\vec{e}_3)$. To do this, we follow the same recipe: ***Encode***, ***Multiply*** and ***Decode***.

For the ***Encode*** step, we need to solve the three systems:

$$\left[\begin{array}{ccc|c} 2 & 0 & 5 & 1 \\ -5 & 3 & 2 & 0 \\ 0 & 1 & -6 & 0 \end{array}\right], \left[\begin{array}{ccc|c} 2 & 0 & 5 & 0 \\ -5 & 3 & 2 & 1 \\ 0 & 1 & -6 & 0 \end{array}\right] \text{ and } \left[\begin{array}{ccc|c} 2 & 0 & 5 & 0 \\ -5 & 3 & 2 & 0 \\ 0 & 1 & -6 & 1 \end{array}\right].$$

But since the coefficient matrices are all the same, we can solve all three ***simultaneously*** using:

$$\left[\begin{array}{ccc|ccc} 2 & 0 & 5 & 1 & 0 & 0 \\ -5 & 3 & 2 & 0 & 1 & 0 \\ 0 & 1 & -6 & 0 & 0 & 1 \end{array}\right].$$

This system should look familiar: it is exactly the same system we use to find the ***inverse*** of a matrix! Applying Gauss-Jordan to this augmented matrix, we find the rref:

$$\left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{20}{65} & -\frac{5}{65} & \frac{15}{65} \\ 0 & 1 & 0 & \frac{30}{65} & \frac{12}{65} & \frac{29}{65} \\ 0 & 0 & 1 & \frac{5}{65} & \frac{2}{65} & -\frac{6}{65} \end{array}\right].$$

Thus, we see $[\vec{e}_1]_B$, $[\vec{e}_2]_B$ and $[\vec{e}_3]_B$ in the three ***columns*** on the right.

Now, we are ready for the ***Multiply*** step. But again, instead of multiplying $[proj_{\Pi}]_B$ by $[\vec{e}_1]_B$, $[\vec{e}_2]_B$ and $[\vec{e}_3]_B$ separately, we can do it ***simultaneously*** with a ***single*** matrix product:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{20}{65} & -\frac{5}{65} & \frac{15}{65} \\ \frac{30}{65} & \frac{12}{65} & \frac{29}{65} \\ \frac{5}{65} & \frac{2}{65} & -\frac{6}{65} \end{bmatrix} = \begin{bmatrix} \frac{4}{13} & -\frac{1}{13} & \frac{3}{13} \\ \frac{6}{13} & \frac{12}{65} & \frac{29}{65} \\ 0 & 0 & 0 \end{bmatrix}.$$

Now, this matrix contains the coordinates of $proj_{\Pi}(\vec{e}_1)$, $proj_{\Pi}(\vec{e}_2)$ and $proj_{\Pi}(\vec{e}_3)$, all with respect to B, in the respective columns. In order to find the coordinates of these three vectors with respect to the ***standard basis***, we must use the 3 coefficients for each column in a linear combination with the 3 vectors of B. But these 3 vectors are found in the coefficient matrix:

$$\begin{bmatrix} 2 & 0 & 5 \\ -5 & 3 & 2 \\ 0 & 1 & -6 \end{bmatrix},$$

which we saw above. Thus, to ***Decode*** these three columns, we again form just ***one matrix product***, and we finally obtain the standard matrix:

$$[proj_\Pi] = \begin{bmatrix} 2 & 0 & 5 \\ -5 & 3 & 2 \\ 0 & 1 & -6 \end{bmatrix} \begin{bmatrix} \frac{4}{13} & -\frac{1}{13} & \frac{3}{13} \\ \frac{6}{13} & \frac{12}{65} & \frac{29}{65} \\ 0 & 0 & 0 \end{bmatrix} = \frac{1}{65} \begin{bmatrix} 40 & -10 & 30 \\ -10 & 61 & 12 \\ 30 & 12 & 29 \end{bmatrix}.$$

We can verify using the ready-made formula in Exercise 26 of Section 2.2 that the matrix we obtained above is correct. Before we leave this Example, let us summarize our computations above in just ***one equation:***

$$[proj_\Pi] = \begin{bmatrix} 2 & 0 & 5 \\ -5 & 3 & 2 \\ 0 & 1 & -6 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{20}{65} & -\frac{5}{65} & \frac{15}{65} \\ \frac{30}{65} & \frac{12}{65} & \frac{29}{65} \\ \frac{5}{65} & \frac{2}{65} & -\frac{6}{65} \end{bmatrix} = [B][proj_\Pi]_B[B]^{-1}.$$

Here, we used the notation $[B]$ to denote the matrix whose ***columns*** are the ***coordinates*** of the vectors in B (with respect to the standard basis for $\mathbb{R}^3$).

We can use these ideas to compute $[refl_\Pi]$ as well. The only difference is that:

$$[refl_\Pi] = Diag(1, 1, -1),$$

because $refl_\Pi(\vec{n}) = -\vec{n}$, while the reflection of any vector on Π is itself. □

We will see in Sections 6.4 and 6.5 that the formula $[proj_\Pi] = [B][proj_\Pi]_B [B]^{-1}$ can be generalized so that we can find the matrix of an operator on a finite-dimensional vector space V with respect to two different bases for V using an analogous product of three matrices.

3.6 Section Summary

Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be an ordered basis for an n-dimensional vector space V. If $\vec{v}$ is any vector in V, we know that $\vec{v}$ can be expressed ***uniquely*** as a linear combination of the vectors of B: $\vec{v} = c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_n\vec{w}_n$.

We call the vector $\langle c_1, c_2, \ldots, c_n\rangle$ the ***coordinate vector*** of $\vec{v}$ with respect to B, written symbolically as $\langle\vec{v}\rangle_B = \langle c_1, c_2, \ldots, c_n\rangle$.

The $n \times 1$ matrix corresponding to $\langle\vec{v}\rangle_B$ is called the ***coordinate matrix*** of $\vec{v}$ with respect to B, written symbolically as $[\vec{v}]_B$.

For any basis B of an n-dimensional vector space V, the function $T : V \to \mathbb{R}^n$ given by $T(\vec{v}) = \langle\vec{v}\rangle_B$, namely, finding the coordinates with respect to B, is a ***linear transformation***. In particular, if $V = \mathbb{R}^n$ and B is a basis for $\mathbb{R}^n$, then T is in fact one-to-one and onto, i.e., an ***isomorphism***.

Let $T : V \to W$ be a linear transformation, where $dim(V) = n$ and $dim(W) = m$. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for V, and let $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be a basis for W. The $m \times n$ matrix whose columns, from left to right, are $[T(\vec{v}_1)]_{B'}$ through $[T(\vec{v}_n)]_{B'}$ is called the ***matrix*** of T ***relative*** to B and B', and written symbolically as:

$$[T]_{B,B'} = [[T(\vec{v}_1)]_{B'} \,|\, [T(\vec{v}_2)]_{B'} \,|\, \cdots \,|\, [T(\vec{v}_n)]_{B'}].$$

Moreover, for any $\vec{v} \in V$: $[T(\vec{v})]_{B'} = [T]_{B,B'}[\vec{v}]_B$.

If $T : V \to V$ is a linear ***operator*** and we use the same basis B for the domain and codomain, we write $[T]_B$ instead of $[T]_{B,B}$.

We can compute $T(\vec{v})$ using $[T]_{B,B'}$ and a three-step process:

1. ***ENCODE:*** Given $\vec{v} \in V$, find $\langle \vec{v} \rangle_B \in \mathbb{R}^n$, and rewrite this into the $n \times 1$ column matrix $[\vec{v}]_B$.
2. ***MULTIPLY:*** Compute the product $[T]_{B,B'}[\vec{v}]_B = [T(\vec{v})]_{B'}$, an $m \times 1$ column matrix.
3. ***DECODE:*** Use the coefficients of $[T(\vec{v})]_{B'}$ and the basis B' to explicitly find $T(\vec{v}) \in W$.

If a finite-dimensional function space $W = Span(B)$, where B is linearly independent, is preserved by the ***derivative*** D, we can construct $[D]_B$ and compute the derivative of a function from W using a matrix product.

The ideas in this Section can be used to find $[proj_L]$, $[refl_L]$, $[proj_\Pi]$, and $[refl_\Pi]$, for any line L passing through the origin in $\mathbb{R}^2$ or $\mathbb{R}^3$, and any plane Π passing through the origin in $\mathbb{R}^3$.

3.6 Exercises

1. Let $B = \{\langle 1,-1, 1\rangle, \langle 1,-1,-1\rangle, \langle 0, 1, 1\rangle\}$.
   a. Find $\langle 3, 5,-8\rangle_B$ using the Gauss-Jordan algorithm.
   b. By looking at the rref of your augmented matrix in (a), explain why B is indeed a basis for $\mathbb{R}^3$.
   c. Find $\langle \vec{e}_2 \rangle_B$, where $\vec{e}_2 = \langle 0, 1, 0\rangle$.
2. Let $B = \left\{2x^3 - 5x^2 + 3x + 7, x^2 - 4x + 9, \frac{1}{2}x + 5, 3\right\}$.
   a. Explain, just by inspection, why B is a basis for $\mathbb{P}^3$.
   b. Find $\langle 3x^3 + 6x^2 - 8x + 2\rangle_B$ by the method of comparing coefficients.
3. Show that the following vectors are members of $W = Span(\{\sin(x), \cos(x)\})$, from the 4th Example of this Section, and find their coordinate vectors with respect to $B = \{\sin(x), \cos(x)\}$.
   a. $\cos(x + \pi/4)$
   b. $\sin(x + \sin^{-1}(3/5))$
   c. $\sin(x + \cos^{-1}(-12/13))$
   d. $\cos(x - \tan^{-1}(20/21))$
4. Let $\vec{a} = \langle -3, 1\rangle$ and $E_{\vec{a}} : \mathbb{P}^2 \to \mathbb{R}^2$, the linear transformation from Exercise 3 in Section 3.5. Let $B = \{1, x, x^2\}$ and $B' = \{\vec{e}_1, \vec{e}_2\}$.
   a. Find $[T]_{B,B'}$.
   b. Use $[T]_{B,B'}$ to recompute $T(-5x^2 + 8x + 3)$.
   c. Use $[T]_{B,B'}$ to compute $T(7x^2 - 5x + 4)$.
5. Let $\vec{a} = \langle -5, 3,-2\rangle$ and $E_{\vec{a}} : \mathbb{P}^3 \to \mathbb{R}^3$, the linear transformation from Exercise 4 in Section 3.5. Let $B = \{1, x, x^2, x^3\}$ and $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$.
   a. Find $[T]_{B,B'}$.
   b. Use $[T]_{B,B'}$ to recompute $T(7x^3 - 4x^2 + 3x - 6)$.
   c. Use $[T]_{B,B'}$ to compute $T(9x^3 - 7x^2 - 2x + 5)$.

6. Let $\vec{a} = \langle -5, 3, -2 \rangle$, and $E_{\vec{a}} : \mathbb{P}^2 \to \mathbb{R}^3$, the linear transformation from Exercise 5 in Section 3.5. Let $B = \{1, x, x^2\}$ and $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$.
   a. Find $[T]_{B,B'}$.
   b. Use $[T]_{B,B'}$ to recompute $T(4x^2 - 5x - 8)$.
   c. Use $[T]_{B,B'}$ to compute $T(3x^2 - 2x + 6)$.
7. Let $T : \mathbb{P}^2 \to \mathbb{R}^3$ be the linear transformation from Exercise 6 in Section 3.5 given by:
$$T(p(x)) = \langle p(1), p'(-2), 2p(3) - 5p'(1) \rangle.$$
   Let $B = \{1, x, x^2\}$ and $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$.
   a. Find $[T]_{B,B'}$.
   b. Use $[T]_{B,B'}$ to recompute $T(-5x^2 + 8x + 3)$.
   c. Use $[T]_{B,B'}$ to compute $T(7x^2 - 5x + 4)$.
8. Let $T : \mathbb{P}^2 \to \mathbb{R}^4$ be the linear transformation from Exercise 7 in Section 3.5:
$$T(p(x)) = \left\langle p(-2), p'(1), p''(x), \int_0^1 p(x)dx \right\rangle.$$
   Let $B = \{1, x, x^2\}$ and $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$.
   a. Find $[T]_{B,B'}$.
   b. Use $[T]_{B,B'}$ to recompute $T(-5x^2 + 8x + 3)$.
   c. Use $[T]_{B,B'}$ to compute $T(7x^2 - 5x + 4)$.
9. Let $T : \mathbb{P}^3 \to \mathbb{P}^1$ be the linear transformation from Exercise 8 in Section 3.5 given by:
$$T(p(x)) = p''(x).$$
   Let $B = \{1, x, x^2, x^3\}$ and $B' = \{1, x\}$.
   a. Find $[T]_{B,B'}$.
   b. Use $[T]_{B,B'}$ to recompute $T(2x^3 + 5x^2 - 4x + 3)$.
   c. Use $[T]_{B,B'}$ to compute $T(7x^3 - 8x^2 + 3x - 6)$.
10. Let $Ind : \mathbb{P}^2 \to \mathbb{P}^3$ be the linear transformation from Exercise 10 in Section 3.5 given by:
$$Ind(p(x)) = \int_0^x p(t)dt.$$
   Let $B = \{1, x, x^2\}$ and $B' = \{1, x, x^2, x^3\}$.
   a. Find $[Ind]_{B,B'}$.
   b. Use $[Ind]_{B,B'}$ to recompute $Ind(3x^2 + 2x - 7)$.
   c. Use $[Ind]_{B,B'}$ to compute $Ind(7x^2 - 5x + 4)$.

For Exercises (11) to (18): The following function spaces were seen in Exercises 11 to 18 of Section 3.5. (a) Use the basis B, in the order given, to find $[D]_B$, and (b) find the derivative of the function $f(x)$ by using a matrix product. Reminder: Encode, Multiply, Decode.

11. $W = Span(B)$, where $B = \{e^{-x}, e^{2x}\}$; $f(x) = 5e^{-x} - 3e^{2x}$.
12. $W = Span(B)$, where $B = \{e^x \sin(x), e^x \cos(x)\}$; $f(x) = 4e^x \sin(x) - 3e^x \cos(x)$.
13. $W = Span(B)$, where $B = (\{e^{-3x} \sin(2x), e^{-3x} \cos(2x)\})$; $f(x) = 5e^{-3x} \sin(2x) - 9e^{-3x} \cos(2x)$.

14. $W = Span(B)$, where $B = (\{xe^{5x}, e^{5x}\})$; $f(x) = -2xe^{5x} + 7e^{5x}$.

15. $W = Span(B)$, where $B = \{x^2e^{-4x}, xe^{-4x}, e^{-4x}\}$; $f(x) = -5x^2e^{-4x} + 2xe^{-4x} - 7e^{-4x}$.

16. $W = Span(B)$, where $B = (\{x^2 \cdot 5^x, x \cdot 5^x, 5^x\})$; $f(x) = -4x^2 \cdot 5^x + 9x \cdot 5^x - 2(5^x)$.

17. $W = \mathbb{P}^3 = Span(B)$, where $B = \{1, x, x^2, x^3\}$; $f(x) = 2x^3 - 8x^2 + 3x + 7$.

18. $W = Span(B)$, where $B = (\{x\sin(2x), x\cos(2x), \sin(2x), \cos(2x)\})$;
$f(x) = 4x\sin(2x) + 9x\cos(2x) - 5\sin(2x) + 8\cos(2x)$.

19. Let $B = \{\sin(mx), \cos(mx)\}$, where $m \in \mathbb{R}$, and let $W = Span(B)$. Consider the differentiation operator: $D : W \to W$.
    a. Explain in general why $D(f(x)) \in W$ for any $f(x) \in W$.
    b. Find $[D]_B$.

20. Let $B = \{e^{k_1x}, e^{k_2x}, \ldots, e^{k_nx}\}$, with $k_1, k_2, \ldots, k_n \in \mathbb{R}$ ***distinct*** real numbers, and let $W = Span(B)$. Consider the differentiation operator: $D : W \to W$.
    a. Explain in general why $D(f(x)) \in W$ for any $f(x) \in W$.
    b. Find $[D]_B$.
    c. What kind of matrix is $[D]_B$?

21. Let $B = \{e^{ax}\sin(bx), e^{ax}\cos(bx)\}$, where $a, b \in \mathbb{R}$, and let $W = Span(B)$. We know from Exercise 21 in Section 3.5 that D preserves W. Find $[D]_B$.

22. Let $B = \{x^2e^{kx}, xe^{kx}, e^{kx}\}$, where $k \in \mathbb{R}$, and $W = Span(B)$. Consider the differentiation operator: $D : W \to W$.
    a. Explain in general why $D(f(x)) \in W$ for any $f(x) \in W$.
    b. Find $[D]_B$.
    c. Find $D(x^ne^{kx})$ for any positive integer n and any $k \in \mathbb{R}$.
    d. Use induction to show that $W = Span(\{x^ne^{kx}, x^{n-1}e^{kx}, \ldots, x^2e^{kx}, xe^{kx}, e^{kx}\})$ is the smallest subspace that contains x^ne^{kx} and is ***preserved*** under D.

23. Let $B = \{\sin(x), \cos(x)\}$ and $W = Span(B)$. Let $T : W \to W$ be the linear operator from Exercise 19 in Section 3.5:

$$T(f(x)) = f''(x) - 3f'(x) + 2f(x).$$

    a. Find $[T]_B$.
    b. Use $[T]_B$ to recompute $T(3\sin(x) + 8\cos(x))$.
    c. Use Exercise 3(b) to compute $T(\sin(x + \sin^{-1}(3/5)))$.

24. Let $S = \{e^{4x}\sin(3x), e^{4x}\cos(3x)\}$ and $U = Span(S)$. Let $T : U \to U$ be the linear operator from Exercise 20 in Section 3.5 given by:

$$T(f(x)) = f''(x) - 3f'(x) + 2f(x).$$

    a. Find $[T]_S$.
    b. Use $[T]_S$ to recompute $T(5e^{4x}\sin(3x) - 9e^{4x}\cos(3x))$.
    c. Use $[T]_S$ to compute $T(-3e^{4x}\sin(3x) + 7e^{4x}\cos(3x))$.

25. Let $T : \mathbb{P}^3 \to \mathbb{P}^2$ be given by:

$$T(p(x)) = p'(x) + (x+1)p''(x) + 2p(-1).$$

a. Directly compute $T(5x^3 - 6x^2 + 4x + 9)$.
b. Explain why, in general, if $p(x) \in \mathbb{P}^3$, then $T(p(x)) \in \mathbb{P}^2$.
c. Show that T is indeed a linear transformation.
d. Find the matrix $[T]_{B,B'}$ with respect to the standard bases $B = \{1, x, x^2, x^3\}$ for $\mathbb{P}^3$ and $B' = \{1, x, x^2\}$ for $\mathbb{P}^2$.
e. Use $[T]_{B,B'}$ to recompute $T(5x^3 - 6x^2 + 4x + 9)$.

26. Let $T : \mathbb{P}^2 \to \mathbb{P}^3$ be given by:

$$T(p(x)) = p(x) \cdot (2x - 5) + p'(x) \cdot (x^2 + 3) - p(-2) \cdot x^3.$$

a. Directly compute $T(6x^2 - 2x + 7)$.
b. Explain why, in general, if $p(x) \in \mathbb{P}^2$, then $T(p(x)) \in \mathbb{P}^3$.
c. Show that T is indeed a linear transformation.
d. Find the matrix of T with respect to the standard bases $B = \{1, x, x^2\}$ for $\mathbb{P}^2$ and $B' = \{1, x, x^2, x^3\}$ for $\mathbb{P}^3$.
e. Use $[T]_{B,B'}$ to recompute $T(6x^2 - 2x + 7)$.

27. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^1$ is a linear transformation whose matrix with respect to the bases $B = \{1, 5 - x, 2 + 3x - x^2\}$ for $\mathbb{P}^2$ and $B' = \{x + 3, 2\}$ for $\mathbb{P}^1$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 4 & -1 & 5 \\ -7 & 0 & -2 \end{bmatrix}.$$

a. Find $\langle 6x^2 - 3x + 8 \rangle_B$.
b. Use (a) to compute $T(6x^2 - 3x + 8)$. Don't forget to perform all three steps.
c. Use the idea of (a) and (b) to compute $T(1)$, $T(x)$ and $T(x^2)$. Computational Hint: You can find the three coordinate vectors at the same time by solving a 3×6 augmented matrix.
d. Use (c) to construct $[T]_{S,S'}$, where $S = \{1, x, x^2\}$ and $S' = \{1, x\}$ are the standard bases for $\mathbb{P}^2$ and $\mathbb{P}^1$, respectively.
e. Use $[T]_{S,S'}$ to recompute $T(6x^2 - 3x + 8)$.

28. Suppose that $T : \mathbb{P}^1 \to \mathbb{P}^2$ is a linear transformation whose matrix with respect to the bases $B = \{1, 2 + x\}$ for $\mathbb{P}^1$ and $B' = \{x^2 - x, x + 1, -1\}$ for $\mathbb{P}^2$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 5 & 3 \\ -1 & 2 \\ 8 & -7 \end{bmatrix}.$$

a. Find $\langle 3x - 5 \rangle_B$.
b. Use (a) to compute $T(3x - 5)$. Don't forget to perform all three steps.
c. Use the idea of (a) and (b) to compute $T(1)$ and $T(x)$. See the Hint in Exercise 27 (c).
d. Use (c) to construct $[T]_{S,S'}$, where $S = \{1, x\}$ and $S' = \{1, x, x^2\}$ are the standard bases for $\mathbb{P}^1$ and $\mathbb{P}^2$, respectively.
e. Use $[T]_{S,S'}$ to recompute $T(3x - 5)$.

29. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^2$ is an operator whose matrix with respect to the basis $B = \{2, 5 - x, 2 + 3x - x^2\}$ is given by:

$$[T]_B = \begin{bmatrix} -2 & 1 & 2 \\ 3 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}.$$

a. Find $\langle 3x^2 + 5x - 7 \rangle_B$.

b. Use (a) to compute $T(3x^2 + 5x - 7)$. Don't forget to perform all three steps.

c. Use the idea of (a) and (b) to compute $T(1)$, $T(x)$ and $T(x^2)$.
See the Hint in Exercise 27 (c).

d. Use (c) to construct $[T]_{B'}$, the matrix of T with respect to the standard basis $B' = \{1, x, x^2\}$.

e. Use $[T]_{B'}$ to recompute $T(3x^2 + 5x - 7)$.

30. If V is a vector space and B is a fixed basis for V, prove that for all vectors $\vec{u} \in V$:

$$\langle c \cdot \vec{u} \rangle_B = c \cdot \langle \vec{u} \rangle_B.$$

This completes the proof that the operation of finding coordinates is a linear transformation. Hint: what is the ***meaning*** of each side of this equation, starting with $\langle \vec{u} \rangle_B$?

31. Consider the plane through the origin, Π, with equation:

$$3x + 7y - 8z = 0.$$

Review the last Example in this Section.

a. Show that $\vec{v}_1 = \langle 7, -3, 0 \rangle$ and $\vec{v}_2 = \langle 8, 0, 3 \rangle$ are two linearly independent vectors on the plane Π.

b. Let $\vec{n} = \langle 3, 7, -8 \rangle$ be the obvious normal for the plane. Show that:

$$B = \{\vec{v}_1, \vec{v}_2, \vec{n}\}$$

is a basis for $\mathbb{R}^3$.

c. Use the geometric description of $proj_\Pi$ to ***explain*** why $[proj_\Pi]_B = Diag(1, 1, 0)$.

d. Use a single matrix (of dimension 3×6) to find $\langle \vec{e}_1 \rangle_B$, $\langle \vec{e}_2 \rangle_B$, and $\langle \vec{e}_3 \rangle_B$.

e. Compute $proj_\Pi(\vec{e}_1)$, $proj_\Pi(\vec{e}_2)$, and $proj_\Pi(\vec{e}_3)$ using (c) and (d).

f. Find the standard matrix $[proj_\Pi]$, using (e).

g. Similarly, show that $[refl_\Pi]_B = Diag(1, 1, -1)$.

h. Find the standard matrix $[refl_\Pi]$. You may use your answers from (d).

i. Let L be the line $Span(\{\vec{n}\})$. Find the standard matrix $[proj_L]$.

For Exercises (32) to (34): Repeat Exercise 31 with the indicated plane. Find vectors $\vec{v}_1$ and $\vec{v}_2$ on Π where either $x = 0$ or $y = 0$ or $z = 0$ in part (a) to slightly simplify the computations.

32. $\Pi : 5x - 3y + 7z = 0$.

33. $\Pi : 2x - y + 5z = 0$.

34. $\Pi : x = \frac{2}{3}z$. Think very carefully about part (a).

35. ***Alternative Formulas for Projection and Reflection Matrices:*** Let us generalize the four previous Exercises. Suppose that $\Pi : ax + by + cz = 0$ is a plane in $\mathbb{R}^3$ passing through the origin. For simplicity, let us assume that ***none*** of the coefficients is zero. Consider the matrix:

$$C = \begin{bmatrix} -b & -c & a \\ a & 0 & b \\ 0 & a & c \end{bmatrix}.$$

a. Explain the relevance of the three columns of C.

b. Explain why C is invertible.

c. Show that:

$$[proj_\Pi] = C \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} C^{-1} \quad \text{and} \quad [refl_\Pi] = C \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} C^{-1}.$$

d. How would you modify the matrix C if Π had equation: $ax + cz = 0$, where $a \neq 0$ and $c \neq 0$?

36. ***The Minimizing Theorem:*** Now that we have coordinate vectors, we can state a general version of The Minimizing Theorem: Suppose that V is a finite dimensional vector space, with basis $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. Suppose that $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is any finite subset of vectors from V. Let $[\vec{w}_1]_B$, $[\vec{w}_2]_B$, $\ldots$, $[\vec{w}_k]_B$ be the respective coordinate matrices. Let us assemble them in the columns of:

$$A = [\,[\vec{w}_1]_B \mid [\vec{w}_2]_B \mid \cdots \mid [\vec{w}_k]_B\,], \ \text{ with rref } R.$$

Suppose that $i_1, i_2, \ldots, i_m$ are the columns of R that contain the ***leading variables***. Prove that the set $S' = \{\vec{w}_{i_1}, \vec{w}_{i_2}, \ldots, \vec{w}_{i_m}\}$, that is, the subset of vectors of S corresponding to the leading columns of A, is a ***linearly independent*** set, and:

$$Span(S) = Span(S').$$

Furthermore, every $\vec{v}_i \in S - S'$, that is, the vectors of S corresponding to the ***free variables*** of R, can be expressed as linear combinations of the vectors of S', using the ***coefficients*** found in the corresponding column of R.

For Exercises (37) to (40): Use the Minimizing Theorem above to find a subset S' of $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ which is linearly independent such that $Span(S) = Span(S')$, and for every $\vec{v}_i \in S - S'$, find a linear combination of the vectors from S' that will add up to $\vec{v}_i$. Use a convenient basis B for the ambient space (if this is given as $Span(B)$, use B itself as a basis). You may use the symbols $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k$ in your answers.

37. $S = \{5 - 4x + 3x^2, 6 - 7x + 2x^2, 2 + 5x + 6x^2, 1 + 2x + 3x^2\} \subset \mathbb{P}^2$

38. $S = \{3 + 5x + x^2 + 4x^3, 4 + 7x + 2x^2 + 3x^3, x + 2x^2 - 7x^3,$
$3 + 4x - x^2 + 2x^3, 7 + 3x - 15x^2 + 7x^3\} \subset \mathbb{P}^3$

39. $S = \{3 - 4x - 3x^2 + x^3 - 5x^4, -1 + 3x + 2x^2 + 4x^3 + 3x^4, 3 + 11x + 6x^2 + 40x^3 + 7x^4,$
$7 + 4x + x^2 + 37x^3 - x^4, -2 + 3x + 2x^2 + x^3 + 3x^2, 1 + 3x + 3x^2 + 6x^3 + 6x^6\} \subset \mathbb{P}^4$

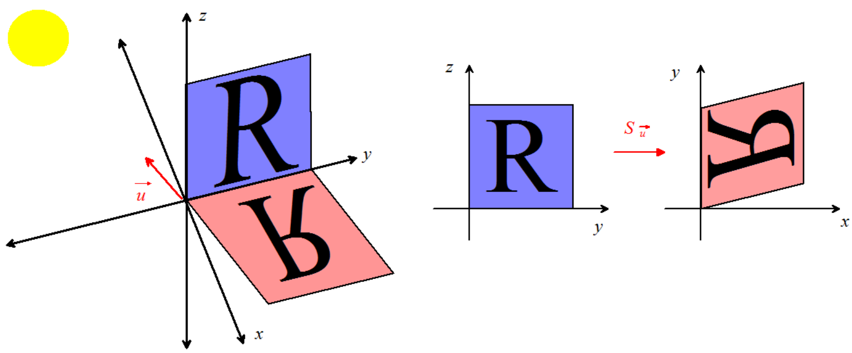
40. $S = \{\, 5e^x + 4e^{3x} + 3e^{4x}, 6e^x + 7e^{3x} + 2e^{4x}, 2e^x - 5e^{3x} + 6e^{4x},$
$e^x - 2e^{3x} + 3e^{4x}, 8e^x + 11e^{3x}\} \subset Span(\{e^x, e^{3x}, e^{4x}\})$

41. Suppose that V and W are vectors spaces with $dim(V) = n$ and $dim(W) = m$. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for V, and let $B^{/} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be a basis for W. Suppose that T_1 and T_2 are both linear transformations with domain V and codomain W. Prove that:

a. $[T_1 + T_2]_{B,B^{/}} = [T_1]_{B,B^{/}} + [T_2]_{B,B^{/}}$.

b. $[k \bullet T_1]_{B,B^{/}} = k \bullet [T_1]_{B,B^{/}}$

42. ***Casting Shadows:*** Let us imagine that the yz-plane is a wall, and a window-pane is formed by the unit vectors $\vec{j}$ and $\vec{k}$. Imagine also that the sun is located infinitely far away, in the direction of $\vec{u} = \langle a, b, c\rangle$. For now, let us assume that $c > 0$ (i.e. the sun is above the horizon), and that light is coming from the sun in parallel rays, in the direction of $-\vec{u}$.



If $\vec{v}$ is an arbitrary vector on the yz-plane, let $S_{\vec{u}}(\vec{v})$ be its shadow on the xy-plane. Thus, we have a function: $S_{\vec{u}} : \mathbb{R}^2 \to \mathbb{R}^2$, where the domain $\mathbb{R}^2$ refers to the yz-plane, with basis $B = \left\{\vec{j}, \vec{k}\right\}$, and the codomain $\mathbb{R}^2$ refers to the xy-plane, with basis $B^{/} = \left\{\vec{i}, \vec{j}\right\}$. Above, on the right, we show the image under $S_{\vec{u}}$ of the unit square, with the letter R inside.

a. Explain why our assumptions imply that the shadow of a triangle is again a triangle, and the shadow of two parallel vectors are again parallel. Thus, $S_{\vec{u}}$ is additive and homogeneous.

b. Find $S_{\vec{u}}\left(\vec{j}\right)$ and $S_{\vec{u}}\left(\vec{k}\right)$.

c. Use (b) to assemble $[S_{\vec{u}}]_{B,B^{/}}$.

d. Find $[S_{\vec{u}}]_{B,B^{/}}$ if $\vec{u} = \langle 3, -2, 5\rangle$, and sketch the effect of $S_{\vec{u}}$ on B.

e. Show that $[S_{\vec{u}}]_{B,B^{/}}$ is undefined if $c = 0$, but it is still defined if $c < 0$.

What would be the physical interpretation of $[S_{\vec{u}}]_{B,B^{/}}$ if $c < 0$?

Demonstrate your answer with $\vec{u} = \langle 3, -2, -5\rangle$.

3.7 One-to-One and Onto Linear Transformations; Compositions of Linear Transformations

Now that we have some understanding of how to compute the action of a linear transformation by constructing a matrix for it, we will go into a deeper exploration of the properties of linear transformations. In particular, we will see how to generalize the idea of a linear transformation being one-to-one or onto, how to use the rref of a matrix in order to find a basis for the kernel and the range and to test for the one-to-one and onto properties, how to compose two transformations, and find a matrix for this composition.

One-to-One Transformations and Onto Transformations

We can use exactly the same definition for one-to-one functions that we saw with Euclidean spaces:

> ***Definition:*** We say that a linear transformation $T : V \to W$ is ***one-to-one*** or ***injective*** if the image of different vectors from the domain are different vectors from the codomain:
> $$\text{if } \vec{v}_1 \neq \vec{v}_2 \text{ then } T(\vec{v}_1) \neq T(\vec{v}_2).$$
> We again say that T is an ***injection*** or an ***embedding***.

As before, we can rephrase this definition in terms of its ***contrapositive:***

> ***Theorem:*** A linear transformation $T : V \to W$ is ***one-to-one*** ***if and only if*** the only way two vectors from the domain have the same image in the codomain is for them to be the same vector to begin with:
> $$\text{if } T(\vec{v}_1) = T(\vec{v}_2) \text{ then } \vec{v}_1 = \vec{v}_2.$$
> In other words, the only solution to $T(\vec{v}_1) = T(\vec{v}_2)$ is $\vec{v}_1 = \vec{v}_2$.

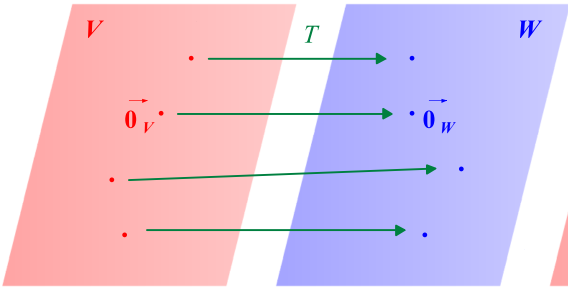
Finally, this condition is once again intimately related to the kernel of T :

> ***Theorem:*** A linear transformation $T : V \to W$ is ***one-to-one*** ***if and only if*** $ker(T) = \left\{ \vec{\mathbf{0}}_V \right\}$.

The proof, of course, is exactly the same as in Chapter 2, thanks to the linearity properties of T. We would also like to point out that all the statements above are true even if V or W is ***infinite dimensional*** (you will notice in the proof in Chapter 2 that there is no mention whatsoever of a ***matrix*** for T). As with one-to-one transformations, we can define onto transformations in exactly the same way as with Euclidean spaces:

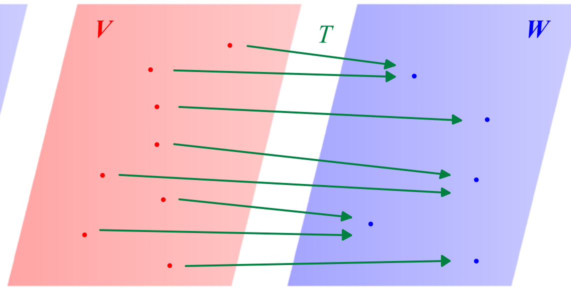
> ***Definition/Theorem:*** We say that a linear transformation $T : V \to W$ is ***onto*** or ***surjective*** if the range of T is ***all*** of W:
> $$range(T) = W.$$
> Since $rank(T) = dim(range(T))$, we can also say that T is ***onto*** ***if and only if*** $rank(T) = dim(W)$, in the case when W is ***finite dimensional***.

We again say that T is a ***surjection*** or a ***covering***. We can visualize these two concepts using essentially the same diagrams from Chapter 2:



T is ***one-to-one*** ***if and only if***

$$ker(T) = \left\{\vec{\mathbf{0}}_V\right\}.$$



T is ***onto*** ***if and only if***

$$range(T) = W.$$

Finding the Kernel and Range Using $[T]_{B,B'}$

Our next task is to determine if a given linear transformation is one-to-one, onto, neither or both. For this, we will need to study its kernel and range. Recall that we can find a basis for the kernel and range of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ by examining the rref of its standard matrix $[T]$. The nullspace of $[T]$ is the same as $ker(T)$, and the original columns of $[T]$ corresponding to the leading 1's in the rref form a basis for $range(T)$, since this subspace is the same as $colspace([T])$. We can apply this idea to $[T]_{B,B'}$ when we are dealing with a linear transformation between abstract vector spaces. We will leave the proof of the following as an Exercise:

Theorem: Suppose that $T : V \to W$ is a linear transformation, with $dim(V) = n$ and $dim(W) = m$, both finite-dimensional vector spaces. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for V, and let $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be a basis for W. Let us construct the $m \times n$ matrix $[T]_{B,B'}$ as we did in the previous Section, and let R be the rref of $[T]_{B,B'}$. Suppose that:

$$\{\vec{z}_1, \vec{z}_2, \ldots, \vec{z}_k\} \subset \mathbb{R}^n$$

is the basis that we obtain for $nullspace\left([T]_{B,B'}\right)$ using R, as we did in Chapter 2.
By the Uniqueness of Representation Property, we know that there exists $\vec{u}_i \in V$ so that $\langle \vec{u}_i \rangle_B = \vec{z}_i$ for every $i = 1 \ldots k$.
We conclude that the set $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\} \subset V$ is a ***basis*** for $ker(T)$.
As usual, if there are no free variables in R, then $nullspace\left([T]_{B,B'}\right) = \left\{\vec{\mathbf{0}}_n\right\}$, and consequently $ker(T) = \left\{\vec{\mathbf{0}}_V\right\}$. Similarly, the set of original columns:

$$\{\vec{c}_{i_1}, \vec{c}_{i_2}, \ldots, \vec{c}_{i_r}\} \subset \mathbb{R}^m$$

from $[T]_{B,B'}$ corresponding to the leading 1's of R form a basis for $columnspace\left([T]_{B,B'}\right)$ as we found in Chapter 2, and there exists $\vec{d}_j \in W$ so that $\left\langle \vec{d}_j \right\rangle_{B'} = \vec{c}_{i_j}$ for every $j = 1 \ldots r$.
We conclude that the set $\left\{\vec{d}_1, \vec{d}_2, \ldots, \vec{d}_r\right\} \subset W$ is a ***basis*** for $range(T)$.
If T is the zero transformation, then $range(T) = \left\{\vec{\mathbf{0}}_W\right\}$.

In other words, the information provided by $[T]_{B,B'}$ and R simply needs to be ***decoded*** with respect to the corresponding basis: we use B to find a basis for $ker(T)$ from a basis for $nullspace\left([T]_{B,B'}\right)$, and we use B' to find a basis for $range(T)$ from a basis for $columnspace\left([T]_{B,B'}\right)$.

Example: Let $T : \mathbb{P}^3 \to \mathbb{P}^2$ be given by:

$$T(p(x)) = p'(x) + 3x \bullet p''(x) - 2p(-1).$$

We leave it the reader to verify that this is indeed a linear transformation. Let us choose the standard bases $B = \{1, x, x^2, x^3\}$ for $\mathbb{P}^3$ and $B' = \{1, x, x^2\}$ for $\mathbb{P}^2$. We compute T on the basis vectors:

$$\begin{aligned} T(1) &= 0 + 0 - 2 \bullet 1 = -2, \\ T(x) &= 1 + 3x \bullet 0 - 2(-1) = 3, \\ T(x^2) &= 2x + 3x \bullet 2 - 2 \bullet 1 = 8x - 2, \text{ and} \\ T(x^3) &= 3x^2 + 3x \bullet 6x - 2(-1) = 21x^2 + 2. \end{aligned}$$

Now, we encode each as a column for $[T]_{B,B'}$:

$$[T]_{B,B'} = \begin{bmatrix} -2 & 3 & -2 & 2 \\ 0 & 0 & 8 & 0 \\ 0 & 0 & 0 & 21 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & -3/2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus, we have one free variable, and the ***coordinates*** with respect to B of the single member of the basis for our kernel are $\langle 3/2, 1, 0, 0\rangle$. Clearing fractions, we can use $\langle 3, 2, 0, 0\rangle$.

Decoding these coordinates with respect to B, the actual polynomial is:

$$p(x) = 3 \bullet 1 + 2 \bullet x + 0 \bullet x^2 + 0 \bullet x^3 = 3 + 2x.$$

We can check that:

$$T(p(x)) = 2 + 3x \bullet 0 - 2(3 - 2) = 0,$$

so $p(x)$ is indeed in the kernel. Thus:

$$ker(T) = Span(\{3 + 2x\}).$$

Since $ker(T) \neq \{z(x)\}$, T is ***not*** one-to-one.

Similarly, we have leading 1's in the 1st, 3rd and 4th columns, so the ***coordinates*** with respect to B' of the members of the basis for our range are found in the ***original*** 1st, 3rd and 4th columns of $[T]_{B,B'}$:

$$\langle -2, 0, 0\rangle, \langle -2, 8, 0\rangle \text{ and } \langle 2, 0, 21\rangle.$$

Decoding these coordinates with respect to B', the actual members of our basis are:

$$\begin{aligned} &\{-2 + 0 \bullet x + 0 \bullet x^2, -2 + 8 \bullet x + 0 \bullet x^2, 2 + 0 \bullet x + 21x^2\} \\ &= \{-2, -2 + 8x, 2 + 21x^2\}. \end{aligned}$$

But notice that $dim(\mathbb{P}^2) = 3$ and our basis above has 3 members, so actually:

$$range(T) = \mathbb{P}^2 = Span(\{-2, -2 + 8x, 2 + 21x^2\}) = Span(\{1, x, x^2\})$$

and T is ***onto***.$\square$

The Dimension Theorem for Abstract Vector Spaces

We will now fully generalize the Dimension Theorem for a linear transformation involving abstract vector spaces:

> ***Theorem — The Dimension Theorem:***
> Let $T : V \to W$ be a linear transformation, and suppose that V is ***finite dimensional*** with $dim(V) = n$. Then, both $ker(T)$ and $range(T)$ are finite dimensional, and we can define:
> $$rank(T) = dim(range(T)), \text{ and}$$
> $$nullity(T) = dim(ker(T)).$$
> Furthermore, these quantities are related by the equation:
> $$rank(T) + nullity(T) = n = dim(V) = dim\left(domain \text{ of } T\right).$$

Proof: Since V is finite dimensional, $ker(T) \trianglelefteq V$ is automatically finite dimensional.

Now, suppose that $nullity(T) = k$, with $k \leq n$. If $k > 0$, that is, T is not one-to-one, suppose that $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a basis for $ker(T)$. Note that if T is one-to-one, then $k = 0$, and there is no basis for $ker(T)$, so we may just assume that S is the empty set.

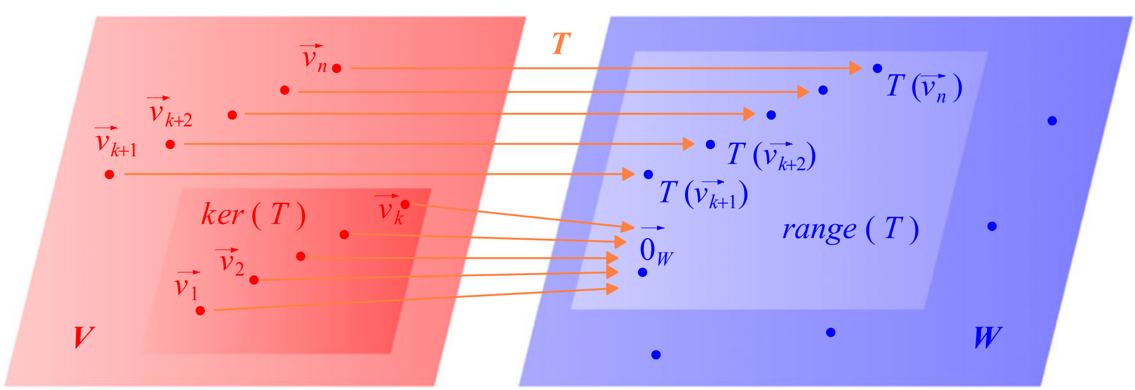
By the ***Extension Theorem***, we can enlarge S, one vector at a time, to a basis $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k, \vec{v}_{k+1}, \vec{v}_{k+2}, \ldots, \vec{v}_n\}$ for V. Thus, any vector $\vec{v} \in V$ can be expressed uniquely as a linear combination:

$$\vec{v} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k + c_{k+1}\vec{v}_{k+1} + c_{k+2}\vec{v}_{k+2} + \cdots + c_n\vec{v}_n.$$

From this, we get:

$$\begin{aligned} T(\vec{v}) &= T(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k + c_{k+1}\vec{v}_{k+1} + c_{k+2}\vec{v}_{k+2} + \cdots + c_n\vec{v}_n) \\ &= T(c_1\vec{v}_1) + T(c_2\vec{v}_2) + \cdots + T(c_k\vec{v}_k) + T(c_{k+1}\vec{v}_{k+1}) + T(c_{k+2}\vec{v}_{k+2}) + \cdots + T(c_n\vec{v}_n) \\ &= c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_kT(\vec{v}_k) + c_{k+1}T(\vec{v}_{k+1}) + c_{k+2}T(\vec{v}_{k+2}) + \cdots + c_nT(\vec{v}_n) \\ &= c_{k+1}T(\vec{v}_{k+1}) + c_{k+2}T(\vec{v}_{k+2}) + \cdots + c_nT(\vec{v}_n), \end{aligned}$$

since $T(\vec{v}_1) = \vec{\mathbf{0}}_W$, $T(\vec{v}_2) = \vec{\mathbf{0}}_W, \ldots,$ $T(\vec{v}_k) = \vec{\mathbf{0}}_W$. This tells us that $range(T)$ is ***Spanned*** by the set $\{T(\vec{v}_{k+1}), T(\vec{v}_{k+2}), \ldots, T(\vec{v}_n)\}$, which we illustrate in the following diagram:



The Dimension Theorem

Since there are $n-k$ vectors in this set, we will complete the Proof by showing that this set is also ***linearly independent***, and thus $rank(T) = n-k$. We construct the ***dependence test equation:***

$$d_{k+1}T(\vec{v}_{k+1}) + d_{k+2}T(\vec{v}_{k+2}) + \cdots + d_nT(\vec{v}_n) = \vec{\mathbf{0}}_W.$$

Reversing the steps above, we get: $T(d_{k+1}\vec{v}_{k+1} + d_{k+2}\vec{v}_{k+2} + \cdots + d_n\vec{v}_n) = \vec{\mathbf{0}}_W$.

This shows that the vector $\vec{v} = d_{k+1}\vec{v}_{k+1} + d_{k+2}\vec{v}_{k+2} + \cdots + d_n\vec{v}_n$ is a member of $ker(T)$. Recall, though, that $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a basis for $ker(T)$. Thus, we can find coefficients $d_1, d_2, \ldots, d_k$ such that:

$$d_{k+1}\vec{v}_{k+1} + d_{k+2}\vec{v}_{k+2} + \cdots + d_n\vec{v}_n = d_1\vec{v}_1 + d_2\vec{v}_2 + \cdots + d_k\vec{v}_k,$$

in other words:

$$-d_1\vec{v}_1 - d_2\vec{v}_2 - \cdots - d_k\vec{v}_k + d_{k+1}\vec{v}_{k+1} + d_{k+2}\vec{v}_{k+2} + \cdots + d_n\vec{v}_n = \vec{\mathbf{0}}_W.$$

But now, since $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k, \vec{v}_{k+1}, \vec{v}_{k+2}, \ldots, \vec{v}_n\}$ is a basis for all of V, this set is ***linearly independent***, and so $d_1 = d_2 = \cdots = d_k = d_{k+1} = \cdots = d_n = 0$.

In particular, this shows that the coefficients d_{k+1} through d_n in our dependence test equation for $T(\vec{v}_{k+1}), T(\vec{v}_{k+2}), \ldots, T(\vec{v}_n)$ above are all zero, and so this set is ***linearly independent***. This completes the Proof.■

Example: In our previous Example, the basis for our kernel had one vector, so $nullity(T) = 1$. However, $range(T) = \mathbb{P}^2$, so $rank(T) = 3$. Thus we verify the Dimension Theorem for this Example: $rank(T) + nullity(T) = 3 + 1 = 4 = dim(\mathbb{P}^3)$.□

The Dimension Theorem also tells us that if V is finite dimensional, then the ***range*** of $T : V \to W$ is also finite dimensional, even if the codomain W is infinite dimensional. Thus, we can also regard T as a linear transformation: $T : V \to range(T)$, and now ***both*** V and $range(T)$ are finite dimensional. This means that we can always construct the matrix of a linear transformation with respect to ***finite*** bases when the ***domain*** is finite dimensional.

Comparing Dimensions

As before, knowledge of the relative dimensions of the domain and codomain can immediately tell us if T is not one-to-one or not onto (proven exactly as in Chapter 2):

> ***Theorem:*** Suppose $T : V \to W$ is a linear transformation of finite dimensional vector spaces. Then:
>
> a) if $dim(V) < dim(W)$, then T ***cannot*** be ***onto***.
>
> b) if $dim(V) > dim(W)$, then T ***cannot*** be ***one-to-one***.

Example: Any linear transformation $T : \mathbb{R}^4 \to \mathbb{P}^7$ cannot be onto since:

$$dim(\mathbb{R}^4) = 4 < 8 = dim(\mathbb{P}^7).$$

However, it may or may not be one-to-one.

Similarly, any linear transformation $T : \mathbb{P}^6 \to \mathbb{R}^6$ may or may not be onto, but it cannot be one-to-one, since:

$$dim(\mathbb{P}^6) = 7 > 6 = dim(\mathbb{R}^6).\ \square$$

Compositions of Linear Transformations

We can compose two linear transformations, as before, as long as the codomain of the first transformation is the same as the domain of the second transformation:

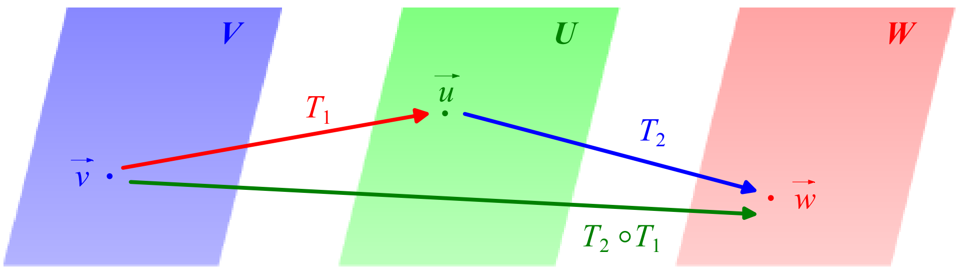
Definition/Theorem: Suppose that $T_1 : V \to U$ and $T_2 : U \to W$ are ***linear transformations***. Their ***composition:***

$$T_2 \circ T_1 : V \to W$$

is also a linear transformation. Its action is given as follows:
Suppose $\vec{v} \in V$, $T_1(\vec{v}) = \vec{u} \in U$, and $T_2(\vec{u}) = \vec{w} \in W$. Then:

$$(T_2 \circ T_1)(\vec{v}) = T_2(T_1(\vec{v})) = T_2(\vec{u}) = \vec{w}.$$

We can visualize the composition of these two transformations using the following diagram:



The Composition of Two Linear Transformations

Again, the linearity of the composition follows from that of the individual transformations, and is left as an easy Exercise.

Example: Let $T_1 : \mathbb{P}^2 \to \mathbb{P}^4$, and $T_2 : \mathbb{P}^4 \to \mathbb{P}^3$ be given by:

$$T_1(p(x)) = (x^2 - x + 3) \cdot p(x), \quad \text{and} \quad T_2(r(x)) = \frac{d}{dx} r(x) = D(r(x)).$$

For instance:

$$\begin{aligned} T_1(3x^2 - 5x + 4) &= (x^2 - x + 3) \cdot (3x^2 - 5x + 4) \\ &= 3x^4 - 8x^3 + 18x^2 - 19x + 12, \quad \text{and} \end{aligned}$$

$$T_2(5x^4 - 3x^2 + 7x - 2) = 20x^3 - 6x + 7.$$

We saw in the Exercises of Section 3.4 that multiplying a polynomial by a ***fixed*** polynomial is a linear transformation, and so is taking a derivative, so both T_1 and T_2 are linear transformations.

Let us demonstrate the composition $T_2 \circ T_1$ on $3x^2 - 5x + 4$:

$$\begin{aligned} (T_2 \circ T_1)(3x^2 - 5x + 4) &= T_2(T_1(3x^2 - 5x + 4)) \\ &= T_2(3x^4 - 8x^3 + 18x^2 - 19x + 12) \quad \text{(from above)} \\ &= 12x^3 - 24x^2 + 36x - 19. \square \end{aligned}$$

Example: Let $D : C^1(I) \to C(I)$ be the differentiation operation, where $I = [a,b]$, and $Ind(f) : C(I) \to C^1(I)$ be the indefinite integral operation. Then:

$$(D \circ Ind)(f) = D\left(\int_a^x f(t)\,dt\right) = \frac{d}{dx}\int_a^x f(t)\,dt = f(x),$$

where the last equation follows from ***The Fundamental Theorem of Calculus***. Thus, $D \circ Ind = \boldsymbol{I}_{C(I)}$, the identity operator on $C(I)$. However, let us see what happens to $f(x) = x^2 - 5x + 3$ under the ***reverse composition*** $Ind \circ D$, where $I = [0,1]$:

$$\begin{aligned}&(Ind \circ D)(x^2 - 5x + 3)\\ &= Ind\left(\frac{d}{dx}(x^2 - 5x + 3)\right) = Ind(2x - 5)\\ &= \int_0^x (2t-5)dt = t^2 - 5t\big|_0^x = x^2 - 5x.\end{aligned}$$

Thus, in this case, $Ind \circ D \neq \boldsymbol{I}_{C^1(I)}$. Notice that in particular, if $f(x) = c$, any constant valued function where $c \neq 0$, then $Ind \circ D(c) = 0 \neq c$. □

In Section 3.8, we will study ***invertible*** linear transformations. Recall that we learned that a left inverse for a linear operator on $\mathbb{R}^n$ must also be a right inverse. But our Example shows that this is not always true in the ***infinite dimensional*** case.

More generally, if we have a (finite) sequence of linear transformations, $T_1, T_2, \ldots, T_n$, where the ***codomain*** of T_i is the ***domain*** of T_{i+1} for all $i = 1 \ldots n-1$, we can once again construct the ***n-fold composition:***

$$T_n \circ T_{n-1} \circ \cdots \circ T_2 \circ T_1,$$

defined inductively in the usual manner as $T_n \circ (T_{n-1} \circ \cdots \circ T_2 \circ T_1)$.

The Matrix of a Composition

It should be no surprise that we can compute the matrix of a composition using a matrix product:

> ***Theorem:*** Let $T_1 : V \to U$ and $T_2 : U \to W$ be linear transformations of finite dimensional vector spaces. Let B be a basis for V, $B^{/}$ a basis for U, and $B^{//}$ a basis for W. Then:
>
> $$[T_2 \circ T_1]_{B,B^{//}} = [T_2]_{B^{/},B^{//}} \cdot [T_1]_{B,B^{/}}.$$
>
> In particular, if $V = U = W$, that is, T_1 and T_2 are ***operators*** on V, then:
>
> $$[T_2 \circ T_1]_B = [T_2]_B \cdot [T_1]_B.$$
>
> Furthermore, if $T_1 = T_2 = T$, the ***self-composition*** $T \circ T = T^2$ has matrix:
>
> $$[T^2]_B = [T]_B^2.$$
>
> We can generalize this formula for the ***r-fold self-composition:***
>
> $$[T^r]_B = [T]_B^r.$$

Proof: Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, $B^{/} = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$, and $B^{//} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be the respective bases. By construction, the matrices we are interested in are:

$$[T_2 \circ T_1]_{B,B^{//}} = Z = [\,[(T_2 \circ T_1)(\vec{v}_1)]_{B^{//}} \,|\, [(T_2 \circ T_1)(\vec{v}_2)]_{B^{//}} \,|\, \cdots \,|\, [(T_2 \circ T_1)(\vec{v}_n)]_{B^{//}}\,],$$
$$[T_2]_{B^/,B^{//}} = X = [\,[T_2(\vec{u}_1)]_{B^{//}} \,|\, [T_2(\vec{u}_2)]_{B^{//}} \,|\, \cdots \,|\, [T_2(\vec{u}_k)]_{B^{//}}\,], \text{ and}$$
$$[T_1]_{B,B^/} = Y = [\,[T_1(\vec{v}_1)]_{B^/} \,|\, [T_1(\vec{v}_2)]_{B^/} \,|\, \cdots \,|\, [T_1(\vec{v}_n)]_{B^/}\,].$$

Our goal is to show that $Z = XY$. All we have to do is ***unravel*** the definitions. The first column of Z is $[(T_2 \circ T_1)(\vec{v}_1)]_{B^{//}}$. We need to show that this equals the first column of XY. But recall from the definition of general matrix products that the first column of XY is $X\vec{y}_1$, where $\vec{y}_1$ is the first column of Y. But $\vec{y}_1$ is $[T_1(\vec{v}_1)]_{B^/}$. Thus, we must show that:

$$[(T_2 \circ T_1)(\vec{v}_1)]_{B^{//}} = X[T_1(\vec{v}_1)]_{B^/} = [T_2]_{B^/,B^{//}}[T_1(\vec{v}_1)]_{B^/}$$

But recall that in general (changing notation slightly to avoid confusion):

$$[T]_{S,S^/}[\vec{v}]_S = [T(\vec{v})]_{S^/}$$

where S is a basis for the domain of T and $S^/$ a basis for the codomain. Thus:

$$[T_2]_{B^/,B^{//}}[T_1(\vec{v}_1)]_{B^/} = [T_2(T_1(\vec{v}_1))]_{B^{//}} = [(T_2 \circ T_1)(\vec{v}_1)]_{B^{//}}.$$

Similarly, the rest of the columns of C are equal to the corresponding columns of AB. ■

Clearly, this idea also applies to the composition of several linear transformations, as long as the compatibility criterion is satisfied.

Example: Let $T_1 : \mathbb{P}^3 \to \mathbb{P}^2$ and $T_2 : \mathbb{P}^2 \to \mathbb{R}^2$ be given by:

$$T_1(p(x)) = 3p^/(x) - 5x \cdot p^{//}(x), \text{ and}$$
$$T_2(q(x)) = \langle q(-2), q(3)\rangle.$$

We will leave it as an Exercise to show that these are indeed linear transformations. Let us find the individual matrices and the matrix of the composition using the standard bases $B = \{1, x, x^2, x^3\}$, $B^/ = \{1, x, x^2\}$, and $B^{//} = \{\vec{e}_1, \vec{e}_2\}$ for $\mathbb{P}^3$, $\mathbb{P}^2$ and $\mathbb{R}^2$ respectively:

$$T_1(1) = 3 \cdot 0 - 5x \cdot 0 = 0,$$
$$T_1(x) = 3 \cdot 1 - 5x \cdot 0 = 3,$$
$$T_1(x^2) = 3 \cdot 2x - 5x \cdot 2 = -4x, \text{ and}$$
$$T_1(x^3) = 3 \cdot 3x^2 - 5x \cdot 6x = -21x^2.$$

Thus we can assemble:

$$[T_1]_{B,B^/} = \begin{bmatrix} 0 & 3 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & -21 \end{bmatrix}$$

Now for T_2:

$$T_2(1) = \langle 1, 1\rangle,$$
$$T_2(x) = \langle -2, 3\rangle, \text{ and}$$
$$T_2(x^2) = \langle 4, 9\rangle.$$

Thus:

$$[T_2]_{B^{/},B^{//}} = \begin{bmatrix} 1 & -2 & 4 \\ 1 & 3 & 9 \end{bmatrix}$$

We get the matrix of the composition using a matrix product:

$$[T_2 \circ T_1]_{B,B^{//}} = [T_2]_{B^{/},B^{//}}[T_1]_{B,B^{/}} = \begin{bmatrix} 1 & -2 & 4 \\ 1 & 3 & 9 \end{bmatrix}\begin{bmatrix} 0 & 3 & 0 & 0 \\ 0 & 0 & -4 & 0 \\ 0 & 0 & 0 & -21 \end{bmatrix} = \begin{bmatrix} 0 & 3 & 8 & -84 \\ 0 & 3 & -12 & -189 \end{bmatrix}.$$

Since we computed T_1 explicitly for the members of B, we can find their values under the composition directly:

$$\begin{aligned} T_2(T_1(1)) &= T_2(0) = \langle 0, 0\rangle, \\ T_2(T_1(x)) &= T_2(3) = \langle 3, 3\rangle, \\ T_2(T_1(x^2)) &= T_2(-4x) = \langle 8, -12\rangle, \text{ and} \\ T_2(T_1(x^3)) &= T_2(-21x^2) = \langle -21 \bullet 4, -21 \bullet 9\rangle = \langle -84, -189\rangle \end{aligned}$$

and we can see that these are indeed the four columns of the matrix product.□

Example: We saw in the previous Section that the function space:

$$W = Span(B), \text{ where } B = \{x^2e^{-3x}, xe^{-3x}, e^{-3x}\},$$

is ***preserved*** by the derivative operation D. In other words, D is an ***operator:***

$$D : W \to W.$$

We also found the matrix of D with respect to B:

$$[D]_B = \begin{bmatrix} -3 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 1 & -3 \end{bmatrix}.$$

Thus, the ***second derivative*** $D \circ D = D^2$ also ***preserves*** W, and:

$$[D^2]_B = [D]_B^2 = \begin{bmatrix} -3 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 1 & -3 \end{bmatrix}^2 = \begin{bmatrix} 9 & 0 & 0 \\ -12 & 9 & 0 \\ 2 & -6 & 9 \end{bmatrix}.$$

We can use this matrix to find the 2nd derivative of $f(x) = 8x^2e^{-3x} - 5xe^{-3x} + 9e^{-3x}$ using the matrix product:

$$\begin{bmatrix} 9 & 0 & 0 \\ -12 & 9 & 0 \\ 2 & -6 & 9 \end{bmatrix}\begin{bmatrix} 8 \\ -5 \\ 9 \end{bmatrix} = \begin{bmatrix} 72 \\ -141 \\ 127 \end{bmatrix}.$$

Thus, $f^{//}(x) = 72x^2e^{-3x} - 141xe^{-3x} + 127e^{-3x}$.□

3.7 Section Summary

We say that a linear transformation $T : V \to W$ is ***one-to-one*** or ***injective*** if the image of different vectors from the domain are different vectors from the codomain: if $\vec{v}_1 \neq \vec{v}_2$ then $T(\vec{v}_1) \neq T(\vec{v}_2)$.
T is ***one-to-one*** ***if and only if*** $ker(T) = \left\{\vec{\mathbf{0}}_V\right\}$.

We say that a linear transformation $T : V \to W$ is ***onto*** or ***surjective*** if $range(T) = W$.

The rref R of $[T]_{B,B'}$ can be used to find a basis for $nullspace\left([T]_{B,B'}\right)$ and $columnspace\left([T]_{B,B'}\right)$, as we did in Chapter 2.
By decoding the basis for $nullspace\left([T]_{B,B'}\right)$ using B, we can find a basis for $ker(T)$. Similarly, by decoding the basis for $columnspace\left([T]_{B,B'}\right)$ using B', we can find a basis for $range(T)$.

The Dimension Theorem for Abstract Vector Spaces: Let $T : V \to W$ be a linear transformation, and suppose that V is ***finite dimensional*** with $dim(V) = n$. Then, both $ker(T)$ and $range(T)$ are finite dimensional, and we define $rank(T) = dim(range(T))$, and $nullity(T) = dim(ker(T))$. Furthermore:

$$rank(T) + nullity(T) = n = dim(V).$$

Let $T : V \to W$ be a linear transformation of ***finite dimensional*** vector spaces. Then:

(a) if $dim(V) < dim(W)$, then T ***cannot*** be ***onto***;

(b) if $dim(V) > dim(W)$, then T ***cannot*** be ***one-to-one***.

Let $T_1 : V \to U$, and $T_2 : U \to W$ be linear transformations. The composition: $T_2 \circ T_1 : V \to W$ is again a linear transformation. Its action is given as follows: suppose $\vec{v} \in V$, $T_1(\vec{v}) = \vec{u} \in U$, and $T_2(\vec{u}) = \vec{w} \in W$. Then: $(T_2 \circ T_1)(\vec{v}) = T_2(T_1(\vec{v})) = T_2(\vec{u}) = \vec{w}$.

Let B be a basis for V, B' a basis for U, and B'' a basis for W. Then:

$$[T_2 \circ T_1]_{B,B''} = [T_2]_{B',B''}[T_1]_{B,B'}.$$

In particular, if $V = U = W$, that is, T_1 and T_2 are ***operators*** on V, then $[T_2 \circ T_1]_B = [T_2]_B \cdot [T_1]_B$.

Furthermore, if $T_1 = T_2 = T$, the ***self-composition*** $T \circ T = T^2$ has matrix: $[T^2]_B = [T]_B^2$.

We can generalize this formula for the ***r-fold self-composition:*** $[T^r]_B = [T]_B^r$.

3.7 Exercises

1. Let $T : \mathbb{P}^2 \to \mathbb{R}^4$ be the linear transformation from Exercise 7 of Section 3.5, and Exercise 11 of Section 3.6, given by: $T(p(x)) = \left\langle p(-2), p'(1), p''(x), \int_0^1 p(x)dx\right\rangle$.

   a. Can we immediately say that T is ***not*** one-to-one? Why or why not?

   b. Can we immediately say that T is ***not*** onto? Why or why not?

   Let $B = \{1, x, x^2\}$ and $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$. The matrix $[T]_{B,B'}$ can be found in the Answer Key for Exercise 11, Section 3.6.

   c. Find the rref of $[T]_{B,B'}$.

   d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.

   e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.

   f. Is T one-to-one? Is T onto?

g. Verify the Dimension Theorem for T.

h. Describe all polynomials $p(x) \in \mathbb{P}^2$, such that:

$$p(-2) = 38,\ p'(1) = 3,\ p''(x) = 10,\ \text{and } \int_0^1 p(x)dx = 13/6.$$

Hint: Solve an augmented system that uses $[T]_{B,B'}$. What should be on the rightmost column?

2. Let $T : \mathbb{P}^3 \to \mathbb{P}^1$ be the linear transformation from Exercise 8 in Section 3.5 and Exercise 12 in Section 3.6, given by: $T(p(x)) = p''(x)$.

 a. Can we immediately say that T is ***not*** one-to-one? Why or why not?

 b. Can we immediately say that T is ***not*** onto? Why or why not?

 Let $B = \{1, x, x^2, x^3\}$ and $B' = \{1, x\}$. The matrix $[T]_{B,B'}$ can be found in the Answer Key for Exercise 12, Section 3.6.

 c. Find the rref of $[T]_{B,B'}$.

 d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.

 e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.

 f. Is T one-to-one? Is T onto?

 g. Verify the Dimension Theorem for T.

3. Let $T : \mathbb{P}^2 \to \mathbb{P}^3$ be the linear transformation from Exercise 10 in Section 3.5 and Exercise 10 in Section 3.6, given by: $T(p(x)) = \int_0^x p(t)dt$.

 a. Can we immediately say that T is ***not*** one-to-one? Why or why not?

 b. Can we immediately say that T is ***not*** onto? Why or why not?

 Let $B = \{1, x, x^2\}$ and $B' = \{1, x, x^2, x^3\}$. The matrix $[T]_{B,B'}$ can be found in the Answer Key for Exercise 10, Section 3.6.

 c. Find the rref of $[T]_{B,B'}$.

 d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.

 e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.

 f. Is T one-to-one? Is T onto?

 g. Verify the Dimension Theorem for T.

4. Let $T : \mathbb{P}^3 \to \mathbb{P}^2$ be the linear transformation from Exercise 25 in Section 3.6, given by:

$$T(p(x)) = p'(x) + (x+1) \cdot p''(x) + 2p(-1).$$

 a. Can we immediately say that T is ***not*** one-to-one? Why or why not?

 b. Can we immediately say that T is ***not*** onto? Why or why not?

 Let $B = \{1, x, x^2, x^3\}$ and $B' = \{1, x, x^2\}$. The matrix $[T]_{B,B'}$ can be found in the Answer Key for Exercise 21, Section 3.6.

 c. Find the rref of $[T]_{B,B'}$.

 d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.

 e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.

 f. Is T one-to-one? Is T onto?

 g. Verify the Dimension Theorem for T.

5. Let $T : \mathbb{P}^2 \to \mathbb{P}^3$ be the linear transformation from Exercise 29 in Section 3.6, given by:

$$T(p(x)) = (2x-5) \cdot p(x) + (x^2+3) \cdot p'(x) - p(-2) \cdot x^3.$$

   a. Can we immediately say that T is ***not*** one-to-one? Why or why not?
   b. Can we immediately say that T is ***not*** onto? Why or why not?
   Let $B = \{1, x, x^2\}$ and $B' = \{1, x, x^2, x^3\}$. The matrix $[T]_{B,B'}$ can be found in the Answer Key for Exercise 22, Section 3.6.
   c. Find the rref of $[T]_{B,B'}$.
   d. Use (c) to find a basis (if possible) for *ker*(T) and state *nullity*(T).
   e. Use (c) to find a basis for *range*(T) and state *rank*(T).
   f. Is T one-to-one? Is T onto?
   g. Verify the Dimension Theorem for T.

6. Let $T : \mathbb{P}^2 \to \mathbb{P}^3$ be the linear transformation given by:

$$T(p(x)) = (x^2-5) \cdot p'(x) + p''(-1) \cdot (x^3+2x-4).$$

   a. Convince yourself mentally that T is indeed a linear transformation from $\mathbb{P}^2$ to $\mathbb{P}^3$.
   b. Can we immediately say that T is ***not*** one-to-one? Why or why not?
   c. Can we immediately say that T is ***not*** onto? Why or why not?
   d. Let $B = \{1, x, x^2\}$ and $B' = \{1, x, x^2, x^3\}$. Find $[T]_{B,B'}$.
   e. Find the rref of $[T]_{B,B'}$.
   f. Use (e) to find a basis (if possible) for *ker*(T) and state *nullity*(T).
   g. Use (e) to find a basis for *range*(T) and state *rank*(T).
   h. Is T one-to-one? Is T onto?
   i. Verify the Dimension Theorem for T.

7. Let $T : \mathbb{P}^3 \to \mathbb{P}^2$ be the linear transformation given by:

$$T(p(x)) = p(-2) \cdot (2x^2 - 10x + 6) + p'(-1) \cdot (3x^2 - 15x + 9).$$

   a. Convince yourself mentally that T is indeed a linear transformation from $\mathbb{P}^3$ to $\mathbb{P}^2$.
   b. Can we immediately say that T is ***not*** one-to-one? Why or why not?
   c. Can we immediately say that T is ***not*** onto? Why or why not?
   d. Let $B = \{1, x, x^2, x^3\}$ and $B' = \{1, x, x^2\}$. Find $[T]_{B,B'}$.
   e. Find the rref of $[T]_{B,B'}$.
   f. Use (e) to find a basis (if possible) for *ker*(T) and state *nullity*(T).
   g. Use (e) to find a basis for *range*(T) and state *rank*(T).
   h. Is T one-to-one? Is T onto?
   i. Verify the Dimension Theorem for T.

8. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^1$ is a linear transformation whose matrix with respect to the bases $B = \{1, 5-x, 2+3x-x^2\}$ for $\mathbb{P}^2$ and $B' = \{x+3, 2\}$ for $\mathbb{P}^1$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 4 & -1 & 5 \\ -7 & 0 & -2 \end{bmatrix}.$$

Note: this was the linear transformation in Section 3.6, Exercise 27.

a. Can we immediately say that T is ***not*** one-to-one? Why or why not?
b. Can we immediately say that T is ***not*** onto? Why or why not?
c. Find the rref of $[T]_{B,B'}$.
d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.
e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.
f. Is T one-to-one? Is T onto?
g. Verify the Dimension Theorem for T.

9. Suppose that $T : \mathbb{P}^1 \to \mathbb{P}^2$ is a linear transformation whose matrix with respect to the bases $B = \{1, 2+x\}$ for $\mathbb{P}^1$ and $B' = \{x^2 - x, x+1, -1\}$ for $\mathbb{P}^2$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 5 & 3 \\ -1 & 2 \\ 8 & -7 \end{bmatrix}.$$

Note: this was the linear transformation in Section 3.6, Exercise 28.

a. Can we immediately say that T is ***not*** one-to-one? Why or why not?
b. Can we immediately say that T is ***not*** onto? Why or why not?
c. Find the rref of $[T]_{B,B'}$.
d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.
e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.
f. Is T one-to-one? Is T onto?
g. Verify the Dimension Theorem for T.

10. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^1$ is a linear transformation whose matrix with respect to the bases $B = \{1, 2+x, x-x^2\}$ for $\mathbb{P}^2$ and $B' = \{x+3, x-1\}$ for $\mathbb{P}^1$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 2 & -1 & 3 \\ -8 & 4 & -12 \end{bmatrix}.$$

a. Can we immediately say that T is ***not*** one-to-one? Why or why not?
b. Can we immediately say that T is ***not*** onto? Why or why not?
c. Find the rref of $[T]_{B,B'}$.
d. Use (a) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.
e. Use (a) to find a basis for $range(T)$ and state $rank(T)$.
f. Is T one-to-one? Is T onto?
g. Verify the Dimension Theorem for T.

11. Suppose that $T : \mathbb{P}^1 \to \mathbb{P}^2$ is a linear transformation whose matrix with respect to the bases $B = \{1, 1-x\}$ for $\mathbb{P}^1$ and $B' = \{x^2+2x, x-1, 1\}$ for $\mathbb{P}^2$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 14 & 10 \\ -21 & -15 \\ 35 & 25 \end{bmatrix}.$$

a. Can we immediately say that T is ***not*** one-to-one? Why or why not?
b. Can we immediately say that T is ***not*** onto? Why or why not?
c. Find the rref of $[T]_{B,B'}$.
d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.
e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.
f. Is T one-to-one? Is T onto?
g. Verify the Dimension Theorem for T.

12. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^2$ is the operator whose matrix with respect to the standard basis $B = \{1, x, x^2\}$ is given by:

$$[T]_B = \begin{bmatrix} 4 & 3 & -6 \\ -1 & 2 & 5 \\ 5 & 12 & 3 \end{bmatrix}.$$

a. Can we immediately say that T is ***not*** one-to-one? Why or why not?
b. Can we immediately say that T is ***not*** onto? Why or why not?
c. Find the rref of $[T]_B$.
d. Use (c) to find a basis (if possible) for $ker(T)$ and state $nullity(T)$.
e. Use (c) to find a basis for $range(T)$ and state $rank(T)$.
f. Is T one-to-one? Is T onto?
g. Verify the Dimension Theorem for T.
h. Find all polynomials $p(x) \in \mathbb{P}^2$, if possible, such that:

$$T(p(x)) = 6 - 7x - 9x^2.$$

Hint: find the rref of a certain augmented matrix.

13. Let $T_1 : \mathbb{P}^2 \to \mathbb{P}^3$ and $T_2 : \mathbb{P}^3 \to \mathbb{P}^2$ be the linear transformations given by:

$$T_1(p(x)) = (x^2 + 3x - 5) \bullet p'(x) + 4p(x), \text{ and}$$
$$T_2(q(x)) = 3q'(x) - 5q''(x).$$

Let $B = \{1, x, x^2\}$ and $B' = \{1, x, x^2, x^3\}$ be the standard bases, respectively, for $\mathbb{P}^2$ and $\mathbb{P}^3$.

a. Convince yourself mentally that T_1 and T_2 are linear transformations.
b. Find $[T_1]_{B,B'}$ and $[T_2]_{B',B}$.
c. Explain why both compositions $T_2 \circ T_1$ and $T_1 \circ T_2$ are defined.
d. Find $[T_2 \circ T_1]_B$ and $[T_1 \circ T_2]_{B'}$.

14. Let $T_1 : \mathbb{P}^2 \to \mathbb{P}^3$ and $T_2 : \mathbb{P}^3 \to \mathbb{R}^3$ be the linear transformations given by:

$$T_1(p(x)) = (2x + 3) \bullet p(x), \text{ and}$$
$$T_2(q(x)) = \langle q(-3), q'(2), q''(-1) \rangle.$$

Let $B = \{1, x, x^2\}$, $B' = \{1, x, x^2, x^3\}$ and $B'' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the standard bases, respectively, for $\mathbb{P}^2$, $\mathbb{P}^3$ and $\mathbb{R}^3$.

a. Convince yourself mentally that T_1 and T_2 are linear transformations.

b. Find $[T_1]_{B,B'}$ and $[T_2]_{B',B''}$.

c. Explain why the composition $T_2 \circ T_1$ is defined, and find $[T_2 \circ T_1]_{B,B''}$ using (b).

d. Is the ***composition*** $T_1 \circ T_2$ defined? Why or why not?

e. Is the ***matrix product*** $[T_1]_{B,B'} \cdot [T_2]_{B',B''}$ defined? Why or why not?

15. Let $T_1 : \mathbb{P}^2 \to \mathbb{P}^3$ and $T_2 : \mathbb{P}^3 \to \mathbb{P}^1$ be linear transformations. Suppose that $B = \{1, 1-x, 2x+x^2\} \subset \mathbb{P}^2$, $B' = \{1, 1+x, 3x-x^2, x^2+x^3\} \subset \mathbb{P}^3$, and $B'' = \{-1, 1+x\} \subset \mathbb{P}^1$. Note that each set contains polynomials of distinct degrees, and each set contains the correct number of vectors, so both are bases for the corresponding space. Now, suppose we are given that:

$$[T_1]_{B,B'} = \begin{bmatrix} 3 & 5 & 1 \\ -2 & -4 & 2 \\ 1 & 0 & 7 \\ 1 & 2 & -1 \end{bmatrix}, \text{ and } [T_2]_{B',B''} = \begin{bmatrix} 3 & -2 & 5 & 1 \\ 2 & 4 & 7 & -3 \end{bmatrix}.$$

a. Does the composition $T_1 \circ T_2$ make sense? Why or why not? If so, what are the domain and codomain of $T_1 \circ T_2$?

b. Does the composition $T_2 \circ T_1$ make sense? Why or why not? If so, what are the domain and codomain of $T_2 \circ T_1$?

c. Compute $T_1(3x^2 - 5x + 2)$

d. Use your work to (c) to compute $(T_2 \circ T_1)(3x^2 - 5x + 2)$.

e. Find $[T_2 \circ T_1]_{B,B''}$.

f. Use your answer to (e) to compute $(T_2 \circ T_1)(3x^2 - 5x + 2)$ directly. You should get the same answer as in part (d).

16. Let $T_1 : \mathbb{P}^1 \to \mathbb{P}^2$ and $T_2 : \mathbb{P}^2 \to \mathbb{P}^1$ be linear transformations. Let $B = \{1, 1-x\} \subset \mathbb{P}^1$, and $B' = \{1, 1+x, 3x-x^2\} \subset \mathbb{P}^2$. Note that each set contains polynomials of distinct degrees, and each set contains the correct number of vectors needed to form a basis, so they are all bases for the corresponding spaces. Now, suppose we are given that:

$$[T_1]_{B,B'} = \begin{bmatrix} 2 & -3 \\ 1 & 4 \\ 0 & 7 \end{bmatrix}, \text{ and } [T_2]_{B',B} = \begin{bmatrix} 4 & 3 & -2 \\ 7 & -1 & 4 \end{bmatrix}$$

a. Is the composition $T_1 \circ T_2$ defined? Explain. If so, what are the domain and codomain of $T_1 \circ T_2$?

b. Is the composition $T_2 \circ T_1$ defined? Explain. If so, what are the domain and codomain of $T_2 \circ T_1$?

c. Compute $T_1(5x - 7)$

d. Use (c) to compute $(T_2 \circ T_1)(5x - 7)$.

e. Find $[T_2 \circ T_1]_B$.

f. Use your answer to (e) to compute $(T_2 \circ T_1)(5x - 7)$ directly. You should get the same answer as in part (d).

g. Compute $T_2(6x^2 + 3x - 4)$.

h. Use (g) to compute $(T_1 \circ T_2)(6x^2 + 3x - 4)$.

i. Find $[T_1 \circ T_2]_{B'}$.

j. Use your answer to (i) to compute $(T_1 \circ T_2)(6x^2 + 3x - 4)$ directly. You should get the same answer as in part (h).

For Exercises (17) to (23): In Section 3.6, Exercises 11 to 18, we found the matrix $[D]_B$ of the derivative operation D on the subspaces $W = Span(B)$. (a) Use your answers in that section to find the matrices of the 2nd and 3rd derivatives, $[D^2]_B$ and $[D^3]_B$; (b) Use these matrices to directly find the 2nd and 3rd derivatives of the indicated function $f(x)$ using a matrix product. (c) Show that D is both one-to-one and onto on W by finding the rref of $[D]_B$ and describing $ker(D)$ and $range(B)$.

17. $W = Span(B)$, where $B = \{e^{-x}, e^{2x}\}$; $f(x) = 5e^{-x} - 3e^{2x}$.

18. $W = Span(B)$, where $B = \{e^x \sin(x), e^x \cos(x)\}$; $f(x) = 4e^x \sin(x) - 3e^x \cos(x)$.

19. $W = Span(B)$, where $B = (\{e^{-3x} \sin(2x), e^{-3x} \cos(2x)\})$; $f(x) = 5e^{-3x} \sin(2x) - 9e^{-3x} \cos(2x)$.

20. $W = Span(B)$, where $B = (\{xe^{5x}, e^{5x}\})$; $f(x) = -2xe^{5x} + 7e^{5x}$.

21. $W = Span(B)$, where $B = \{x^2e^{-4x}, xe^{-4x}, e^{-4x}\}$; $f(x) = -5x^2e^{-4x} + 2xe^{-4x} - 7e^{-4x}$.

22. $W = Span(B)$, where $B = (\{x^2 5^x, x5^x, 5^x\})$; $f(x) = -4x^2 5^x + 9x5^x - 2(5^x)$.

23. $W = Span(B)$, where $B = (\{x\sin(2x), x\cos(2x), \sin(2x), \cos(2x)\})$;
$f(x) = 4x\sin(2x) + 9x\cos(2x) - 5\sin(2x) + 8\cos(2x)$.

24. In Exercise 21 of Section 3.6, we constructed the matrix $[D]_B$ of the derivative operator D on $W = Span(B)$, where $B = \{e^{ax}\sin(bx), e^{ax}\cos(bx)\}$:

$$[D]_B = \begin{bmatrix} a & -b \\ b & a \end{bmatrix}.$$

a. Find $[D^2]_B$ and $[D^3]_B$. Observe how the four entries are related to each other in two pairs.

b. Use Induction to show that for any positive integer k: $[D^k]_B = \begin{bmatrix} a_k & -b_k \\ b_k & a_k \end{bmatrix}$,

for some real numbers a_k and b_k.

25. Suppose that $T_1 : V \to U$, and $T_2 : U \to W$ are linear transformations of vector spaces. Prove that $T_2 \circ T_1$ is also a linear transformation. In other words, prove that $T_2 \circ T_1$ is ***additive*** and ***homogeneous***.

26. Prove that if $T : V \to W$ is ***one-to-one*** and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a set of ***linearly independent*** vectors from V, then $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_k)\}$ is a set of ***linearly independent*** vectors from W.

27. Let $a_1, a_2, \ldots, a_n, a_{n+1} \in \mathbb{R}$ be $n+1$ distinct real numbers, and construct $\vec{a} = \langle a_1, a_2, \ldots, a_n, a_{n+1} \rangle \in \mathbb{R}^{n+1}$. Prove that the evaluation homomorphism:

$$E_{\vec{a}} : \mathbb{P}^n \to \mathbb{R}^{n+1}, \quad \text{where: } E_{\vec{a}}(p(x)) = \langle p(a_1), p(a_2), \ldots, p(a_n), p(a_{n+1}) \rangle,$$

as defined in Section 3.5, is ***one-to-one***. Hint: See Exercise 24 in Section 3.5.

28. Suppose that $T : V \to W$ is a linear transformation, with $dim(V) = n$ and $dim(W) = m$. Prove the following statements:

   a. If $n \leq m$, then: T is ***one-to-one*** ***if and only if*** for ***any*** basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for V, the image set $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_n)\}$ is linearly independent. Hint: think of $ker(T)$.

   b. If $n \geq m$, then: T is ***onto*** ***if and only if*** there exists a linearly ***independent*** subset $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}$ from V, such that the image set $\{T(\vec{v}_1), T(\vec{v}_2), \ldots, T(\vec{v}_m)\}$ is also linearly ***independent***. Note that there are only m vectors in these sets. Hint: T is onto ***if and only if*** $rank(T) = m$.

   c. Bonus: show that (a) is still true if the phrase "for any basis" is replaced with "for at least one basis."

29. ***Decoding the Kernel and Range:*** The purpose of this Exercise is to show that we can obtain a basis for $ker(T)$ and $range(T)$ by decoding the information found in any matrix for T. Suppose that $T : V \to W$ is a linear transformation, with $dim(V) = n$ and $dim(W) = m$, both finite-dimensional vector spaces. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for V, and let $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be a basis for W. Let us construct the $m \times n$ matrix $[T]_{B,B'}$, and let R be the rref of $[T]_{B,B'}$.

   a. Suppose that $\vec{z} \in nullspace\left([T]_{B,B'}\right) \trianglelefteq \mathbb{R}^{\boldsymbol{n}}$. By the Uniqueness of Representation Property, we know that there exists $\vec{u} \in V$ such that $\langle \vec{u} \rangle_B = \vec{z}$. Show that $\vec{u} \in ker(T)$.

   b. Conversely, suppose that $\vec{u} \in ker(T)$. Show that $\langle \vec{u} \rangle_B \in nullspace\left([T]_{B,B'}\right)$.

   c. Now suppose that $\vec{b} \in colspace\left([T]_{B,B'}\right) \trianglelefteq \mathbb{R}^{\boldsymbol{m}}$. By the Uniqueness of Representation Property, we know that there exists $\vec{d} \in W$ such that $\left\langle \vec{d} \right\rangle_B = \vec{b}$. Show that $\vec{d} \in range(T)$.

   d. Conversely, suppose that $\vec{d} \in range(T)$. Show that $\left\langle \vec{d} \right\rangle_B \in colspace\left([T]_{B,B'}\right)$.

   e. Use (a) and (b) to prove that $ker(T) = \left\{\vec{\mathbf{0}}_V\right\}$ ***if and only if*** $nullspace\left([T]_{B,B'}\right) = \left\{\vec{\mathbf{0}}_{\boldsymbol{n}}\right\}$.

   f. Similarly, use (c) and (d) to prove that $range(T) = \left\{\vec{\mathbf{0}}_W\right\}$ ***if and only if*** $colspace\left([T]_{B,B'}\right) = \left\{\vec{\mathbf{0}}_{\boldsymbol{m}}\right\}$.

   Parts (e) and (f) handle the trivial cases where either $ker(T)$ or $range(T)$ is the zero-subspace. Thus, we finish the problem by assuming that ***neither*** space is the zero-subspace:

   g. Let $\{\vec{z}_1, \vec{z}_2, \ldots, \vec{z}_k\} \subset \mathbb{R}^{\boldsymbol{n}}$ be the basis that we obtain for $nullspace\left([T]_{B,B'}\right)$ using R, as we did in Chapter 2. By (a), the corresponding vectors $\vec{u}_i$ such that $\langle \vec{u}_i \rangle_B = \vec{z}_i$ for every $i = 1 \ldots k$ form a subset $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ for $ker(T)$.

   Prove that $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is a ***basis*** for $ker(T)$. Reminder: this means you have to prove two properties: linear independence and Spanning.

   In particular, this tells us that $dim(ker(T)) = dim\left(nullspace\left([T]_{B,B'}\right)\right)$ for any matrix $[T]_{B,B'}$ representing T.

   h. Finally, let $\{\vec{c}_{i_1}, \vec{c}_{i_2}, \ldots, \vec{c}_{i_r}\} \subset \mathbb{R}^{\boldsymbol{m}}$ be the original columns of $[T]_{B,B'}$ that correspond to the leading 1's found in R, so that this set forms a basis for $columnspace\left([T]_{B,B'}\right)$ as we found in Chapter 2. By (b), the corresponding vectors $\vec{d}_j \in W$ such that $\left\langle \vec{d}_j \right\rangle_{B'} = \vec{c}_{i_j}$ for every $j = 1 \ldots r$ form a subset $\left\{\vec{d}_1, \vec{d}_2, \ldots, \vec{d}_r\right\}$ for $range(T)$.

Prove that $\left\{\vec{d}_1, \vec{d}_2, \ldots, \vec{d}_r\right\}$ is a ***basis*** for $range(T)$. Again, you have to show both linear independence and Spanning properties.

In particular, this tells us that $dim(range(T)) = dim\left(colspace\left([T]_{B,B'}\right)\right)$ for any matrix $[T]_{B,B'}$ representing T.

30. ***The Kernel and Range of a Composition:*** The purpose of this Exercise is to generalize Exercise 23 in Section 2.5. We will investigate the kernel and range of the composition of two linear transformations. Suppose that:

$$T_1 : V \to U, \quad \text{and} \quad T_2 : U \to W,$$

are linear transformations of vector spaces. We do ***not*** have to assume that any of these spaces is finite-dimensional.

a. Write down the general definition of the ***kernel*** of ***any*** linear transformation $T : X \to Y$, where X and Y are arbitrary vector spaces.

b. Adapt the definition in part (a) to write down the definition of $ker(T_1)$, $ker(T_2)$ and $ker(T_2 \circ T_1)$, as set-up above. There should be ***three*** separate definitions. Make sure that you precisely use the symbols V, U, W, $\vec{\mathbf{0}}_V$, $\vec{\mathbf{0}}_U$ and $\vec{\mathbf{0}}_W$, where appropriate.

c. Two out of the three subspaces that you defined in (b) are subspaces of the same vector space. Which of the two kernels live in which same vector space?

d. Use your definitions to prove that $ker(T_1) \trianglelefteq ker(T_2 \circ T_1)$.

   Hint: This means that you must show that every member $\vec{v}$ of $ker(T_1)$ is also a member of $ker(T_2 \circ T_1)$. Note that in Section 2.5, we did not know the general concept of a subspace. We can now say that $ker(T_1)$ is a ***subspace*** of $ker(T_2 \circ T_1)$, and not just a subset.

e. Use part (d) to prove that if $T_2 \circ T_1$ is one-to-one, then T_1 is also one-to-one.

f. Write down the ***contrapositive*** of the statement in (e).

   Now, we will repeat the steps above for the ***range:***

g. Write down the general definition of the ***range*** of $T : X \to Y$, as set-up in (a).

h. Adapt the definition in part (g) to write down the definition of $range(T_1)$, $range(T_2)$, and $range(T_2 \circ T_1)$, as set up above. There should be ***three*** separate definitions. Make sure that you precisely use the symbols V, U, and W, where appropriate.

i. Two out of the three subspaces that you defined in (h) are subspaces of the same vector space. Which of the two ranges live in which same vector space?

j. Use your definitions to prove that $range(T_2 \circ T_1) \trianglelefteq range(T_2)$.

   Hint: This means that you must show that every member $\vec{w}$ of $range(T_2 \circ T_1)$ is also a member of $range(T_2)$. Again, we can now say that $range(T_2 \circ T_1)$ is a ***subspace*** of $range(T_2)$, and not merely a subset.

k. Use part (j) to prove that if $T_2 \circ T_1$ is onto, then T_2 is also onto. Do you notice the difference with part (e)?

l. Write down the ***contrapositive*** of the statement in (k).

3.8 Isomorphisms

Now that we understand the nature of one-to-one and onto linear transformations, we will put them together, as before, to generalize a very special kind of linear transformation:

Definition: If V and W are vector spaces, we say that a linear transformation $T : V \to W$ is an ***isomorphism*** if T is both ***one-to-one*** and ***onto***. We also say that T is ***invertible***, T is ***bijective***, and that V and W are ***isomorphic*** to each other.
If $V = W$, an isomorphism $T : V \to V$ is also called an ***automorphism*** or self-isomorphism.

The existence of an isomorphism between two spaces forces their dimensions to be equal:

Theorem: Let $T : V \to W$ be an isomorphism of ***finite dimensional*** vector spaces. Then:

$$dim(V) = dim(W).$$

Proof: If $dim(V) < dim(W)$, then T cannot be onto, and if $dim(V) > dim(W)$, then T cannot be one-to-one. Thus we must have $dim(V) = dim(W)$. ■

This Theorem says that we do not have to bother asking if a linear transformation is an isomorphism if the domain and the codomain do not have the same dimension. (In fact, this Theorem is true even if both V and W are infinite dimensional, but the proof needs the Axiom of Choice). Likewise, the converse of this Theorem is also true and will be proven in the Exercises:

Theorem: If V and W are ***finite dimensional*** vector spaces and $dim(V) = dim(W)$, then ***there exists*** an ***isomorphism*** $T : V \to W$.

Again, this Theorem is true even if both spaces are infinite dimensional, but the Axiom of Choice is needed as well. Thus, the two statements above can be combined in full generality as:

Theorem: Two vector spaces V and W are ***isomorphic*** to each other ***if and only if:***

$$dim(V) = dim(W).$$

Now, we come to a Theorem that is often taken by some textbooks to be the ***definition*** of what an isomorphism of vector spaces is:

Definition/Theorem: A linear transformation $T : V \to W$ is an ***isomorphism*** of vector spaces ***if and only if*** there exists another linear transformation:

$$T^{-1} : W \to V,$$

called the ***inverse*** of T, which is ***also*** an ***isomorphism***, such that if $\vec{v} \in V$ and $T(\vec{v}) = \vec{w} \in W$, then $T^{-1}(\vec{w}) = \vec{v}$, and thus:

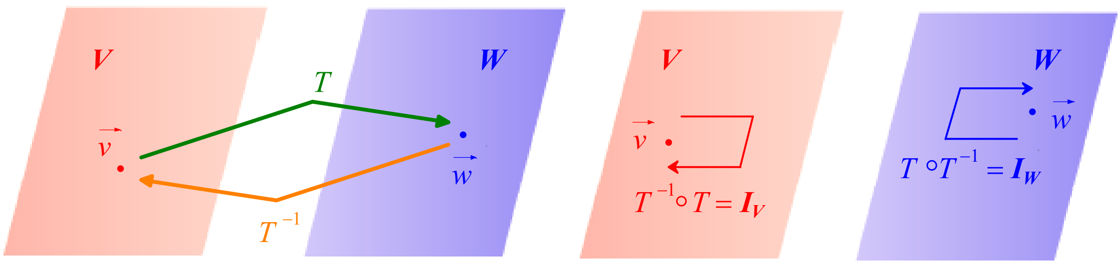
$$(T^{-1} \circ T)(\vec{v}) = \vec{v} \quad \text{and} \quad (T \circ T^{-1})(\vec{w}) = \vec{w}.$$

In other words, T^{-1} is also a ***one-to-one*** and ***onto*** linear transformation.

Furthermore, T^{-1} is ***unique***, and T and T^{-1} possess the ***cancellation properties:***

$$T^{-1} \circ T = \boldsymbol{I}_V \quad \text{and} \quad T \circ T^{-1} = \boldsymbol{I}_W,$$

where $\boldsymbol{I}_V$ and $\boldsymbol{I}_W$ are the ***identity*** operators on V and W, respectively.
In particular, if T is an ***automorphism***, we get: $T^{-1} \circ T = \boldsymbol{I}_V = T \circ T^{-1}$.



The Composition of T with T^{-1} $\qquad$ $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$

Since the roles of T and T^{-1} can be reversed, this Theorem tells us that it is indeed appropriate that we say that V and W are ***isomorphic to each other***. We can also say that "V is ***isomorphic*** to W" and "W is ***isomorphic*** to V."

Proof: ($\Rightarrow$) Suppose T is an isomorphism, which according to our definition means that T is both one-to-one and onto. We must show that we can construct a linear transformation $T^{-1} : W \to V$, which is also an isomorphism, such that $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$.

Suppose that $\vec{w} \in W$. We must define $T^{-1}(\vec{w})$. Since T is ***onto***, we can find ***at least one*** member $\vec{v} \in V$ so that $T(\vec{v}) = \vec{w}$. However, T is also ***one-to-one***, so there is ***at most one*** such vector $\vec{v}$. Thus, we can define:

$$T^{-1}(\vec{w}) = \vec{v}, \quad \text{where} \quad T(\vec{v}) = \vec{w},$$

and this can be done in ***exactly one way***. Thus, T^{-1} is a well-defined ***function***. Furthermore, with this notation, we immediately get the last part of the Theorem:

$$T(T^{-1}(\vec{w})) = T(\vec{v}) = \vec{w} \text{ for all } \vec{w} \in W, \quad \text{and}$$
$$T^{-1}(T(\vec{v})) = T^{-1}(\vec{w}) = \vec{v}.$$

Now, let us show that T is a ***linear transformation***, that is, it enjoys the two linearity properties:

Additivity: Suppose $\vec{w}_1, \vec{w}_2 \in W$. We must show that:

$$T^{-1}(\vec{w}_1 + \vec{w}_2) = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2).$$

But $T^{-1}(\vec{w}_1) = \vec{v}_1$ and $T^{-1}(\vec{w}_2) = \vec{v}_2$, where these two vectors satisfy the defining conditions: $T(\vec{v}_1) = \vec{w}_1$ and $T(\vec{v}_2) = \vec{w}_2$. But this means that:

$$T(\vec{v}_1 + \vec{v}_2) = T(\vec{v}_1) + T(\vec{v}_2) = \vec{w}_1 + \vec{w}_2,$$

since T itself is additive. By our previous reasoning, this means that $\vec{v}_1 + \vec{v}_2$ is the ***unique*** member of V with image $\vec{w}_1 + \vec{w}_2$. Thus:

$$T^{-1}(\vec{w}_1 + \vec{w}_2) = \vec{v}_1 + \vec{v}_2 = T^{-1}(\vec{w}_1) + T^{-1}(\vec{w}_2),$$

and thus T^{-1} is additive.

Homogeneity: Suppose $k \in \mathbb{R}$. We must show that:

$$T^{-1}(k \cdot \vec{w}) = k \cdot T^{-1}(\vec{w}).$$

Again, we know that $T^{-1}(\vec{w}) = \vec{v}$, with the defining condition that $T(\vec{v}) = \vec{w}$.

But this means that:

$$T(k \cdot \vec{v}) = k \cdot T(\vec{v}) = k \cdot \vec{w},$$

since T itself is homogeneous. Thus, $k \cdot \vec{v}$ is the ***unique*** member of V with image $k \cdot \vec{w}$, and so:

$$T^{-1}(k \cdot \vec{w}) = k \cdot \vec{v} = k \cdot T^{-1}(\vec{w}).$$

Hence, T^{-1} is also homogeneous.

Finally, we also have to show that T^{-1} is both one-to-one and onto:

T^{-1} is ***one-to-one:*** Suppose $T^{-1}(\vec{w}_1) = T^{-1}(\vec{w}_2)$. Using our notation above, we find $T^{-1}(\vec{w}_1) = \vec{v}_1$ and $T^{-1}(\vec{w}_2) = \vec{v}_2$. Thus $\vec{v}_1 = \vec{v}_2$. But now $T(\vec{v}_1) = T(\vec{v}_2)$, and this means $\vec{w}_1 = \vec{w}_2$. Thus T^{-1} is one-to-one.

T^{-1} is ***onto:*** Suppose $\vec{v} \in V$. We have to show that we can find $\vec{w} \in W$ such that $T^{-1}(\vec{w}) = \vec{v}$. But:

$$T^{-1}(T(\vec{v})) = \vec{v},$$

and thus $T(\vec{v}) = \vec{w} \in W$ is a vector mapped by T^{-1} to $\vec{v}$. Thus T^{-1} is onto.

Thus, T^{-1} is also a one-to-one and onto linear transformation, with the property that $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$. This proves one direction.

($\Longleftarrow$) Conversely, suppose that we can construct a linear transformation $T^{-1} : W \to V$ that is both one-to-one and onto, such that $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$. We must show that T itself is both one-to-one and onto.

T is ***one-to-one:*** Suppose $T(\vec{v}_1) = T(\vec{v}_2)$. Then $T^{-1}(T(\vec{v}_1)) = T^{-1}(T(\vec{v}_2))$. But $T^{-1} \circ T = \boldsymbol{I}_V$, and thus $\boldsymbol{I}_V(\vec{v}_1) = \boldsymbol{I}_V(\vec{v}_2)$, hence $\vec{v}_1 = \vec{v}_2$. Thus T is one-to-one.

T is ***onto:*** Suppose $\vec{w} \in W$. We must show that we can find $\vec{v} \in V$ such that $T(\vec{v}) = \vec{w}$. This time, we will use the composition $T \circ T^{-1}$, which our hypotheses says is the same as $\boldsymbol{I}_W$. Thus:

$$\vec{w} = \boldsymbol{I}_W(\vec{w}) = T \circ T^{-1}(\vec{w}) = T(T^{-1}(\vec{w})),$$

and so $T^{-1}(\vec{w}) = \vec{v} \in V$ is a vector mapped by T to $\vec{w}$. Thus, T is onto.■

Notice that T^{-1} exists whether or not V and W are finite dimensional. In this special case, though, we can find its matrix in a natural way:

Theorem: Suppose $T : V \to W$ is an isomorphism of ***finite dimensional*** vector spaces. By the previous Theorems, we know that $dim(V) = dim(W) = n$, say, and there exists $T^{-1} : W \to V$ such that $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$. If B is a ***basis*** for V and B' is a ***basis*** for W, then $[T]_{B,B'}$ is an ***invertible*** $n \times n$ matrix, and:

$$[T^{-1}]_{B',B} = [T]_{B,B'}^{-1}.$$

In particular, if $T : V \to V$ is an ***automorphism***, then:

$$[T^{-1}]_B = [T]_B^{-1}.$$

Proof: We know from the previous Section that the matrix of the composition of two transformations is the product of the matrices of each transformation, in the same order, using the appropriate bases. Thus:

$$[T^{-1}]_{B',B} \cdot [T]_{B,B'} = [T^{-1} \circ T]_{B,B} = [\boldsymbol{I_V}]_B \quad \text{and}$$
$$[T]_{B,B'} \cdot [T^{-1}]_{B',B} = [T \circ T^{-1}]_{B',B'} = [\boldsymbol{I_W}]_{B'}.$$

However, the matrix of the identity operator on ***any*** finite-dimensional vector space with respect to ***any*** of its bases is always the identity matrix, since by definition, $\boldsymbol{I_V}(\vec{v}) = \vec{v}$ for any $\vec{v} \in V$. Since $dim(V) = dim(W) = n$, we get:

$$[\boldsymbol{I_V}]_B = \boldsymbol{I_n} = [\boldsymbol{I_W}]_{B'},$$

and thus:

$$[T]_{B,B'} \cdot [T^{-1}]_{B',B} = \boldsymbol{I_n} = [T^{-1}]_{B',B} \cdot [T]_{B,B'}.$$

This shows that $[T]_{B,B'}$ is invertible, with inverse $[T^{-1}]_{B',B}$. ■

Example: Let $T : \mathbb{P}^3 \to \mathbb{R}^4$ be the linear transformation given by:

$$T(p(x)) = \langle\, p(-2),\, p(3),\, p'(-1),\, p''(1) \rangle.$$

It is easy to check that T is a linear transformation. Let us find its matrix with respect to the standard bases $B = \{1, x, x^2, x^3\}$ for $\mathbb{P}^3$ and similarly $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ for $\mathbb{R}^4$. Since we need to evaluate $p(x)$, $p'(x)$ and $p''(x)$ at the indicated points, we first create a table:

$p(x)$	$p'(x)$	$p''(x)$	$p(-2)$	$p(3)$	$p'(-1)$	$p''(1)$
1	0	0	1	1	0	0
x	1	0	-2	3	1	0
x^2	$2x$	2	4	9	-2	2
x^3	$3x^2$	$6x$	-8	27	3	6

We can now compute:

$$T(1) = \langle 1, 1, 0, 0 \rangle,$$
$$T(x) = \langle -2, 3, 1, 0 \rangle,$$
$$T(x^2) = \langle 4, 9, -2, 2 \rangle, \quad \text{and}$$
$$T(x^3) = \langle -8, 27, 3, 6 \rangle.$$

We assemble $[T]_{B,B'}$, column by column:

$$[T]_{B,B'} = \begin{bmatrix} 1 & -2 & 4 & -8 \\ 1 & 3 & 9 & 27 \\ 0 & 1 & -2 & 3 \\ 0 & 0 & 2 & 6 \end{bmatrix}.$$

Using technology, we find that this 4×4 matrix is invertible, and its inverse, which is $[T^{-1}]_{B',B}$, is:

$$[T]^{-1}_{B,B'} = [T^{-1}]_{B',B} = \frac{1}{50}\begin{bmatrix} 54 & -4 & 120 & 30 \\ -18 & 18 & -40 & -85 \\ -6 & 6 & -30 & -20 \\ 2 & -2 & 10 & 15 \end{bmatrix}.$$

We can now use this inverse matrix to accomplish the following:

> Find a polynomial $p(x)$, of degree at most 3, such that $p(-2) = 5$, $p(3) = -10$, $p'(-1) = 2$, and $c = 1$ is an inflection point.

This is equivalent to finding a polynomial $p(x)$ such that:

$$T(p(x)) = \langle 5, -10, 2, 0 \rangle.$$

(The second derivative of a cubic is a linear polynomial, so we are guaranteed a sign change whenever there is a zero; conversely, quadratics and linear functions do not have inflection points.) To find the coordinates of $p(x)$ with respect to B, we multiply:

$$\frac{1}{50}\begin{bmatrix} 54 & -4 & 120 & 30 \\ -18 & 18 & -40 & -85 \\ -6 & 6 & -30 & -20 \\ 2 & -2 & 10 & 15 \end{bmatrix}\begin{bmatrix} 5 \\ -10 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 11 \\ -7 \\ -3 \\ 1 \end{bmatrix}.$$

Decoding these coordinates with respect to B, we get:

$$p(x) = 11 - 7x - 3x^2 + x^3.$$

We can check algebraically that $p(-2) = 5$ and $p(3) = -10$. We get the derivatives: $p'(x) = -7 - 6x + 3x^2$ and $p''(x) = -6 + 6x$, and likewise see that $p'(-1) = 2$ and $p''(1) = 0$, and the 2nd derivative indeed experiences a sign change at $c = 1$, so this is indeed an inflection point. □

Applications in Calculus and Ordinary Differential Equations

We saw in the two previous Sections that in some cases, we can restrict the derivative transformation D to a finite dimensional function space $W = Span(B)$, and the derivatives are again in W. We say in this case that D ***preserves*** W, and so D is an ***operator:***

$$D : W \to W.$$

If $[D]_B$ is an invertible matrix, then D is an invertible operator. However, this means that D^{-1} gives an an ***antiderivative*** for every function $f(x)$ from W.

Example: We saw in the previous Section that the function space:

$$W = Span(B), \text{ where } B = \{x^2e^{-3x}, xe^{-3x}, e^{-3x}\},$$

is ***preserved*** by the derivative operation D, in other words, D is an ***operator:***

$$D : W \to W.$$

We also found the matrix of D with respect to B:

$$[D]_B = \begin{bmatrix} -3 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 1 & -3 \end{bmatrix}.$$

Since this lower triangular matrix is invertible, D is an isomorphism on W. In other words, D is an ***automorphism***. We can find $[D^{-1}]_B$ via:

$$[D^{-1}]_B = [D]_B^{-1} = \begin{bmatrix} -3 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 1 & -3 \end{bmatrix}^{-1} = -\frac{1}{27}\begin{bmatrix} 9 & 0 & 0 \\ 6 & 9 & 0 \\ 2 & 3 & 9 \end{bmatrix}$$

Thus, we want to find $\int(5x^2e^{-3x} - 8xe^{-3x} + 2e^{-3x})dx$, we encode the coefficients $\langle 5, -8, 2\rangle$ in a column matrix, perform the matrix product:

$$-\frac{1}{27}\begin{bmatrix} 9 & 0 & 0 \\ 6 & 9 & 0 \\ 2 & 3 & 9 \end{bmatrix}\begin{bmatrix} 5 \\ -8 \\ 2 \end{bmatrix} = \begin{bmatrix} -\frac{5}{3} \\ \frac{14}{9} \\ -\frac{4}{27} \end{bmatrix}$$

and decode the coordinates for our answer (adding an arbitrary constant C, as usual), to get:

$$\int(5x^2e^{-3x} - 8xe^{-3x} + 2e^{-3x})dx = -\frac{5}{3}x^2e^{-3x} + \frac{14}{9}xe^{-3x} - \frac{4}{27}e^{-3x} + C.$$

You might recall that to find this antiderivative directly, we would need to apply a technique called ***Integration by Parts***. This Example shows how to do it instead using a matrix!□

Let us extend this idea to solve a special kind of ***differential equation***, that is, an equation involving one or more derivatives:

Definition: Let x be an independent variable, and y a variable that depends on x. An ***ordinary linear differential equation*** is an equation of the form:

$$c_n y^{(n)} + \cdots + c_2 y^{(2)} + c_1 y' + c_0 y = g(x),$$

for some positive integer n, scalars $c_0, c_1, \ldots, c_n$, and function $g(x)$.

A ***solution*** to such an equation is a function $y = f(x)$ defined on some interval I, that satisfies the differential equation. The word "ordinary" refers to the appearance of only ordinary derivatives from basic Calculus (as opposed to partial derivatives that appear in Multi-Variable Calculus; an equation involving partial derivatives is naturally referred to as a ***partial differential equation*** or P.D.E.).

STEM majors often take a separate course on Differential Equations, but we will see below how to use the concept of a linear transformation, and in particular an invertible transformation, in order to solve these kinds of differential equations.

Example: Let us consider the differential equation:

$$2y'' - 5y' + 4y = 185x^2e^{-3x} - 281xe^{-3x} - 188e^{-3x}.$$

We want to find ***one*** solution to this equation (the process of finding ***all*** solutions to this differential equation is more difficult, and is treated more appropriately in a full-term course in Differential Equations). Since the function on the right is a member of the vector space:

$$W = Span(B), \text{ where } B = \{x^2e^{-3x}, xe^{-3x}, e^{-3x}\},$$

it is natural to ***guess*** that we will find a solution to this o.d.e. also in W. Since W is preserved by D, and thus by D^2 as well, this further gives us hope that W contains at least one solution.

Now that we have a guess for the space to work with, we can think of the left side of this equation as the ***operator:***

$$T : W \to W, \quad \text{given by:}$$

$$T(f(x)) = 2f''(x) - 5f'(x) + 4f(x).$$

Notice that $T = 2D^2 - 5D + 4\boldsymbol{I}_W$, a ***linear combination*** of the operators $\boldsymbol{I}_W$, D and D^2, and thus we are certain that T is in fact a linear transformation. Thus, we can find $[T]_B$ using $\boldsymbol{I}_3$, $[D]_B$ and $[D^2]_B = [D]_B^2$ which we found in an Example in Section 3.7:

$$[T]_B = 2[D^2]_B - 5[D]_B + 4\boldsymbol{I}_3$$

$$= 2\begin{bmatrix} 9 & 0 & 0 \\ -12 & 9 & 0 \\ 2 & -6 & 9 \end{bmatrix} - 5\begin{bmatrix} -3 & 0 & 0 \\ 2 & -3 & 0 \\ 0 & 1 & -3 \end{bmatrix} + 4\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 37 & 0 & 0 \\ -34 & 37 & 0 \\ 4 & -17 & 37 \end{bmatrix}.$$

Notice that the final matrix is again lower triangular. It is invertible, with inverse:

$$[T^{-1}]_B = [T]_B^{-1} = \frac{1}{50653}\begin{bmatrix} 1369 & 0 & 0 \\ 1258 & 1369 & 0 \\ 430 & 629 & 1369 \end{bmatrix}.$$

Since we want $T(f(x)) = 185x^2e^{-3x} - 281xe^{-3x} - 188e^{-3x}$, we get:

$$f(x) = T^{-1}(185x^2e^{-3x} - 281xe^{-3x} - 188e^{-3x}).$$

Using the inverse matrix and the desired output encoded in a column matrix, we compute:

$$\frac{1}{50653}\begin{bmatrix} 1369 & 0 & 0 \\ 1258 & 1369 & 0 \\ 430 & 629 & 1369 \end{bmatrix}\begin{bmatrix} 185 \\ -281 \\ -188 \end{bmatrix} = \frac{1}{50653}\begin{bmatrix} 253,265 \\ -151,959 \\ -354,571 \end{bmatrix} = \begin{bmatrix} 5 \\ -3 \\ -7 \end{bmatrix}.$$

Thus, one solution to our differential equation is:

$$f(x) = 5x^2e^{-3x} - 3xe^{-3x} - 7e^{-3x}. \square$$

You might be wondering — doesn't this computation show that the function we found is the ***only*** solution to this differential equation? The answer is ***no***, because we looked for a solution in only one function space W. To find all solutions, we need to extend our transformation above to the entire space of twice differentiable functions defined on all real numbers:

$$T : C^2(\mathbb{R}) \to C^0(\mathbb{R}),$$

and thus there may be ***other*** solutions that are members of this $C^2(\mathbb{R})$, and not just the subspace W that we considered. However, as before, any two solutions will differ by a member of the ***kernel*** of T, and finding the members of this kernel is one of the tasks of a full-term course in Differential Equations.

It is also crucial in this process that we pick the appropriate function space W to work with. The function $g(x)$ on the right side of the differential equation should point us in the right direction. It is important in the process that W is preserved by the first, and possibly second and higher derivatives.

Polynomial Curve Fitting

We can use the idea of a linear transformation to find polynomials (or possibly other functions for that matter, but there are no guarantees for functions that are not polynomials) that pass through certain points.

We know from basic algebra that two distinct points determine a unique line. Since we want our lines to represent a polynomial ***function***, though, we will insist that all the points that we deal with have ***distinct*** x-coordinates. Similarly, three non-collinear points will determine a unique parabolic function $p(x) = ax^2 + bx + c$. If the points are collinear, we get a "degenerate" quadratic $p(x) = bx + c$ or a constant polynomial $p(x) = c$, but notice that all these polynomials are still members of $\mathbb{P}^2$. Continuing with this analogy, ***four points*** with distinct x-coordinates will determine a unique polynomial of ***at most third degree***, in other words, a member of $\mathbb{P}^3$, and so on. In fact, we proved in Exercise 28 of Section 3.7 that the ***evaluation homomorphism:***

$$E_{\vec{a}} : \mathbb{P}^n \to \mathbb{R}^{n+1}, \quad \text{where:}$$

$$E_{\vec{a}}(p(x)) = \langle p(a_1), p(a_2), \ldots, p(a_n), p(a_{n+1}) \rangle,$$

is a ***one-to-one*** function, if the a_i are distinct numbers. By the Dimension Theorem, and the fact that $dim(\mathbb{P}^n) = n + 1 = dim(\mathbb{R}^{n+1})$, $E_{\vec{a}}$ is also ***onto***, and thus is ***invertible***.

Example: Let us find a cubic polynomial that passes through the points:

$$(-4, -198),\ (-1, 102),\ (2, -48) \text{ and } (3, -58).$$

We will do this by constructing the evaluation transformation:

$$E_{\vec{a}} : \mathbb{P}^3 \to \mathbb{R}^4,$$

for the vector $\vec{a} = \langle -4, -1, 2, 3 \rangle$, and then we look for the unique member of $\mathbb{P}^3$ whose image under $E_{\vec{a}}$ is $\langle -198, 102, -48, -58 \rangle$.

We can use the standard basis $B = \{1, x, x^2, x^3\}$ for $\mathbb{P}^3$ and similarly the standard basis $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ for $\mathbb{R}^4$. We compute:

$$\begin{aligned} E_{\vec{a}}(1) &= \langle 1, 1, 1, 1 \rangle, \\ E_{\vec{a}}(x) &= \langle -4, -1, 2, 3 \rangle, \\ E_{\vec{a}}(x^2) &= \langle 16, 1, 4, 9 \rangle, \text{ and} \\ E_{\vec{a}}(x^3) &= \langle -64, -1, 8, 27 \rangle. \end{aligned}$$

We assemble the matrix:

$$[E_{\vec{a}}]_{B,B'} = \begin{bmatrix} 1 & -4 & 16 & -64 \\ 1 & -1 & 1 & -1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{bmatrix}.$$

Our discussion preceding this Example tells us that this should be an invertible matrix. Indeed, its inverse is:

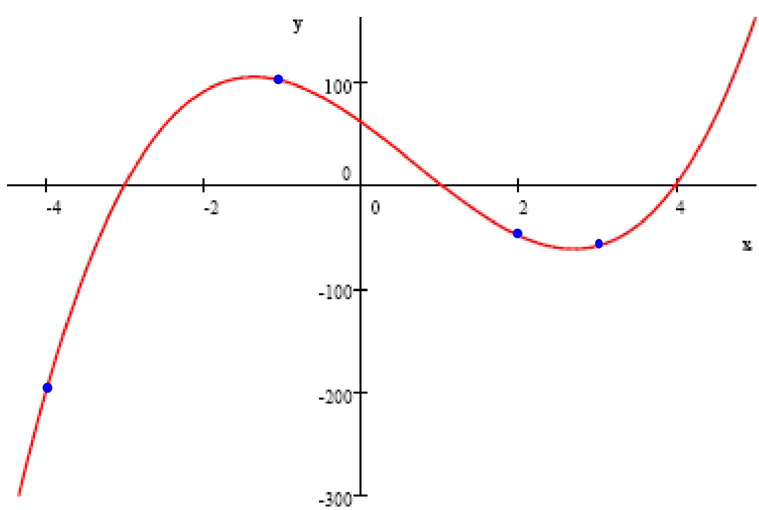
$$[E_{\vec{a}}^{-1}]_{B',B} = \frac{1}{252}\begin{bmatrix} -12 & 168 & 168 & -72 \\ -2 & -98 & 154 & -54 \\ 8 & -7 & -28 & 27 \\ -2 & 7 & -14 & 9 \end{bmatrix},$$

which we can find using technology. Thus, there is exactly one cubic polynomial that passes through the given four points. We can find it by assembling the desired y-coordinates in a coordinate matrix and performing the matrix product with this inverse:

$$\frac{1}{252}\begin{bmatrix} -12 & 168 & 168 & -72 \\ -2 & -98 & 154 & -54 \\ 8 & -7 & -28 & 27 \\ -2 & 7 & -14 & 9 \end{bmatrix}\begin{bmatrix} -198 \\ 102 \\ -48 \\ -58 \end{bmatrix} = \begin{bmatrix} 62 \\ -55 \\ -10 \\ 5 \end{bmatrix}.$$

Using our ordered basis for $\mathbb{P}^3$ to decode this coordinate matrix, the actual polynomial we are looking for is:

$$p(x) = 5x^3 - 10x^2 - 55x + 62.$$



We can see from its graph above that $p(x)$ passes through the points $(-4,-198)$, $(-1,102)$, $(2,-48)$ and $(3,-58)$. □

3.8 Section Summary

We say that $T : V \to W$ is an ***isomorphism*** if T is both ***one-to-one*** and ***onto***. We also say that T is ***invertible***, T is ***bijective***, and the vector spaces V and W are ***isomorphic*** to each other.

Two ***finite dimensional*** vector spaces V and W are isomorphic to each other ***if and only if*** $dim(V) = dim(W)$. The converse will be proven in the Exercises.

A linear transformation $T : V \to W$ is an isomorphism of vector spaces ***if and only if*** there exists a linear transformation $T^{-1} : W \to V$, which is also an isomorphism, such that: $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$, where $\boldsymbol{I}_V$ and $\boldsymbol{I}_W$ are the identity operators on V and W, respectively.

Suppose $T : V \to W$ is an isomorphism of finite dimensional vector spaces. By the Theorems above, we know that $dim(V) = dim(W) = n$, say, and there exists a linear transformation $T^{-1} : W \to V$ such that $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$. If B is a basis for V and B' is a basis for W, then $[T]_{B,B'}$ is an ***invertible*** $n \times n$ matrix, and $[T^{-1}]_{B',B} = [T]^{-1}_{B,B'}$.

Isomorphisms can be used to solve ordinary linear differential equations or to find polynomials that pass through given points or possess certain attributes (when applicable).

3.8 Exercises

1. Let $T : \mathbb{P}^2 \to \mathbb{R}^3$ be the linear transformation given by:

$$T(p(x)) = \langle p(-3), p(5), p'(2) \rangle,$$

and let $B = \{1, x, x^2\} \subset \mathbb{P}^2$, and $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ be the standard bases for $\mathbb{P}^2$ and $\mathbb{R}^3$.

   a. Verify that T is indeed a linear transformation and find $[T]_{B,B'}$.
   b. Prove that T is an isomorphism by finding the inverse of this matrix.
   c. Use (b) to find a polynomial $p(x)$ of degree at most 2 that passes through $(-3, 75)$ and $(5, 99)$, and with $p'(2) = 13$.

2. Let $T : \mathbb{P}^3 \to \mathbb{R}^4$ be the linear transformation given by:

$$T(p(x)) = \langle p(-4), p(1), p(3), p'(-1) \rangle,$$

and let $B = \{1, x, x^2, x^3\} \subset \mathbb{P}^3$, and $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ the standard bases for $\mathbb{P}^3$ and $\mathbb{R}^4$.

   a. Verify that T is indeed a linear transformation and find $[T]_{B,B'}$.
   b. Prove that T is an isomorphism by finding the inverse of this matrix.
   c. Use (b) to find a polynomial of degree at most 3 passing through $(-4, -247)$, $(1, -7)$ and $(3, 19)$, and with $p'(-1) = 23$.

For Exercises (3) to (9): adapt the ideas in Exercises (1) and (2) in order to construct a linear transformation $T : \mathbb{P}^n \to \mathbb{R}^{n+1}$ (for an appropriate $\mathbb{P}^n$), construct the matrix for T with respect to the standard basis of each space, find the inverse of this matrix, and use it to find the polynomials with the indicated properties. Use technology if allowed by your instructor to invert the 4×4 matrices.

3. Find a polynomial $p(x)$ of degree at most 2 that passes through the points:
   a. $(-2, 52)$ and $(4, 58)$, and with $p'(3) = 21$.
   b. $(-2, -13)$ and $(4, -25)$, and with $p'(3) = -14$.

4. Find a polynomial $p(x)$ of degree at most 2 that passes through the point:
   a. $(3,83)$, with $p'(-4) = -77$, and $\int_0^1 p(x)dx = 35/2$.
   b. $(3,-106)$, with $p'(-4) = 45$, and $\int_0^1 p(x)dx = 65/6$.

5. Find a polynomial $p(x)$ of degree at most 2 that passes through the point:
   a. $(5,-58)$, with $p'(-2) = 25$ and $p'(7) = -47$.
   b. $(5,324)$, with $p'(-2) = -68$, and $p'(7) = 202$.
   c. $(5,53)$, with $p'(-2) = 12$, and $p'(7) = 12$ also. Explain what happened.

6. Find a polynomial $p(x)$ of degree at most 3 that passes through the points:
   a. $(-5,851)$, $(-2,89)$, and $(3,-61)$, and with $p'(2) = -31$.
   b. $(-5,-349)$, $(-2,-55)$, and $(3,-85)$, and with $p'(2) = -45$.

7. Find a polynomial $p(x)$ of degree at most 3 that passes through the points:
   a. $(-3,152)$ and $(2,47)$, with $p'(-4) = -269$ and $p'(5) = -161$.
   b. $(-3,-532)$ and $(2,148)$, with $p'(-4) = 868$ and $p'(5) = 1237$.

8. Find a polynomial $p(x)$ of degree at most 3 that passes through the points:
   a. $(-4,815)$ and $(7,-2474)$, with $p'(-6) = -1133$ and $p''(9) = -460$.
   b. $(-4,-188)$ and $(7,1275)$, with $p'(-6) = 417$ and $p''(9) = 216$.

9. Find a polynomial $p(x)$ of degree at most 3 that passes through the point:
   a. $(6,2185)$, with $p'(-8) = 1616$, $p''(-3) = -148$, and $\int_0^1 p(x)dx = -77/12$.
   b. $(6,2277)$, with $p'(-8) = 2094$, $p''(-3) = -198$, and $\int_0^1 p(x)dx = 11/4$.

10. Find a polynomial $p(x)$ of degree at most 3, with:
   a. $p'(0) = -11$, $p''(7) = 10$, $\int_0^1 p(x)dx = \dfrac{26}{3}$, and $\int_{-2}^0 p(x)dx = \dfrac{88}{3}$.
   b. $p'(0) = 9$, $p''(7) = -\dfrac{1001}{2}$, $\int_0^1 p(x)dx = -\dfrac{11}{12}$, and $\int_{-2}^0 p(x)dx = \dfrac{86}{3}$.

For Exercises (11) to (16): We saw the following sets of functions B and spaces $W = Span(B)$ in the Exercises of Sections 3.5 and 3.6. We saw in Section 3.5 that the derivative operator D preserves W, and we constructed $[D]_B$ in Section 3.6. The corresponding Exercise numbers are indicated for your reference. (a) Check your homework solutions and the Answer Key for $[D]_B$;
(b) Show that D is invertible on W by finding $[D]_B^{-1}$;
(c) Use (b) to find the indicated antiderivative.

11. $B = \{e^{-3x}\sin(2x), e^{-3x}\cos(2x)\}$; Exercise 13, Section 3.5 and Exercise 13, Section 3.6.
    Find the antiderivative: $\int (-11e^{-3x}\sin(2x) + 29e^{-3x}\cos(2x))\,dx$.

12. $B = \{xe^{5x}, e^{5x}\}$; Exercise 14, Section 3.5 and Exercise 14, Section 3.6.
    Find the antiderivative: $\int (15xe^{5x} + 43e^{5x})\,dx$.

13. $B = \{x^2e^{-4x}, xe^{-4x}, e^{-4x}\}$; Exercise 15, Section 3.5 and Exercise 15, Section 3.6.
    Find the antiderivative: $\int (-16x^2e^{-4x} + 44xe^{-4x} + 3e^{-4x})\,dx$.

14. $B = \{x^2 \cdot 5^x, x \cdot 5^x, 5^x\}$; Exercise 16, Section 3.5, and Exercise 16 of Section 3.6.
    Find the general antiderivative: $\int (7x^2 \cdot 5^x - 4x \cdot 5^x + 9 \cdot 5^x)\,dx$.

15. $B = \{x\sin(2x),\ x\cos(2x),\ \sin(2x),\ \cos(2x)\}$;
Exercise 18, Section 3.5, and Exercise 18 of Section 3.6.
Find the general antiderivative: $\int(4x\sin(2x) + 9x\cos(2x) - 5\sin(2x) + 8\cos(2x))\,dx$.

16. Let $B = \{e^{ax}\sin(bx),\ e^{ax}\cos(bx)\}$, where a and b are ***non-zero*** scalars.
Exercise 21, Section 3.5, and Exercise 21, Section 3.6.
Find the two general antiderivatives: $\int e^{ax}\sin(bx)\,dx$ and $\int e^{ax}\cos(bx)\,dx$.

17. Use the inverse of the matrix in the subsection on Applications in Calculus and Ordinary Differential Equations to find a solution to the differential equation:

$$2y'' - 5y' + 4y = -74x^2e^{-3x} + 364xe^{-3x} - 33e^{-3x}.$$

For Exercises (18) to (31): For each of the following linear ordinary linear differential equations: (a) choose an appropriate function space $W = Span(B)$ where you are likely to find a solution to the equation; (b) construct a linear transformation $T : W \to W$ that represents T; (c) find $[T]_B$; (d) verify that $[T]_B$ is invertible, and find its inverse; (e) use the inverse to find a solution $y = f(x)$ to the equation. Note that we already have the matrix for D and its powers for some of the W that appear below in Exercises 17 to 23 of Section 3.7. Consult the Answer Key.

18. $-2y'' + 5y' + 3y = -\dfrac{49}{3} + \dfrac{29}{3}x + 5x^2$

19. $-2y'' + 5y' + 3y = 1215 - 4158x + 189x^2 + 486x^3$

20. $3y'' + 8y' - 7y = 64\sin(x) - 166\cos(x)$

21. $-2y''' - 4y'' + 3y' + 8y = 25\sin(x) + 109\cos(x)$

22. $3y'' + 8y' - 7y = 319\sin(2x) + 186\cos(2x)$

23. $-2y''' - 4y'' + 3y' + 8y = 319\sin(2x) + 186\cos(2x)$

24. $-9y'' + 5y' + 4y = -2250e^{-3x}\sin(2x) + 1390e^{-3x}\cos(2x)$

25. $3y''' + 7y'' + 2y' - 6y = 134e^{-3x}\sin(2x) + 390e^{-3x}\cos(2x)$

26. $2y'' - 9y' + 4y = 36xe^{5x} - 19e^{5x}$

27. $2y''' - 7y'' - 3y' + 4y = -576xe^{5x} + 139e^{5x}$

28. $3y'' + 11y' + 8 = 36x^2e^{-4x} + 6xe^{-4x} - 13e^{-4x}$

29. $3y''' + 4y'' - 8y' + 11 = -170x^2e^{-4x} - 349xe^{-4x} + 213e^{-4x}$

30. $4y'' + 9y' - 8 = 23\sinh(3x) + 32\cosh(3x)$.
Recall that $D(\sinh(x)) = \cosh(x)$ and $D(\cosh(x)) = \sinh(x)$. Don't forget the Chain Rule.

31. $3y'' + 4y' + 6 = 2x\sin(2x) - 86x\cos(2x) - 64\sin(2x) + 4\cos(2x)$

32. Let $B = \{e^{ax}\sin(bx),\ e^{ax}\cos(bx)\}$, and $W = Span(B)$. Suppose that D is the derivative operator on W. In Exercise 24 of Section 3.7, we showed that for all positive integers k:

$$[D^k]_B = \begin{bmatrix} a_k & -b_k \\ b_k & a_k \end{bmatrix},$$

for some real numbers a_k and b_k.

a. Show that a matrix which has the form above is always invertible if ***either*** a_k or b_k is non-zero.

b. Show that any linear combination of the matrices $[D]_B$, $[D^2]_B$, ..., $[D^k]_B$ also has the same form as the matrix above.

c. Use (b) to show that the linear differential equation:

$$c_n y^{(n)} + c_{n-1} y^{(n-1)} + \cdots + c_2 y^{\prime\prime} + c_1 y^{\prime} + c_0 y = d_1 e^{ax}\sin(bx) + d_2 e^{ax}\cos(bx)$$

always has a solution in W, for any real numbers d_1, d_2, c_0, c_1, ..., c_n, as long as transformation representing the left side of the equation is not the zero transformation on W.

d. Is it true in general that the linear combination of invertible matrices is again invertible, as long as the sum is a non-zero matrix?

33. Use the inverse of the matrix in the subsection on Applications in Polynomial Curve Fitting to find a cubic polynomial passing through the four points:

$$(-4,-116),\ (-1,-31/2),\ (2,-14)\ \text{and}\ (3,-43/2).$$

34. Find a quadratic polynomial $p(x)$ passing through the three points:

$$(-3,-147/2),\ (1,29/2),\ \text{and}\ (2,29).$$

35. Find a quadratic polynomial $p(x)$ passing through the three points:

$$(-4,185),\ (-1,32),\ \text{and}\ (5,158).$$

For Exercises (36) to (38): Use technology to find the inverse of $E_{\vec{a}}$ in order to find $p(x)$, if allowed by your instructor.

36. Find a cubic polynomial $p(x)$ passing through the four points:

$$(-3,-187),\ (1,5),\ (2,33)\ \text{and}\ (4,359).$$

37. Find a cubic polynomial $p(x)$ passing through the four points:

$$(-4,435),\ (-1,6),\ (2,-45)\ \text{and}\ (3,-174).$$

38. Find a quartic (degree 4) polynomial $p(x)$ passing through the five points:

$$(-5,-5176),\ (-2,-169),\ (3,-504),\ (6,-8289),\ (8,-26639).$$

39. Suppose that $T : \mathbb{P}^2 \to \mathbb{R}^3$ is a linear transformation, and the matrix of T with respect to the bases $B = \{2, 3-x, 5+7x+x^2\}$ for $\mathbb{P}^2$ and $B^{\prime} = \{\langle 1,-2,5\rangle, \langle 0,-1,4\rangle, \langle 0,0,-3\rangle\}$ for $\mathbb{R}^3$ is:

$$[T]_{B,B^{\prime}} = Diag(3, 1/2, -5).$$

a. Explain why T is an isomorphism.

b. Compute $T(4-8x+3x^2)$. Express your answer as a vector in $\mathbb{R}^3$. Reminder: Encode, Multiply, Decode.

c. Find $[T^{-1}]_{B^{\prime},B}$.

d. Use (c) to find $T^{-1}(\langle 5,-4,7\rangle)$.

40. Repeat the previous Exercise if: $[T]_{B,B^{\prime}} = \begin{bmatrix} -2 & -5 & 7 \\ 0 & 1/3 & 2 \\ 0 & 0 & -1 \end{bmatrix}$.

You may use part of your solutions to (b) and (d) from the previous Exercise.

41. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^2$ is an operator whose matrix with respect to the basis $B = \{2, 5 - x, 2 + 3x - x^2\}$ is given by:

$$[T]_B = \begin{bmatrix} -2 & 1 & 2 \\ 3 & 2 & -1 \\ -1 & -1 & 3 \end{bmatrix}.$$

a. Prove that T is invertible by finding the inverse of this matrix.

b. Use the inverse of this matrix to find a quadratic polynomial $p(x)$, such that:

$$T(p(x)) = 81 - 134x + 33x^2.$$

Reminder: Encode, Multiply, Decode.

42. ***From Kansas to Oz . . . and Back!*** Show that: $T : \mathbb{R} \to \mathbb{R}^+$, given by: $T(x) = e^x$ is an isomorphism. What is the name of its inverse?

43. Let V be any vector space (it can be infinite dimensional). Show that V is isomorphic to ***itself***. Hint: state the definition of any isomorphism $T : V \to W$ and rewrite it to say what it means if $V = W$.

44. Let V, U and W be any three vector spaces (they can be infinite dimensional). Suppose that $T_1 : V \to U$ is a linear transformation, and $T_2 : U \to W$ is also a linear transformation.

a. Prove that if T_1 and T_2 are both one-to-one, then $T_2 \circ T_1$ is also one-to-one.

b. Prove that if T_1 and T_2 are both onto, then $T_2 \circ T_1$ is also onto.

c. Prove that if T_1 and T_2 are both isomorphisms, then $T_2 \circ T_1$ is also an isomorphism.

d. Conclude that if V is isomorphic to U and U is isomorphic to W, then V is isomorphic to W. Thus, all three spaces are isomorphic to each other.

45. Show that $T : Mat(m,n) \to Mat(n,m)$, where: $T(A) = A^\top$, is an isomorphism.

46. Let B be a ***fixed*** basis for a finite-dimensional vector space V. Show that:

$$T : V \to \mathbb{R}^n, \quad \text{given by:} \quad T(\vec{v}) = \langle \vec{v} \rangle_B,$$

that is, the coordinates of $\vec{v}$ with respect to B, is an ***isomorphism***. Note: since we already know from Section 3.6 that T is a linear transformation, we only have to show that T is both ***one-to-one*** and ***onto***.

47. Suppose that V and W are vector spaces, with $dim(V) = n$ and $dim(W) = m$, and suppose that B is a fixed basis for V and B' is a fixed basis for W. Consider now the vector spaces:

$$X = Hom(V, W) = \{ T \mid T : V \to W \text{ is a linear transformation} \}, \text{ and}$$
$$Y = Mat(m,n) = \{ A \mid A \text{ is an } m \times n \text{ matrix} \}.$$

a. Show that the function: $S : X \to Y$, given by: $S(T) = [T]_{B,B'}$,

which takes a linear transformation T and creates its matrix with respect to B and B', is a linear transformation.

b. Show that S is in fact an isomorphism, and thus: $Hom(V, W) \cong Mat(m,n)$.

c. Explain in words what this isomorphism means.

48. We know that if $T : V \to V$ is an ***operator*** on a ***finite-dimensional*** vector space V, then T is ***one-to-one*** ***if and only if*** T is ***onto***, thanks to the Dimension Theorem. The purpose of this

Exercise is to show that this statement may be ***false*** if V is an infinite-dimensional vector space. Consider the vector space $\mathbb{P}$ consisting of ***all polynomials***. We know from Section 3.4 that $dim(\mathbb{P}) = \aleph_0$. Now, consider the function:

$$T : \mathbb{P} \to \mathbb{P}, \quad \text{given by:} \quad T(p(x)) = x \cdot p(x).$$

a. Show that T is indeed a linear transformation.

b. Show that T is one-to-one.

c. Show that T is ***not*** onto. Hint: this means that there exists at least one polynomial $q(x)$ which is not in $range(T)$.

d. Now, let D be the ***derivative*** operator. Show that $\mathbb{P}$ is preserved by D.

e. Show that D is ***onto***.

f. Show that D is ***not*** one-to-one. Hint: what is $ker(D)$?

49. The objective of this Exercise is to prove the Theorem: Two ***finite dimensional*** vectors spaces V and W are isomorphic to each other ***if and only if*** $dim(V) = dim(W)$.

a. Warm-up: Construct a simple isomorphism from $\mathbb{R}^4$ to $\mathbb{P}^3$. Hint: use the standard bases.

b. Use the Dimension Theorem to prove that if $T : V \to W$ is an isomorphism, then $\dim(V) = \dim(W)$.

For the converse: Suppose $dim(V) = dim(W) = n$. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be an ordered basis for V, and let $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be an ordered basis for W. Define a function:

$$T : B \to B' \quad \text{given by:} \quad T(\vec{v}_i) = \vec{w}_i \ \text{ for } i = 1..n.$$

c. Show using linearity that T can be extended to a linear transformation:

$$T : V \to W,$$

on the entire vector space V, and not just the basis vectors. In other words, show how to define $T(\vec{v})$ for an arbitrary vector $\vec{v} \in V$. Hint: use Uniqueness of Representation.

d. Show that the linear transformation T in (c) is both ***one-to-one*** and ***onto***, and thus T is an ***isomorphism***.

50. Use the previous Exercise to prove the following:

a. Any two ***lines*** L_1 and L_2 passing through the origin of $\mathbb{R}^2$ are isomorphic to each other. Use a similar argument for any two lines passing through the origin in $\mathbb{R}^3$.

b. More generally, if L_1 is a line in $\mathbb{R}^n$ passing through the origin, and L_2 is a line in $\mathbb{R}^m$ passing through the origin, then L_1 and L_2 are isomorphic to each other (even though they are in different ambient spaces).

c. Any two ***planes*** Π_1 and Π_2 passing through the origin of $\mathbb{R}^3$ are isomorphic to each other.

d. The vector space $Diag(n)$ of ***diagonal*** $n \times n$ matrices, and Euclidean ***n-space*** $\mathbb{R}^n$, are isomorphic to each other.

e. The vector space $Upper(n)$ of ***upper triangular*** $n \times n$ matrices and the space $Lower(n)$ of ***lower triangular*** $n \times n$ matrices are isomorphic to each other. What would be a simple isomorphism connecting these two spaces?

f. Both $Upper(n)$ and $Lower(n)$, as in the previous part, and the space $Sym(n)$ of ***symmetric*** $n \times n$ matrices, are all isomorphic to each other.

g. Suppose that $n = 2k - 1$, where k is a positive integer. Show that the vector space $Bisym(n)$ of ***bisymmetric*** $n \times n$ matrices, and $Mat(k,k)$, the space of ***all*** $k \times k$ matrices, are isomorphic to each other.

A Summary of Chapter 3

We call $(V, \oplus, \odot)$ a ***vector space*** if V is a non-empty set, and the operations $\oplus$ and $\odot$ satisfy the Ten Axioms for a Vector Space: for all $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$ and all $r, s \in \mathbb{R}$, $(V, \oplus, \odot)$ satisfies:

1. $\vec{u} \oplus \vec{v} \in V$; 2. $r \odot \vec{u} \in V$; 3. $\vec{u} \oplus \vec{v} = \vec{v} \oplus \vec{u}$; 4. $(\vec{u} \oplus \vec{v}) \oplus \vec{w} = \vec{u} \oplus (\vec{v} \oplus \vec{w})$;
5. There exists $\vec{\mathbf{0}}_V \in V$, such that: $\vec{\mathbf{0}}_V \oplus \vec{v} = \vec{v} = \vec{v} \oplus \vec{\mathbf{0}}_V$;
6. There exists $-\vec{v} \in V$ such that: $\vec{v} \oplus (-\vec{v}) = \vec{\mathbf{0}}_V = (-\vec{v}) \oplus \vec{v}$;
7. $(r+s) \odot \vec{v} = (r \odot \vec{v}) \oplus (s \odot \vec{v})$; 8. $r \odot (\vec{u} \oplus \vec{v}) = (r \odot \vec{u}) \oplus (r \odot \vec{v})$;
9. $r \odot (s \odot \vec{v}) = s \odot (r \odot \vec{v}) = (rs) \odot \vec{v}$; 10. $1 \odot \vec{v} = \vec{v}$.

When the operations $\oplus$ and $\odot$ are understood, we write V instead of $(V, \oplus, \odot)$.

The ***Span*** of a ***finite*** set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset (V, \oplus, \odot)$ is the set of all possible ***linear combinations*** from S: ***Span***$(S) = \{(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) \,|\, c_1, c_2, \ldots, c_n \in \mathbb{R}\}$.

S is ***linearly independent*** if the only solution to: $(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) = \vec{\mathbf{0}}_V$, the ***dependence test equation***, is the ***trivial solution*** $c_1 = 0, c_2 = 0, \ldots, c_n = 0$.

Suppose that $S = \{\vec{v}_i \,|\, i \in I\} \subset (V, \oplus, \odot)$ is an ***infinite*** set of vectors, where $I \subset \mathbb{R}$ is a non-empty ***indexing set***. A ***linear combination*** of vectors from S can be constructed in the following way:

1. Choose a ***finite subset*** of vectors: $\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_n}\} \subset S$, where $i_1 < i_2 < \cdots < i_n \in I$.
2. Choose a finite list of scalars $c_1, c_2, \ldots, c_n \in \mathbb{R}$.
3. Form the vector expression: $(c_1 \odot \vec{v}_{i_1}) \oplus (c_2 \odot \vec{v}_{i_2}) \oplus \cdots \oplus (c_n \odot \vec{v}_{i_n})$.

Span(S) is the set of ***all possible linear combinations*** of vectors from ***all finite subsets*** of S.

S is ***linearly independent*** if every ***finite*** subset of S is linearly independent. In other words, the only solution to the ***dependence test equation:*** $c_1 \odot \vec{v}_{i_1} \oplus c_2 \odot \vec{v}_{i_2} \oplus \cdots \oplus c_n \odot \vec{v}_{i_n} = \vec{\mathbf{0}}_V$,
is the trivial solution: $c_1 = 0, c_2 = 0, \ldots, c_n = 0$, for ***all*** indices $i_1 < i_2 < \cdots < i_n \in I$.

The Elimination Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset (V, \oplus, \odot)$ be ***linearly dependent***, with ***dependence equation:*** $(c_1 \odot \vec{v}_1) \oplus (c_2 \odot \vec{v}_2) \oplus \cdots \oplus (c_n \odot \vec{v}_n) = \vec{\mathbf{0}}_V$, where $c_n \neq 0$.
Then: $Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}) = Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-1}\})$.

The Extension Theorem: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\} \subset (V, \oplus, \odot)$ be ***linearly independent***, and let $\vec{w} \in V$ be a vector which is ***not*** in $Span(S)$.
Then: the ***enlarged*** set $S' = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n, \vec{w}\}$ is ***still*** linearly independent.

A non-empty subset $W \subset V$ is a ***subspace*** of V if W is ***closed*** under $\oplus$ and $\odot$. We write: $W \trianglelefteq V$.

Suppose S is a non-empty subset of a vector space V. Then: $Span(S)$ is a ***subspace*** of V.

A set of vectors $B \subset V$ is a ***basis*** for V if it is ***linearly independent*** and ***Spans*** V.

Every vector space V has a basis B. V is ***finite dimensional*** if we can find a ***finite basis*** B for V, otherwise we say that V is ***infinite dimensional***.

The Dependent/Independent Sets from Spanning Sets Theorem: Suppose we have a set of n vectors, $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\} \subset (V, \oplus, \odot)$, and we form $W = Span(S)$. Suppose now we randomly choose a set of m vectors from W to form a new set: $L = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$.
Then, we can conclude that: ***if*** $m > n$, ***then*** L is linearly ***dependent***.
Consequently, the ***contrapositive*** states that: ***if*** L is linearly ***independent***, ***then*** $m \leq n$.

Any two bases for a ***finite-dimensional*** vector space V have exactly the ***same*** number of vectors.

We call this common number the ***dimension*** of V.

Let W be a subspace of a ***finite-dimensional*** vector space V. Then: $dim(W) \le dim(V)$. Furthermore, $dim(W) = dim(V)$ ***if and only if*** $W = V$ (this can be false if V is infinite dimensional).

A ***linear transformation*** $T : (V, \oplus_V, \odot_V) \to (W, \oplus_W, \odot_W)$ is a function that assigns a ***unique*** member $\vec{w} \in W$ to every vector $\vec{v} \in V$, such that T satisfies for all $\vec{u}, \vec{v} \in V$ and all scalars c:

$$T(\vec{u} \oplus_V \vec{v}) = T(\vec{u}) \oplus_W T(\vec{v}), \text{ and } T(c \odot_V \vec{u}) = c \odot_W T(\vec{u}).$$

We call V the ***domain*** of T, and W the ***codomain*** of T. If $T : V \to V$, we call T an ***operator***.

The ***kernel*** of T is the subspace: $ker(T) = \left\{ \vec{v} \in V \,|\, T(\vec{v}) = \vec{\mathbf{0}}_W \right\} \trianglelefteq V$.

The ***range*** of T is the subspace: $range(T) = \left\{ \vec{w} \in W \,|\, \vec{w} = T(\vec{v}) \text{ for some } \vec{v} \in V \right\} \trianglelefteq W$.

Uniqueness of Representation Property: Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be an ordered ***basis*** for a ***finite dimensional*** vector space V. If $\vec{v} \in V$, then $\vec{v}$ can be expressed ***uniquely*** as a linear combination of the vectors of B: $\vec{v} = (c_1 \odot \vec{w}_1) \oplus (c_2 \odot \vec{w}_2) \oplus \cdots \oplus (c_n \odot \vec{w}_n)$.

The vector $\langle c_1, c_2, \ldots, c_n \rangle$ is the ***coordinate vector*** of $\vec{v}$ with respect to B, written as: $\langle \vec{v} \rangle_B = \langle c_1, c_2, \ldots, c_n \rangle$. The $n \times 1$ column matrix corresponding to $\langle \vec{v} \rangle_B$ is the ***coordinate matrix*** of $\vec{v}$ with respect to B, written as $[\vec{v}]_B$.

Let $T : V \to W$ be a linear transformation, where $dim(V) = n$ and $dim(W) = m$.
Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for V, and let $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be a basis for W.
The $m \times n$ matrix whose columns, from left to right, are $[T(\vec{v}_1)]_{B'}$ through $[T(\vec{v}_n)]_{B'}$ is the ***matrix*** of T ***relative*** to B and B', and written as: $[T]_{B,B'} = \left[[T(\vec{v}_1)]_{B'} \mid [T(\vec{v}_2)]_{B'} \mid \cdots \mid [T(\vec{v}_n)]_{B'} \right]$.

If $\vec{v} \in V$, then $[T(\vec{v})]_{B'} = [T]_{B,B'} \cdot [\vec{v}]_B$. If $T : V \to V$ is an ***operator***, and we use the same basis B for the domain and codomain, we write $[T]_B$ instead of $[T]_{B,B}$.

Let $T_1 : V \to U$ and $T_2 \to U \to W$ be linear transformations of finite dimensional vector spaces. Let B be a basis for V, B' a basis for U, and B'' a basis for W. Then: $[T_2 \circ T_1]_{B,B''} = [T_2]_{B',B''} \cdot [T_1]_{B,B'}$.

A linear transformation $T : V \to W$ is ***one-to-one*** or ***injective*** if the image of different vectors from the domain are different vectors from the codomain: if $\vec{v}_1 \neq \vec{v}_2$, then $T(\vec{v}_1) \neq T(\vec{v}_2)$.

T is ***one-to-one*** ***if and only if*** $ker(T) = \left\{ \vec{0}_V \right\}$.

T is ***onto*** or ***surjective*** ***if and only if*** $range(T) = W$.

The Dimension Theorem for Abstract Vector Spaces: Let $T : V \to W$ be a linear transformation, and suppose that V is ***finite dimensional***, with $dim(V) = n$. Then, both $ker(T)$ and $range(T)$ are finite dimensional, and if we define $\boldsymbol{rank}(T) = dim(\boldsymbol{range}(T))$, and $\boldsymbol{nullity}(T) = dim(\boldsymbol{ker}(T))$, then:

$$\boldsymbol{rank}(T) + \boldsymbol{nullity}(T) = n = dim(V).$$

We say that $T : V \to W$ is an ***isomorphism*** if T is both ***one-to-one*** and ***onto***. We also say that T is ***invertible***, T is ***bijective***, and that V and W are ***isomorphic*** to each other. Two ***finite dimensional*** vector spaces V and W are isomorphic ***if and only if*** $dim(V) = dim(W)$.

$T : V \to W$ is an ***isomorphism*** of vector spaces ***if and only if*** there exists a linear transformation $T^{-1} : W \to V$, which is ***also*** an isomorphism, such that: $T^{-1} \circ T = \boldsymbol{I}_V$ and $T \circ T^{-1} = \boldsymbol{I}_W$.

Suppose $T : V \to W$ is an isomorphism of ***finite dimensional*** vector spaces. We know that $dim(V) = dim(W)$, and there exists $T^{-1} : W \to V$ such that $T^{-1} \circ T = I_V$ and $T \circ T^{-1} = I_W$. In this case, for any basis B for V, and any basis B' for W, $[T]_{B,B'}$ is an ***invertible*** square matrix, and $[T^{-1}]_{B',B} = [T]_{B,B'}^{-1}$.

Chapter 4

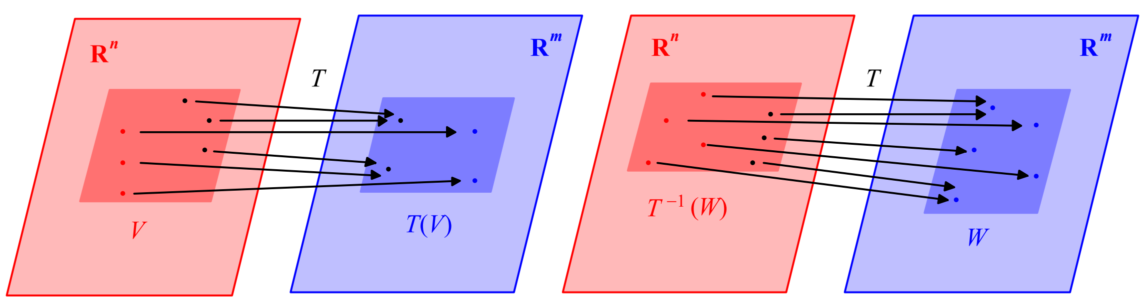
Peeling the Onion:

The Subspace Structure of Vector Spaces

We defined Linear Algebra as the study of vector spaces, their structure, and the linear transformations that map one vector space to another. In this Chapter, we will explore at a deeper level the ***structure*** of vector spaces — that is, how the subspaces of a vector space interact with each other, and how subspaces behave with respect to a linear transformation.

We will see first how to make two subspaces V and W of some ambient space U interact using two basic operations: the ***join*** and the ***intersection*** operations, denoted by $V \vee W$ and $V \cap W$, in order to produce new subspaces of U. We will find algorithms in order to find a basis for these resulting subspaces if we are given bases for V and W.

The ***Preservation of Subspaces Theorem*** tells us that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ will take a subspace V of the domain $\mathbb{R}^n$ and map it into a subspace $T(V)$ of the codomain $\mathbb{R}^m$. Similarly, if W is a subspace of $\mathbb{R}^m$ and we look for all the vectors in $\mathbb{R}^n$ that get mapped by T into W, the resulting set of vectors $T^{-1}(W)$ is a subspace of the domain $\mathbb{R}^n$. We will see how to explicitly find a basis for these ***image*** and ***pre-image*** subspaces.



The Image $T(V)$ of V, and the Preimage $T^{-1}(W)$ of W

We will take a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ and ***restrict*** it to a subspace V of $\mathbb{R}^n$. In particular, we will see that restricting a linear transformation to the ***rowspace*** of $[T]$ always results in a ***one-to-one linear*** transformation, and as such we can construct an ***inverse*** to this restriction.

We will construct the ***quotient space*** V/W where W is a subspace of V, and see the ***Isomorphism Theorems*** of Amalie Emie Noether. We know that a linear transformation $T : V \to W$ must be one-to-one and onto in order to be an isomorphism. However, the First Isomorphism Theorem says that ***any*** linear transformation will ***induce*** an isomorphism between $V/ker(T)$ and $range(T)$. The Second Isomorphism Theorem refers to an isomorphism concerning nested subspaces, and the Third Isomorphism Theorem, also known as ***The Diamond Isomorphism Theorem***, refers to the join and intersection of two subspaces that were constructed in the first section.

4.1 The Join and Intersection of Two Subspaces

We know from basic Set Theory as seen in Chapter Zero that if we have two subsets A and B of some universal set U, we can find the ***union*** and the ***intersection*** of these two sets, written as:

$$A \cup B = \{x \in U \mid x \in A \textbf{ or } x \in B\}, \quad \text{and}$$
$$A \cap B = \{x \in U \mid x \in A \textbf{ and } x \in B\}$$

We can picture these operations with the Venn diagrams from Chapter Zero:

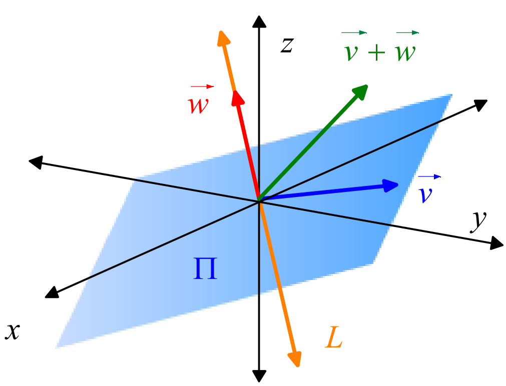
The Union and Intersection of two Sets A and B

Our goal in this Section is to describe what happens when A and B are ***subspaces*** of a vector space U. For simplicity, we will focus our examples on subspaces of $\mathbb{R}^n$. However, if $dim(U) = n$ and B is a fixed basis for U, the process of finding coordinates with respect to U will produce an ***isomorphism:***

$$T : U \to \mathbb{R}^n, \quad \text{where} \quad T(\vec{u}) = \langle \vec{u} \rangle_B.$$

Thus, we can perform whatever computation we show below in $\mathbb{R}^n$ using these coordinate vectors, and us B to decode our answers back to vectors in V. We will see some Exercises where we will apply this concept. Whenever possible, we will state our Theorems in terms of abstract vector spaces instead of just Euclidean spaces.

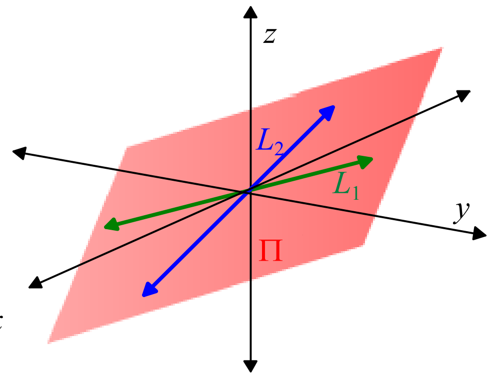
Unfortunately, if V and W are arbitrary subspaces of some ambient space U, their union $V \cup W$ is usually ***not*** a ***subspace*** of U. For example, in $\mathbb{R}^3$, the proper subspaces are lines and planes through the origin, aside from the two trivial subspaces. However, the union of a plane Π and a line L that is not on the plane is certainly not a subspace of $\mathbb{R}^3$:



The Union of a Plane Π and a Line L is ***Not*** a Subspace of $\mathbb{R}^3$

This is because if we add a non-zero vector $\vec{v}$ from Π to a non-zero vector $\vec{w}$ from L, the sum $\vec{v}+\vec{w}$ is ***neither*** on Π nor L.

However, let's take a look at ***two distinct lines*** L_1 and L_2 that pass through the origin. Just by themselves, these two lines do not form a subspace of $\mathbb{R}^3$ either, but the unique ***plane*** Π that contains these two line ***is*** a ***subspace*** of $\mathbb{R}^3$:



The Plane Π Containing Two Lines is a Subspace

Moreover, it is the ***smallest subspace*** of $\mathbb{R}^3$ that ***contains*** these two lines. But recall from Chapter 1 that every point on this plane is a ***sum*** of two vectors, one from each line. We can generalize this construction using the following:

Definition/Theorem: Let V and W be two ***subspaces*** of some ambient vector space U. We define the ***join*** of these two subspaces as the set of all vectors of U that can be written as the ***sum*** of one vector from V and one vector from W, and it is denoted:

$$V \vee W = \left\{ \vec{u} \in U \,|\, \vec{u} = \vec{v} + \vec{w} \text{ for some } \vec{v} \in V, \text{ and some } \vec{w} \in W \right\}.$$

Then: $V \vee W$ is a ***subspace*** of U, and if B is a basis for V and B' is a basis for W, then $B \cup B'$ ***Spans*** $V \vee W$. Consequently, $V \vee W$ is the ***smallest*** subspace of U that contains $V \cup W$, and:

$$dim(V \vee W) \leq dim(V) + dim(W).$$

We pronounce $V \vee W$ as "V ***join*** W." It also follows from the commutative property of vector addition that $V \vee W = W \vee V$.

Proof: We have to show that $V \vee W$ is closed under addition and scalar multiplication. If $\vec{u}_1$ and $\vec{u}_2$ are members of $V \vee W$, then we can write:

$$\vec{u}_1 = \vec{v}_1 + \vec{w}_1 \quad \text{and} \quad \vec{u}_2 = \vec{v}_2 + \vec{w}_2,$$

for some $\vec{v}_1$ and $\vec{v}_2$ from V, and some $\vec{w}_1$ and $\vec{w}_2$ from W. But then:

$$\vec{u}_1 + \vec{u}_2 = (\vec{v}_1 + \vec{w}_1) + (\vec{v}_2 + \vec{w}_2) = (\vec{v}_1 + \vec{v}_2) + (\vec{w}_1 + \vec{w}_2),$$

by the commutative property. Since V and W are closed under addition, $\vec{u}_1 + \vec{u}_2$ is again a member of $V \vee W$. Similarly, we can easily show that $k \cdot \vec{u}_1$ is a member of $V \vee W$.

Now, if $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for V and $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is a basis for W, then every member of V is a linear combination of the members of B, and every member of W is a linear combination of the members of B'. Thus, every member of $V \vee W$ can be written as:

$$\vec{u} = \vec{v} + \vec{w} = c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_n\vec{v}_n + d_1\vec{w}_1 + d_2\vec{w}_2 + \cdots + d_m\vec{w}_m,$$

which looks like a linear combination of the vectors of $B \cup B'$. Thus, $B \cup B'$ ***Spans*** $V \vee W$. However, $B \cup B'$ is ***not*** necessarily linearly ***independent***. However, we can use ***The Minimizing Theorem*** to obtain a subset $S \subseteq B \cup B'$ that ***is*** linearly independent so that:

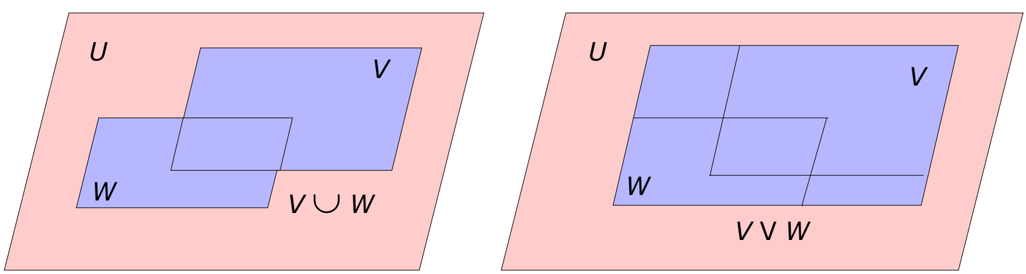
$$Span(S) = Span(B \cup B').$$

Since the number of members of $B \cup B'$ is at most the number of members of B plus the number of members of B', we get:

$$dim(V \vee W) \leq n + m = dim(V) + dim(W).$$

Next, to show that $V \vee W$ is the ***smallest*** subspace that contains $V \cup W$, we have to show that if X is any subspace of U that contains $V \cup W$, then X also contains $V \vee W$. So let us suppose that X is a subspace of U that contains $V \cup W$. Certainly X contains $B \cup B'$, since this is a subset of $V \cup W$. But since X is a subspace of U, it is also ***closed*** under vector addition and scalar multiplication. Thus X contains all linear combinations from the set $B \cup B'$. But we saw above that this set Spans $V \vee W$, so all the members of $V \vee W$ are also members of X. Thus X contains $V \vee W$. ■

We can visualize $V \cup W$ and $V \vee W$ using the following diagram:



The Union $V \cup W$ versus The Join $V \vee W$

Now we can think about constructing a basis for $V \vee W$. Since we know that $B \cup B'$ Spans $V \vee W$, we can find a linearly independent subset of $B \cup B'$, as we did in ***The Minimizing Theorem*** of Chapter 1, which still Spans $V \vee W$.

Example: Let $B = \{\langle 7, -1, 4, 1\rangle, \langle -1, -3, 6, 7\rangle\}$ and $B' = \{\langle -3, 1, -4, -5\rangle, \langle 4, -3, 7, 5\rangle\}$. We can easily see that the two vectors in each set are not parallel to each other, so they are respectively bases for $V = Span(B)$ and $W = Span(B')$. The join $V \vee W$ is ***Spanned*** by the four vectors of:

$$B \cup B' = \{\langle 7, -1, 4, 1\rangle, \langle -1, -3, 6, 7\rangle, \langle -3, 1, -4, -5\rangle, \langle 4, -3, 7, 5\rangle\}.$$

We do not see any obvious dependence relationships among these four vectors, so we assemble the four vectors into the ***columns*** of a matrix:

$$\begin{bmatrix} 7 & -1 & -3 & 4 \\ -1 & -3 & 1 & -3 \\ 4 & 6 & -4 & 7 \\ 1 & 7 & -5 & 5 \end{bmatrix}, \quad \text{with rref} \quad \begin{bmatrix} 1 & 0 & 0 & 5/6 \\ 0 & 1 & 0 & 5/6 \\ 0 & 0 & 1 & 1/3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, the first three columns of our original matrix form a basis for the join:

$$V \vee W = Span(\{\langle 7, -1, 4, 1\rangle, \langle -1, -3, 6, 7\rangle, \langle -3, 1, -4, -5\rangle\}),$$

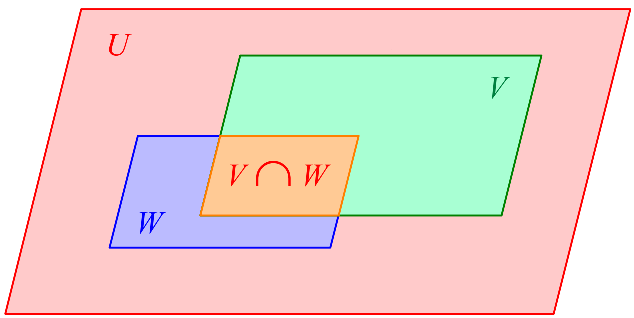
and therefore $V \vee W$ is a 3-dimensional subspace of $\mathbb{R}^4$. $\square$

The Intersection of Two Subspaces

Unlike the union of two subspaces, we are guaranteed that the ***intersection*** of two subspaces is again a subspace:

> ***Theorem:*** Let V and W be two ***subspaces*** of some ambient vector space U. Then, the ***intersection*** of these two subspaces:
>
> $$V \cap W = \{\vec{u} \in U \mid \vec{u} \in V \textbf{ and } \vec{u} \in W\}$$
>
> is a ***subspace*** of U.

We can visualize the intersection of two subspaces below:



The Intersection of two Subspaces $V \cap W$

Proof: Both subspaces contain $\vec{\mathbf{0}}_U$, so $V \cap W$ is not empty. Next, we have to show that $V \cap W$ is closed under vector addition. Suppose $\vec{u}_1$ and $\vec{u}_2$ are members of $V \cap W$. Thus $\vec{u}_1$ and $\vec{u}_2$ are both members of V and W. But each of these is a subspace of U, and thus each is closed under vector addition. Thus $\vec{u}_1 + \vec{u}_2$ is a member of V as well as W, and thus $\vec{u}_1 + \vec{u}_2 \in V \cap W$. As usual, we leave it as an exercise to show that $V \cap W$ is closed under scalar multiplication. $\blacksquare$

Now, if we are given a basis $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for V and a basis $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ W, how would we find a basis for $V \cap W$? Unfortunately, this is not a straightforward process — it is not a simple matter of finding $B \cap B'$. Let us motivate its construction, once again, by considering the familiar $\mathbb{R}^3$.

Example: Let us find a basis for the intersection of ***two planes*** through the origin:

$$V = Span(\{\langle -4, 1, 0\rangle, \langle 0, 1, -2\rangle\}), \text{ and}$$
$$W = Span(\{\langle -5, 1, 0\rangle, \langle 0, 3, 5\rangle\}).$$

We saw how to find the intersection of two planes in Chapter 1 if they were given in terms of their more natural ***Cartesian equations***. Unfortunately, we are given these two rather inconvenient bases instead. However, it is not that difficult to find a normal vector for each plane. Although the cross product would give us a normal easily, we will instead find the ***orthogonal complements*** of V and W, that is, finding the ***nullspaces*** of the two matrices:

$$\begin{bmatrix} -4 & 1 & 0 \\ 0 & 1 & -2 \end{bmatrix} \text{ and } \begin{bmatrix} -5 & 1 & 0 \\ 0 & 3 & 5 \end{bmatrix}.$$

The rref of these two matrices are:

$$\begin{bmatrix} 1 & 0 & -\frac{1}{2} \\ 0 & 1 & -2 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & \frac{1}{3} \\ 0 & 1 & \frac{5}{3} \end{bmatrix}.$$

Thus, we have the normals $\langle 1/2, 2, 1 \rangle$ and $\langle -1/3, -5/3, 1 \rangle$, or more conveniently, $\langle 1, 4, 2 \rangle$ and $\langle 1, 5, -3 \rangle$. Now that we have the normal vectors, we want to find the ***intersection*** of the two planes:

$$x + 4y + 2z = 0, \quad \text{and}$$
$$x + 5y - 3z = 0.$$

Now all we have to do is find the ***nullspace*** of a second matrix:

$$\begin{bmatrix} 1 & 4 & 2 \\ 1 & 5 & -3 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 22 \\ 0 & 1 & -5 \end{bmatrix}.$$

Thus, the intersection of these two planes, as expected, is a ***line*** given by:

$$V \cap W = Span(\{\langle -22, 5, 1 \rangle\}).\ \square$$

Surprisingly, the algorithm to find the intersection of two arbitrary subspaces V and W of some ambient space $\mathbb{R}^k$ is exactly the same as our example above.

Theorem: Let V and W be two subspaces of some ambient space $\mathbb{R}^k$. Then:

$$V \cap W = (V^{\perp} \vee W^{\perp})^{\perp}.$$

Furthermore, suppose $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for V and $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ a basis for W. The following algorithm finds a basis for $V \cap W$:

Step 1. Form the matrix C, with ***rows*** $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and D, with ***rows*** $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m$.

Step 2. Find a basis for the ***nullspace*** of each of these matrices using their rrefs.

Step 3. Assemble these two sets of basis vectors ***together*** as the rows of a third matrix E.

Step 4. Find a basis for the ***nullspace*** of this final matrix using its rref.

The basis for the nullspace in Step 4 is also a basis for $V \cap W$.

We note that the equation above can be rewritten as:

$$(V \cap W)^{\perp} = V^{\perp} \vee W^{\perp},$$

which is analogous to ***De Morgan's Law*** from Chapter Zero.

Proof: Recall that every vector in V is orthogonal to every vector in $V^{\perp}$, and vice-versa. Similarly, every vector in W is orthogonal to every vector in $W^{\perp}$, and vice-versa. Thus, the vectors in $V \cap W$ are exactly those vectors that are orthogonal to ***both*** $V^{\perp}$ ***and*** $W^{\perp}$. But the vectors which are orthogonal to both of these spaces are precisely those vectors in the orthogonal complement of the space Spanned by both $V^{\perp}$ and $W^{\perp}$, and this is computed precisely by assembling a matrix whose rows are formed by a basis for $V^{\perp}$ and a basis for $W^{\perp}$ and finding the nullspace of this matrix. This also tells us that $V \cap W = (V^{\perp} \vee W^{\perp})^{\perp}$. ■

Example: Let us look again at our previous example, with:

$$B = \{\langle 7, -1, 4, 1\rangle, \langle -1, -3, 6, 7\rangle\}, \quad B' = \{\langle -3, 1, -4, -5\rangle, \langle 4, -3, 7, 5\rangle\},$$

$V = Span(B)$ and $W = Span(B')$. This time, let us find a basis for $V \cap W$. We assemble the two bases ***separately*** as the ***rows*** of two matrices:

$$C = \begin{bmatrix} 7 & -1 & 4 & 1 \\ -1 & -3 & 6 & 7 \end{bmatrix} \quad \text{and} \quad D = \begin{bmatrix} -3 & 1 & -4 & -5 \\ 4 & -3 & 7 & 5 \end{bmatrix}.$$

We individually find the rref of these matrices, which are:

$$\begin{bmatrix} 1 & 0 & 3/11 & -2/11 \\ 0 & 1 & -23/11 & -25/11 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 1 & -1 & 1 \end{bmatrix}.$$

From these rref, we sight-read a basis for the ***nullspaces*** of C and D:

$$\{\langle -3, 23, 11, 0\rangle, \langle 2, 25, 0, 11\rangle\} \quad \text{and} \quad \{\langle -1, 1, 1, 0\rangle, \langle -2, -1, 0, 1\rangle\}.$$

We could take a minute to check mentally that both of the vectors from each set are orthogonal to the corresponding two vectors from the original bases using the dot product, so we are confident that these are indeed bases for the orthogonal complements. Now, for the final phase, we assemble these final four vectors into the ***rows***, once again, of a third matrix:

$$\begin{bmatrix} -3 & 23 & 11 & 0 \\ 2 & 25 & 0 & 11 \\ -1 & 1 & 1 & 0 \\ -2 & -1 & 0 & 1 \end{bmatrix}, \quad \text{with rref} \quad \begin{bmatrix} 1 & 0 & 0 & -3/4 \\ 0 & 1 & 0 & 1/2 \\ 0 & 0 & 1 & -5/4 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We have a single free variable, and thus $V \cap W$ is 1-dimensional, and by clearing fractions, we get:

$$V \cap W = Span(\{\langle 3, -2, 5, 4\rangle\}).$$

To check this answer, we will verify that the single vector in our basis is a member of ***both*** V and W by expressing it as a linear combination of both B and B':

$$\langle 3, -2, 5, 4\rangle = \frac{1}{2}\langle 7, -1, 4, 1\rangle + \frac{1}{2}\langle -1, -3, 6, 7\rangle, \quad \text{and}$$

$$\langle 3, -2, 5, 4\rangle = -\frac{1}{5}\langle -3, 1, -4, -5\rangle + \frac{3}{5}\langle 4, -3, 7, 5\rangle. \square$$

This has certainly not been a simple process. Notice that we had to perform the Gauss-Jordan algorithm on ***three*** matrices. There is one special case, though, where the intersection is quite simple:

Theorem: Let W be any ***subspace*** of some Euclidean space $\mathbb{R}^n$. Then:

$$W \cap W^{\perp} = \left\{\vec{\mathbf{0}}_n\right\}.$$

In other words, the only vector common to a subspace W and its orthogonal complement $W^{\perp}$ is the zero vector.

Proof: This proof is best described as ***magical***. Let $\vec{w} \in W \cap W^{\perp}$. By definition, every member of W is orthogonal to every member of $W^{\perp}$, and vice versa. So in particular, $\vec{w}$ is orthogonal to ***itself***, that is:

$$\vec{w} \circ \vec{w} = 0.$$

But recall that $\vec{w} \circ \vec{w} = \|\vec{w}\|^2$, and the only vector whose length is 0 is the zero vector. Thus $\vec{w} = \vec{\mathbf{0}}_n$, and $W \cap W^{\perp} = \left\{\vec{\mathbf{0}}_n\right\}$. ■

On the other hand, we will find out in Chapter 7 that $W \vee W^{\perp} = \mathbb{R}^n$, for any subspace W of $\mathbb{R}^n$.

In a special case, the join of two subspaces is known by another name:

Definition: Let V and W be two ***subspaces*** of some ambient vector space U. Suppose $V \cap W = \left\{\vec{\mathbf{0}}_U\right\}$. Then the join of V and W is written as: $V \vee W = V \oplus W$, and is called the ***direct sum*** of V and W.

The Relationship Between $V \vee W$ and $V \cap W$

We will find out in Section 4.5, at the end of this Chapter, that the join and the intersection of two subspaces are connected by ***The Diamond Isomorphism Theorem***. The proof of the following consequence regarding their dimensions is an Exercise in that Section, and easily follows from the aforementioned Theorem:

Theorem — The Dimension Theorem for the Join and Intersection:
Let V and W be ***finite-dimensional subspaces*** of a vector space U. Then: $V \vee W$ is also finite dimensional, and:

$$dim(V \vee W) = dim(V) + dim(W) - dim(V \cap W).$$

Example: In our previous Example, with:

$$B = \{\langle 7,-1, 4, 1\rangle, \langle -1,-3, 6, 7\rangle\},\ \ B' = \{\langle -3, 1,-4,-5\rangle, \langle 4,-3, 7, 5\rangle\},$$

and $V = Span(B)$, $W = Span(B')$, we saw that:

$$dim(V \vee W) = 3,\ \ dim(V) = 2,\ \ dim(W) = 2,\ \ \text{and}\ \ dim(V \cap W) = 1.$$

Thus, indeed:

$$dim(V \vee W) = 3 = 2 + 2 - 1 = dim(V) + dim(W) - dim(V \cap W). \square$$

4.1 Section Summary

Let V and W be two subspaces of some ambient vector space U. We define the ***join*** of these two subspaces as the set of all vectors of U that can be written as the sum of one vector from V and one vector from W, and it is denoted:

$$V \vee W = \left\{ \vec{u} \in U \mid \vec{u} = \vec{v} + \vec{w} \text{ for some } \vec{v} \in V \text{ and some } \vec{w} \in W \right\}.$$

Then: $V \vee W$ is a ***subspace*** of U, and if B is a basis for V and B' is a basis for W, then $B \cup B'$ ***Spans*** $V \vee W$. Consequently, $V \vee W$ is the ***smallest*** subspace of U that contains $V \cup W$, and:

$$dim(V \vee W) \leq dim(V) + dim(W).$$

The ***intersection*** of these two subspaces:

$$V \cap W = \{ \vec{u} \in U \mid \vec{u} \in V \textbf{ and } \vec{u} \in W \}$$

is a ***subspace*** of U.

Suppose $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for V and $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is a basis for W, and these are subspaces of some $\mathbb{R}^k$. Form the matrix C, with rows $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and D, with rows $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m$. Next, find a basis for the nullspace of each of these two matrices. Assemble these two sets of basis vectors ***together*** as the rows of a third matrix. The nullspace of this final matrix is $V \cap W$.

Let W be any subspace of some Euclidean space $\mathbb{R}^n$. Then $W \cap W^\perp = \left\{ \vec{\mathbf{0}}_n \right\}$.

The Dimension Theorem for the Join and Intersection:

Let V and W be ***finite-dimensional subspaces*** of a vector space U. Then: $V \vee W$ is also finite dimensional, and:

$$dim(V \vee W) = dim(V) + dim(W) - dim(V \cap W).$$

4.1 Exercises

For Exercises (1) to (5): For the following subspaces V and W for their respective $\mathbb{R}^n$: (a) find a basis for $V \vee W$ and state its dimension; (b) find a basis for $V \cap W$ and state its dimension; (c) check that each basis member that you found in (b) is in fact a linear combination of the members from the basis for V as well as W; (d) verify that the Dimension Theorem for the Join and Intersection is true.

It is highly recommended that technology be used to find the rref of the matrices involved.

1. $V, W \trianglelefteq \mathbb{R}^4$; $V = Span(\{\langle 1, -1, -12, 6\rangle, \langle 11, -16, 13, 1\rangle\})$,
   $W = Span(\{\langle 1, 1, -16, 10\rangle, \langle 7, -11, 5, 1\rangle\})$.
2. $V, W \trianglelefteq \mathbb{R}^4$; $V = Span(\{\langle 3, 5, -2, 4\rangle, \langle 1, 2, 7, -3\rangle\})$,
   $W = Span(\{\langle 0, 2, 1, -5\rangle, \langle 2, -3, 1, 6\rangle\})$.
3. $V, W \trianglelefteq \mathbb{R}^4$; $V = Span(\{\langle -3, -2, 7, -4\rangle, \langle -2, 13, -12, -2\rangle, \langle -2, 3, -5, 1\rangle\})$,
   $W = Span(\{\langle -3, -5, 6, -11\rangle, \langle -1, 16, -8, 8\rangle, \langle 1, -3, 2, -4\rangle\})$
4. $V, W \trianglelefteq \mathbb{R}^5$; $V = Span(\{\langle -3, 4, -1, 4, 6\rangle, \langle -6, 8, 5, 15, -13\rangle, \langle 1, -2, 0, -5, 3\rangle\})$,
   $W = Span(\{\langle 1, 3, -2, 7, 2\rangle, \langle -4, -1, 7, -7, -6\rangle\})$
5. $V = Span(\{\langle -1, 7, 5, -6, 6\rangle, \langle -1, -8, 2, -4, 2\rangle, \langle 1, 0, 3, -4, 3\rangle, \langle 5, 3, -2, 7, -4\rangle\}) \trianglelefteq \mathbb{R}^5$,
   $W = Span(\{\langle -6, 9, -2, 0, 0\rangle, \langle -5, 1, -3, -3, -2\rangle, \langle -3, 2, -1, -2, 0\rangle\}) \trianglelefteq \mathbb{R}^5$.

For Exercises (6) to (8): For the following subspaces $V = Span(B)$ and $W = Span(B^{/})$ of the respective $\mathbb{P}^n$, perform parts (a) through (d) in the instructions for Exercises (1) to (5). Use the standard basis $\{1, x, x^2, \ldots, x^n\}$ to ***encode*** the polynomials into matrices. Exercise 6 is started for you.

6. $B = \{6 - x + 2x^2 + 10x^3, 11 - 3x + 6x^2 + 2x^3\}$, and
$B^{/} = \{3 + 17x + 5x^2 + 4x^3, 3 + 11x + 4x^2 + 2x^3\} \subseteq \mathbb{P}^3$.

$$C = \begin{bmatrix} 6 & -1 & 2 & 10 \\ 11 & -3 & 6 & 2 \end{bmatrix}; \quad D = \begin{bmatrix} 3 & 17 & 5 & 4 \\ 3 & 11 & 4 & 2 \end{bmatrix}.$$

7. $B = \{2 + 5x - 10x^2 + 5x^3, 6 - 7x - 4x^2 + x^3, -8 + 14x - 16x^2 + 3x^3\}$, and
$B^{/} = \{2 + 3x - 19x^2 + 13x^3, 7 - 7x + 11x^2 - 12x^3\} \subseteq \mathbb{P}^3$.

8. $B = \{-3 - 2x + 4x^2 + x^4, 6 - 3x^2 + 5x^3 - 5x^4, -7 - 7x + 8x^2 + 2x^3 + 8x^4,$
$-5 - 2x + 7x^2 + x^3 - x^4\}$, and
$B^{/} = \{1 - 6x + 3x^2 - 2x^3 - 4x^4, 5 - 14x + 7x^2 + x^3 - 12x^4,$
$-9x + 3x^2 + 2x^4, -3 + 6x + 3x^3 + 2x^4\} \subseteq \mathbb{P}^4$.

9. Let V and W be subspaces of some ambient vector space U. Prove that $V \vee W$ as well as $V \cap W$ are closed under scalar multiplication.

10. Suppose that V and W are subspaces of $\mathbb{R}^8$ with $dim(V) = 4$ and $dim(W) = 6$. Prove that $2 \leq dim(V \cap W) \leq 4$.

11. Suppose that V and W are subspaces of $\mathbb{R}^{12}$ with $dim(V) = 8$ and $dim(W) = 9$. State and prove a compound inequality analogous to the one in the previous Exercise.

12. Suppose V and W are subspaces of some ambient space U, and $V \cap W = \left\{\vec{0}_U\right\}$. Prove that every vector $\vec{u} \in V \vee W$ can be expressed ***uniquely*** in the form $\vec{v} + \vec{w}$, where $\vec{v} \in V$ and $\vec{w} \in W$. In other words:

$$\text{If } \vec{u} = \vec{v}_1 + \vec{w}_1 \text{ and } \vec{u} = \vec{v}_2 + \vec{w}_2, \text{ where } \vec{v}_1, \vec{v}_2 \in V$$
$$\text{and } \vec{w}_1, \vec{w}_2 \in W, \text{ then } \vec{v}_1 = \vec{v}_2 \text{ and } \vec{w}_1 = \vec{w}_2.$$

13. Suppose V is a 5-dimensional subspace of $\mathbb{R}^7$ and W is a 3-dimensional subspace of $\mathbb{R}^7$, and you computed that $V \vee W$ is 5-dimensional. What must be the relationship between V and W? Prove your answer. Hint: Think about the process of finding the basis for $V \vee W$. Don't peek at the answer key until you have your own answer.

14. Similarly, suppose that V is a 6-dimensional subspace of $\mathbb{R}^8$ and W is a 4-dimensional subspace of $\mathbb{R}^8$. What can you conclude if you computed that $V \cap W$ is 4-dimensional? Again, prove your answer.

15. Put together and generalize the ideas behind the last two Exercises into a single Theorem by completing the following ***Theorem***, and prove it: Let V and W be two finite-dimensional subspaces of a vector space U. Then, W is a subspace of V if and only if ***either*** . . .

4.2 Restricting Linear Transformations and the Role of the Rowspace

In Algebra, Trigonometry and Calculus, we encounter functions that are usually defined on an interval, and perhaps even the set of all real numbers. However, we also often study these functions when they are restricted to a smaller domain. We see this, for example, when we study a ***continuous*** function on a closed interval $[a, b]$. In Calculus, the ***Extreme Value Theorem*** tells us that such a restricted continuous function has an absolute ***maximum*** and ***minimum*** on $[a, b]$. In this section, we will analogously see how to restrict a linear transformation T to a subspace of the domain, and the surprising role that the rowspace has when we want a certain desirable quality to be manifested by this restriction.

Restricting a Linear Transformation

In Linear Algebra, we can take a linear transformation $T : V \to W$ and restrict it to a subspace U of V. Again, for simplicity, we will focus our examples on linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$, but we will state our theorems in terms of abstract vector spaces instead of just Euclidean spaces.

Definition/Theorem: Let $T : V \to W$ be a linear transformation, and U a subspace of the domain V. The ***restriction*** of T to U, denoted $T|_U$ and pronounced "T ***restricted*** to U," is the linear transformation:

$$T|_U \ : \ U \to W, \text{ given by}$$
$$T|_U(\vec{u}) = T(\vec{u}) \ \text{ for all } \vec{u} \in U.$$

The only issue here is that of linearity, but the additivity and homogeneity properties are ***inherited*** by the restriction from T. In other words, since they are valid for the vectors of V, they are also valid for the vectors of the subspace U.

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be the linear transformation given by the 4×3 matrix:

$$[T] = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix}.$$

Let us consider the restriction of T to the ***plane*** through the origin, U, given by:

$$3x - 5y + 2z = 0.$$

It is easy to see that the set $B = \{\langle 5, 3, 0\rangle, \langle 0, 2, 5\rangle\}$ is a basis for U. Let us find the matrix of $T|_U$ with respect to the bases B for U and the standard basis $B' = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ for $\mathbb{R}^4$. Thus $[T|_U]$ will be a 4×2 matrix. We compute:

$$T(\langle 5, 3, 0\rangle) = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 5 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 21 \\ -13 \\ -23 \\ 8 \end{bmatrix}, \text{ and}$$

$$T(\langle 0, 2, 5\rangle) = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \\ 5 \end{bmatrix} = \begin{bmatrix} 29 \\ -7 \\ 23 \\ 22 \end{bmatrix}.$$

Since we are using the standard basis for $\mathbb{R}^4$, these images are already ***encoded***. Thus:

$$[T|_U]_{B,B'} = \begin{bmatrix} 21 & 29 \\ -13 & -7 \\ -23 & 23 \\ 8 & 22 \end{bmatrix}. \square$$

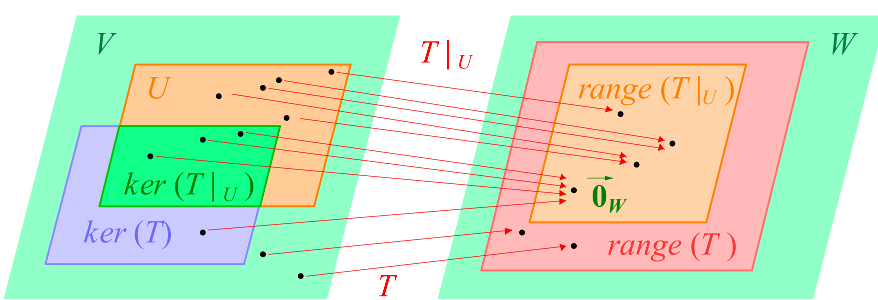
What can we say about the kernel and range of a restricted transformation? We have the following Theorem, which will be proven in the Exercises:

Theorem: Let $T : V \to W$ be a linear transformation, and U a subspace of V. Let us define:

$$ker(T|_U) = \left\{ \vec{u} \in U \mid T(\vec{u}) = \vec{0}_W \right\}, \text{ and}$$

$$range(T|_U) = \left\{ \vec{w} \in W \mid \vec{w} = T(\vec{u}) \text{ for some } \vec{u} \in U \right\}.$$

Then: $ker(T|_U) = ker(T) \cap U$ and $range(T|_U) \trianglelefteq range(T)$.



The Kernel and Range of the Restriction of T to U

This Theorem tells us that instead of studying $T : V \to W$, we can study the restricted transformation:

$$T|_U : U \to range(T|_U),$$

and the kernel of this restriction is: $ker(T|_U) = ker(T) \cap U$.

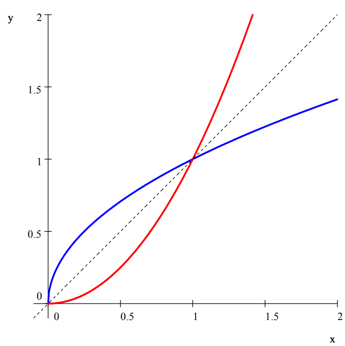
The Role of the Rowspace

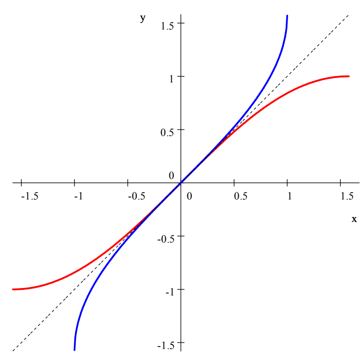
In Algebra and Trigonometry, we see many important functions that are ***not*** one-to-one, such as the quadratic and sine functions. However, we would very much like to construct an "inverse" for such a function. To do so, we would need to first ***restrict*** it to an interval where it is ***one-to-one***. For the inverse to be as general as possible, we want this interval to be the biggest interval possible, such that the function is still one-to-one on this interval (although for some functions such as the secant, the domain could consist of two or more intervals).

For example, we traditionally restrict $y = x^2$ on $x \in [0, \infty)$ and $y = \sin(x)$ on $[-\pi/2, \pi/2]$, thus making these functions one-to-one on these domains. From these, we construct the inverse functions: $y = \sqrt{x}$ and $y = \sin^{-1}(x)$. These functions have the familiar ***Cancellation Properties:***

$$\sqrt{x^2} = x \text{ if } x \geq 0, \text{ and } (\sqrt{x})^2 = x \text{ if } x \geq 0;$$
$$\sin(\sin^{-1}(x)) = x \text{ if } x \in [-1, 1], \text{ and } \sin^{-1}(\sin(x)) = x \text{ if } x \in [-\pi/2, \pi/2].$$

Let us show each pair of functions, where the original function is in red, and the inverse function is in blue:





$y = x^2$ versus $y = \sqrt{x}$ $\qquad$ $y = \sin(x)$ versus $y = \sin^{-1}(x)$

In the same way, if a linear transformation $T : V \to W$ is ***not*** one-to-one on V, then perhaps we can ***restrict*** T on a subspace U of V so that the transformation will be ***one-to-one*** on U. We would also like to have U "as big as possible." Unfortunately, we do not generally have a graph of a linear transformation to help us visually check ***where*** a transformation is one-to-one, in the same way that we have the Horizontal Line Test in precalculus. And how would we know if this domain is as big as possible, even if we were to visualize it?

In the case when $T : \mathbb{R}^n \to \mathbb{R}^m$, the best choice for U turns out to be a subspace closely related to $[T]$:

> ***Theorem:*** Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with standard matrix $[T]$.
> Let $U = \textit{rowspace}([T]) \trianglelefteq \mathbb{R}^n$. Then: the restriction $T|_U : U \to \mathbb{R}^m$ is ***one-to-one***.
> Furthermore, for any subspace W of $\mathbb{R}^n$ such that $dim(W) > \dim(U) = rank(T)$, the restriction $T|_X : X \to \mathbb{R}^m$ is ***not*** one-to-one.

Proof: We know that the ***rowspace*** of $[T]$ is the ***orthogonal complement*** of the ***nullspace*** of $[T]$, which is also known as $ker(T)$. But we also saw from Section 4.1 that for any $W \trianglelefteq \mathbb{R}^n$, $W \cap W^{\perp} = \left\{\vec{\mathbf{0}}_n\right\}$. This means that:

$$rowspace([T]) \cap ker(T) = \left\{ \vec{\mathbf{0}}_n \right\}.$$

But this means that if $\vec{u} \in U = rowspace([T])$ and $T(\vec{u}) = \vec{\mathbf{0}}_m$, then $\vec{u} = \vec{\mathbf{0}}_n$. Thus, $ker(T|_U) = \left\{ \vec{\mathbf{0}}_n \right\}$ and T is one-to-one on U.

Now, let $rank(T) = k$. Then $nullity(T) = n - k$. Suppose $X \trianglelefteq \mathbb{R}^n$ such that $dim(X) > k$. If we apply The Dimension Theorem to the restricted transformation:

$$T|_X : X \to \mathbb{R}^m,$$

then we will get:

$$rank(T|_X) + nullity(T|_X) = dim(X).$$

Suppose that $T|_X$ is ***one-to-one***. Then $nullity(T|_X) = 0$. By the Dimension Theorem, we get:

$$rank(T|_X) = dim(X) > k.$$

However, we also know from the previous Theorem that:

$$range(T|_X) \trianglelefteq range(T),$$

and so we get $rank(T|_X) \leq rank(T) = k$. We get a contradiction. Thus, $T|_X$ is ***not*** one-to-one. ■

Note: It is possible for X to be a subspace with $dim(X) = rank(T)$, but X is ***not*** the rowspace of $[T]$, and $T|_X$ is still ***one-to-one***. Thus, although the rowspace has the biggest possible dimension where the restriction of T is one-to-one, it is not necessarily a ***unique*** subspace of biggest possible dimension where the restriction of T is one-to-one. Notice that we have analogous phenomena in algebra — we can also restrict $y = x^2$ to $(-\infty, 0]$, where this function is one-to-one, but we prefer $\sqrt{x}$ to be positive.

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be from our previous Example, given by:

$$[T] = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix}, \quad \text{with rref} \quad \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Notice that there is a free variable, so T is ***not*** one-to-one on all of $\mathbb{R}^3$. A basis for the rowspace is:

$$B = \{\langle 1, 0, -3\rangle, \langle 0, 1, 7\rangle\}.$$

and T is one-to-one when restricted to the rowspace $U = Span(B)$. Notice that since U is 2-dimensional, it is a ***plane*** through the origin in $\mathbb{R}^3$. The cross product of the two vectors in B is easily computed to be $\langle 3, -7, 1\rangle$, and so U has Cartesian equation:

$$3x - 7y + z = 0.$$

However, looking back to the previous Example, we saw there that if T is restricted to the plane:

$$3x - 5y + 2z = 0,$$

then T is ***also*** one-to-one on this other plane. Thus the rowspace is not the only subspace of $\mathbb{R}^3$ where T is one-to-one. □

As a bonus, we can now find the inverse of a restriction once it is one-to-one on the smaller domain. The proof of the following is straightforward and is left as an Exercise:

Theorem: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and let U be ***any*** subspace of $\mathbb{R}^n$ such that $T|_U$ is ***one-to-one***. Then:

$$T|_U : U \to range(T|_U)$$

is an ***isomorphism***. Moreover, if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_j\}$ is a ***basis*** for U, then $B' = \{T(\vec{u}_1), T(\vec{u}_2), \ldots, T(\vec{u}_j)\}$ is a ***basis*** for $range(T|_U)$.

In particular, if $U = rowspace([T])$, or U is any subspace of such that $dim(U) = rank(T)$ and T is ***one-to-one*** when restricted to U, then $range(T|_U) = range(T)$.

Example: Let us find a matrix for the restricted linear transformation $T|_U$ in our previous Example, as well as its inverse. Notice that $[T]$ is 4×3, so it is definitely ***not*** invertible.

We already found a basis for the rowspace U, namely: $B = \{\langle 1, 0, -3\rangle, \langle 0, 1, 7\rangle\}$.

Since T is one-to-one on U, $range(T|_U) = range(T) = colspace([T])$. We can see from the rref for $[T]$ in the previous Example that the first two columns of $[T]$ form a basis for B' for $colspace([T])$:

$$B' = \{\langle 3, -2, -4, 1\rangle, \langle 2, -1, -1, 1\rangle\}.$$

Now, let us find $[T]_{B,B'}$. We compute:

$$T(\langle 1, 0, -3\rangle) = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -3 \end{bmatrix} = \begin{bmatrix} -12 \\ 1 \\ -19 \\ -11 \end{bmatrix}, \text{ and}$$

$$T(\langle 0, 1, 7\rangle) = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 7 \end{bmatrix} = \begin{bmatrix} 37 \\ -8 \\ 34 \\ 29 \end{bmatrix}.$$

Now, to find the coordinates of both of these vectors with respect to B' at the same time, we assemble the augmented matrix:

$$\left[\begin{array}{cc|cc} 3 & 2 & -12 & 37 \\ -2 & -1 & 1 & -8 \\ -4 & -1 & -19 & 34 \\ 1 & 1 & -11 & 29 \end{array}\right] \text{ with rref } \left[\begin{array}{cc|cc} 1 & 0 & 10 & -21 \\ 0 & 1 & -21 & 50 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Thus:

$$[T|_U]_{B,B'} = \begin{bmatrix} 10 & -21 \\ -21 & 50 \end{bmatrix}.$$

Notice that the matrix of the restriction is now square ***and*** invertible. Its inverse is:

$$[T|_U^{-1}]_{B',B} = \frac{1}{59}\begin{bmatrix} 50 & 21 \\ 21 & 10 \end{bmatrix}. \square$$

More generally, if $T : V \to W$ is a linear transformation with $dim(V) = n$ and $dim(W) = m$, we can use a basis B for V and a basis $B^{/}$ for W to construct $[T]_{B,B^{/}}$. If we find the rref of this matrix and ***decode*** the resulting non-zero rows into a basis S for a subspace U of V, then T will be one-to-one when restricted to U as before. Notice that the rowspace itself is meaningless without this decoding process. From this point, we can find a basis $S^{/}$ for $T(U)$, the most convenient of which would be $S^{/} = T(S)$. If $rank(T) = k$, we can finally construct the $k \times k$ matrix $[T]_{S,S^{/}}$.

Before we leave this Section, we want to point out that we actually saw the concept of a restricted linear transformation. In Chapter 3, we saw the linear transformation $T : \mathbb{P}^3 \to \mathbb{R}^4$ given by:

$$T(p(x)) = \langle p(-2), p(3), p^{/}(-1), p^{//}(1) \rangle.$$

This is the restriction of the linear transformation $T : C^2(I) \to \mathbb{R}^4$ on the space $\mathbb{P}^3$, with exactly the same formula, where $p(x) \in C^2(I)$ for any open interval I containing $[-2, 3]$. Thus, we can find $T(e^x)$ and $T(\sin(x))$ and the image of any other function whose second derivative is continuous. Of course, it helps to restrict T on a finite-dimensional vector space if we want to find its matrix.

4.2 Section Summary

Let $T : V \to W$ be a linear transformation, and U a subspace of the domain V. The ***restriction*** of T to U, denoted $T|_U$ and pronounced "T ***restricted*** to U," is the linear transformation:

$$T|_U : U \to W, \text{ given by}$$
$$T|_U(\vec{u}) = T(\vec{u}) \text{ for all } \vec{u} \in U.$$

Let us define:

$$ker(T|_U) = \left\{ \vec{u} \in U \mid T(\vec{u}) = \vec{0}_W \right\}, \text{ and}$$
$$range(T|_U) = \left\{ \vec{w} \in W \mid \vec{w} = T(\vec{u}) \text{ for some } \vec{u} \in U \right\}.$$

Then: $ker(T|_U) = ker(T) \cap U$ and $range(T|_U) \trianglelefteq range(T)$.

Thus, we can regard $T|_U$ as a linear transformation: $T|_U : U \to range(T|_U)$.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with standard matrix $[T]$.

Let $U =$ ***rowspace***$([T]) \trianglelefteq \mathbb{R}^n$. Then: the restriction $T|_U : U \to \mathbb{R}^m$ is ***one-to-one***.

Furthermore, for any subspace W of $\mathbb{R}^n$ such that $dim(W) > \dim(U) = rank(T)$, the restriction $T|_W : W \to \mathbb{R}^m$ is ***not*** one-to-one.

Let U be ***any*** subspace of $\mathbb{R}^n$ such that $T|_U$ is ***one-to-one***. Then:

$$T|_U : U \to range(T|_U)$$

is an ***isomorphism***. Moreover, if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_j\}$ is a basis for U, then $B^{/} = \{T(\vec{u}_1), T(\vec{u}_2), \ldots, T(\vec{u}_j)\}$ is a basis for $range(T|_U)$.

In particular, if $U = rowspace([T])$, or U is any subspace of such that $dim(U) = rank(T)$ and T is ***one-to-one*** when restricted to U, then $range(T|_U) = range(T)$.

4.2 Exercises

For Exercises (1) to (16): For the following linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ with standard matrices $[T]$ and corresponding rref R: (a) find a basis B for the rowspace of $[T]$ using its rref, (b) find a basis B' for the range of T consisting of original columns of $[T]$, (c) find $[T|_U]_{B,B'}$, where $U = rowspace([T])$, and (d) show that the matrix you found in (c) is indeed invertible by computing its inverse.

1. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 3 & 2 & -2 \\ 5 & 3 & -1 \\ 4 & 2 & 2 \\ -1 & -1 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

2. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 2 & -6 & -7 \\ -3 & 9 & -1 \\ -4 & 12 & 9 \\ 5 & -15 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

3. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 2 & -6 & -7 \\ -3 & 9 & -1 \\ -4 & 12 & 9 \\ 5 & -5 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

   Note that only one entry in $[T]$ was changed from the previous Exercise.

4. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 3 & 5 & 4 & -1 \\ 2 & 3 & 2 & -1 \\ -2 & -1 & 2 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & -2 & -2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

   Note that this is the transpose of the matrix from Exercise 1.

5. $T : \mathbb{R}^3 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 3 & 5 & -1 \\ 2 & 3 & -1 \\ -2 & -1 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

   Note that this matrix was obtained from the previous Exercise by deleting the 3rd column.

6. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 2 & 10 & 5 & -7 \\ 3 & 15 & 7 & -9 \\ -4 & -20 & -9 & 11 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 4 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

7. $T:\mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 2 & 10 & 5 & -7 \\ 3 & 15 & 7 & -9 \\ -4 & -20 & -9 & 8 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

Note that only one entry in $[T]$ was changed from the previous Exercise.

8. $T:\mathbb{R}^5 \to \mathbb{R}^4$, with

$$[T] = \begin{bmatrix} -5 & -20 & -4 & 1 & 2 \\ 3 & 12 & -2 & -27 & 12 \\ 2 & 8 & 3 & 8 & -5 \\ -4 & -16 & 1 & 26 & -11 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & 4 & 0 & -5 & 2 \\ 0 & 0 & 1 & 6 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

9. $T:\mathbb{R}^4 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} -5 & -20 & -4 & 1 \\ 3 & 12 & -2 & -27 \\ 2 & 8 & 3 & 8 \\ -4 & -16 & 1 & 26 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 4 & 0 & -5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

Note that this matrix was obtained from the previous Exercise by deleting the last column.

10. $T:\mathbb{R}^5 \to \mathbb{R}^4$, with:

$$[T] = \begin{bmatrix} -5 & 10 & -30 & -3 & 5 \\ 2 & -4 & 12 & -3 & -23 \\ -3 & 6 & -18 & 2 & 22 \\ 4 & -8 & 24 & -5 & -41 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & -2 & 6 & 0 & -4 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

11. $T:\mathbb{R}^5 \to \mathbb{R}^4$, with:

$$[T] = \begin{bmatrix} -2 & 6 & -3 & -5 & -27 \\ 3 & -9 & 7 & 4 & 6 \\ 4 & -12 & -1 & -2 & -2 \\ -5 & 15 & 2 & 3 & 3 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & -3 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

12. $T:\mathbb{R}^5 \to \mathbb{R}^4$, with:

$$[T] = \begin{bmatrix} -2 & 6 & -3 & -5 & -7 \\ 3 & -9 & 7 & 4 & 6 \\ 4 & -12 & -1 & -2 & -2 \\ -5 & 15 & 2 & 3 & 3 \end{bmatrix}, \quad R = \begin{bmatrix} 1 & -3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Note that only one entry in $[T]$ was changed from the previous Exercise.

13. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} 2 & 10 & -1 & -10 \\ -4 & -20 & 1 & 14 \\ 3 & 15 & 1 & 0 \\ 5 & 25 & -3 & -28 \\ -6 & -30 & 2 & 24 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & -2 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

14. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} 2 & 10 & -1 & -9 \\ -4 & -20 & 1 & 14 \\ 3 & 15 & 1 & 0 \\ 5 & 25 & -3 & -28 \\ -6 & -30 & 2 & 24 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

    Note that only one entry in $[T]$ was changed from the previous Exercise.

15. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} -2 & 1 & 1 & 2 \\ 5 & -1 & -1 & -8 \\ 1 & 1 & -1 & 6 \\ -2 & -2 & 1 & -7 \\ -1 & 1 & 1 & 0 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

16. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} -2 & 1 & 1 & 3 \\ 5 & -1 & -1 & -8 \\ 1 & 1 & -1 & 6 \\ -2 & -2 & 1 & -7 \\ -1 & 1 & 1 & 0 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

    Note that only one entry in $[T]$ was changed from the previous Exercise.

17. Let $T : V \to W$ be a linear transformation, and U any subspace of the domain V. Prove that:
    a. $ker(T|_U) = ker(T) \cap U$. Reminder: to show that two sets are equal, you must show that the first is a subset of the second, and the second is a subset of the first.
    b. $range(T|_U) \trianglelefteq range(T)$.

18. Let U be ***any*** subspace of $\mathbb{R}^n$ such that $T|_U$ is ***one-to-one***.
    a. Show that $T|_U : U \to range(T|_U)$ is an ***isomorphism***.
    b. Prove that if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_j\}$ is a basis for U, then $B' = \{T(\vec{u}_1), T(\vec{u}_2), \ldots, T(\vec{u}_j)\}$ is a basis for $range(T|_U)$. Hint: this is essentially the same idea as Exercise 26 in Section 3.7.

19. Prove that if a linear transformation $T : V \to W$ is ***one-to-one***, and U is any subspace of the domain V, then $T|_U$ is also ***one-to-one***.

20. Is the statement in the previous Exercise still true if both occurrences of the phrase "one-to-one" are replaced with the word "onto"? If so, prove this new statement, but if not, give a counterexample.

4.3 The Image and Preimage of Subspaces

In this Section, we will see that a linear transformation T essentially ***preserves*** subspaces. This means that subspaces of the domain are transformed by T into subspaces of the codomain. Similarly, vectors in the domain that are sent to a subspace of the codomain form a subspace of the domain:

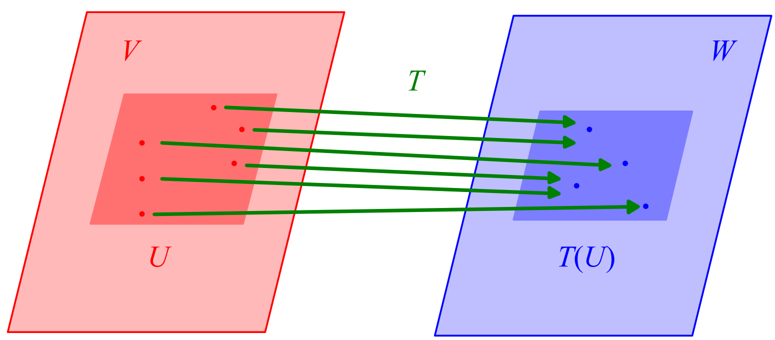
Preservation of Subspaces

The Additivity and Homogeneity Properties of linear transformations allow us to prove that linear transformations on abstract vector spaces map a subspace of the domain into a subspace of the codomain, and vice versa:

Theorem — The Preservation of Subspaces Theorem:
Suppose that $T : V \to W$ is a linear transformation. Then, for any ***subspace*** $U \trianglelefteq V$:

$$T(U) = \left\{ \vec{w} \in W \,|\, \vec{w} = T(\vec{u}\,) \text{ for some } \vec{u} \in U \right\},$$

called the ***image*** of U under T, is a ***subspace*** of the ***codomain*** W.

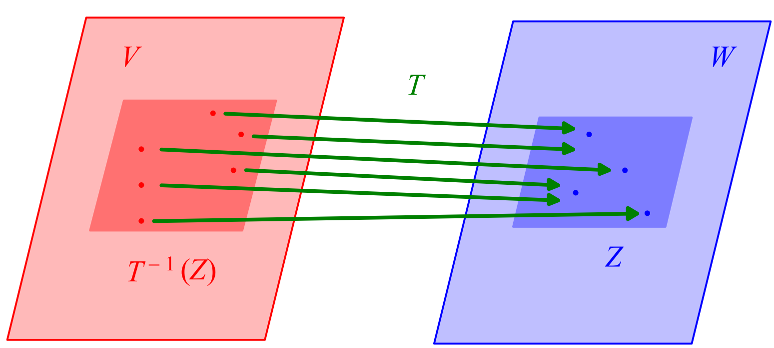


$T(U)$ is the ***Image*** Under T of a Subspace U of V

Similarly, for any subspace $Z \trianglelefteq W$:

$$T^{-1}(Z) = \left\{ \vec{v} \in V \,|\, T(\vec{v}\,) = \vec{z} \text{ for some } \vec{z} \in Z \right\},$$

called the ***pre-image*** of Z under T, is a ***subspace*** of the ***domain*** V.



$T^{-1}(Z)$ is the ***Pre-Image*** Under T of a Subspace Z of W

We note that $T(U)$ and $T^{-1}(Z)$, in general, are both ***sets*** of vectors, not just single vectors. In fact, as soon as one of them contains a non-zero vector, it becomes an ***infinite*** set of vectors.

Also, notice that the two diagrams are virtually identical, except for the labelling. However, $T(U)$ represents ***where U goes to in W***, and $T^{-1}(Z)$ represents ***which vectors in V go to Z***.

Proof: We will show that $T(U)$ is a subspace of $\mathbb{R}^m$ if U is a subspace of $\mathbb{R}^n$, and leave $T^{-1}(Z)$ as an Exercise. We must show that if $\vec{w}_1$ and $\vec{w}_2$ are members of $T(U)$ and k is any scalar, then:

$$\vec{w}_1 + \vec{w}_2 \in T(U), \quad \text{and} \quad k\vec{w}_1 \in T(U)$$

Both properties will follow directly from the linearity property of T and the closure property of a subspace. By definition, $\vec{w}_1 = T(\vec{v}_1)$ and $\vec{w}_2 = T(\vec{v}_2)$ for some vectors $\vec{v}_1, \vec{v}_2 \in V$. Thus:

$$\vec{w}_1 + \vec{w}_2 = T(\vec{v}_1) + T(\vec{v}_2) = T(\vec{v}_1 + \vec{v}_2).$$

Since U is a subspace of V, $\vec{v}_1 + \vec{v}_2 \in U$, and therefore $\vec{w}_1 + \vec{w}_2 \in T(U)$ as well.

Similarly, since $k\vec{v}_1 \in U$:

$$k\vec{w}_1 = kT(\vec{v}_1) = T(k\vec{v}_1) \in T(U). \blacksquare$$

In Chapters 2 and 3, we saw two simple examples of the image and pre-image of subspaces. If $T : V \to W$ is any linear transformation, then we call $T(V)$ the ***range*** of T. Similarly, $T^{-1}\left(\left\{\vec{\mathbf{0}}_W\right\}\right)$ is the ***kernel*** of T. In the case where V and W are finite-dimensional vector spaces, we already saw how to find a basis for these two subspaces using the matrix of T with respect to a basis B for V and a basis $B^{/}$ for W. Let us now see how to do this for more general subspaces.

The Image of a Subspace

The following Theorem tells us how to find a basis for $T(U)$ given a basis for U when T is a transformation of Euclidean spaces. The idea can be generalized if the domain or codomain are finite dimensional abstract vector spaces by using coordinate vectors and a basis for each space.

> ***Theorem:*** Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and suppose that U is any subspace of $\mathbb{R}^n$. Suppose that $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is a basis for U.
> Then: $T(B) = \{T(\vec{u}_1), T(\vec{u}_2), \ldots, T(\vec{u}_k)\}$ ***Spans*** $T(U)$.
>
> Thus, the output of ***The Minimizing Theorem*** applied to $T(B)$ will be a basis for $T(U)$.

The proof of this Theorem is easy and is left as an Exercise. Again, we are not insisting that $T(B)$ is a basis for $T(U)$, but by our Theorem in Section 4.2, this is true if T is ***one-to-one*** when restricted to U.

Example: Let us consider $T : \mathbb{R}^3 \to \mathbb{R}^4$ from the previous Section, given by:

$$[T] = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix}.$$

Let $U = Span(B)$, where $B = \{\langle 2,-5,-4\rangle, \langle 0, 1,-14\rangle\}$.

Clearly, B is linearly independent. We compute:

$$T(\langle 2,-5,-4\rangle) = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 2 \\ -5 \\ -4 \end{bmatrix} = \begin{bmatrix} -24 \\ 5 \\ -23 \\ -19 \end{bmatrix}, \text{ and}$$

$$T(\langle 0, 1,-2\rangle) = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ -1 \\ -14 \end{bmatrix} = \begin{bmatrix} -72 \\ 15 \\ -69 \\ -57 \end{bmatrix}.$$

Thus $T(U)$ is Spanned by $\{\langle -24, 5,-23,-19\rangle, \langle -72, 15,-69,-57\rangle\}$. A quick check reveals that these two vectors are in fact ***parallel*** (by a factor of 3), so $T(U)$ is only 1-dimensional, and:

$$T(U) = Span(\{\langle -24, 5,-23,-19\rangle\}). \square$$

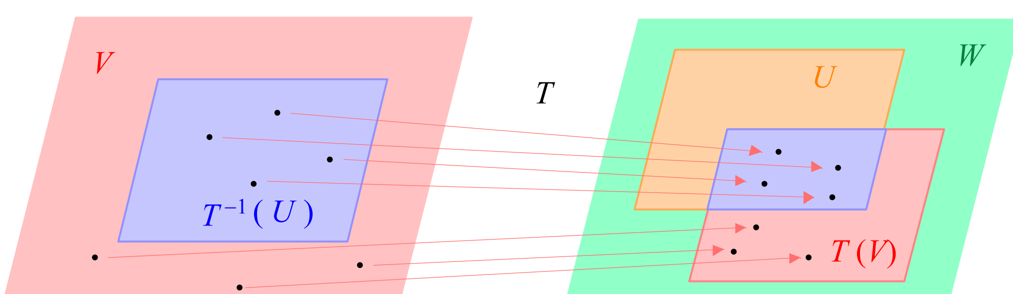
The Preimage of a Subspace

We know that if $T : V \to W$ is a linear transformation, and if U is a subspace W, then the preimage $T^{-1}(U)$ is a subspace of V. However, it is not an easy task to find a basis for the preimage of U, even if we are given a basis for U. Let us look at the issues involved in this task and see how we can go about finding a basis for this preimage.

Let us look at an arbitrary member $\vec{u} \in U$. We must find all vectors $\vec{v} \in V$ such that $T(\vec{v}) = \vec{u}$. But then, by definition, $T(\vec{v}) \in T(V) = range(T)$ also, and thus $\vec{u} = T(\vec{v}) \in T(V)$ also. Thus, we only need to think about the vectors in the intersection $U \cap T(V)$. It is therefore natural to focus our attention on the restricted transformation:

$$T|_{T^{-1}(U)} : T^{-1}(U) \to U \cap T(V),$$

as illustrated by the following diagram:

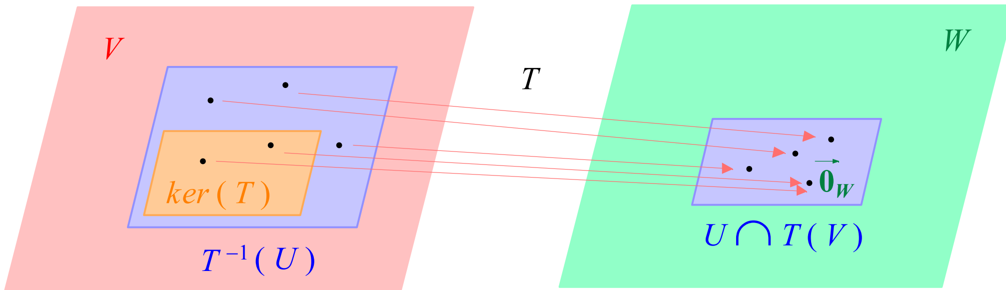


The Image Under T of $T^{-1}(U)$ is the Intersection of U and $T(V)$

Next, remember that the zero vector is a member of any subspace, and thus $\vec{\mathbf{0}}_W \in U \cap T(V)$. Thus

$ker(T) = T^{-1}\left(\left\{\vec{\mathbf{0}}_W\right\}\right)$ must be a subspace of $T^{-1}(U)$. We can find a basis $\left\{\vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ for $ker(T)$, where $s = nullity(T)$, and we reiterate that this is a subset of $T^{-1}(U)$.

We note that if T is ***one-to-one***, then $s = 0$ and this step in the algorithm just gives us the empty set.



$T^{-1}(U)$ Contains The Kernel of T

Thus, this basis for $ker(T)$ will be part of our basis for $T^{-1}(U)$. Obviously, there could be other members of $T^{-1}(U)$ that are ***not*** members of $ker(T)$, as shown above, so we must think some more about how to find a full basis for $T^{-1}(U)$.

The Dimension Theorem for Linear Transformations say that:

$$rank\left(T|_{T^{-1}(U)}\right) + nullity\left(T|_{T^{-1}(U)}\right) = dim(T^{-1}(U)).$$

Since $ker(T) \trianglelefteq T^{-1}(U)$, $nullity\left(T|_{T^{-1}(U)}\right) = nullity(T)$. Also, by definition, $T|_{T^{-1}(U)}$ is ***onto*** $U \cap T(V)$, and thus we have:

$$dim(U \cap T(V)) + nullity(T) = dim(T^{-1}(U)).$$

We saw in the previous Section how to find a basis for the intersection of two subspaces, so suppose $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ is a basis for $U \cap T(V)$, where $r = dim(U \cap T(V))$. By definition, for every $\vec{u}_i \in B$, we can find a vector $\vec{v}_i \in T^{-1}(U)$ such that $T(\vec{v}_i) = \vec{u}_i$. Hence we produce a set of vectors $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\} \subseteq T^{-1}(U)$.

We note that if $U \cap T(V) = \left\{\vec{\mathbf{0}}_W\right\}$, then $r = 0$ and this step just gives us the empty set.

Let us put the two sets we constructed together and consider:

$$B' = \left\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\} \subseteq T^{-1}(U).$$

This set has $r + s = dim(U \cap T(V)) + nullity(T) = dim(T^{-1}(U))$ members, and thus we have a good candidate for a basis for $T^{-1}(U)$. We note that if $r = s = 0$, then we get that $T^{-1}(U) = \left\{\vec{\mathbf{0}}_V\right\}$ so $T^{-1}(U)$ has no basis.

By the ***Two-for-One Theorem***, we can prove that B' is a ***basis*** for $T^{-1}(U)$ by showing that it is ***linearly independent***. Consider the dependence test equation:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_r\vec{v}_r + d_1\vec{k}_1 + d_2\vec{k}_2 + \cdots + d_s\vec{k}_s = \vec{\mathbf{0}}_V.$$

We have to show that all of the coefficients are zero. Let us apply T to both sides of this equation:

$$T\left(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_r\vec{v}_r + d_1\vec{k}_1 + d_2\vec{k}_2 + \cdots + d_s\vec{k}_s\right) = T\left(\vec{\mathbf{0}}_V\right) = \vec{\mathbf{0}}_W, \quad \text{so}$$

$$c_1T(\vec{v}_1) + c_2T(\vec{v}_2) + \cdots + c_rT(\vec{v}_r) + T\left(d_1\vec{k}_1 + d_2\vec{k}_2 + \cdots + d_s\vec{k}_s\right) = \vec{\mathbf{0}}_W.$$

But $d_1\vec{k}_1 + d_2\vec{k}_2 + \cdots + d_s\vec{k}_s \in ker(T)$ and $T(\vec{v}_i) = \vec{u}_i$, so this last equation can be rewritten as:

$$c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_r\vec{u}_r = \vec{\mathbf{0}}_W.$$

However, $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ is a ***basis*** for $U \cap T(V)$, so c_1 through c_r must all be zero. Thus, our original dependence test equation reduces to:

$$d_1\vec{k}_1 + d_2\vec{k}_2 + \cdots + d_s\vec{k}_s = \vec{\mathbf{0}}_V.$$

But this time, since $\left\{\vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ is a ***basis*** for $ker(T)$, d_1 through d_s must all be zero. Thus B' is linearly independent and is a basis for the preimage of U.

We summarize this discussion in the following Theorem, and to simplify our computations, we focus on the case when the domain and codomain are Euclidean spaces. Again the ideas can be generalized to arbitrary vector spaces with the use of coordinate vectors:

Theorem: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and $U \trianglelefteq \mathbb{R}^m$.
The following algorithm will produce a basis for $T^{-1}(U)$, given as its input the standard matrix $[T]$ and a basis B for U:

Step 1. Find the rref of $[T]$.
Step 2. Use the rref of $[T]$ to find a basis for $T(\mathbb{R}^n) = colspace([T])$
and a basis $\left\{\vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ for $ker(T)$.
Step 3. Find a basis $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ for $U \cap T(\mathbb{R}^n)$ using the techniques from Section 4.1.
Step 4. For each $\vec{u}_i$, find any vector $\vec{v}_i \in T^{-1}(U)$ such that $T(\vec{v}_i) = \vec{u}_i$.
We can accomplish this simultaneously by solving the system:

$$[\,[T] \mid \vec{u}_1 \mid \vec{u}_2 \mid \ldots \mid \vec{u}_r\,].$$

Step 5. The combined set $B' = \left\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ is a basis for $T^{-1}(U)$.
We note that if $r = 0$ or $s = 0$, the corresponding set is empty.

Example: Let us go back to our previous Example, $T : \mathbb{R}^3 \to \mathbb{R}^4$, given by:

$$[T] = \begin{bmatrix} 3 & 2 & 5 \\ -2 & -1 & -1 \\ -4 & -1 & 5 \\ 1 & 1 & 4 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Let us find a basis for the preimage of $U = Span(S) \trianglelefteq \mathbb{R}^4$, where:

$$S = \{\langle 2, 1, 7, 5\rangle, \langle 0, 3, 11, 7\rangle\}.$$

S is clearly a basis because the two vectors are not parallel to each other.

We see from the rref of $[T]$ that $T(\mathbb{R}^3) = range(T)$ has basis consisting of the first two columns of $[T]$:

$$\{\langle 3, -2, -4, 1\rangle, \langle 2, -1, -1, 1\rangle\},$$

and a basis for $ker(T)$ is $\{\langle 3, -7, 1\rangle\}$.

Next, we must find a basis for $U \cap T(\mathbb{R}^3)$, so we assemble separately into the ***rows*** of two matrices our two bases above:

$$C = \begin{bmatrix} 2 & 1 & 7 & 5 \\ 0 & 3 & 11 & 7 \end{bmatrix} \text{ and } D = \begin{bmatrix} 3 & -2 & -4 & 1 \\ 2 & -1 & -1 & 1 \end{bmatrix}.$$

Their rrefs are:

$$\begin{bmatrix} 1 & 0 & \frac{5}{3} & \frac{4}{3} \\ 0 & 1 & \frac{11}{3} & \frac{7}{3} \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & 5 & 1 \end{bmatrix}.$$

From these, the two nullspaces are:

$$Span,(\{\langle -5,-11,3,0\rangle, \langle -4,-7,0,3\rangle\}) \text{ and}$$
$$Span(\{\langle -2,-5,1,0\rangle, \langle -1,-1,0,1\rangle\}).$$

We assemble these two bases into the ***rows*** of a single matrix:

$$\begin{bmatrix} -5 & -11 & 3 & 0 \\ -4 & -7 & 0 & 3 \\ -2 & -5 & 1 & 0 \\ -1 & -1 & 0 & 1 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 0 & -4/3 \\ 0 & 1 & 0 & 1/3 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus $U \cap T(\mathbb{R}^3)$ is only 1-dimensional, with basis $\{\langle 4,-1,3,3\rangle\}$. To find a preimage for this single vector, we need to solve the augmented system using $[T]$:

$$\left[\begin{array}{ccc|c} 3 & 2 & 5 & 4 \\ -2 & -1 & -1 & -1 \\ -4 & -1 & 5 & 3 \\ 1 & 1 & 4 & 3 \end{array}\right] \text{ with rref } \left[\begin{array}{ccc|c} 1 & 0 & -3 & -2 \\ 0 & 1 & 7 & 5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

Thus, one particular preimage could be $\langle -2,5,0\rangle$.

Finally, a basis for $T^{-1}(U)$ consists of this final vector, and our single basis vector for the kernel:

$$T^{-1}(U) = Span(\{\langle -2,5,0\rangle, \langle 3,-7,1\rangle\}),$$

Thus, $T^{-1}(U)$ is a 2-dimensional subspace of $\mathbb{R}^3$. □

4.3 Section Summary

Theorem — The Preservation of Subspaces Theorem:

Suppose that $T : V \to W$ is a linear transformation. Then, for any subspace $U \trianglelefteq V$:

$$T(U) = \left\{ \vec{w} \in W \mid \vec{w} = T(\vec{u}) \text{ for some } \vec{u} \in U \right\},$$

called the ***image*** of U under T, is a subspace of the codomain W. Similarly, for any subspace $Z \trianglelefteq W$:

$$T^{-1}(Z) = \left\{ \vec{v} \in V \mid T(\vec{v}) = \vec{z} \text{ for some } \vec{z} \in Z \right\},$$

called the ***pre-image*** of Z under T, is a subspace of the domain V.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and suppose that V is any subspace of $\mathbb{R}^n$. Suppose that B is a basis for V. Then $T(B)$ ***Spans*** $T(V)$. Thus, the output of ***The Minimizing Theorem*** applied to $T(B)$ will be a basis for $T(V)$.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and $U \trianglelefteq \mathbb{R}^m$.

The following algorithm will produce a basis for $T^{-1}(U)$, given as its input the standard matrix $[T]$ and a basis B for U:

Step 1. Find the rref of $[T]$.

Step 2. Use the rref of $[T]$ to find a basis for $T(\mathbb{R}^n) = colspace([T])$
and a basis $\left\{\vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ for $ker(T)$.

Step 3. Find a basis $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ for $U \cap T(\mathbb{R}^n)$ using the techniques from the previous Section.

Step 4. For each $\vec{u}_i$, find any vector $\vec{v}_i \in T^{-1}(U)$ such that $T(\vec{v}_i) = \vec{u}_i$.

We can accomplish this simultaneously by solving the system:

$$[\,[T] \mid \vec{u}_1 \mid \vec{u}_2 \mid \ldots \mid \vec{u}_r\,].$$

Step 5. The combined set $B' = \left\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ is a basis for $T^{-1}(U)$.

We note that if $r = 0$ or $s = 0$, the corresponding set is empty.

4.3 Exercises

For Exercises (1) to (15): The following linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$, each with indicated standard matrices $[T]$ and rref R, are the same ones in the corresponding Exercises from Section 4.2. For each of them: (a) find a basis for $T(V)$, where V is the indicated subspace of $\mathbb{R}^n$, and (b) find a basis for $T^{-1}(U)$, where U is the indicated subspace of $\mathbb{R}^m$. Aside from the basis for the range of T that you found in Section 4.2, you will also need to find a basis for $ker(T)$. It is highly recommended that technology be used to find the rrefs needed in the computations.

1. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 3 & 2 & -2 \\ 5 & 3 & -1 \\ 4 & 2 & 2 \\ -1 & -1 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$,

   $V = Span(\{\langle -2,7,9\rangle, \langle -3,7,5\rangle\})$, and $U = Span(\{\langle 5,8,6,-2\rangle\})$.

2. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 2 & -6 & -7 \\ -3 & 9 & -1 \\ -4 & 12 & 9 \\ 5 & -15 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$,

   $V = Span(\{\langle 2,-1,0\rangle, \langle 6,2,5\rangle\})$, and $U = Span(\{\langle -3,-7,1,13\rangle\})$.

3. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 2 & -6 & -7 \\ -3 & 9 & -1 \\ -4 & 12 & 9 \\ 5 & -5 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$,

$V = Span(\{\langle 2,-1,0\rangle, \langle 6,2,5\rangle\})$, and $U = Span(\{\langle -5,-4,5,8\rangle, \langle 0,0,0,1\rangle\})$.

4. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 3 & 5 & 4 & -1 \\ 2 & 3 & 2 & -1 \\ -2 & -1 & 2 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & -2 & -2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

$V = Span(\{\langle 3,-1,1,-1\rangle, \langle 3,-2,-4,-2\rangle, \langle 1,0,1,1\rangle\})$, and $U = Span(\{\langle 2,1,1\rangle, \langle 11,7,-5\rangle\})$.

5. $T : \mathbb{R}^3 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 3 & 5 & -1 \\ 2 & 3 & -1 \\ -2 & -1 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$,

$V = Span(\{\langle 3,-1,-1\rangle, \langle 1,0,0\rangle\})$, and $U = Span(\{\langle 2,1,1\rangle, \langle 1,0,0\rangle\})$.

6. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 2 & 10 & 5 & -7 \\ 3 & 15 & 7 & -9 \\ -4 & -20 & -9 & 11 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 4 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

$V = Span(\{\langle -2,3,0,2\rangle, \langle -5,1,2,-6\rangle, \langle -2,0,2,-1\rangle\})$, and $U = Span(\{\langle 3,4,-5\rangle\})$.

7. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 2 & 10 & 5 & -7 \\ 3 & 15 & 7 & -9 \\ -4 & -20 & -9 & 8 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$,

$V = Span(\{\langle 3,-1,1,-1\rangle, \langle 4,-3,4,1\rangle, \langle 3,2,5,4\rangle\})$, and $U = Span(\{\langle 3,4,-5\rangle, \langle 2,-1,7\rangle\})$.

8. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with

$[T] = \begin{bmatrix} -5 & -20 & -4 & 1 & 2 \\ 3 & 12 & -2 & -27 & 12 \\ 2 & 8 & 3 & 8 & -5 \\ -4 & -16 & 1 & 26 & -11 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 4 & 0 & -5 & 2 \\ 0 & 0 & 1 & 6 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$,

$V = Span(\{\langle 2,5,-6,-4,2\rangle, \langle -1,4,4,1,0\rangle, \langle 6,4,-7,-10,5\rangle\})$, and $U = Span(\{\langle 1,-5,1,5\rangle\})$.

9. $T : \mathbb{R}^4 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} -5 & -20 & -4 & 1 \\ 3 & 12 & -2 & -27 \\ 2 & 8 & 3 & 8 \\ -4 & -16 & 1 & 26 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 4 & 0 & -5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

$V = Span(\{\langle -3,5,1,1\rangle, \langle 6,4,-7,-10\rangle, \langle -2,5,-5,2\rangle\})$, and $U = Span(\{\langle -9,1,5,-3\rangle\})$.

10. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with

$$[T] = \begin{bmatrix} -5 & 10 & -30 & -3 & 5 \\ 2 & -4 & 12 & -3 & -23 \\ -3 & 6 & -18 & 2 & 22 \\ 4 & -8 & 24 & -5 & -41 \end{bmatrix}, \; R = \begin{bmatrix} 1 & -2 & 6 & 0 & -4 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$V = Span(\{\langle 2,-4,12,3,7\rangle, \langle -2,15,-35,-14,-20\rangle\})$, and $U = Span(\{\langle -2,5,-5,9\rangle\})$.

11. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with

$$[T] = \begin{bmatrix} -2 & 6 & -3 & -5 & -27 \\ 3 & -9 & 7 & 4 & 6 \\ 4 & -12 & -1 & -2 & -2 \\ -5 & 15 & 2 & 3 & 3 \end{bmatrix}, \; R = \begin{bmatrix} 1 & -3 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$V = Span(\{\langle -1,3,1,1,1\rangle, \langle 1,3,-3,8,0\rangle, \langle 4,-2,-1,-1,-1\rangle\})$, and $U = Span(\{\langle -2,5,-5,9\rangle\})$.

12. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with

$$[T] = \begin{bmatrix} -2 & 6 & -3 & -5 & -7 \\ 3 & -9 & 7 & 4 & 6 \\ 4 & -12 & -1 & -2 & -2 \\ -5 & 15 & 2 & 3 & 3 \end{bmatrix}, R = \begin{bmatrix} 1 & -3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$V = Span(\{\langle 2,-2,3,7,4\rangle, \langle 4,7,-3,5,2\rangle, \langle 5,-1,3,7,4\rangle\})$, and

$U = Span(\{\langle 1,0,0,0\rangle, \langle 0,1,0,0\rangle, \langle 0,0,0,1\rangle\})$.

13. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} 2 & 10 & -1 & -10 \\ -4 & -20 & 1 & 14 \\ 3 & 15 & 1 & 0 \\ 5 & 25 & -3 & -28 \\ -6 & -30 & 2 & 24 \end{bmatrix}, \; R = \begin{bmatrix} 1 & 5 & 0 & -2 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$

$V = Span(\{\langle 3,2,0,-1\rangle, \langle 3,5,-6,-1\rangle, \langle 2,0,-5,7\rangle, \langle -4,0,13,4\rangle\})$, and

$U = Span(\{\langle 3,-7,7,7,-10\rangle\})$.

14. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} 2 & 10 & -1 & -9 \\ -4 & -20 & 1 & 14 \\ 3 & 15 & 1 & 0 \\ 5 & 25 & -3 & -28 \\ -6 & -30 & 2 & 24 \end{bmatrix}, \; R = \begin{bmatrix} 1 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$

$V = Span(\{\langle -10,2,3,2\rangle, \langle 15,-3,8,5\rangle, \langle 0,0,4,3\rangle\})$, and

$U = Span(\{\langle 3,-5,2,8,-8\rangle, \langle -11,16,2,-34,28\rangle, \langle -2,1,1,-3,2\rangle\})$.

15. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} -2 & 1 & 1 & 2 \\ 5 & -1 & -1 & -8 \\ 1 & 1 & -1 & 6 \\ -2 & -2 & 1 & -7 \\ -1 & 1 & 1 & 0 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$,

$V = Span(\{\langle 3,-3,5,-1\rangle, \langle 2,-2,5,4\rangle, \langle 2,-2,6,-1\rangle\})$, and

$U = Span(\{\langle -1,5,1,-2,-1\rangle, \langle 1,0,1,-2,1\rangle, \langle 1,1,-1,1,1\rangle\})$.

16. Suppose that $T : V \to W$ is a linear transformation. Prove that for any subspace $Z \trianglelefteq W$, the preimage of Z under T:

$$T^{-1}(Z) = \left\{\vec{v} \in V \mid T(\vec{v}) = \vec{z} \text{ for some } \vec{z} \in Z\right\},$$

is a subspace of the domain V.

17. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and suppose that U is any subspace of $\mathbb{R}^n$. Suppose that B is a basis for U. Then $T(B)$ ***Spans*** $T(U)$. Thus, the output of The Minimizing Theorem applied to $T(B)$ will be a basis for $T(U)$.

18. Let us construct the following subspaces of $\mathbb{R}^n$:

$$\begin{aligned} W_1 &= Span(\{\vec{e}_1\}), \\ W_2 &= Span(\{\vec{e}_1, \vec{e}_2\}), \ \ldots, \\ W_k &= Span(\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_k\}), \ \ldots, \\ W_n &= Span(\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}) = \mathbb{R}^n. \end{aligned}$$

These subspaces form an ***ascending sequence*** of subspaces or ***flag:***

$$W_1 \subseteq W_2 \subseteq W_3 \subseteq \cdots \subseteq W_n.$$

Show that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation whose matrix is upper triangular ***if and only if*** $T(W_k) \subseteq W_k$ for all $k = 1..n$. In other words, the condition says that $T(\vec{e}_k)$ can be expressed as a linear combination of only $\vec{e}_1$ through $\vec{e}_k$, without involving $\vec{e}_{k+1}$ through $\vec{e}_n$.

19. This Exercise is the analog of the previous one for lower triangular matrices. We construct the following subspaces of $\mathbb{R}^n$:

$$\begin{aligned} V_n &= Span(\{\vec{e}_n\}), \\ V_{n-1} &= Span(\{\vec{e}_{n-1}, \vec{e}_n\}), \ \ldots, \\ V_k &= Span(\{\vec{e}_k, \ldots, \vec{e}_{n-1}, \vec{e}_n\}), \ \ldots, \\ V_1 &= Span(\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_{n-1}, \vec{e}_n\}) = \mathbb{R}^n. \end{aligned}$$

These subspaces form a ***decreasing sequence*** of subspaces or ***reverse flag:***

$$V_1 \supseteq V_2 \supseteq V_3 \supseteq \ldots \supseteq V_{n-1} \supseteq V_n.$$

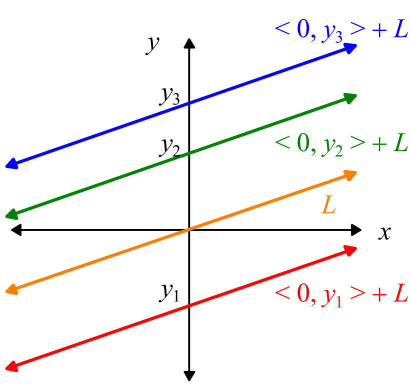
Show that $T : \mathbb{R}^n \to \mathbb{R}^n$ is a linear transformation whose matrix is lower triangular ***if and only if*** $T(V_k) \subseteq (V_k)$ for all $k = 1..n$. In other words, the condition says that $T(\vec{e}_k)$ can be expressed as a linear combination of only $\vec{e}_k$ through $\vec{e}_n$, without involving $\vec{e}_1$ through $\vec{e}_{k-1}$.

4.4 Cosets and Quotient Spaces

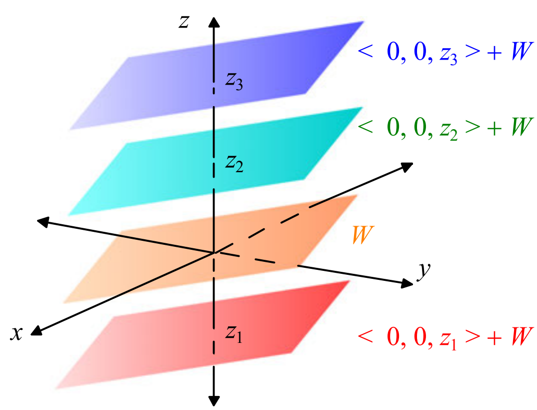
This Section gives the background material that will be used in the next Section where we will see ***The Three Isomorphism Theorems***. All these Theorems refer to what are called ***quotient spaces***, and the members of these spaces are called ***cosets***, so we begin with the following:

> ***Definition:*** Let W be a subspace of a vector space V. A ***coset*** X of W is another word for a ***translate*** of W:
>
> $$X = \vec{v} + W = \{\, \vec{v} + \vec{w} \mid \vec{w} \in W \,\},$$
>
> for some ***fixed*** vector $\vec{v} \in V$. We call $\vec{v}$ a ***representative*** of the coset $\vec{v} + W$, or say that the coset $\vec{v} + W$ is ***represented*** by $\vec{v}$.

Cosets, therefore, are nothing new to us. Cosets appear most commonly as any line in the Cartesian plane or Cartesian space, or any plane in Cartesian space: they are all translates of lines or planes that pass through the origin, which are of course the non-trivial subspaces of $\mathbb{R}^2$ and $\mathbb{R}^3$.



The Cosets of a Line L in $\mathbb{R}^2$
Which Passes through the Origin



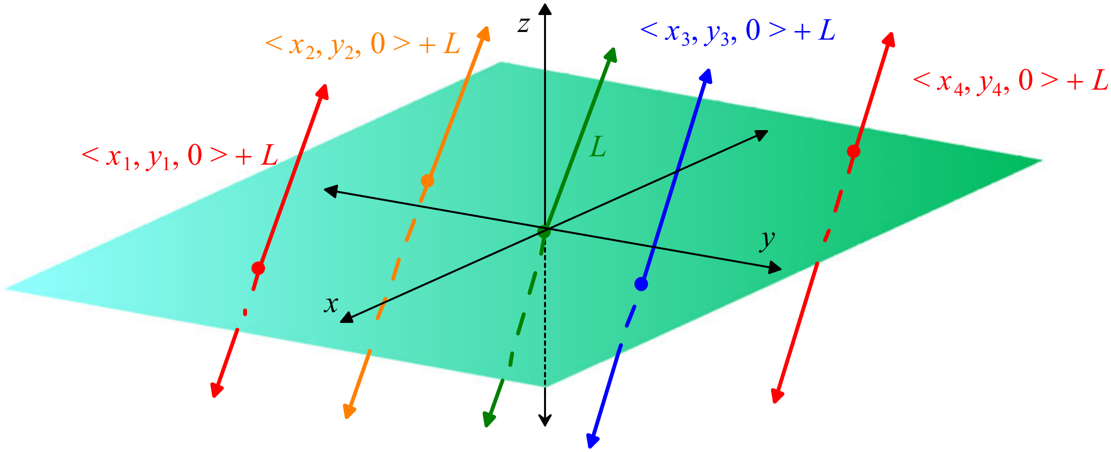
The Cosets of a Plane W in $\mathbb{R}^3$
Which Passes through the Origin

Notice also that since $\vec{\mathbf{0}}_V \in W$, then $\vec{v} + \vec{\mathbf{0}}_V = \vec{v} \in X$, so a ***representative*** for X must be a ***member*** of X. The coset $\vec{\mathbf{0}}_V + W$ is also written as W.

Suppose that $A\vec{x} = \vec{b}$ is a non-homogeneous system of linear equations, where A is the $m \times n$ coefficient matrix of the system. Cosets appear as the set of all solutions X to this system:

$$X = \vec{x}_p + nullspace(A),$$

where $nullspace(A)$ is a subspace of $\mathbb{R}^n$, and $\vec{x}_p$ is a ***particular solution*** to this system. This means that any solution $\vec{x}$ to this system must have the form $\vec{x} = \vec{x}_p + \vec{x}_h$, where $A\vec{x}_h = \vec{\mathbf{0}}_n$.



The Cosets of a Line L in $\mathbb{R}^3$ Which Passes through the Origin

We can find $\vec{x}_p$ by using the rref of the augmented matrix $\left[A \mid \vec{b}\right]$ and setting all the free variables to 0. However, there are many ways to describe all the solutions: if $\vec{x}_2 = \vec{x}_p + \vec{x}_h$ is another solution to the non-homogeneous system, as described above, then we can also say $X = \vec{x}_2 + nullspace(A)$. But notice that $\vec{x}_2 - \vec{x}_p = \vec{x}_0 \in nullspace(A)$. This is exactly the idea we need to decide when two cosets are actually the same set:

Theorem: Let $X = \vec{v} + W$ and $Y = \vec{u} + W$ be cosets of $W \trianglelefteq V$. Then:

$$X = Y \textit{ if and only if } \vec{v} - \vec{u} \in W.$$

In particular:

$$\vec{v} + W = \vec{\mathbf{0}}_V + W = W \textit{ if and only if } \vec{v} \in W.$$

We will refer to this last statement as ***The Absorption Rule***.

Proof: ($\Rightarrow$) Suppose $X = Y$. This means that every member of X is also a member of Y, and vice versa. Thus, a member $\vec{v} + \vec{w}_1$ of X must also be of the form $\vec{u} + \vec{w}_2$, where $\vec{w}_1$ and $\vec{w}_2$ are two (possibly different) members of W. In other words:

$$\vec{v} + \vec{w}_1 = \vec{u} + \vec{w}_2.$$

However, the closure property of W tells us that:

$$\vec{v} - \vec{u} = \vec{w}_2 - \vec{w}_1 \in W.$$

($\Leftarrow$) Suppose $\vec{v} - \vec{u} = \vec{w}_1 \in W$. Then $\vec{v} = \vec{u} + \vec{w}_1$, and thus:

$$\begin{aligned} X &= \vec{v} + W \\ &= \{\vec{v} + \vec{w} \mid \vec{w} \in W\} \\ &= \{\vec{u} + \vec{w}_1 + \vec{w} \mid \vec{w} \in W\} \\ &= \{\vec{u} + \vec{w}_2 \mid \vec{w}_2 \in W\} \\ &= Y, \end{aligned}$$

where again, we used the closure property of W by ***renaming*** the sum $\vec{w}_1 + \vec{w}$ as the vector $\vec{w}_2 \in W$. Notice that since $\vec{w}$ is any vector in W, then $\vec{w}_1 + \vec{w}$ is likewise any vector in W. ■

Notice also that we deliberately used subscripted notation to avoid confusion: the $\vec{w}$ in the definition of the coset X is allowed to be ***any*** member of W but the $\vec{w}_1$ in the equation $\vec{v} - \vec{u} = \vec{w}_1$ is a ***fixed*** member $\vec{w}_1 \in W$.

Example: Let W be the plane in $\mathbb{R}^3$ with Cartesian equation $3x - 5y + 2z = 0$. W passes through the origin, so indeed it is a subspace of $\mathbb{R}^3$. As mentioned above, every coset:

$$X = \langle a, b, c \rangle + W$$

is a translate of W, in this case a plane ***parallel*** to W that passes through the point (a, b, c) in Cartesian space. Thus the coset:

$$X = \langle 0, 0, -6 \rangle + W$$

is the plane parallel to W with z-intercept at -6. Its equation must be:

$$3x - 5y + 2z = 0 - 0 + 2(-6) = -12.$$

Notice that the x-intercept of this plane is $-12/3 = -4$, so this coset must be exactly the same as:

$$X = \langle -4, 0, 0 \rangle + W.$$

In other words, shifting W ***down*** by 6 units yields exactly the same plane as shifting W ***left*** by 4 units. By our Theorem above, we must have:

$$\langle 0, 0, -6 \rangle - \langle -4, 0, 0 \rangle = \langle 4, 0, -6 \rangle \in W,$$

and we can indeed check that:

$$3(4) - 5(0) + 2(-6) = 0,$$

proving that $\langle 4, 0, -6 \rangle$ is a vector on the plane $3x - 5y + 2z = 0$. □

This last Theorem also gives us a democratic consequence, whose proof we leave as an Exercise:

> ***Theorem:*** Let W be a subspace of a vector space V. Then:
>
> $$\vec{x} \in \vec{v} + W \textbf{\textit{ if and only if }} \vec{x} + W = \vec{v} + W.$$

In other words, ***any*** member of $\vec{v} + W$ can serve as a representative for $\vec{v} + W$, and any representative of $\vec{v} + W$ must also be a member of $\vec{v} + W$.

Quotient Spaces

The Three Isomorphism Theorems, as they apply to Linear Algebra, all involve what are called ***quotient spaces:***

> ***Definition:*** Let W be a ***subspace*** of a vector space V. The ***quotient space*** $\boldsymbol{V/W}$, pronounced "***V modulo W***," is the set of ***all cosets*** of W:
>
> $$\boldsymbol{V/W} = \{X = \vec{v} + W \mid \vec{v} \in V\}.$$

We saw in our Example above that a single coset can be written in many different ways. Let us see how we can, in practice, systematically understand all the cosets of a single subspace:

Example: Let us generalize our previous Example, and suppose that W is any plane in $\mathbb{R}^3$ through the origin but does not the z-axis. Equivalently, its Cartesian equation has the form:

$$ax + by + cz = 0 \quad \text{where} \quad c \neq 0.$$

Every coset of this plane is a plane ***parallel*** to it, so it has Cartesian equation:

$$ax + by + cz = d.$$

This plane has a unique z-intercept, namely $(0, 0, d/c)$. Since $d/c = z_0 \in \mathbb{R}$, we can write V/W as:

$$V/W = \{\langle 0, 0, z_0 \rangle + W \mid z_0 \in \mathbb{R}\}.$$

In other words, we can say that every coset of V/W is ***represented*** by a unique z –intercept. We can also say that V/W is ***parametrized*** by the z-axis. Of course, there is nothing special about the z-axis. As long as L is any line that is ***not*** on W and passes through the origin, say with direction vector $\vec{d}$, then we can write:

$$V/W = \left\{ k\vec{d} + W \mid k \in \mathbb{R} \right\},$$

and so L likewise parametrizes V/W. □

Example: Let V be any vector space. We know that V has two trivial subspaces, namely V itself and $\left\{\vec{\mathbf{0}}_V\right\}$, the subspace consisting only of the zero vector. Let us see what the two corresponding quotient spaces look like:

$$V/V = \{\vec{v} + V \mid \vec{v} \in V\}.$$

But since $\vec{v} \in V$, we get $\vec{v} + V = V$ by the Absorption Principle. Thus there is exactly ***one coset***, and $V/V = \{V\}$.

On the other hand: $V/\left\{\vec{\mathbf{0}}_V\right\} = \left\{\vec{v} + \left\{\vec{\mathbf{0}}_V\right\} \mid \vec{v} \in V\right\}$.

Since the subspace $\left\{\vec{\mathbf{0}}_V\right\}$ contains only one vector: $\vec{v} + \left\{\vec{\mathbf{0}}_V\right\} = \{\vec{v}\}$.

Thus, each coset also contains only one vector, and we can rewrite our quotient space above as:

$$V/\left\{\vec{\mathbf{0}}_V\right\} = \{\{\vec{v}\} \mid \vec{v} \in V\}.$$

Thus, it "looks like" $V/\left\{\vec{\mathbf{0}}_V\right\}$ is just V. We will revisit this Example in Exercise 30 of Section 4.5 and phrase this conclusion more precisely. □

As the name implies, and as you have probably suspected, the quotient space V/W is a ***vector space***, and indeed there is a natural way to define the necessary vector arithmetic:

> ***Theorem:*** Let W be any subspace of a vector space V. The quotient space V/W is also a ***vector space***, where we define addition of ***cosets*** and scalar multiplication as follows:
> If $X = \vec{u} + W$ and $Y = \vec{v} + W$ are two cosets of V/W and $k \in \mathbb{R}$, then:
>
> $$X + Y = (\vec{u} + W) + (\vec{v} + W) = (\vec{u} + \vec{v}) + W, \quad \text{and}$$
> $$kX = k(\vec{u} + W) = k\vec{u} + W.$$

You might be thinking that the equations above don't quite look right, that is, we should get $(\vec{u}+\vec{v})+2W$ and $k\vec{u}+kW$ respectively, if the "normal rules" of arithmetic prevail. However, recall that W is a ***subspace*** of V, so we can think of $2W$ and kW as ***scalar multiples*** of the vectors of W, which we know by ***closure*** to be W (unless of course $k = 0$).

Proof of the Theorem: Before we even begin to show that the Ten Axioms of a vector space are satisfied, we first have to show that both addition and scalar multiplication are ***well-defined***. This is because there are many ways that we can express a coset, as we have seen above, so we must make sure that these operations do not depend on the particular ***choice*** of a representative. Thus, suppose:

$$X = \vec{u}_1 + W = \vec{u}_2 + W, \quad \text{and} \quad Y = \vec{v}_1 + W = \vec{v}_2 + W.$$

By our Theorem on the equality of cosets, we must have:

$$\vec{u}_1 - \vec{u}_2 = \vec{w}_1 \in W, \quad \text{and} \quad \vec{v}_1 - \vec{v}_2 = \vec{w}_2 \in W.$$

In other words, we have:

$$\vec{u}_1 = \vec{u}_2 + \vec{w}_1 \quad \text{and} \quad \vec{v}_1 = \vec{v}_2 + \vec{w}_2.$$

Now let us compute $X+Y$ in two different ways:

$$\begin{aligned} X+Y &= (\vec{u}_1 + W) + (\vec{v}_1 + W) \\ &= (\vec{u}_1 + \vec{v}_1) + W. \end{aligned}$$

But also:

$$\begin{aligned} X+Y &= (\vec{u}_1 + W) + (\vec{v}_1 + W) \\ &= (\vec{u}_2 + \vec{w}_1 + W) + (\vec{v}_2 + \vec{w}_2 + W) \\ &= (\vec{u}_2 + W) + (\vec{v}_2 + W) && \text{(by the Absorption Rule)} \\ &= (\vec{u}_2 + \vec{v}_2) + W. \end{aligned}$$

Thus, $(\vec{u}_1 + \vec{v}_1) + W = X+Y = (\vec{u}_2 + \vec{v}_2) + W$, so addition is well defined. Similarly, scalar multiplication is well defined, as you will prove in the Exercises.

Verifying the Ten Axioms is now fairly easy. The two ***closure*** properties are satisfied by our computations above, and the ***commutative*** and ***associative*** properties are ***inherited*** from the properties of addition in V. The ***zero vector*** of V/W should naturally be the coset:

$$\vec{\mathbf{0}}_{V/W} = \vec{\mathbf{0}}_V + W = W,$$

since we get:

$$(\vec{u} + W) + \left(\vec{\mathbf{0}}_V + W\right) = \left(\vec{u} + \vec{\mathbf{0}}_V\right) + W = \vec{u} + W.$$

for all $\vec{u} \in V$. Similarly, the ***additive inverse*** of $\vec{u} + W$ is naturally $-\vec{u} + W$. The rest of the properties are also inherited from the vector operations of V. ■

Example: Let us go back to our generic example of a plane W in $\mathbb{R}^3$ that does not pass through the z –axis. Using the z –intercepts as our representatives, we can "add" two planes using our definition:

$$(\langle 0, 0, z_1 \rangle + W) + (\langle 0, 0, z_2 \rangle + W) = (\langle 0, 0, z_1 \rangle + \langle 0, 0, z_2 \rangle) + W = \langle 0, 0, z_1 + z_2 \rangle + W.$$

Geometrically, this operation makes perfect sense, assuming that z_1 and z_2 are positive: if we translate W ***up*** by z_1, then ***up*** again by z_2, then we effectively translate W ***up*** by $z_1 + z_2$.

We can of course generalize this interpretation for negative intercepts. Similarly:

$$k(\langle 0, 0, z_1 \rangle + W) = k\langle 0, 0, z_1 \rangle + W = \langle 0, 0, kz_1 \rangle + W.$$

This time, the geometric effect is to ***scale*** the z-intercept z_1 by k. □

Basis and Dimension for V/W

Now that we know that V/W is actually also a vector space, a natural question to ask would be: how can we find a basis for V/W, and consequently, determine its dimension? Let us motivate this process with our example above, where W is any plane in $\mathbb{R}^3$ that contains the origin but does not pass through the z-axis. Every translate of W is determined uniquely by the z-intercept z_0. Thus, there is exactly one coset for one real number, so it would seem that V/W is 1-dimensional. That is indeed the case as we shall see below.

We will restrict our analysis to finite-dimensional vector spaces. Recall that if V has dimension n, then by repeated applications of The Extension Theorem, any linearly independent set S can be extended to a basis for V by including vectors one at a time while still maintaining independence (by choosing the next vector to be outside the Span of the previous vectors). Thus we have:

Theorem: Let W be an m-dimensional subspace of an n-dimensional vector space V, where $0 < m < n$. Suppose that $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is a basis for W (where B is empty if $W = \left\{\vec{\mathbf{0}}_V\right\}$). Since B is linearly independent, it can be ***extended*** to a basis $B^/$ for all of V, say:

$$B^/ = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m, \vec{v}_{m+1}, \vec{v}_{m+2}, \ldots, \vec{v}_n\},$$

using the ***Extension Theorem***. Then, the set of cosets:

$$B^{//} = \{\vec{v}_{m+1} + W, \vec{v}_{m+2} + W, \ldots, \vec{v}_n + W\}$$

forms a ***basis*** for V/W, and thus $dim(V/W) = n - m = dim(V) - dim(W)$.

In the case when $W = V$, and so $m = n$, the coset space $V/W = V/V$ consists of the single coset V, and thus $dim(V/V) = 0$.

The formula above for $dim(V/W)$ is therefore true in general.

Proof: Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be a basis for W, and extend B to a basis:

$$B^/ = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m, \vec{v}_{m+1}, \vec{v}_{m+2}, \ldots, \vec{v}_n\}$$

for V, as indicated above. (If we allow W to be the zero subspace, B is the empty set.) We must show that the set of cosets:

$$B^{//} = \{\vec{v}_{m+1} + W, \vec{v}_{m+2} + W, \ldots, \vec{v}_n + W\}$$

forms a basis for V/W, that is, this set ***Spans*** V/W and is ***linearly independent*** in V/W.

Let us prove the Spanning property: Suppose $\vec{v} + W$ is a coset in V/W. Since $B^/$ is a basis for V, we can write:

$$\vec{v} = c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_m\vec{w}_m + c_{m+1}\vec{v}_{m+1} + c_{m+2}\vec{v}_{m+2} + \cdots + c_n\vec{v}_n.$$

Now, using the definitions of coset addition and scalar multiplication, we have:

$$\begin{aligned}\vec{v} + W &= (c_1\vec{w}_1 + c_2\vec{w}_2 + \cdots + c_m\vec{w}_m + c_{m+1}\vec{v}_{m+1} + c_{m+2}\vec{v}_{m+2} + \cdots + c_n\vec{v}_n) + W \\ &= c_1(\vec{w}_1 + W) + c_2(\vec{w}_2 + W) + \cdots + c_m(\vec{w}_m + W) + \\ &\quad c_{m+1}(\vec{v}_{m+1} + W) + c_{m+2}(\vec{v}_{m+2} + W) + \cdots + c_n(\vec{v}_n + W) \\ &= (c_{m+1}\vec{v}_{m+1} + c_{m+2}\vec{v}_{m+2} + \cdots + c_n\vec{v}_n) + W \\ &= c_{m+1}(\vec{v}_{m+1} + W) + c_{m+2}(\vec{v}_{m+2} + W) + \cdots + c_n(\vec{v}_n + W)\end{aligned}$$

by the Absorption Rule and the definitions of coset addition and scalar multiplication in V/W. Thus $B^{//}$ indeed Spans V/W.

Similarly, we prove linear independence in V/W by considering the dependence test equation:

$$c_{m+1}(\vec{v}_{m+1} + W) + c_{m+2}(\vec{v}_{m+2} + W) + \cdots + c_n(\vec{v}_n + W) = \vec{\mathbf{0}}_{V/W} = W.$$

Then, again using the definitions of coset addition and scalar multiplication in V/W, we have:

$$\begin{aligned}W &= c_{m+1}(\vec{v}_{m+1} + W) + c_{m+2}(\vec{v}_{m+2} + W) + \cdots + c_n(\vec{v}_n + W) \\ &= (c_{m+1}\vec{v}_{m+1} + c_{m+2}\vec{v}_{m+2} + \cdots + c_n\vec{v}_n) + W,\end{aligned}$$

which would imply that $\vec{v} = c_{m+1}\vec{v}_{m+1} + c_{m+2}\vec{v}_{m+2} + \cdots + c_n\vec{v}_n \in W$ by the Absorption Rule. Thus, we can express $\vec{v}$ as a linear combination from B:

$$\vec{v} = c_{m+1}\vec{v}_{m+1} + c_{m+2}\vec{v}_{m+2} + \cdots + c_n\vec{v}_n = d_1\vec{w}_1 + d_2\vec{w}_2 + \cdots + d_m\vec{w}_m.$$

However, this now yields the equation:

$$-d_1\vec{w}_1 - d_2\vec{w}_2 - \cdots - d_m\vec{w}_m + c_{m+1}\vec{v}_{m+1} + c_{m+2}\vec{v}_{m+2} + \cdots + c_n\vec{v}_n = \vec{\mathbf{0}}_V,$$

and since $B^{/}$ is a basis for V, ***all*** the coefficients have to be 0. In particular, $c_{m+1}, c_{m+2}, \ldots, c_n$ must all be 0. Thus $B^{//}$ is linearly independent, and is a basis for V/W. ■

Note: We use the notation V/W because this is the notation that is used in ***Group Theory***, the area of mathematics where the ***Isomorphism Theorems*** are stated in full generality. However, the fact that $dim(V/W) = dim(V) - dim(W)$ tells us that a more suitable notation in the context of Linear Algebra would be $\boldsymbol{V - W}$ instead of $\boldsymbol{V/W}$, and this should be called the ***difference space*** instead of the ***quotient space***. However, even this context, we still refer to V/W as a quotient space.

Example: Suppose that W is ***any*** plane in $\mathbb{R}^3$ through the origin, with no other restrictions. Let us apply the theorem above to this situation. Since W is 2-dimensional, we can construct a basis for W:

$$B = \{\vec{w}_1, \vec{w}_2\},$$

consisting of any two non-zero vectors on W that are ***not parallel*** to each other, as we saw back in Chapter 1. Since we already have two vectors, the algorithm in the Theorem says that all we have to do is include one other vector that is ***not*** in W. The most natural vector to include would be a ***normal*** vector $\vec{n}$ for this plane. Thus:

$$B^{/} = \{\vec{w}_1, \vec{w}_2, \vec{n}\}$$

is a basis for $\mathbb{R}^3$. But according to our theorem, the set:

$$B^{//} = \{\vec{n} + W\}$$

consisting of this single coset, is a ***basis*** for $\mathbb{R}^3/W$, hence the quotient space is 1-dimensional. This should not come as a surprise because we already mentioned that V/W is parametrized by any line L that is not on W and passes through the origin, and we know that a line is 1-dimensional. □

4.4 Section Summary

Let W be a ***subspace*** of a vector space V. A ***coset*** X of W is a ***translate*** of W:

$$X = \vec{v} + W = \{\vec{v} + \vec{w} \mid \vec{w} \in W\}$$

for some fixed vector $\vec{v} \in V$. We call $\vec{v}$ a ***representative*** of the coset $\vec{v} + W$, or say that the coset $\vec{v} + W$ is ***represented*** by $\vec{v}$.

Equality of cosets: Let $X = \vec{v} + W$ and $Y = \vec{u} + W$ be cosets of $W \trianglelefteq V$. Then:

$$X = Y \ \textbf{\textit{if and only if}}\ \vec{v} - \vec{u} \in W.$$

The Absorption Rule: $\vec{v} + W = W = \vec{\mathbf{0}}_V + W$ ***if and only if*** $\vec{v} \in W$.

Membership in a coset is equivalent to representing the same coset:

$$\vec{x} \in \vec{v} + W \ \textbf{\textit{if and only if}}\ \vec{x} + W = \vec{v} + W.$$

The ***quotient space*** $\boldsymbol{V/W}$, pronounced "***V mod W***," is the set of ***all cosets*** of W:

$$\boldsymbol{V/W} = \{X = \vec{v} + W \mid \vec{v} \in V\}.$$

V/W is also a ***vector space***, where we define addition of cosets and scalar multiplication as follows: If $X = \vec{u} + W$ and $Y = \vec{v} + W$ are two cosets of V/W and $k \in \mathbb{R}$, then:

$$X + Y = (\vec{u} + W) + (\vec{v} + W) = (\vec{u} + \vec{v}) + W, \text{ and } kX = k(\vec{u} + W) = k\vec{u} + W.$$

Basis for $\boldsymbol{V/W}$: Let W be an m-dimensional subspace of an n-dimensional vector space V, where $0 < m < n$. Suppose that:$B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is a basis for W (where B is empty if $W = \left\{\vec{\mathbf{0}}_V\right\}$). Since B is linearly independent, it can be ***extended*** to a basis $B^{/}$ for all of V, say:

$$B^{/} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m, \vec{v}_{m+1}, \vec{v}_{m+2}, \ldots, \vec{v}_n\}.$$

Then, the set of cosets: $B^{//} = \{\vec{v}_{m+1} + W, \vec{v}_{m+2} + W, \ldots, \vec{v}_n + W\}$ is a ***basis*** for V/W.

Thus: $dim(V/W) = n - m = dim(V) - dim(W)$.

In the special case when $W = V$, the coset space $V/W = V/V$ consists of the single coset V, and thus $dim(V/V) = 0$. The formula above for $dim(V/W)$ is therefore true in general.

4.4 Exercises

For Exercises (1) to (5): Use the Absorption Rule to decide whether or not each indicated coset $\vec{v} + W$ is equal to the subspace W.

1. $\langle 10, 6\rangle + W$, where W is the line $y = \frac{3}{5}x$ in $\mathbb{R}^2$.
2. $\langle 21, 12\rangle + W$, where W is the line $y = -\frac{4}{7}x$ in $\mathbb{R}^2$.
3. $\langle 6, -1, 4\rangle + W$, where W is the plane $3x - 2y - 5z = 0$ in $\mathbb{R}^3$.
4. $\langle -1, 3, 7\rangle + W$, where W is the plane $2x + 7y - 3z = 0$ in $\mathbb{R}^3$.
5. $\langle -2, -8, 5, 14\rangle + W$, where:

$$W = Span(\{\langle 4, -3, 2, 7\rangle, \langle 0, -1, 3, 5\rangle, \langle 3, 2, 0, -1\rangle\}) \trianglelefteq \mathbb{R}^4$$

For Exercises (6) to (10): Decide whether or not $\vec{v} + W$ and $\vec{u} + W$ are equal.

6. $\langle 7, 2\rangle + W$ and $\langle -3, -5\rangle + W$, where W is the line $y = \frac{3}{5}x$ in $\mathbb{R}^2$.

7. $\langle 3, 12\rangle + W$ and $\langle 17, 4\rangle + W$, where W is the line $y = -\frac{4}{7}x$ in $\mathbb{R}^2$.

8. $\langle 6,-4, 9\rangle + W$ and $\langle 27,-10,-6\rangle + W$, where W is the plane $3x - 2y + 5z = 0$ in $\mathbb{R}^3$.

9. $\langle 3, 0, 7\rangle + W$ and $\langle 8,-4, 0\rangle + W$, where W is the plane $2x + 7y - 3z = 0$ in $\mathbb{R}^3$.

10. $\langle -2,-5, 1, 4\rangle + W$ and $\langle 9,-10, 8, 22\rangle + W$, where:

$$W = Span(\{\langle 4,-3, 2, 7\rangle, \langle 0,-1, 3, 5\rangle, \langle 3, 2, 0,-1\rangle\}) \trianglelefteq \mathbb{R}^4.$$

11. Find the values of x_0 and z_0 such that the coset $\langle 0,-6, 0\rangle + W$ is the same as the cosets $\langle x_0, 0, 0\rangle + W$ and $\langle 0, 0, z_0\rangle + W$, where W is the plane $2x + 7y - 3z = 0$ in $\mathbb{R}^3$.

12. Find the values of x_0, y_0 and z_0 such that the coset $\langle 7,-2, 13\rangle + W$ is the same as the cosets $\langle x_0, 0, 0\rangle + W$, $\langle 0, y_0, 0\rangle + W$ and $\langle 0, 0, z_0\rangle + W$, where W is the plane $3x - 2y + 5z = 0$ in $\mathbb{R}^3$.

For Exercises (13) to (17): For each indicated subspace W of the corresponding $\mathbb{R}^n$: (a) Use the standard basis, in the order $\vec{e}_1, \ldots, \vec{e}_n$, to extend the given basis for W to a basis for $\mathbb{R}^n$;
(b) Use (a) to find a basis for $\mathbb{R}^n/W$; (c) State the dimension of $\mathbb{R}^n/W$;
(d) Verify that $dim(\mathbb{R}^n/W) = dim(\mathbb{R}^n) - dim(W)$.

13. $W = Span(\{\langle 3,-1, 2, 0\rangle\}) \trianglelefteq \mathbb{R}^4$;

14. $W = Span(\{\langle 3, 5, 2,-2\rangle, \langle -2, 1, 2,-2\rangle\}) \trianglelefteq \mathbb{R}^4$;

15. $W = Span(\{\langle 3, 0,-2, 0, 7\rangle\}) \trianglelefteq \mathbb{R}^5$;

16. $W = Span(\{\langle 2, 0, 7, 3, 0\rangle, \langle 0, 5,-14,-6, 0\rangle\}) \trianglelefteq \mathbb{R}^5$;

17. $W = Span(\{\langle 4,-3, 0, 0, 5\rangle, \langle 2,-3, 0, 0, 5\rangle, \langle 2, 1, 0, 0, 5\rangle\}) \trianglelefteq \mathbb{R}^5$;

18. Let W be a subspace of a vector space V. Prove that $\vec{x} \in \vec{v} + W$ ***if and only if*** $\vec{x} + W = \vec{v} + W$.

19. Prove that the operation of scalar multiplication is ***well defined*** on V/W.
This means that if $X = \vec{u}_1 + W = \vec{u}_2 + W$, and $k \in \mathbb{R}$, then $k\vec{u}_1 + W = k\vec{u}_2 + W$.
Hint: you must directly apply the definition of equality of cosets ***twice***.

20. Verify that the last four Axioms of a vector space are satisfied by the quotient space V/W.

21. Let Π be the plane $ax + by + cz = 0$ in $\mathbb{R}^3$, with normal vector $\vec{n} = \langle a, b, c\rangle$.

   a. Prove that every coset of Π has the form: $X = k\langle a, b, c\rangle + \Pi$, for some $k \in \mathbb{R}$.
   Hint: any coset of Π has the form $ax + by + cz = d$. Show how d is related to k.

   b. Use (a) to show that $T_1 : \mathbb{R}^3/\Pi \to \mathbb{R}$, where: $T_1(k\langle a, b, c\rangle + \Pi) = k$, is an ***isomorphism*** from $\mathbb{R}^3/\Pi$ to $\mathbb{R}$.

   c. Conclude that $dim(\mathbb{R}^3/\Pi) = 1$.

22. Analogously, let L be the line $Span(\{\vec{n}\})$, where $\vec{n} = \langle a, b, c\rangle$, as in the previous Exercise, is a normal to the plane Π with Cartesian equation $ax + by + cz = 0$.

   a. Prove that every coset of L can be written in the form: $Y = \langle x_0, y_0, z_0\rangle + L$,
   where (x_0, y_0, z_0) is a point on Π. Hint: any coset of L has the form $\langle x_1, y_1, z_1\rangle + L$ where (x_1, y_1, z_1) is ***any*** point in Cartesian space. Show how to obtain a point (x_0, y_0, z_0) on Π using (x_1, y_1, z_1). It will be useful to think in terms of parametric equations.

   b. In the notation of part (a), show that $T_2 : \mathbb{R}^3/L \to \Pi$, where:

$$T_2(\langle x_0, y_0, z_0\rangle + L) = \langle x_0, y_0, z_0\rangle,$$

   is an ***isomorphism*** from $\mathbb{R}^3/L$ to Π.

   c. Conclude that $dim(\mathbb{R}^3/L) = 2$.

4.5 The Three Isomorphism Theorems

We will now see a capstone topic of Linear Algebra: The Three Isomorphism Theorems by arguably the greatest female mathematician of all time, ***Amalie Emmy Noether*** (1882-1935).

Through most of her professional career in her native Germany, Noether faced discrimination both for being a woman and being a Jew, working for several years without pay or recognition. Despite this, she is known for many major results in various fields, including Noether's Theorem from theoretical physics and the development of ***Module Theory*** in Abstract algebra, where she developed the structures that will later be known as ***Noetherian Rings*** (a vector space is an example of a module). She was eventually dismissed from her professorship at the University of Göttingen in 1933, during the Nazi era. She went to Bryn Mawr College in Pennsylvania (whose undergraduate population was and remains all-female) to continue her work. Sadly, she would die two years later at the age of 53 due to complications from surgery to remove an ovarian cyst. This Section is a tribute to her memory and her mathematical legacy.

The Isomorphism Theorems are stated in general using ***Group Theory***, a subject that all mathematics, physics and lately, electrical engineering and computer science majors, must be familiar with. However, we can state them in the language of Linear Algebra using the concept of a ***quotient space*** that we saw in the previous Section.

The First Isomorphism Theorem

In a sense, the First Isomorphism Theorem is a stronger version of The Dimension Theorem. Let us recall what the latter says: Suppose that $T : V \to U$ is a linear transformation where V is a ***finite-dimensional*** vector space. Then:

$$rank(T) + nullity(T) = dim(V) = dim\big(domain\ of\ T\big).$$

We can rewrite this equation as: $dim(V) - dim(ker(T)) = dim(range(T))$.

However, we know from the previous Section that: $dim(V/ker(T)) = dim(V) - dim(ker(T))$.

Thus the Dimension Theorem allows us to conclude that: $dim(V/ker(T)) = dim(range(T))$.

Since two vector spaces are ***isomorphic*** ***if and only if*** they have the ***same dimension***, we see that:

$$V/ker(T) \cong range(T).$$

But the bonus is that the First Isomorphism Theorem explicitly tells us ***how*** to construct an isomorphism between these two vector spaces using T:

The First Isomorphism Theorem:
Let $T : V \to U$ be a linear transformation of vector spaces, with V a ***finite-dimensional*** vector space. Then: $ker(T)$ is a finite-dimensional subspace of V, $range(T)$ is a finite dimensional subspace of U, and:

$$V/ker(T) \cong range(T),$$

with an isomorphism ***induced*** by T via:

$$\widetilde{T} : V/ker(T) \to range(T), \quad \text{given by:} \quad \widetilde{T}(\vec{v} + ker(T)) = T(\vec{v}).$$

Note: The squiggly symbol above T is called a "tilde", so $\widetilde{T}$ is pronounced "***T til-deh***." Also, notice that the codomain U could be infinite dimensional, but under our assumptions, $range(T)$ is a finite dimensional subspace of U. We say that T ***induces*** $\widetilde{T}$ because we are using T to create a new linear transformation on the related quotient space $V/ker(T)$.

Proof of the Theorem: Before we verify that $\widetilde{T}$ is additive and homogeneous, we first need to verify that $\widetilde{T}$ as described above is actually ***well-defined,*** just as we first verified that coset addition and scalar multiplication were well defined in the previous Section. In other words, we have to check that:

$$\text{If } \vec{v}_1 + ker(T) = \vec{v}_2 + ker(T), \text{ then } \widetilde{T}(\vec{v}_1 + ker(T)) = \widetilde{T}(\vec{v}_2 + ker(T)).$$

But if $\vec{v}_1 + ker(T) = \vec{v}_2 + ker(T)$, then $\vec{v}_1 - \vec{v}_2 = \vec{z} \in ker(T)$. Thus $\vec{v}_1 = \vec{v}_2 + \vec{z}$, hence:

$$\begin{aligned}
\widetilde{T}(\vec{v}_1 + ker(T)) &= T(\vec{v}_1) = T(\vec{v}_2 + \vec{z}) \\
&= T(\vec{v}_2) + T(\vec{z}) \\
&= T(\vec{v}_2) + \vec{\mathbf{0}}_U \qquad (\text{since } \vec{z} \in ker(T)) \\
&= T(\vec{v}_2) = \widetilde{T}(\vec{v}_2 + ker(T)).
\end{aligned}$$

Thus $\widetilde{T}$ is well defined. Now, it is easy to verify the linearity properties of $\widetilde{T}$, as they are ***inherited*** from the linearity properties of T:

$$\begin{aligned}
&\widetilde{T}((\vec{u} + ker(T)) + (\vec{v} + ker(T))) \\
&= \widetilde{T}((\vec{u} + \vec{v}) + ker(T)) && \text{(by definition of coset addition)} \\
&= T(\vec{u} + \vec{v}) && \text{(by definition of } \widetilde{T}) \\
&= T(\vec{u}) + T(\vec{v}) && \text{(by the additivity of } T) \\
&= \widetilde{T}(\vec{u} + ker(T)) + \widetilde{T}(\vec{v} + ker(T)) && \text{(by definition of coset addition).}
\end{aligned}$$

Similarly, $T(k \cdot (\vec{v} + ker(T))) = k \cdot \widetilde{T}(\vec{v} + ker(T))$ will be proven in the Exercises. Thus $\widetilde{T}$ is linear.

Next, we have to prove that $\widetilde{T}$ is ***one-to-one***, that is, $ker(\widetilde{T})$ consists only of the zero vector of $V/ker(T)$, which is the subspace $ker(T)$.

So suppose that $\widetilde{T}(\vec{v} + ker(T)) = \vec{\mathbf{0}}_U$. But $\widetilde{T}(\vec{v} + ker(T)) = T(\vec{v})$, so therefore $\vec{v} + ker(T) \in ker(\widetilde{T})$ ***if and only if*** $T(\vec{v}) = \vec{\mathbf{0}}_U$, that is, if and only if $\vec{v} \in ker(T)$. Thus the only member of $ker(\widetilde{T})$ is the single coset $ker(T)$. Thus $\widetilde{T}$ is one-to-one.

Finally, we have to prove that $\widetilde{T}$ is ***onto***, that is, $range\left(\widetilde{T}\right) = range(T)$.

But again by definition, $\widetilde{T}(\vec{v} + ker(T)) = T(\vec{v})$, and since $\vec{v}$ can be any member of V, $T(\vec{v})$ is a member of $range(T)$. Thus, $range\left(\widetilde{T}\right) = range(T)$ and $\widetilde{T}$ is onto.

We can therefore conclude that $\widetilde{T}$ is an ***isomorphism***. ■

Example: Let $T : \mathbb{R}^4 \to \mathbb{R}^3$ be given by:

$$[T] = \begin{bmatrix} 3 & -6 & -2 & -3 \\ 2 & -4 & 3 & 11 \\ -1 & 2 & 5 & 14 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & -2 & 0 & 1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus x_2 and x_4 are free variables, and $ker(T)$ has basis:

$$B = \{\langle 2, 1, 0, 0\rangle, \langle -1, 0, -3, 1\rangle\}.$$

The leading 1's are in columns 1 and 3 of the rref, so a basis for $range(T) = colspace([T])$ is:

$$S = \{\vec{c}_1, \vec{c}_3\} = \{\langle 3, 2, -1\rangle, \langle -2, 3, 5\rangle\}.$$

Thus $range(T)$ is a plane Π in $\mathbb{R}^3$, and we get:

$$\mathbb{R}^4/ker(T) \cong \Pi.$$

Let us use the First Isomorphism Theorem to make this isomorphism ***explicit***. We will first need to construct a basis for the quotient space $\mathbb{R}^4/ker(T)$ using the ideas from the previous Section. To do this, we will extend the basis B for $ker(T)$ above to a basis B' for all of $\mathbb{R}^4$. What two vectors can we add to B? Recall that $\widetilde{T}(\vec{v} + ker(T)) = T(\vec{v})$, so it would be nice if we can find $\vec{v}_1$ and $\vec{v}_2$ so that:

$$T(\vec{v}_1) = \vec{c}_1 \text{ and}$$
$$T(\vec{v}_2) = \vec{c}_3.$$

But from the definition of ***matrix multiplication***, we need $\vec{v}_1 = \vec{e}_1$ and $\vec{v}_2 = \vec{e}_3$. Thus, let us extend B to $B' = \{\langle 2, 1, 0, 0\rangle, \langle -1, 0, -3, 1\rangle, \vec{e}_1, \vec{e}_3\}$ and verify that:

$$\begin{bmatrix} 2 & -1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & -3 & 0 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \text{ has rref } \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \boldsymbol{I}_4.$$

Thus $\mathbb{R}^4/ker(T)$ has basis:

$$B'' = \{\vec{e}_1 + ker(T), \vec{e}_3 + ker(T)\},$$

according to the last Theorem of the previous Section. By the ***First Isomorphism Theorem***, we have:

$$\widetilde{T}(\vec{e}_1 + ker(T)) = T(\vec{e}_1) = \vec{c}_1 \text{ and}$$
$$\widetilde{T}(\vec{e}_3 + ker(T)) = T(\vec{e}_3) = \vec{c}_3,$$

giving us an elegant and explicit ***correspondence*** between the basis for $\mathbb{R}^4/ker(T)$ and the basis for $range(T)$. □

Let us examine further our clever choices for vectors to extend B to B'. We will show that it is ***not*** a coincidence that $\vec{e}_1$ and $\vec{e}_3$ completed our basis for $\mathbb{R}^4$. Since x_2 and x_4 are free variables, we find a basis for $ker(T)$ by solving for x_1 and x_3 in terms of x_2 and x_4. Thus, notice that our two basis vectors:

$$B = \left\{\left\langle 2, \boxed{1}, 0, \boxed{0}\right\rangle, \left\langle -1, \boxed{0}, -3, \boxed{1}\right\rangle\right\}$$

have matching pairs of 0's and 1's in the $\mathbf{2}^{nd}$ and $\mathbf{4}^{th}$ coordinates respectively, which we boxed above. Now, if we assemble the matrix:

$$\begin{bmatrix} 2 & -1 & \boxed{1} & 0 \\ \boxed{1} & 0 & 0 & 0 \\ 0 & -3 & 0 & \boxed{1} \\ 0 & \boxed{1} & 0 & 0 \end{bmatrix},$$

with $\vec{e}_1$ and $\vec{e}_3$ in the last two columns, notice that we have a boxed $\boxed{1}$ in every row and in every column. We can use Type 3 ***column*** operations to obtain the ***column equivalent*** matrix $\left[\vec{e}_2\,\vec{e}_4\,\vec{e}_1\,\vec{e}_3\right]$. Thus this matrix is ***invertible*** and we successfully complete a basis for $\mathbb{R}^4$. We can generalize this argument and so we state the following:

Theorem — Addendum to the First Isomorphism Theorem:
Let $T : \mathbb{R}^{\boldsymbol{n}} \to \mathbb{R}^{\boldsymbol{m}}$ be given by the $m \times n$ matrix $[T] = [\vec{c}_1|\vec{c}_2|\ldots|\vec{c}_m]$, with rref R. Suppose the leading columns of R are columns $i_1, i_2, \ldots, i_r$, where $r = rank(T)$. Then: we can find a basis for $\mathbb{R}^{\boldsymbol{n}}/ker(T)$ using only standard basis vectors of $\mathbb{R}^{\boldsymbol{n}}$, via:

$$B = \{\vec{e}_{i_1} + ker(T), \vec{e}_{i_2} + ker(T), \ldots, \vec{e}_{i_r} + ker(T)\}.$$

Furthermore, under the induced transformation $\widetilde{T}$ in the First Isomorphism Theorem:

$$\widetilde{T}\left(\vec{e}_{i_j} + ker(T)\right) = \vec{c}_{i_j},$$

for all $j = 1..r$.

Example: Let us clarify this Theorem by considering a linear transformation:

$$T : \mathbb{R}^{\mathbf{5}} \to \mathbb{R}^{\mathbf{8}},$$

with 8×5 standard matrix $[T] = \left[\vec{c}_1\,\vec{c}_2\,\vec{c}_3\,\vec{c}_4\,\vec{c}_5\right]$ whose rref is R. Suppose the leading 1's of R are in columns 1, 3 and 4. Then $rank(T) = 3$ and $nullity(T) = 2$, so $ker(T)$ is 2-dimensional and $dim(\mathbb{R}^{\mathbf{5}}/ker(T)) = 5 - 2 = 3$. By the Theorem above, a basis for $\mathbb{R}^{\mathbf{5}}/ker(T)$ is given by:

$$B = \{\vec{e}_1 + ker(T), \vec{e}_3 + ker(T), \vec{e}_4 + ker(T)\}.$$

Under T, we have:

$$T(\vec{e}_1 + ker(T)) = \vec{c}_1,$$
$$T(\vec{e}_3 + ker(T)) = \vec{c}_3, \text{ and}$$
$$T(\vec{e}_4 + ker(T)) = \vec{c}_4. \square$$

The Second Isomorphism Theorem

There is no general agreement as to which of the next two Isomorphism Theorems is the Second and which is the Third, so we choose to present them in the following order:

The Second or Double-Quotient Isomorphism Theorem:
Let V be a finite dimensional vector space, and let U and W be ***nested*** subspaces of V, that is:

$$U \trianglelefteq W \trianglelefteq V.$$

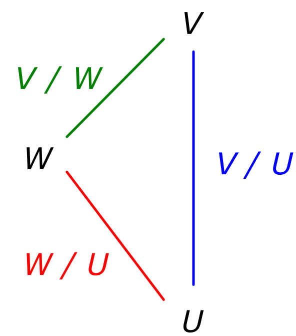
Then: W/U is a ***subspace*** of V/U, and V/W is ***isomorphic*** to the ***double-quotient space:***

$$(V/U)/(W/U) \cong V/W.$$

Notice that the conclusion is similar to the rule for dividing fractions:

$$(a/c)/(b/c) = a/b, \quad \text{where } b, c \neq 0.$$

In the Theorem, though, there is nothing stopping U or W or both from being $\left\{\vec{\mathbf{0}}_V\right\}$. We traditionally visualize this Theorem by the following diagram:



The Double Quotient Isomorphism Theorem: $(V/U)/(W/U) \cong V/W$

Notice that lines are used to indicate that U is a subspace of W as well as V, and W is a subspace of V. These lines are labeled with the corresponding quotient space. Not shown in this diagram is the double quotient space $(V/U)/(W/U)$. However, it is there in spirit because we will show that it is isomorphic to V/W.

Proof of the Theorem: It is easy to show that W/U is a ***subspace*** of V/U. The members of W/U are cosets of the form $\vec{w}+U$ for some $\vec{w} \in W$, and since $W \subset V$, $\vec{w}+U \in V/U$ as well. Thus W/U is a ***subset*** of V/U. It is not empty because it contains the coset U. Now we prove closure under vector addition. Suppose $\vec{w}_1 + U$ and $\vec{w}_2 + U \in W/U$. Since W is a subspace of V, we get:

$$(\vec{w}_1 + U) + (\vec{w}_2 + U) = (\vec{w}_1 + \vec{w}_2) + U \in W/U.$$

Thus W/U is closed under addition. Similarly it is easy to see it is closed under scalar multiplication and we leave that as an Exercise.

Now for the rest of the Theorem. The First Isomorphism Theorem says that if $T : V \to W$, then $V/ker(T) \cong range(T)$. The idea is to apply this in a creative way to a linear transformation T. We look at the desired ***conclusion*** in order to find the right T:

$$(V/U)/(W/U) \cong V/W.$$

The left side tells us that the ***domain*** of T should be V/U and the ***kernel*** of T should be W/U. The right side tells us that the ***range*** of T should be V/W. In other words, we need to construct a linear transformation:

$$T : V/U \to V/W$$

such that T is ***onto***, and the ***kernel*** of T is W/U. Now, the members of V/U are the cosets $\vec{v}+U$, where $\vec{v} \in V$. Similarly, the members of V/W are the cosets $\vec{v}+W$, where $\vec{v} \in V$ as well. Thus, the linear transformation that we need appears to be:

$$T(\vec{v}+U) = \vec{v}+W.$$

(Note that even though T leaves $\vec{v}$ untouched, this is ***not*** the identity transformation, since $\vec{v}+U$ and $\vec{v}+W$ are cosets of ***different*** subspaces U and W.) However, as before, we first need to verify that this linear transformation is ***well defined***, that is:

$$\text{If } \vec{v}_1 + U = \vec{v}_2 + U, \text{ then } \vec{v}_1 + W = \vec{v}_2 + W \text{ as well.}$$

Suppose $\vec{v}_1 + U = \vec{v}_2 + U$. Then $\vec{v}_1 = \vec{v}_2 + \vec{u}$ for some vector $\vec{u} \in U$. But since $U \trianglelefteq W$, we also have $\vec{u} \in W$, and thus:

$$\vec{v}_1 + W = \vec{v}_2 + \vec{u} + W = \vec{v}_2 + W$$

by the ***Absorption Rule***, so our linear transformation T is well defined.

Now, T is clearly ***onto*** because if $\vec{v} + W \in V/W$, then $\vec{v} + U \in V/U$, so there exists a vector in V/U whose image under T is $\vec{v} + W$. Thus, $range(T) = V/W$.

Finally, let us look at the ***kernel*** of T. We want to find all cosets $\vec{v} + U$ such that:

$$T(\vec{v} + U) = \vec{\mathbf{0}}_{V/W}.$$

In other words, $\vec{v} + W = W$.

But according to the ***Absorption Rule***, $\vec{v} + W = W$ ***if and only if*** $\vec{v} \in W$. Thus $\vec{v} + U \in ker(T)$ ***if and only if*** $\vec{v} + U \in W/U$. We can therefore conclude that:

$$ker(T) = W/U.$$

Since we saw above that $range(T) = V/W$, by the ***First Isomorphism Theorem***, we get:

$$(V/U)/ker(T) \cong range(T), \quad \text{or}$$
$$(V/U)/(W/U) \cong V/W. \blacksquare$$

Note: The members of V/U have the form $\vec{v} + U$, and therefore the members of the double-quotient space $(V/U)/(W/U)$ have the form $(\vec{v} + U) + W/U$. This expression ***cannot*** be simplified any further.

Example: Let $V = \mathbb{R}^4$ and let:

$$S_1 = \{\vec{v}_1\} = \{\langle 1, 5, -2, 3\rangle\}, \quad \text{and}$$
$$S_2 = \{\vec{v}_1, \vec{v}_2\} = \{\langle 1, 5, -2, 3\rangle, \langle -2, 1, 4, -6\rangle\}.$$

Let $U = Span(S_1)$ and $W = Span(S_2)$.

Clearly $U \trianglelefteq W \trianglelefteq V$, $dim(U) = 1$, $dim(W) = 2$ and $dim(V) = 4$. Let us first investigate the dimensions of the quotient spaces involved:

$$dim(V/U) = dim(V) - dim(U) = 4 - 1 = 3,$$
$$dim(W/U) = dim(W) - dim(U) = 2 - 1 = 1, \text{ and}$$
$$dim(V/W) = dim(V) - dim(W) = 4 - 2 = 2.$$

The dimension of our double-quotient space is:

$$dim((V/U)/(W/U)) = dim((V/U)) - dim(W/U) = 3 - 1 = 2,$$

which equals $dim(V/W)$. Thus, $(V/U)/(W/U)$ should indeed be isomorphic to V/W.

Now, let us explicitly construct the linear transformation $T : V/U \to V/W$ and the induced isomorphism $\tilde{T}$. We start by finding a basis for $ker(T) = W/U$. Notice that $\vec{v}_1$ is in both S_1 and S_2, so clearly:

$$W/U = Span(\{\vec{v}_2 + U\}).$$

Next, let us find a basis for $range(T) = V/W$. We need to extend our basis S_2 for W to a basis for $V = \mathbb{R}^4$. To use the Extension Theorem most efficiently, let us assemble the set of vectors $\{\vec{v}_1, \vec{v}_2, \vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$ into the ***columns*** of a matrix:

$$\begin{bmatrix} 1 & -2 & 1 & 0 & 0 & 0 \\ 5 & 1 & 0 & 1 & 0 & 0 \\ -2 & 4 & 0 & 0 & 1 & 0 \\ 3 & -6 & 0 & 0 & 0 & 1 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 0 & 2/11 & 0 & 1/33 \\ 0 & 1 & 0 & 1/11 & 0 & -5/33 \\ 0 & 0 & 1 & 0 & 0 & -1/3 \\ 0 & 0 & 0 & 0 & 1 & 2/3 \end{bmatrix}.$$

Notice that $\vec{e}_2 \in W$. The leading 1's in the rref are in columns 1, 2, 3 and 5, and therefore we obtain a basis for $\mathbb{R}^4$:

$$S_3 = \{\vec{v}_1, \vec{v}_2, \vec{e}_1, \vec{e}_3\}.$$

This basis is useful in ***three*** ways. According to the last Theorem of Section 4.4, we have the following bases for the corresponding quotient spaces:

$$\begin{aligned} V/U &= Span(\{\vec{v}_2 + U, \vec{e}_1 + U, \vec{e}_3 + U\}), \\ V/W &= Span(\{\vec{e}_1 + W, \vec{e}_3 + W\}), \text{ and} \\ (V/U)/(W/U) &= Span(\{(\vec{e}_1 + U) + W/U, (\vec{e}_3 + U) + W/U\}). \end{aligned}$$

The linear transformation we want, $T : V/U \to V/W$, is given by its action on the basis vectors of V/U:

$$\begin{aligned} T(\vec{v}_2 + U) &= \vec{v}_2 + W = W, \\ T(\vec{e}_1 + U) &= \vec{e}_1 + W, \quad \text{and} \\ T(\vec{e}_3 + U) &= \vec{e}_3 + W, \end{aligned}$$

and extends by linearity to all of V/U.

Notice that since $\vec{v}_2 \in W$, we have $\vec{v}_2 + W = W = \vec{\mathbf{0}}_{V/W}$. Thus $\vec{v}_2 + U \in ker(T)$, and the ***isomorphism*** $\widetilde{T}$ that we are looking for is:

$$\begin{aligned} &\widetilde{T} : (V/U)/(W/U) \to V/W, \quad \text{given by:} \\ &\widetilde{T}((\vec{e}_1 + U) + W/U) = \vec{e}_1 + W \quad \text{and} \\ &\widetilde{T}((\vec{e}_3 + U) + W/U) = \vec{e}_3 + W, \end{aligned}$$

and extends by linearity to all of $(V/U)/(W/U)$. Again, $\vec{e}_1$ and $\vec{e}_3$ are essentially ***unchanged***, but $\widetilde{T}$ is not the identity operator because we have ***different coset spaces***. We also remark that there is nothing special about $\vec{e}_1$ and $\vec{e}_3$. They can easily be replaced by any two vectors $\vec{v}_3$ and $\vec{v}_4$ such that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is a basis for $\mathbb{R}^4$.□

The Third Isomorphism Theorem

The final Isomorphism Theorem involves the construction of the ***join*** and ***intersection*** of two subspaces that we saw in Section 4.1. Recall that if V and W are subspaces of a ***finite dimensional*** vector space U, then:

$$V \cap W = \left\{ \vec{u} \in U \mid \vec{u} \in V \text{ and } \vec{u} \in W \right\} \quad \text{and}$$

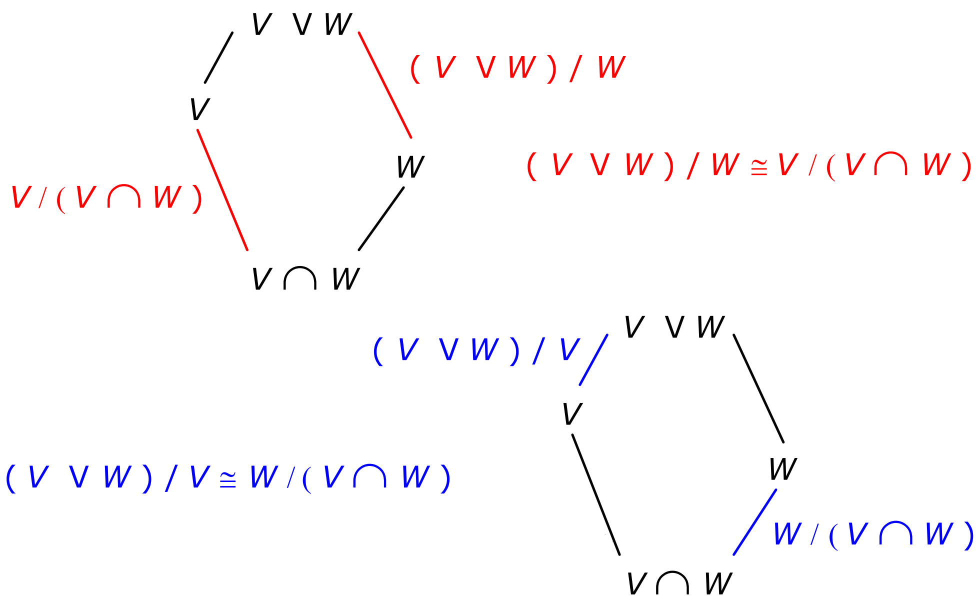
$$V \vee W = \left\{ \vec{u} \in U \mid \vec{u} = \vec{v} + \vec{w}, \text{ for some } \vec{v} \in V \text{ and some } \vec{w} \in W \right\}$$

are both subspaces of U. Thus we come full circle and end this Chapter with the same topics we started with, linked together by the concept of quotient spaces, as stated in the following jewel:

The Third or Diamond Isomorphism Theorem:

Let V and W be ***finite dimensional*** subspaces of a vector space U. Then:

$$(V \vee W)/W \cong V/(V \cap W), \quad \text{and} \quad (V \vee W)/V \cong W/(V \cap W).$$



The Diamond Isomorphism Theorem

Notice that the conclusions of this Theorem do not mention U at all, but only the two subspaces V and W. In other words, we only use U as an ***ambient space*** for these two subspaces. Since U is closed under addition and scalar multiplication, the join $V \vee W$ is also a subspace of U. Furthermore, $V \vee W$ is finite dimensional.

Proof of the Theorem: For the sake of brevity and clarity of notation, let us name:

$$X = V \cap W, \quad \text{and} \quad Y = V \vee W.$$

Thus, our first goal is to prove that $Y/V \cong W/X$. The proof that $Y/W \cong V/X$ follows by exchanging the roles of V and W.

Again, the idea is to creatively use the First Isomorphism Theorem by constructing a linear transformation T. This time, the ***domain*** of T should be Y, the ***kernel*** of T should be V, the ***codomain*** should be W/X, and T should again be ***onto***. So we must construct:

$$T : Y \to W/X$$

with the above conditions. Although this is a good beginning of our analysis, it is hardly clear exactly ***what*** T should be computing. Let us try to make a smart guess. Since $Y = V \vee W$, the members of Y are of the form $\vec{v} + \vec{w}$ for some $\vec{v} \in V$ and some $\vec{w} \in W$. From this, we want $T(\vec{v} + \vec{w})$ to be from the ***coset space*** W/X. But the members of this coset space have the form $\vec{w} + X$ for some $\vec{w} \in W$. Thus, it appears that we should compute T as:

$$T(\vec{v} + \vec{w}) = \vec{w} + X.$$

In other words, we ***ignore*** $\vec{v}$ and only use $\vec{w}$.

Once again, we need to verify next that this function is ***well-defined***, that is:

$$\text{If } \vec{v}_1 + \vec{w}_1 = \vec{v}_2 + \vec{w}_2, \text{ then } \vec{w}_1 + X = \vec{w}_2 + X.$$

But notice that the first equation above can be rearranged as:

$$\vec{v}_1 - \vec{v}_2 = \vec{w}_2 - \vec{w}_1.$$

But since V and W are subspaces of U, the left side of this equation is a member of V, and the right side is a member of W. Thus, this (single) vector is a member of the ***intersection*** $V \cap W = X$. Thus we can now see that:

$$\begin{aligned}\vec{w}_2 + X &= (\vec{v}_1 - \vec{v}_2 + \vec{w}_1) + X \\ &= ((\vec{v}_1 - \vec{v}_2) + X) + (\vec{w}_1 + X) \\ &= X + (\vec{w}_1 + X) \qquad \text{(by the Absorption Rule)} \\ &= \vec{w}_1 + X\end{aligned}$$

so T is well-defined.

The two linearity properties are easily verified and are left as Exercises.

Since $\vec{w}$ can be any vector of W, T is clearly ***onto*** W/X. The last thing we have to check is that $ker(T) = V$. So we need to ask: for what vectors $\vec{v} + \vec{w}$ do we get:

$$T(\vec{v} + \vec{w}) = \vec{w} + X = X = \vec{\mathbf{0}}_{W/X}$$

But again by the Absorption Law, this is true if and only if $\vec{w} \in X = V \cap W$. This means that $\vec{w}$ is a member of ***both*** V and W. But this means that $\vec{v} + \vec{w}$ is a member of V. Thus $ker(T) = V$, and by The First Isomorphism Theorem, we get:

$$Y/V \cong W/X. \blacksquare$$

Example: We saw an Example in Section 4.1 where we went through many computations before we found a basis for the intersection of two subspaces. Let us consider a much simpler example. Suppose $U = \mathbb{R}^4$, and:

$$\begin{aligned} V &= Span(\{\langle 3, 5, -2, 7\rangle, \langle -4, 0, 2, 1\rangle\}), \text{ and} \\ W &= Span(\{\langle 3, 5, -2, 7\rangle, \langle 2, -1, 5, -3\rangle\}).\end{aligned}$$

Clearly V and W are 2-dimensional subspaces. They have the vector $\langle 3, 5, -2, 7\rangle$ in common, so $V \cap W$ is ***at least*** 1-dimensional. However, by assembling the three distinct vectors above into the ***columns*** of a matrix:

$$\begin{bmatrix} 3 & -4 & 2 \\ 5 & 0 & -1 \\ -2 & 2 & 5 \\ 7 & 1 & -3 \end{bmatrix} \text{ whose rref is } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix},$$

we see that these three vectors are ***linearly independent***. Thus $dim(V \vee W) = 3$, and:

$$V \vee W = Span(\{\langle 3, 5, -2, 7\rangle, \langle -4, 0, 2, 1\rangle, \langle 2, -1, 5, -3\rangle\}).$$

By the ***Dimension Theorem for the Join and Intersection:***

$$dim(V \vee W) = dim(V) + dim(W) - dim(V \cap W), \text{ so we get:}$$
$$3 = 2 + 2 - dim(V \cap W),$$
and thus $dim(V \cap W) = 1$. Therefore:
$$V \cap W = Span(\{\langle 3, 5, -2, 7\rangle\}).$$
Let us examine the isomorphism: $(V \vee W)/W \cong V/(V \cap W)$. We have:
$$dim((V \vee W)/W) = dim(V \vee W) - dim(W) = 3 - 2 = 1, \text{ and}$$
$$dim(V/(V \cap W)) = dim(V) - dim(V \cap W) = 2 - 1 = 1.$$
As expected, these dimensions are equal. Let us recap the bases that we have for our spaces so far:
$$V \vee W = Span(\{\langle 3, 5, -2, 7\rangle, \langle -4, 0, 2, 1\rangle, \langle 2, -1, 5, -3\rangle\}),$$
$$V = Span(\{\langle 3, 5, -2, 7\rangle, \langle -4, 0, 2, 1\rangle\}),$$
$$W = Span(\{\langle 3, 5, -2, 7\rangle, \langle 2, -1, 5, -3\rangle\}), \text{ and}$$
$$V \cap W = Span(\{\langle 3, 5, -2, 7\rangle\}).$$
According to the last Theorem of Section 4.4, we have the following bases for the corresponding quotient spaces:
$$(V \vee W)/W = Span(\{\langle -4, 0, 2, 1\rangle + W\}), \text{ and}$$
$$V/(V \cap W) = Span(\{\langle -4, 0, 2, 1\rangle + (V \cap W)\})$$
Thus, we obtain the isomorphism:
$$\widetilde{T} : (V \vee W)/W \to V/(V \cap W), \quad \text{given by:}$$
$$\widetilde{T}(\langle -4, 0, 2, 1\rangle + W) = \langle -4, 0, 2, 1\rangle + (V \cap W).$$
Again, the vector $\langle -4, 0, 2, 1\rangle$ is essentially unchanged, although as before, $\widetilde{T}$ is ***not*** the identity operator.□

4.5 Section Summary

The First Isomorphism Theorem: Let $T : V \to U$ be a linear transformation of vector spaces, with V a ***finite-dimensional*** vector space. Then $ker(T)$ is a finite-dimensional subspace of V, $range(T)$ is a finite dimensional subspace of U, and $V/ker(T) \cong range(T)$, with an isomorphism ***induced*** by T via:
$$\widetilde{T} : V/ker(T) \to range(T), \quad \text{given by:}$$
$$\widetilde{T}(\vec{v} + ker(T)) = T(\vec{v}).$$

The Second or "Double-Quotient" Isomorphism Theorem: Let V be a ***finite dimensional*** vector space, and let U and W be ***nested*** subspaces of V, that is: $U \trianglelefteq W \trianglelefteq V$. Then: W/U is a ***subspace*** of V/U, and V/W is isomorphic to the ***double-quotient space:*** $(V/U)/(W/U) \cong V/W$.

The Third or "Diamond" Isomorphism Theorem: Let V and W be ***subspaces*** of a ***finite dimensional*** vector space U. Then:
$$(V \vee W)/W \cong V/(V \cap W), \quad \text{and}$$
$$(V \vee W)/V \cong W/(V \cap W).$$

Addendum to the First Isomorphism Theorem: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be given by the $m \times n$ matrix $[T] = [\vec{c}_1 | \vec{c}_2 | \ldots | \vec{c}_m]$, with rref R. Suppose the leading columns of R are columns $i_1, i_2, \ldots, i_r$, where $r = rank(T)$. Then we can find a basis for $\mathbb{R}^n/ker(T)$ using only standard basis vectors of $\mathbb{R}^n$, via:

$$B = \{\vec{e}_{i_1} + ker(T), \vec{e}_{i_2} + ker(T), \ldots, \vec{e}_{i_r} + ker(T)\}$$

Furthermore, under the induced transformation $\widetilde{T}$ in the First Isomorphism Theorem: $\widetilde{T}\left(\vec{e}_{i_j} + ker(T)\right) = \vec{c}_{i_j}$ for all $j = 1..r$.

4.5 Exercises

For Exercise (1) to (14): The following linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$ were seen in the Exercises of Sections 4.2 and 4.3 (but not necessarily with the same Exercise number). For each of them: (a) find a basis for $range(T)$, (b) find a basis for $ker(T)$, (c) use the Addendum to the First Isomorphism Theorem to find a basis for $\mathbb{R}^n/ker(T)$ using only the standard basis vectors for $\mathbb{R}^n$, and (d) explicitly construct the isomorphism $\widetilde{T} : \mathbb{R}^n/ker(T) \to range(T)$ in the First Isomorphism Theorem by specifying it on the basis vectors from (c).

You may of course use work from Section 4.2 and 4.3 to answer parts (a) and (b).

1. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 3 & 2 & -2 \\ 5 & 3 & -1 \\ 4 & 2 & 2 \\ -1 & -1 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -7 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

2. $T : \mathbb{R}^3 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} 2 & -6 & -7 \\ -3 & 9 & -1 \\ -4 & 12 & 9 \\ 5 & -15 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.

3. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 3 & 5 & 4 & -1 \\ 2 & 3 & 2 & -1 \\ -2 & -1 & 2 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & -2 & -2 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

4. $T : \mathbb{R}^3 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 3 & 5 & -1 \\ 2 & 3 & -1 \\ -2 & -1 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.

5. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 2 & 10 & 5 & -7 \\ 3 & 15 & 7 & -9 \\ -4 & -20 & -9 & 11 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 4 \\ 0 & 0 & 1 & -3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

6. $T : \mathbb{R}^4 \to \mathbb{R}^3$, with $[T] = \begin{bmatrix} 2 & 10 & 5 & -7 \\ 3 & 15 & 7 & -9 \\ -4 & -20 & -9 & 8 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$.

7. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} -5 & -20 & -4 & 1 & 2 \\ 3 & 12 & -2 & -27 & 12 \\ 2 & 8 & 3 & 8 & -5 \\ -4 & -16 & 1 & 26 & -11 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 4 & 0 & -5 & 2 \\ 0 & 0 & 1 & 6 & -3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

8. $T : \mathbb{R}^4 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} -5 & -20 & -4 & 1 \\ 3 & 12 & -2 & -27 \\ 2 & 8 & 3 & 8 \\ -4 & -16 & 1 & 26 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 4 & 0 & -5 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

9. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} -5 & 10 & -30 & -3 & 5 \\ 2 & -4 & 12 & -3 & -23 \\ -3 & 6 & -18 & 2 & 22 \\ 4 & -8 & 24 & -5 & -41 \end{bmatrix}$, $R = \begin{bmatrix} 1 & -2 & 6 & 0 & -4 \\ 0 & 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

10. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} -2 & 6 & -3 & -5 & -27 \\ 3 & -9 & 7 & 4 & 6 \\ 4 & -12 & -1 & -2 & -2 \\ -5 & 15 & 2 & 3 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & -3 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$.

11. $T : \mathbb{R}^5 \to \mathbb{R}^4$, with $[T] = \begin{bmatrix} -2 & 6 & -3 & -5 & -7 \\ 3 & -9 & 7 & 4 & 6 \\ 4 & -12 & -1 & -2 & -2 \\ -5 & 15 & 2 & 3 & 3 \end{bmatrix}$, $R = \begin{bmatrix} 1 & -3 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$.

12. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} 2 & 10 & -1 & -10 \\ -4 & -20 & 1 & 14 \\ 3 & 15 & 1 & 0 \\ 5 & 25 & -3 & -28 \\ -6 & -30 & 2 & 24 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & -2 \\ 0 & 0 & 1 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

13. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} 2 & 10 & -1 & -9 \\ -4 & -20 & 1 & 14 \\ 3 & 15 & 1 & 0 \\ 5 & 25 & -3 & -28 \\ -6 & -30 & 2 & 24 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

14. $T : \mathbb{R}^4 \to \mathbb{R}^5$, with $[T] = \begin{bmatrix} -2 & 1 & 1 & 2 \\ 5 & -1 & -1 & -8 \\ 1 & 1 & -1 & 6 \\ -2 & -2 & 1 & -7 \\ -1 & 1 & 1 & 0 \end{bmatrix}$, $R = \begin{bmatrix} 1 & 0 & 0 & -2 \\ 0 & 1 & 0 & 3 \\ 0 & 0 & 1 & -5 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$.

For Exercises (15) to (21): For the indicated nested subspaces U and W of the corresponding $V = \mathbb{R}^n$, find a basis for: (a) W/U, (b) V/W, (c) V/U, and (d) $(V/U)/(W/U)$. Next: (e) explicitly construct the isomorphism:

$$\widetilde{T} : (V/U)/(W/U) \to V/W$$

in the Second Isomorphism Theorem. Use the standard basis vectors in the order $\vec{e}_1, \ldots, \vec{e}_n$ to extend the basis for W to a basis for $\mathbb{R}^n$. Notice the subtle differences from one Exercise to the next.

15. $U, W \trianglelefteq \mathbb{R}^3$; $U = Span(\{\langle 1,-1,1\rangle\})$; $W = Span(\{\langle 1,-1,1\rangle, \langle -2,-1,1\rangle\})$.

16. $U, W \trianglelefteq \mathbb{R}^4$; $U = Span(\{\langle 1,-1,1,2\rangle\})$; $W = Span(\{\langle 1,-1,1,2\rangle, \langle 1,1,1,2\rangle\})$.

17. $U, W \trianglelefteq \mathbb{R}^4$; $U = Span(\{\langle 1,-1,1,2\rangle\})$;
$W = Span(\{\langle 1,-1,1,2\rangle, \langle 1,1,1,2\rangle, \langle 3,-1,1,2\rangle\})$.

18. $U, W \trianglelefteq \mathbb{R}^4$; $U = Span(\{\langle 1,-1,1,2\rangle, \langle 1,1,1,2\rangle\})$;
$W = Span(\{\langle 1,-1,1,2\rangle, \langle 1,1,1,2\rangle, \langle 3,-1,1,2\rangle\})$.

19. $U, W \trianglelefteq \mathbb{R}^5$; $U = Span(\{\langle 1,-1,1,2,-3\rangle\})$;
$W = Span(\{\langle 1,-1,1,2,-3\rangle, \langle 1,1,1,2,-3\rangle, \langle 3,-1,1,2,-3\rangle\})$.

20. $U, W \trianglelefteq \mathbb{R}^5$; $U = Span(\{\langle 1,-1,1,2,-3\rangle, \langle 1,1,1,2,-3\rangle\})$;
$W = Span(\{\langle 1,-1,1,2,-3\rangle, \langle 1,1,1,2,-3\rangle, \langle 3,-1,1,2,-3\rangle\})$.

21. $U, W \trianglelefteq \mathbb{R}^5$; $U = Span(\{\langle 1,-1,1,2,-3\rangle, \langle 1,1,1,2,-3\rangle\})$;
$W = Span(\{\langle 1,-1,1,2,-3\rangle, \langle 1,1,1,2,-3\rangle, \langle 3,-1,1,2,-3\rangle, \langle 3,-1,1,-1,-3\rangle\})$.

For Exercises (22) to (26): The following subspaces for the corresponding $\mathbb{R}^n$ are the same ones from Exercise 1 through 5 of Section 4.1. Use your answers from this Section in order to find a basis for: (a) $V \vee W$, (b) $V \cap W$, (c) $(V \vee W)/W$, (d) $V/(V \cap W)$, (e) $(V \vee W)/V$, and (f) $W/(V \cap W)$. ***The Minimization Theorem*** will be useful for (d) and (f). Next, explicitly construct the isomorphisms (g):

$$\widetilde{T}_1 : (V \vee W)/W \to V/(V \cap W),$$

and (h):

$$\widetilde{T}_2 : (V \vee W)/V \to W/(V \cap W),$$

by specifying their action on the basis vectors of the domains.

22. $V,\ W \trianglelefteq \mathbb{R}^4;\ V = Span(\{\langle 1, -1, -12, 6\rangle, \langle 11, -16, 13, 1\rangle\}),$
$W = Span(\{\langle 1, 1, -16, 10\rangle, \langle 7, -11, 5, 1\rangle\}).$

23. $V,\ W \trianglelefteq \mathbb{R}^4;\ V = Span(\{\langle 3, 5, -2, 4\rangle, \langle 1, 2, 7, -3\rangle\}),$
$W = Span(\{\langle 0, 2, 1, -5\rangle, \langle 2, -3, 1, 6\rangle\}).$

24. $V,\ W \trianglelefteq \mathbb{R}^4;\ V = Span(\{\langle -3, -2, 7, -4\rangle, \langle -2, 13, -12, -2\rangle, \langle -2, 3, -5, 1\rangle\}),$
$W = Span(\{\langle -3, -5, 6, -11\rangle, \langle -1, 16, -8, 8\rangle, \langle 1, -3, 2, -4\rangle\}).$

25. $V,\ W \trianglelefteq \mathbb{R}^5;\ V = Span(\{\langle -3, 4, -1, 4, 6\rangle, \langle -6, 8, 5, 15, -13\rangle, \langle 1, -2, 0, -5, 3\rangle\}),$
$W = Span(\{\langle 1, 3, -2, 7, 2\rangle, \langle -4, -1, 7, -7, -6\rangle\}).$

26. $V = Span(\{\langle -1, 7, 5, -6, 6\rangle, \langle -1, -8, 2, -4, 2\rangle, \langle 1, 0, 3, -4, 3\rangle, \langle 5, 3, -2, 7, -4\rangle\}) \trianglelefteq \mathbb{R}^5,$
$W = Span(\{\langle -6, 9, -2, 0, 0\rangle, \langle -5, 1, -3, -3, -2\rangle, \langle -3, 2, -1, -2, 0\rangle\}) \trianglelefteq \mathbb{R}^5.$

27. Complete the proof of the First Isomorphism Theorem by showing that:

$$T(k \bullet (\vec{v} + ker(T))) = k \bullet \widetilde{T}(\vec{v} + ker(T)),$$

in the notation used in the proof.

28. Complete the proof of the Second Isomorphism Theorem by showing that W/U is closed under scalar multiplication, and thus it is a ***subspace*** of V/U.

29. Complete the proof of the Diamond Isomorphism Theorem by showing that the linear transformation T defined in the proof is both additive and homogeneous.

30. In one of the Examples of Section 4.4, we said the for any vector space V, the coset space $V/\left\{\vec{\mathbf{0}}_V\right\}$ "looks like" V. Use the ***First Isomorphism Theorem*** to prove that indeed:

$$V/\left\{\vec{\mathbf{0}}_V\right\} \cong V.$$

Hint: Use the identity transformation $\boldsymbol{I}_V$ for T.

31. Use the ***First Isomorphism Theorem*** to prove that:

$$V/V \cong \left\{\vec{\mathbf{0}}_V\right\}.$$

Which (very trivial) operator $T : V \to V$ should you use?

32. Use the Diamond Isomorphism Theorem to prove ***The Dimension Theorem for Joins and Intersections:*** Let V and W be finite-dimensional subspaces of a vector space U. Then: $V \vee W$ is also finite dimensional, and:

$$dim(V \vee W) = dim(V) + dim(W) - dim(V \cap W).$$

Hint: use the formula for the dimension of a quotient space from the previous Section.

A Summary of Chapter 4

Let V and W be two ***subspaces*** of some ambient vector space U. We define the ***join*** of these two subspaces as the set of all vectors of U that can be written as the sum of one vector from V and one vector from W, and it is denoted: $V \vee W = \left\{ \vec{u} \in U \,|\, \vec{u} = \vec{v} + \vec{w} \text{ for some } \vec{v} \in V \text{ and some } \vec{w} \in W \right\}$.

$V \vee W$ is a ***subspace*** of U, and if B is a basis for V and B' is a basis for W, then $B \cup B'$ ***Spans*** $V \vee W$.

The ***intersection*** of V and W: $V \cap W = \left\{ \vec{u} \in U \,|\, \vec{u} \in V \textbf{ and } \vec{u} \in W \right\}$, is likewise a ***subspace*** of U.

Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ be a basis for V and let $B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ be a basis for W, where V and W are ***subspaces*** of some $\mathbb{R}^k$. Form the matrix C, with rows $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and D, with rows $\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m$. Next, find a basis for the ***nullspace*** of each matrix. Assemble these two sets of basis vectors ***together*** as the rows of a matrix G. The ***nullspace*** of G is $V \cap W$.

Let W be any subspace of some Euclidean space $\mathbb{R}^n$. Then, $W \cap W^{\perp} = \left\{ \vec{\mathbf{0}}_n \right\}$.

Let $T : V \to W$ be a linear transformation, and U a subspace of the domain V. The ***restriction*** of T to U, denoted $T|_U$ and pronounced "T ***restricted*** to U," is the linear transformation:

$$T|_U : U \to W, \text{ given by } T|_U(\vec{u}) = T(\vec{u}) \text{ for all } \vec{u} \in U.$$

Let us define:

$$ker(T|_U) = \left\{ \vec{u} \in U | T(\vec{u}) = \vec{0}_W \right\}, \text{ and } range(T|_U) = \left\{ \vec{w} \in W | \vec{w} = T(\vec{u}) \text{ for some } \vec{u} \in U \right\}.$$

Then: $ker(T|_U) = ker(T) \cap U$ and $range(T|_U) \trianglelefteq range(T)$.

Thus, we can regard $T|_U$ as a linear transformation: $T|_U : U \to range(T|_U)$.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with standard matrix $[T]$.

Let $U = \boldsymbol{rowspace}([T]) \trianglelefteq \mathbb{R}^n$. Then: the restriction $T|_U : U \to \mathbb{R}^m$ is ***one-to-one***.

Furthermore, for any subspace W of $\mathbb{R}^n$ such that $dim(W) > \dim(U) = rank(T)$, the restriction $T|_W : W \to \mathbb{R}^m$ is ***not*** one-to-one.

Let U be ***any*** subspace of $\mathbb{R}^n$ such that $T|_U$ is ***one-to-one***. Then: $T|_U : U \to range(T|_U)$ is an ***isomorphism***. Moreover, if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_j\}$ is a basis for U, then $B' = \{T(\vec{u}_1), T(\vec{u}_2), \ldots, T(\vec{u}_j)\}$ is a basis for $range(T|_U)$.

In particular, if $U = rowspace([T])$, or U is any subspace of such that $dim(U) = rank(T)$ and T is ***one-to-one*** when restricted to U, then $range(T|_U) = range(T)$.

The Preservation of Subspaces Theorem: Suppose that $T : V \to W$ is a linear transformation.

Then, for any subspace $U \trianglelefteq V$: $T(U) = \left\{ \vec{w} \in W \,|\, \vec{w} = T(\vec{u}) \text{ for some } \vec{u} \in U \right\}$,

called the ***image*** of U under T, is a ***subspace*** of the codomain W.

Similarly, for any subspace $Z \trianglelefteq W$: $T^{-1}(Z) = \left\{ \vec{v} \in V \,|\, T(\vec{v}) = \vec{z} \text{ for some } \vec{z} \in Z \right\}$,

called the ***pre-image*** of Z under T, is a ***subspace*** of the domain V.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and suppose that V is any subspace of $\mathbb{R}^n$. Suppose that B is a basis for V. Then $T(B)$ ***Spans*** $T(V)$. Thus, the output of ***The Minimizing Theorem*** applied to $T(B)$ will be a basis for $T(V)$.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, and $U \trianglelefteq \mathbb{R}^m$. The following algorithm will produce a basis for $T^{-1}(U)$, given as its input the standard matrix $[T]$ and a basis B for U:

1. Find the rref of $[T]$, and use it to:
2. Find a basis for $colspace([T])$ and a basis $\left\{\vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ for $ker(T)$.
3. Find a basis $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ for $U \cap T(\mathbb{R}^n)$ using the techniques from Section 4.1.
4. For each $\vec{u}_i$, find any vector $\vec{v}_i \in T^{-1}(U)$ such that $T(\vec{v}_i) = \vec{u}_i$.
   We can accomplish this simultaneously by solving the system: $[\,[T]\,|\,\vec{u}_1\,|\,\vec{u}_2\,|\ldots|\,\vec{u}_r\,]$.
5. The combined set $B' = \left\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{k}_1, \vec{k}_2, \ldots, \vec{k}_s\right\}$ is a basis for $T^{-1}(U)$.

Let $W \trianglelefteq V$. A ***coset*** X of W is a ***translate*** of W: $X = \vec{v} + W = \{\vec{v} + \vec{w}\,|\,\vec{w} \in W\}$, for some fixed vector $\vec{v} \in V$. We call $\vec{v}$ a ***representative*** of the coset $\vec{v} + W$, or say that $\vec{v} + W$ is ***represented*** by $\vec{v}$.

Equality of cosets: Let $X = \vec{v} + W$ and $Y = \vec{u} + W$. Then: $X = Y$ ***if and only if*** $\vec{v} - \vec{u} \in W$.

The Absorption Rule: $\vec{v} + W = W = \vec{\mathbf{0}}_V + W$ ***if and only if*** $\vec{v} \in W$.

Membership in a coset is equivalent to representing the same coset:

$$\vec{x} \in \vec{v} + W \textit{ if and only if } \vec{x} + W = \vec{v} + W.$$

The ***quotient space*** $\boldsymbol{V/W}$, ("***V mod W***") is the set of ***all cosets*** of W: $\boldsymbol{V/W} = \{X = \vec{v} + W\,|\,\vec{v} \in V\}$.

V/W is also a ***vector space***, where we define addition of cosets and scalar multiplication as follows: If $X = \vec{u} + W$ and $Y = \vec{v} + W$ are two cosets of V/W and $k \in \mathbb{R}$, then:

$$X + Y = (\vec{u} + W) + (\vec{v} + W) = (\vec{u} + \vec{v}) + W, \quad \text{and} \quad kX = k(\vec{u} + W) = k\vec{u} + W.$$

Basis for $\boldsymbol{V/W}$: Let W be an m-dimensional subspace of an n-dimensional vector space V, where $0 < m < n$. Suppose that $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m\}$ is a basis for W (where B is empty if $W = \left\{\vec{\mathbf{0}}_V\right\}$). Since B is linearly independent, it can be ***extended*** to a basis B' for all of V, say:

$$B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_m, \vec{v}_{m+1}, \vec{v}_{m+2}, \ldots, \vec{v}_n\}.$$

Then, the set of cosets: $B'' = \{\vec{v}_{m+1} + W, \vec{v}_{m+2} + W, \ldots, \vec{v}_n + W\}$ is a ***basis*** for V/W.

Thus: $dim(V/W) = n - m = dim(V) - dim(W)$.

In the special case when $W = V$, the coset space $V/W = V/V$ consists of the single coset V, and thus $dim(V/V) = 0$. The formula above for $dim(V/W)$ is therefore true in general.

The First Isomorphism Theorem: Let $T : V \to U$ be a linear transformation of vector spaces, with V a ***finite-dimensional*** vector space. Then $ker(T)$ is a finite-dimensional subspace of V, $range(T)$ is a finite dimensional subspace of U, and $V/ker(T) \cong range(T)$, with an isomorphism ***induced*** by T via:

$$\widetilde{T} : V/ker(T) \to range(T), \quad \text{given by:} \quad \widetilde{T}(\vec{v} + ker(T)) = T(\vec{v}).$$

The Second or "Double-Quotient" Isomorphism Theorem: Let V be a ***finite dimensional*** vector space, and let U and W be ***nested*** subspaces of V, that is: $U \trianglelefteq W \trianglelefteq V$. Then: W/U is a ***subspace*** of V/U, and V/W is isomorphic to the ***double-quotient space:*** $(V/U)/(W/U) \cong V/W$.

The Third or "Diamond" Isomorphism Theorem: Let V and W be ***subspaces*** of a ***finite dimensional*** vector space U. Then: $(V \vee W)/W \cong V/(V \cap W)$, and $(V \vee W)/V \cong W/(V \cap W)$.

Addendum to the First Isomorphism Theorem: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be given by the $m \times n$ matrix $[T] = [\vec{c}_1\,|\,\vec{c}_2\,|\ldots|\,\vec{c}_m]$, with rref R. Suppose the leading columns of R are columns $i_1, i_2, \ldots, i_r$, where $r = rank(T)$. Then we can find a basis for $\mathbb{R}^n/ker(T)$ using only standard basis vectors of $\mathbb{R}^n$, via: $B = \{\vec{e}_{i_1} + ker(T), \vec{e}_{i_2} + ker(T), \ldots, \vec{e}_{i_r} + ker(T)\}$.

Furthermore, under the induced transformation $\widetilde{T}$ in the First Isomorphism Theorem:
$\widetilde{T}\left(\vec{e}_{i_j} + ker(T)\right) = \vec{c}_{i_j}$ for all $j = 1..r$.

Chapter 5

From Square to Scalar:

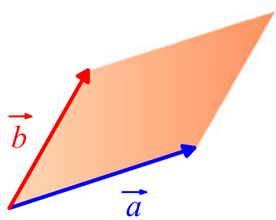
Permutation Theory and Determinants

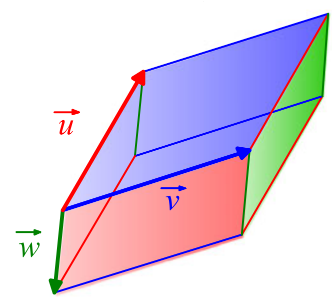
In this Chapter, we will concentrate our attention on $n \times n$ or ***square*** matrices A, and analogously, on ***operators*** $T : \mathbb{R}^n \to \mathbb{R}^n$.

We will study the ***determinant*** function, that returns a ***scalar*** $det(A)$ that depends on the $n \times n$ matrix A. Consequently, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is an operator, we can also compute a determinant for T. The determinant has a nice geometric significance in the case of 2×2 and 3×3 matrices. If $\vec{a}$ and $\vec{b}$ are ***non-parallel*** vectors from $\mathbb{R}^2$, and we assemble the 2×2 matrix:

$$A = \left[\begin{array}{cc} \vec{a} & \vec{b} \end{array} \right],$$

then $|det(A)|$ gives us the ***area*** of the ***parallelogram*** determined by $\vec{a}$ and $\vec{b}$ in standard position:





Similarly, if $\vec{u}$, $\vec{v}$ and $\vec{w}$ are ***non-coplanar*** vectors from $\mathbb{R}^3$, and we assemble the 3×3 matrix:

$$B = \left[\begin{array}{ccc} \vec{u} & \vec{v} & \vec{w} \end{array} \right],$$

then $|det(B)|$ gives us the ***volume*** of the ***parallelepiped*** determined by $\vec{u}$, $\vec{v}$ and $\vec{w}$. These formulas remain valid if we assemble the vectors into the ***rows*** of A or B instead of columns.

The most important property of the determinant function is its ability to test for the ***invertibility*** of a square matrix: an $n \times n$ matrix A is ***invertible*** ***if and only if*** $det(A)$ is ***not*** 0.

We are used to linear transformations being additive. However, we will see that the determinant function is instead ***multiplicative:*** $det(AB) = det(A) \cdot det(B)$, where A and B are both $n \times n$ matrices.

In order to motivate and construct the determinant function, though, we first need to introduce the concept of ***permutations*** or rearrangements of the numbers 1 through n. The determinant formula is basically a summation involving all the permutations of the rows or columns of A, where each term involves an entry from every row and every column of A.

We will see practical strategies for computing the determinant of A using ***row*** and ***column*** operations, and a technique called ***cofactor expansion***. The idea behind cofactors will also give us an alternative, albeit impractical formula for the ***inverse*** of A. We will also use determinants to develop ***Cramer's Rule***, an alternative method to solve an invertible square system of equations.

5.1 Permutations and The Determinant Concept

In Section 2.6, we studied 2×2 matrices and found that the matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

is ***invertible*** ***if and only if*** the number $ad - bc$ is non-zero. This expression is important and goes by a special name:

Definition: Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a 2×2 matrix. The ***determinant*** of A is defined by:

$$det(A) = ad - bc.$$

Other common notations for $det(A)$ are $|A|$ or $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

Example: Let $A = \begin{bmatrix} 5 & -3 \\ 4 & 2 \end{bmatrix}$. Then:

$$det(A) = 5(2) - (-3)4 = 10 + 12 = 22. \square$$

We can rephrase our Theorem from Section 2.6 as:

Theorem: A 2×2 matrix A is ***invertible*** ***if and only if*** $det(A) \neq 0$.

Naturally, we want to generalize this Theorem by defining the determinant of a 3×3 matrix and for square matrices of any dimension.

We begin this generalization process by looking at the 2×2 determinant in terms of the ***row*** and ***column*** numbers of the four individual entries:

If $A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$, then:

$$det(A) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}.$$

The first observation is that there are ***two terms***, and each term has ***two factors***. Furthermore, one term has a positive coefficient, and one term has a negative coefficient. Notice also that we listed the factors in the order of the ***row number***, so we see a factor from row 1 followed by a factor from row 2 in both terms. In other words, the row numbers are in ***ascending order***.

However, the column numbers are "1" followed by "2" in the first term, and "2" followed by "1" in the second term. Notice that $2 > 1$. This is called an ***inversion***, wherein the column number of the first factor is ***bigger*** than the column number of the second factor. It is because of this inversion that the second factor $a_{1,2}a_{2,1}$ has a ***negative coefficient***.

Let us use the observations above to define the determinant of a 3×3 matrix. Every term will now contain ***three factors***. Again, we will list the factors in such a way that the ***row numbers*** are in ***ascending order***, so a typical term will look like:

$$\pm\, a_{1,\_}\,a_{2,\_}\,a_{3,\_}\,.$$

We will fill in the column numbers with the numbers 1, 2 and 3, in any order that we wish. There are ***three ways*** to decide the first column number, and for ***each*** of these three ways, there are ***two ways*** remaining to decide the second column number, and once these have been decided, there is only ***one way*** remaining to fill the third column number. Thus, there are $3 \cdot 2 \cdot 1 = 6$ terms in our determinant.

Now we need to decide if a term will have a positive or a negative coefficient. This will be determined by the ***total number of inversions*** that appear in the list of column numbers. Since there are three columns involved, we will say that an inversion occurs every time a column number on the ***left*** is ***bigger*** than a column number on its ***right***. If the number of inversions is ***even***, the resulting term will have a ***positive coefficient***. If the number of inversions is odd, the resulting term will have a ***negative coefficient***.

We list our six terms in a table below, and decide which have a positive coefficient and which have a negative coefficient. Notice that we get 3 ***positive*** coefficients and 3 ***negative*** coefficients:

The Six Terms of a 3×3 Determinant

Term	Columns	Inversions	Number of Inversions	Coefficient	Final Term
$\pm a_{1,1}a_{2,2}a_{3,3}$	$1,2,3$	none	0	$+$	$+a_{1,1}a_{2,2}a_{3,3}$
$\pm a_{1,1}a_{2,3}a_{3,2}$	$1,3,2$	$3 > 2$	1	$-$	$-a_{1,1}a_{2,3}a_{3,2}$
$\pm a_{1,2}a_{2,1}a_{3,3}$	$2,1,3$	$2 > 1$	1	$-$	$-a_{1,2}a_{2,1}a_{3,3}$
$\pm a_{1,2}a_{2,3}a_{3,1}$	$2,3,1$	$2 > 1;\ 3 > 1$	2	$+$	$+a_{1,2}a_{2,3}a_{3,1}$
$\pm a_{1,3}a_{2,1}a_{3,2}$	$3,1,2$	$3 > 1;\ 3 > 2$	2	$+$	$+a_{1,3}a_{2,1}a_{3,2}$
$\pm a_{1,3}a_{2,2}a_{3,1}$	$3,2,1$	$3 > 2;\ 3 > 1;\ 2 > 1$	3	$-$	$-a_{1,3}a_{2,2}a_{3,1}$

We are now ready to write out the formula for a 3×3 determinant:

Definition: If $A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$, then:

$$det(A) = a_{1,1}a_{2,2}a_{33} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2}$$
$$- a_{1,3}a_{2,2}a_{3,1} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3}.$$

These six terms can also be visualized below:

$$\begin{bmatrix} \boxed{a_{1,1}} & a_{1,2} & a_{1,3} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} \\ a_{3,1} & a_{3,2} & \boxed{a_{3,3}} \end{bmatrix} \quad \begin{bmatrix} a_{1,1} & \boxed{a_{1,2}} & a_{1,3} \\ a_{2,1} & a_{2,2} & \boxed{a_{2,3}} \\ \boxed{a_{3,1}} & a_{3,2} & a_{3,3} \end{bmatrix} \quad \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} \\ \boxed{a_{2,1}} & a_{2,2} & a_{2,3} \\ a_{3,1} & \boxed{a_{3,2}} & a_{3,3} \end{bmatrix}$$

$$+a_{1,1}a_{2,2}a_{3,3} \qquad +a_{1,2}a_{2,3}a_{3,1} \qquad +a_{1,3}a_{2,1}a_{3,2}$$

$$\begin{bmatrix} \boxed{a_{1,1}} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & \boxed{a_{2,3}} \\ a_{3,1} & \boxed{a_{3,2}} & a_{3,3} \end{bmatrix} \quad \begin{bmatrix} a_{1,1} & \boxed{a_{1,2}} & a_{1,3} \\ \boxed{a_{2,1}} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & \boxed{a_{3,3}} \end{bmatrix} \quad \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} \\ \boxed{a_{3,1}} & a_{3,2} & a_{3,3} \end{bmatrix}$$

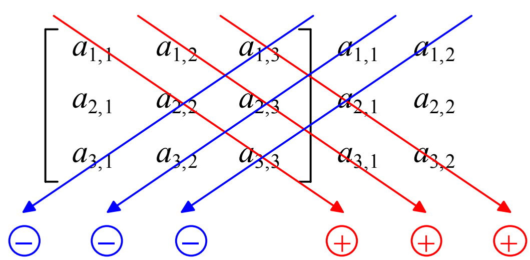
$$-a_{1,1}a_{2,3}a_{3,2} \qquad -a_{1,2}a_{2,1}a_{3,3} \qquad -a_{1,3}a_{2,2}a_{3,1}$$

Notice that each term contains exactly one factor from every ***row***, and from every ***column***. This is a general property of the general determinant formula, which we will see very soon.

There is an easy way to remember how to compute a 3×3 determinant. ***Copy*** the first two columns of the given matrix on the right side of the matrix:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix} \begin{matrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{matrix}$$

Next, multiply the entries diagonally downward, in triples, as if we were weaving a basket:



The products going left to right contribute positively, whereas the products going right to left contribute negatively. Unfortunately, this pattern ***does not generalize*** to larger matrices.

Example: Let us find the determinant of $A = \begin{bmatrix} 5 & 2 & -3 \\ 7 & 4 & 6 \\ -8 & -5 & 2 \end{bmatrix}$.

Copying the first two columns on the right, we get:

$$\left[\begin{array}{rrr} 5 & 2 & -3 \\ 7 & 4 & 6 \\ -8 & -5 & 2 \end{array}\right] \begin{array}{rr} 5 & 2 \\ 7 & 4 \\ -8 & -5 \end{array}$$

Thus:

$$\begin{aligned} det(A) &= 5 \cdot 4 \cdot 2 + 2 \cdot 6 \cdot (-8) + (-3) \cdot 7 \cdot (-5) \\ &\quad - (-3) \cdot 4 \cdot (-8) - 5 \cdot 6 \cdot (-5) - 2 \cdot 7 \cdot 2 \\ &= 40 - 96 + 105 - 96 + 150 - 28 = 75. \square \end{aligned}$$

Permutation Theory

We saw above that the rearrangement of the ***column*** numbers appearing in each term of our determinant formula is crucial in determining whether the term has a positive or a negative coefficient. These rearrangements are very important in their own right in Mathematics, so we now give them special attention, establish some notation, and discover some of their properties:

Definitions: A ***permutation*** of the set of integers $\{1, 2, \ldots, n\}$, is an ordered list consisting of these numbers, with each number appearing ***exactly once***. In other words, a permutation is a ***rearrangement*** of these numbers. We will label permutations with lowercase Greek letters such as σ or τ, and write them as:

$$\sigma = (i_1, i_2, \ldots, i_n).$$

We call i_k the ***k^{th} component*** of σ.

Examples: There are exactly ***two*** permutations of $\{1, 2\}$, and these are:

$$\tau_1 = (1, 2) \text{ and } \tau_2 = (2, 1).$$

There are exactly ***six*** permutations of $\{1, 2, 3\}$, and these are:

$$\sigma_1 = (1, 2, 3),\ \sigma_2 = (1, 3, 2),\ \sigma_3 = (2, 1, 3),$$

$$\sigma_4 = (2, 3, 1),\ \sigma_5 = (3, 1, 2), \text{ and } \sigma_6 = (3, 2, 1). \square$$

More generally, we can make the following easy observation:

Theorem: The number of permutations of $\{1, 2, \ldots, n\}$ is:

$$n! = n \cdot (n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1.$$

Example: The number of permutations of the set $\{1, 2, 3, 4, 5, 6, 7\}$ is exactly $7!$ or 5040.

In order to create a completely general definition of the determinant function, we will introduce the following terminology:

Definition: An ***inversion*** occurs in a permutation σ every time a component on the left is ***bigger*** than a component to its right.

We say that σ is ***even*** if there are an even number of inversions in σ, and σ is ***odd*** if there are an odd number of inversions in σ.

We define the ***sign*** of σ, denoted $sgn(\sigma)$, to be $+1$ if σ is ***even***, and -1 if σ is ***odd***.

Examples: Let us consider the permutation:

$$\sigma = (4, 7, 5, 6, 2, 3, 1).$$

This is a permutation of the set $\{1, 2, 3, 4, 5, 6, 7\}$, of which there are $7! = 5040$ such permutations. Clearly there is nothing to gain by listing all of these permutations. Instead, let us find all the inversions in σ, by keeping one finger fixed at an entry and scanning to its ***right*** to find all numbers that are ***smaller*** than the number on our finger:

$$4 > 2,\ 3 \text{ and } 1,$$
$$7 > 5,\ 6,\ 2,\ 3 \text{ and } 1,$$
$$5 > 2,\ 3 \text{ and } 1,$$
$$6 > 2,\ 3 \text{ and } 1,$$
$$2 > 1, \text{ and}$$
$$3 > 1.$$

Thus, there are $3 + 5 + 3 + 3 + 1 + 1 = 16$ inversions, σ is ***even***, and $sgn(\sigma) = +1$.

The permutation $\iota = (1, 2, 3, \ldots, n-1, n)$, the ***identity permutation***, has no inversions whatsoever (ι is the Greek letter ***iota***). Thus, ι is an ***even*** permutation, and $sgn(\iota) = +1$. $\square$

Permutations as Bijections

Let us think of a permutation σ of the set $S = \{1, 2, \ldots, n\}$ as a ***function*** $\sigma : S \to S$, by defining $\sigma(i)$ to be the i^{th} component of σ. However, since the numbers $1, 2, \ldots, n$ appear ***exactly once*** in σ, we can view σ as a ***one-to-one*** and ***onto*** function. Thus, σ is a ***bijection*** of S. As such, we can construct the ***inverse permutation*** σ^{-1}, which is also a bijection of S, and $\sigma \circ \sigma^{-1} = \sigma^{-1} \circ \sigma = \iota$.

Example: Suppose $\sigma = (3,5,6,1,4,2)$. We can see the action of σ in the table on the left:

i	$\xrightarrow{\sigma}$	$\sigma(i)$		$\sigma(i)$	$\xleftarrow{\sigma^{-1}}$	i
1	$\to$	3		1	$\leftarrow$	3
2	$\to$	5		2	$\leftarrow$	5
3	$\to$	6		3	$\leftarrow$	6
4	$\to$	1		4	$\leftarrow$	1
5	$\to$	4		5	$\leftarrow$	4
6	$\to$	2		6	$\leftarrow$	2

The Actions of σ and σ^{-1}

The ***inverse*** of σ, denoted σ^{-1}, has the usual property that if $\sigma(x) = y$, then $\sigma^{-1}(y) = x$. By simply ***reversing*** the arrows, we can see the action of σ^{-1} on the table on the right, although i is not in ascending order. We list the values of $\sigma^{-1}(i)$ in the (correct) ascending order for i:

$$\sigma^{-1}(1) = 4,\ \sigma^{-1}(2) = 6,\ \sigma^{-1}(3) = 1,\ \sigma^{-1}(4) = 5,\ \sigma^{-1}(5) = 2, \text{ and } \sigma^{-1}(6) = 3.$$

Thus we obtain the inverse permutation $\sigma^{-1} = (4, 6, 1, 5, 2, 3)$. □

Now, here's something interesting: Let us count the inversions in σ and σ^{-1}. We get:

$$2 + 3 + 3 + 0 + 1 = 9 \quad \text{for } \sigma \text{ and}$$

$$3 + 4 + 0 + 2 + 0 = 9 \quad \text{for } \sigma^{-1} \text{ as well.}$$

This is not a coincidence, and we will investigate it next by detecting inversions in a different way.

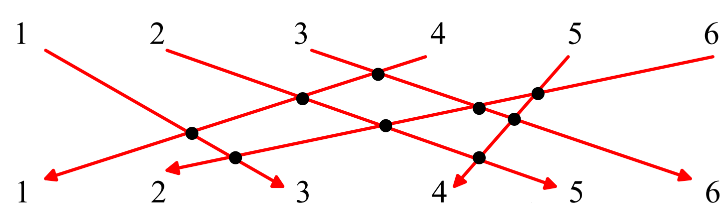
Counting Inversions Using Bipartite Graphs

The following construction is due to Dr. Lyman Chaffee, my friend and colleague at Pasadena City College, who has my sincerest respect and gratitude.

We can visualize a permutation as a ***directed bipartite graph***. A ***graph*** is a set of ***vertices,*** some pairs of which are connected by a set of ***edges***. To represent a permutation of $\{1, 2, \ldots, n\}$, we will use two copies of this set, one on top of the other, to serve as our vertices. If $\sigma(i) = j$, we will have a ***directed edge*** (an arrow) from i on the top row to j on the bottom row.

This is called a ***bipartite graph*** because edges only go from the top vertices to the bottom vertices (there are no edges connecting a top vertex to another top vertex, and similarly there are no edges from one bottom vertex to another). Let us illustrate this idea using our previous Example:

Example: For the permutation $\sigma = (3, 5, 6, 1, 4, 2)$, we have the equivalent bipartite graph:



The Permutation $\sigma = (3, 5, 6, 1, 4, 2)$ as a Bipartite Graph

For instance, since $\sigma(3) = 6$, we have an arrow $3 \to 6$, and so on. □

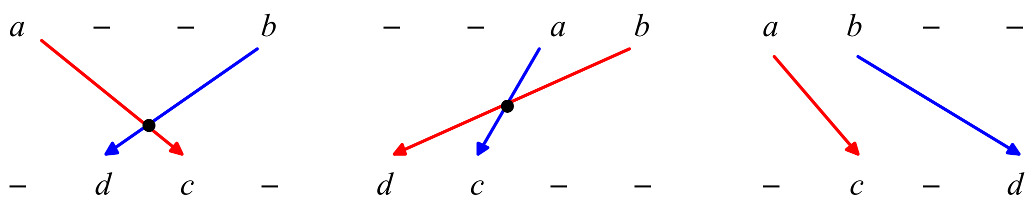
Now, here's the key benefit of this diagram: We have shown with dots how two arrows ***intersect*** each other. Each of these dots represents an ***inversion***, and we can verify that there are 9 dots in the graph corresponding to our 9 inversions. In general:

Theorem: Suppose we represent σ, a permutation of $S = \{1, 2, \ldots, n\}$, as a directed graph in the convention shown above. If $a < b \in S$, $\sigma(a) = c$, and $\sigma(b) = d$, then c is an ***inversion*** with respect to d (that is, $c > d$) ***if and only if*** the edge $a \to c$ ***intersects*** the edge $b \to d$ between our two lines of numbers.

Proof: Suppose that $a < b$ and $c > d$. We must show that the edge $a \to c$ intersects $b \to d$.

Case 1: If $a \leq c$, then $a \to c$ will pass over the vertex d on the bottom row. Since $b > a$, the edge $b \to d$ must intersect $a \to c$.

Case 2: If $a > c$, then $b > c$ as well. Thus, the edge $b \to d$ must pass over the vertex c on the bottom row. Consequently $a \to c$ must intersect $b \to d$.



Case 1 ($a \leq c$) Case 2 ($a > c$) No Intersection ($c < d$)

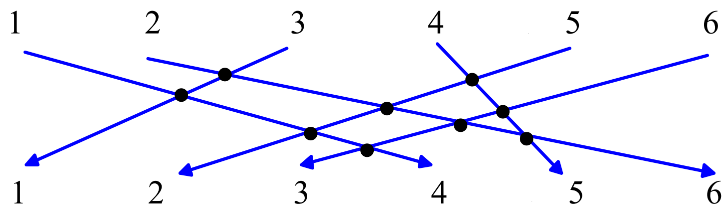
Conversely, if $a < b$ and $c < d$. then $a \to c$ will not intersect $b \to d$ between the two lines (if we extend these arrows, they will intersect ***above*** the top line, or ***below*** the bottom line, but that is irrelevant). ■

We must warn, though, that we do not want more than two edges intersect at one point. As a simple example, consider the permutation $(3, 2, 1)$:

$\sigma = (3, 2, 1)$ with Overlapping Intersections and with Distinct Intersections

It appears that there is only ***one*** point of intersection among the three edges, but in actuality, there are ***three*** inversions: $3 > 2$, $3 > 1$ and $2 > 1$. By repositioning the vertices slightly, we can see that there are indeed 3 distinct pairwise intersection points.

And now for the punch line: in order to draw the graph for σ^{-1}, all we have to do is ***reverse*** the arrows and ***flip*** the graph upside-down! In our previous example, we saw that $\sigma^{-1} = (4, 6, 2, 5, 1, 3)$:



The Permutation $\sigma^{-1} = (4, 6, 1, 5, 2, 3)$ as a Bipartite Graph

Since the number of intersections among the edges of σ and among the edges of σ^{-1} are exactly the same, this construction immediately proves the following:

Theorem: If σ is a permutation of $\{1, 2, \ldots, n\}$, then σ and σ^{-1} have the ***same number of inversions***. Hence, σ and σ^{-1} have the same ***sign*** — they are either both ***even*** or both ***odd***.

The Effect of a Switch

A funny thing happens when you exchange the location of any two numbers in a permutation:

Theorem: Let σ be a permutation of $\{1, 2, 3, .., n\}$, and let σ' be the permutation obtained from σ by exchanging ***any*** two components of σ. Then:

$$sgn(\sigma') = -sgn(\sigma).$$

Proof: Let us begin the proof by looking at the case where the two components are ***adjacent*** to each other. To see the idea behind the proof, let us look at the permutation:

$$\sigma = \left(8, 5, 3, \boxed{2}, \boxed{7}, 1, 4, 6\right)$$

of $\{1, 2, \ldots, 8\}$. Suppose we exchange the 4^{th} and 5^{th} components, which contain 2 and 7, as shown. We obtain the permutation:

$$\sigma' = \left(8, 5, 3, \boxed{7}, \boxed{2}, 1, 4, 6\right).$$

Now, $2 < 7$, so there was no inversion involving both 2 and 7 in σ, but there is now an inversion $7 > 2$ in σ'. However, any inversion in σ involving 2 but not 7 (such as $8 > 2$ and $2 > 1$) is also found in σ', and vice versa. Similarly, any inversion in σ involving 7 but not 2 (such as $8 > 7$ and $7 > 4$) is also found in σ'. And of course if an inversion in σ does not involve 7 or 2, such as $8 > 1$ or $5 > 4$, it is still found in σ'. Thus, σ' has ***one more*** inversion than σ.

The general argument follows from this demonstration: suppose we exchange the entries in component i and component $i + 1$ of σ to produce the permutation σ'. If $\sigma(i) < \sigma(i + 1)$, then σ' will have one new inversion. All other inversions that were found in σ will again be found in σ', and vice versa. Thus the number of inversions ***increases*** by 1. Similarly, if $\sigma(i) > \sigma(i + 1)$, the number of inversions ***decreases*** by 1. In either case, $sgn(\sigma') = -sgn(\sigma)$.

Now, let us generalize further to the case when the components are ***not adjacent*** to each other. Let us illustrate the idea with an example. Consider the same permutation:

$$\sigma = \left(8, \boxed{5}, 3, 2, 7, \boxed{1}, 4, 6\right).$$

This time, let us switch the 2^{nd} and 6^{th} components, as shown, to get a new permutation:

$$\sigma' = \left(8, \boxed{1}, 3, 2, 7, \boxed{5}, 4, 6\right).$$

But we can produce σ' by performing consecutive exchanges, starting with σ, involving ***only adjacent*** components:

$$\begin{aligned}
\sigma &= \left(8, \boxed{5}, 3, 2, 7, \boxed{1}, 4, 6\right) && \text{switch 5 and 3:} \\
&\to \left(8, 3, \boxed{5}, 2, 7, \boxed{1}, 4, 6\right) && \text{switch 5 and 2:} \\
&\to \left(8, 3, 2, \boxed{5}, 7, \boxed{1}, 4, 6\right) && \text{switch 5 and 7:} \\
&\to \left(8, 3, 2, 7, \boxed{5}, \boxed{1}, 4, 6\right) && \text{switch 5 and 1:} \\
&\to \left(8, 3, 2, 7, \boxed{1}, \boxed{5}, 4, 6\right) && \text{switch 1 and 7 (5 is now in the right place):} \\
&\to \left(8, 3, 2, \boxed{1}, 7, \boxed{5}, 4, 6\right) && \text{switch 1 and 2:} \\
&\to \left(8, 3, \boxed{1}, 2, 7, \boxed{5}, 4, 6\right) && \text{switch 1 and 3:} \\
&\to \left(8, \boxed{1}, 3, 2, 7, \boxed{5}, 4, 6\right) = \sigma'.
\end{aligned}$$

Note that there are a total of 7 exchanges involving adjacent components: 4 exchanges to bring $\boxed{5}$ to the 6^{th} component, and 3 exchanges to bring $\boxed{1}$ to the 2^{nd} component. Since 7 is an ***odd*** number, $(-1)^7 = -1$, so again, $sgn(\sigma') = -sgn(\sigma)$.

The general argument can thus be stated as follows: if we want to exchange $\sigma(i_1)$ and $\sigma(i_2)$, where $i_1 < i_2$, then we will need $k = i_2 - i_1$ exchanges involving only adjacent components to bring $\sigma(i_1)$ to the component i_2. Going backwards, we will need $k - 1$ exchanges involving only adjacent components to bring $\sigma(i_2)$ — which is now in component $i_2 - 1$ — to the component i_1. Since $k + (k-1) = 2k - 1$ is an ***odd*** number, $sgn(\sigma') = (-1)^{2k-1} sgn(\sigma) = -sgn(\sigma)$. ■

The Balance of Even and Odd Permutations

We also saw in the formula for the 3×3 determinant that there were 3 terms with coefficient $+1$ and also 3 terms with coefficient -1. This is not a coincidence either:

Theorem: Exactly ***half*** of the $n!$ permutations of $\{1, 2, 3, .., n\}$ are ***even***, and ***half*** are ***odd***.

Proof: Let $\sigma = (i_1, i_2, i_3, \ldots, i_n)$ be a permutation. If we switch the first two components, we get a new permutation:

$$\sigma' = (i_2, i_1, i_3, \ldots, i_n)$$

From the previous Theorem, we know that $sgn(\sigma) = -sgn(\sigma^{/})$. Thus, one of them is even, and one of them is odd. These are the ***only*** permutations that contain i_3 through i_n in the last $n-2$ components. If we change the values of i_3 through i_n and again put them in the last $n-2$ components, then there will again be exactly two permutations that look like this, and one will be even and the other odd. Thus, the $n!$ permutations come in ***pairs***, where one permutation in each pair is even, and the other permutation in the pair is odd. ■

To help you understand this proof better, we list in the table below all the even permutations of $\{1, 2, 3\}$ on the left column, then we switched the first two components to produce the corresponding odd permutation on the right column:

Even Permutations	Odd Permutations
$(1, 2, 3)$	$(2, 1, 3)$
$(2, 3, 1)$	$(3, 2, 1)$
$(3, 1, 2)$	$(1, 3, 2)$

5.1 Section Summary

If $A = \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix}$, then: $det(A) = a_{1,1}a_{2,2} - a_{1,2}a_{2,1}$.

If $A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$, then: $det(A) = \begin{matrix} a_{1,1}a_{2,2}a_{3,3} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{3,2} \\ -a_{1,3}a_{2,2}a_{3,1} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3}. \end{matrix}$

An ordered list consisting of the numbers $1, 2, \ldots, n$, with each number appearing exactly once, is called a ***permutation*** of the set $\{1, 2, 3, .., n\}$. There are $n!$ permutations of $\{1, 2, 3, .., n\}$. We denote a permutation by a lowercase Greek letter like σ or τ.

An ***inversion*** occurs in σ if a number on the left is bigger than a number to its right.

A permutation σ is ***even*** if it has an even number of inversions, and σ is ***odd*** if it has an odd number of inversions. The ***sign*** of σ, denoted $sgn(\sigma)$, is $+1$ if σ is even, and -1 if σ is odd.

A permutation σ is also a ***bijection*** of the set $\{1, 2, 3, .., n\}$, and as such possesses an ***inverse***, σ^{-1}.

Permutations can be represented by a directed bipartite graph. We will use two copies of the set $\{1, 2, 3, .., n\}$, one on top of the other, to serve as our vertices. If $\sigma(i) = j$, we will have a directed edge (an arrow) from i on the top row to j on the bottom row. An inversion occurs when two of these edges intersect.

The graph representing σ^{-1} is the same as that of σ, only with the arrows reversed. Thus, the number of ***inversions*** in σ and σ^{-1} are the ***same***, so $sgn(\sigma) = sgn(\sigma^{-1})$.

If $\sigma^{/}$ is obtained from σ by ***exchanging*** any two components, then $sgn(\sigma^{/}) = -sgn(\sigma)$.

Consequently, ***half*** of the $n!$ permutations of $\{1, 2, 3, .., n\}$ are ***even***, and ***half*** are ***odd***.

5.1 Exercises

For Exercises (1) to (12): Compute the determinant using the 2×2 and 3×3 formulas:

1. $\begin{vmatrix} 7 & 5 \\ 5 & 4 \end{vmatrix}$

2. $\begin{vmatrix} -2 & 5 \\ 3 & 4 \end{vmatrix}$

3. $\begin{vmatrix} -1/2 & 5/3 \\ 3/2 & 7/3 \end{vmatrix}$

4. $\begin{vmatrix} \sqrt{6} & -7 \\ -\sqrt{3} & \sqrt{2} \end{vmatrix}$

5. $\begin{vmatrix} 4 & \ln(3) \\ -7 & \ln(2) \end{vmatrix}$

6. $\begin{vmatrix} \sin\left(\frac{\pi}{8}\right) & \cos\left(\frac{\pi}{8}\right) \\ -\sin\left(\frac{\pi}{24}\right) & \cos\left(\frac{\pi}{24}\right) \end{vmatrix}$

7. $\begin{vmatrix} 3 & 8 & 2 \\ -1 & 5 & -3 \\ 4 & -7 & 6 \end{vmatrix}$

8. $\begin{vmatrix} -5 & -2 & 3 \\ 2 & 1 & 3 \\ 3 & 8 & -7 \end{vmatrix}$

9. $\begin{vmatrix} -3/2 & -1 & 3/2 \\ 5/2 & 1/4 & 7/2 \\ 7/2 & 2 & -1/2 \end{vmatrix}$

10. $\begin{vmatrix} -5/3 & 2/3 & 4/3 \\ 2/5 & -3/5 & -7/5 \\ 7/2 & 9/2 & -3/2 \end{vmatrix}$

11. $\begin{vmatrix} 5\sqrt{2} & 2\sqrt{3} & 3\sqrt{6} \\ 7\sqrt{2} & -4\sqrt{3} & \sqrt{6} \\ -\sqrt{2} & 9\sqrt{3} & -5\sqrt{6} \end{vmatrix}$

12. $\begin{vmatrix} 5\ln(2) & 8\ln(3) & 7\ln(5) \\ 3 & 1 & -6 \\ -2 & -3 & 4 \end{vmatrix}$

For Exercises (13) to (21): (a) Find the determinants of the following matrices in terms of a, b, θ and ϕ, where appropriate, (b) determine under what conditions the matrix is invertible, and (c) when the matrix is invertible, find the inverse of the matrix.

13. $\begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$

14. $\begin{bmatrix} a & b \\ -b & a \end{bmatrix}$

15. $\begin{bmatrix} a & b \\ b & a \end{bmatrix}$

16. $\begin{bmatrix} a & a \\ -b & b \end{bmatrix}$

17. $\begin{bmatrix} 1 & a \\ 1 & b \end{bmatrix}$

18. $\begin{bmatrix} e^{2a} & e^{-a} \\ -e^{2a} & e^{-a} \end{bmatrix}$

19. $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$

20. $\begin{bmatrix} \cosh(a) & \sinh(a) \\ \sinh(a) & \cosh(a) \end{bmatrix}$

21. $\begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \cos(\phi) & \sin(\phi) \end{bmatrix}$

For Exercises (22) to (26): Sketch the directed bipartite graph representing each permutation. Use it to find the ***inverse*** of the permutation, and the number of ***inversions*** in the permutation.

22. $(3, 1, 4, 2)$
23. $(5, 3, 2, 4, 1)$
24. $(4, 6, 2, 5, 1, 3)$
25. $(6, 3, 7, 2, 5, 1, 4)$
26. $(6, 8, 3, 5, 1, 7, 2, 4)$

For Exercises (27) to (31): These are the same permutations that appear in (22) to (26). For each of these: (a) exchange the two boxed components, (b) count the number of inversions in the new permutation, and (c) check that the new permutation has the opposite sign as the original permutation.

27. $\left(\boxed{3}, 1, 4, \boxed{2}\right)$

28. $\left(\boxed{5}, 3, \boxed{2}, 4, 1\right)$

29. $\left(4, \boxed{6}, 2, 5, \boxed{1}, 3\right)$

30. $\left(6, 3, \boxed{7}, 2, 5, 1, \boxed{4}\right)$

31. $\left(6, \boxed{8}, 3, 5, 1, 7, \boxed{2}, 4\right)$

32. Suppose that $\vec{v} = \langle a, b \rangle$ is a ***unit vector*** in $\mathbb{R}^2$, L is the line $Span(\{\vec{v}\})$, and $L^\perp$ is the orthogonal complement of L. We found the matrices of $proj_L$, $proj_{L^\perp}$ and $refl_L$ in Section 2.2. Find their determinants.

   a. $[proj_L] = \begin{bmatrix} a^2 & ab \\ ab & b^2 \end{bmatrix}$

   b. $[proj_{L^\perp}] = \begin{bmatrix} b^2 & -ab \\ -ab & a^2 \end{bmatrix}$

   c. $[refl_L] = \begin{bmatrix} a^2 - b^2 & 2ab \\ 2ab & b^2 - a^2 \end{bmatrix}$

   Give an explanation as to why some of these matrices are invertible, but some are not.

33. Suppose that $\vec{n} = \langle a, b, c \rangle$ is a ***unit vector*** in $\mathbb{R}^3$, Π is the plane in $\mathbb{R}^3$ with equation $ax + by + cz = 0$, and L is the line $Span(\{\vec{n}\})$, the normal line to Π. We found the matrices of $proj_L$, $proj_\Pi$ and $refl_\Pi$ in Section 2.2. Find their determinants.

   a. $[proj_L] = \begin{bmatrix} a^2 & ab & ac \\ ab & b^2 & bc \\ ac & bc & c^2 \end{bmatrix}$

   b. $[proj_\Pi] = \begin{bmatrix} 1 - a^2 & -ab & -ac \\ -ab & 1 - b^2 & -bc \\ -ac & -bc & 1 - c^2 \end{bmatrix}$

   c. $[refl_\Pi] = \begin{bmatrix} 1 - 2a^2 & -2ab & -2ac \\ -2ab & 1 - 2b^2 & -2bc \\ -2ac & -2bc & 1 - 2c^2 \end{bmatrix}$

34. Find a formula for the determinant of the matrix: $\begin{bmatrix} 1 & a & a^2 \\ 1 & b & b^2 \\ 1 & c & c^2 \end{bmatrix}$.

Show that the determinant can be factored into the product of three ***linear*** factors.

35. What is the maximum number of inversions that can happen in a permutation σ of $\{1, 2, \ldots, n\}$? What σ will produce the maximum number of inversions?

36. List all the $4! = 24$ permutations of $\{1,2,3,4\}$ in a table. On the left side, write the even permutations. Then switch the first two components of each permutation and write the new permutation on the right side. Check that the new permutation is odd. There should be 12 even permutations on the left and 12 even permutations on the right.

37. ***Revisiting The Cross Product:*** If $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$ are vectors from $\mathbb{R}^3$, we defined the vector:

$$\vec{u} \times \vec{v} = (u_2 v_3 - u_3 v_2)\vec{i} - (u_1 v_3 - u_3 v_1)\vec{j} + (u_1 v_2 - u_2 v_1)\vec{k},$$

as the ***cross product*** of $\vec{u}$ with $\vec{v}$ in Exercise 14 of Section 1.3. In that Exercise, we showed that $\vec{u} \times \vec{v}$ is ***orthogonal*** to both $\vec{u}$ and $\vec{v}$.

a. Show that we can write the formula above as a determinant:

$$\vec{u} \times \vec{v} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix}$$

(Strictly speaking, since $\vec{i}$, $\vec{j}$ and $\vec{k}$ are ***not*** scalars, we call this a ***formal*** determinant.)

b. Prove that $\|\vec{u} \times \vec{v}\|^2 = \|\vec{u}\|^2 \bullet \|\vec{v}\|^2 - (\vec{u} \circ \vec{v})^2$. Note: the left side is just a messy expansion.

c. Use the previous part to prove that $\|\vec{u} \times \vec{v}\| = \|\vec{u}\| \, \|\vec{v}\| \sin(\theta)$, where as usual, θ is the angle between $\vec{u}$ and $\vec{v}$. Hint: recall the formula for $\vec{u} \circ \vec{v}$ that involves θ.

d. Use (c) to prove that $\vec{u}$ and $\vec{v}$ are ***parallel*** to each other ***if and only if*** $\vec{u} \times \vec{v} = \vec{\mathbf{0}}_3$. Don't forget to include the case when one of the vectors is already $\vec{\mathbf{0}}_3$.

e. Prove that if $\vec{u}$ and $\vec{v}$ are ***orthogonal unit vectors***, then $\vec{u} \times \vec{v}$ is also a ***unit vector***.

f. Draw the parallelogram in space formed by $\vec{u}$ and $\vec{v}$ in standard position, and use (c) to prove that $\|\vec{u} \times \vec{v}\|$ is the ***area*** of the parallelogram that they determine.

g. Now, suppose $\vec{a} = \langle a_1, a_2 \rangle$ and $\vec{b} = \langle b_1, b_2 \rangle$ are vectors from $\mathbb{R}^2$. Construct the vectors $\vec{u} = \langle a_1, a_2, 0 \rangle$ and $\vec{v} = \langle b_1, b_2, 0, \rangle$, and the 2×2 matrix $A = \begin{bmatrix} \vec{a} & \vec{b} \end{bmatrix}$.

Prove that $\|\vec{u} \times \vec{v}\| = |det(A)|$. Explain why this formula, together with part (e), tells us that $|det(A)|$ is the area of the parallelogram determined by the vectors $\vec{a}$ and $\vec{b}$ in standard position, as shown on the cover page of Chapter 5.

h. Suppose $\vec{u}$, $\vec{v}$ and $\vec{w}$ are from $\mathbb{R}^3$. Form the matrix B with these vectors as rows ***in this order*** (the result is the same if we use the vectors as columns, but the standard definition of the cross product is expressed in rows). Prove that $(\vec{u} \times \vec{v}) \circ \vec{w} = det(B)$. Using the formula that the ***volume*** of a ***parallelepiped*** is its ***base area*** multiplied its ***height***, and the properties above, show that $|det(B)|$ is the volume of the parallelepiped determined by $\vec{u}$, $\vec{v}$ and $\vec{w}$. Hint: you will need both the statement and the idea behind part (e).

Parts (f) and (h) are illustrated on the cover page of Chapter 5.

5.2 A General Determinant Formula

We are now ready to present a general formula for the determinant of any $n \times n$ matrix. Our definition will extend the pattern that we saw for 2×2 and 3×3 matrices.

There will be $n!$ terms in the determinant formula for an $n \times n$ matrix. Each term will correspond to a permutation σ of $\{1, 2, \ldots, n\}$. Each term in our formula will contain ***n factors***. As before, we will write the typical term with the row numbers in the correct order, so the typical term will look like:

$$\pm a_{1,\_}a_{2,\_}a_{3,\_}\cdots a_{n,\_}.$$

For the term corresponding to σ, the column numbers will be $\sigma(1)$, $\sigma(2)$, …, $\sigma(n)$. The term will have a positive coefficient if the permutation is even, and it will have a negative coefficient if the permutation is odd. We summarize this in the following:

Definition: Let A be an $n \times n$ matrix with entry $a_{i,j}$ in row i, column j, as usual Then:

$$det(A) = \sum_{\text{all permutations } \sigma \text{ of } \{1,2,\ldots,n\}} sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{n-1,\sigma(n-1)} \cdot a_{n,\sigma(n)}.$$

Recall that we can also write $|A|$ instead of $det(A)$. We will also replace the matrix brackets with vertical bars to denote a determinant instead of a matrix.

Example: Let us pretend that we are finding the determinant of a 5×5 matrix. There are $5! = 120$ terms in this determinant. Let us consider only the term:

$$\pm a_{1,3}a_{2,5}a_{3,4}a_{4,1}a_{5,2}.$$

This single term can be visualized like we did for 3×3 matrices:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} & a_{1,5} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} & \boxed{a_{2,5}} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} & a_{3,5} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} & a_{4,5} \\ a_{5,1} & \boxed{a_{5,2}} & a_{5,3} & a_{5,4} & a_{5,5} \end{bmatrix}.$$

The columns numbers were produced by the permutation:

$$\sigma = (3, 5, 4, 1, 2),$$

The inversions in σ are:

$$3 > 1,\ 3 > 2,\ 5 > 4,\ 5 > 1,\ 5 > 2,\ 4 > 1, \text{ and } 4 > 2,$$

so there are 7 inversions. Thus, σ is ***odd*** and the coefficient of this term is -1. The final signed term is:

$$-a_{1,3}a_{2,5}a_{3,4}a_{4,1}a_{5,2}. \square$$

The previous diagram shows an important property of each term in the determinant formula:

Every term contains ***exactly one factor*** from each ***row***, and from each ***column***.

Clearly, we do not want to do this 120 times, each time computing the product of five numbers after which we have to add them all together. Soon we will find more practical methods to compute the determinant. Before that, though, there are some easy properties we can show.

Basic Properties of $det(A)$

The first property tells us when we can compute $det(A)$ with almost no effort:

Theorem: Let A be an $n \times n$ matrix with a ***row*** of ***zeroes***. Then $det(A) = 0$.

Proof: Suppose row k contains all zeroes. The determinant formula says that each term has exactly one factor from row 1, one factor from row 2, and so on. Thus we will have one factor from row k, and thus the entire product will be 0. Since all the terms are 0, $det(A) = 0$. ■

We defined the determinant by requiring that the n factors that appear for each term have their row numbers in the correct ascending order. Why didn't we require the ***column*** numbers to be in ascending order instead, and use the sign of the resulting permutation on the ***rows***? The answer is . . . for no particular reason. The truth is, we could also have defined the determinant with the columns in the correct order, and we still would have gotten the exact same answer. Let us see why:

Example: Consider a 5×5 determinant and the term we saw above:

$$-a_{1,3}a_{2,5}a_{3,4}a_{4,1}a_{5,2}.$$

Let us rearrange the factors so that the ***columns*** are in ***ascending order***:

$$a_{4,1}a_{5,2}a_{1,3}a_{3,4}a_{2,5}.$$

The resulting permutation of the ***rows*** is $\tau = (4,5,1,3,2)$. But notice that this is the ***inverse permutation*** of the original permutation $\sigma = (3,5,4,1,2)$. Since a permutation and its inverse have the same sign, this term will likewise have a coefficient of -1. □

This is indeed true in general for any of the terms. Thus, we also get the formula:

$$det(A) = \sum_{\substack{\text{all permutations} \\ \sigma \text{ of } \{1,2,\ldots,n\}}} sgn(\sigma) \cdot a_{\sigma(1),1} \cdot a_{\sigma(2),2} \cdot \cdots \cdot a_{\sigma(n-1),n-1} \cdot a_{\sigma(n),n}.$$

This observation leads us to the following:

Theorem: Let A be an $n \times n$ matrix. Then $det(A) = det(A^\top)$.

Proof: Let us demonstrate the idea behind the proof on a 4×4 matrix A:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}, \text{ with } A^{\top} = \begin{bmatrix} a_{1,1} & a_{2,1} & a_{3,1} & \boxed{a_{4,1}} \\ a_{1,2} & \boxed{a_{2,2}} & a_{3,2} & a_{4,2} \\ \boxed{a_{1,3}} & a_{2,3} & a_{3,3} & a_{4,3} \\ a_{1,4} & a_{2,4} & \boxed{a_{3,4}} & a_{4,4} \end{bmatrix}.$$

For convenience and to avoid confusing notation, let us call $A^{\top}$ simply B.

Consider the term from $det(A)$, whose factors we boxed above:

$$a_{1,3}a_{2,2}a_{3,4}a_{4,1},$$

This corresponds to the permutation $\sigma = (3,2,4,1)$. Its inversions are:

$$3 > 2,\ \ 3 > 1,\ \ 2 > 1,\ \text{ and }\ 4 > 1.$$

Thus, σ is ***even***. We have also boxed the identical ***numbers*** in B which appear in the corresponding term for $det(B)$. They are found in ***different entries*** of B, though, so we must understand the permutation that produces this term from $det(B)$. Since $a_{i,j} = b_{j,i}$ for all i, j, we have:

$$a_{1,3}a_{2,2}a_{3,4}a_{4,1} = b_{3,1}b_{2,2}b_{4,3}b_{1,4} = b_{1,4}b_{2,2}b_{3,1}b_{4,3}.$$

But the permutation corresponding to this term from $det(B)$ results once again from the ***inverse*** permutation $\sigma^{-1} = (4,2,1,3)$. Since $sgn(\sigma) = sgn(\sigma^{-1})$, the coefficient of this term is also $+1$.

More generally, this idea tells us that if $B = A^{\top}$, and:

$$sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{n-1,\sigma(n-1)} \cdot a_{n,\sigma(n)}$$

is a typical term from $det(A)$ corresponding to the permutation σ, then the same entries in A can be found in B as $b_{\sigma(1),1}$, $b_{\sigma(2),2}$, $\ldots$, $b_{\sigma(n),n}$. These terms are in increasing ***column*** order. But these terms can be listed instead so that the ***rows*** are in increasing order, as $b_{1,\sigma^{-1}(1)}$, $b_{2,\sigma^{-1}(2)}$, $\ldots$, $b_{n,\sigma^{-1}(n)}$. Since $sgn(\sigma^{-1}) = sgn(\sigma)$, this term from $det(B)$ will have the same sign as the original term from $det(A)$. Since all the terms in $det(A)$ are in a one-to-one correspondence with an equal term in $det(B)$, the two determinants are equal.■

This Theorem has the effect of enabling us to state any future observations in terms of ***rows*** as well as ***columns***. Thus, for example, we can also say:

Theorem: Let A be an $n \times n$ matrix with a ***column*** of zeroes. Then $det(A) = 0$.

We also have another useful, though not so obvious result:

Theorem: Let A be an $n \times n$ matrix with two ***proportional rows*** (or, in particular, two ***equal rows***). Then $det(A) = 0$. Similarly, a matrix with ***proportional columns*** also has zero determinant.

Proof: Again, let us first see the idea behind the proof by looking at a 4×4 matrix where the second row is a multiple of the fourth row:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ ka_{4,1} & \boxed{ka_{4,2}} & ka_{4,3} & ka_{4,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}$$

Once again, we have boxed the term corresponding to the ***even*** permutation $\sigma = (3,2,4,1)$. However, another term appearing in $det(A)$ corresponds to the boxed entries below:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ \boxed{ka_{4,1}} & ka_{4,2} & ka_{4,3} & ka_{4,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} \\ a_{4,1} & \boxed{a_{4,2}} & a_{4,3} & a_{4,4} \end{bmatrix}$$

This term has ***exactly the same factors*** as the original term, but now corresponds to the permutation $\sigma' = (3,1,4,2)$, that is, where the 2nd and 4th entries of σ are ***switched***. But we know that $sgn(\sigma') = -sgn(\sigma)$, so this second term now corresponds to an ***odd*** permutation. Therefore, these two terms will add up to ***zero***. Notice that the 2nd and 4th rows are ***parallel***, and we switched the 2nd and 4th ***components*** of σ.

Let us now generalize this argument. Suppose that *row j* of A is k times *row i* of A (and assume without loss of generality that $i < j$). Thus, $a_{j,\sigma(j)} = k \bullet a_{i,\sigma(j)}$. Notice that it is the ***column*** numbers that are the same. A typical term in the determinant formula for A will look like:

$$\begin{aligned} & sgn(\sigma) \bullet a_{1,\sigma(1)} \bullet a_{2,\sigma(2)} \bullet \cdots \bullet a_{i,\sigma(i)} \bullet \cdots \bullet a_{j,\sigma(j)} \bullet \cdots \bullet a_{n-1,\sigma(n-1)} \bullet a_{n,\sigma(n)} \\ = {} & sgn(\sigma) \bullet a_{1,\sigma(1)} \bullet a_{2,\sigma(2)} \bullet \cdots \bullet a_{i,\sigma(i)} \bullet \cdots \bullet k \bullet a_{i,\sigma(j)} \bullet \cdots \bullet a_{n-1,\sigma(n-1)} \bullet a_{n,\sigma(n)} \\ = {} & k \bullet sgn(\sigma) \bullet a_{1,\sigma(1)} \bullet a_{2,\sigma(2)} \bullet \cdots \bullet a_{i,\sigma(i)} \bullet \cdots \bullet a_{i,\sigma(j)} \bullet \cdots \bullet a_{n-1,\sigma(n-1)} \bullet a_{n,\sigma(n)} \end{aligned}$$

However, if we consider the permutation σ' obtained from σ by exchanging the i and j coordinates, we get a second term that appears in the determinant, namely:

$$\begin{aligned} & sgn(\sigma') \bullet a_{1,\sigma'(1)} \bullet a_{2,\sigma'(2)} \bullet \cdots \bullet a_{i,\sigma'(i)} \bullet \cdots \bullet a_{j,\sigma'(j)} \bullet \cdots \bullet a_{n-1,\sigma'(n-1)} \bullet a_{n,\sigma'(n)} \\ = {} & -sgn(\sigma) \bullet a_{1,\sigma'(1)} \bullet a_{2,\sigma'(2)} \bullet \cdots \bullet a_{i,\sigma'(i)} \bullet \cdots \bullet k \bullet a_{i,\sigma'(j)} \bullet \cdots \bullet a_{n-1,\sigma'(n-1)} \bullet a_{n,\sigma'(n)} \\ = {} & -k \bullet sgn(\sigma) \bullet a_{1,\sigma'(1)} \bullet a_{2,\sigma'(2)} \bullet \cdots \bullet a_{i,\sigma'(i)} \bullet \cdots \bullet a_{i,\sigma'(j)} \bullet \cdots \bullet a_{n-1,\sigma'(n-1)} \bullet a_{n,\sigma'(n)} \\ = {} & -k \bullet sgn(\sigma) \bullet a_{1,\sigma(1)} \bullet a_{2,\sigma(2)} \bullet \cdots \bullet a_{i,\sigma(j)} \bullet \cdots \bullet a_{i,\sigma(i)} \bullet \cdots \bullet a_{n-1,\sigma(n-1)} \bullet a_{n,\sigma(n)}. \end{aligned}$$

Notice that all the factors in the original term appear, except for the order of appearance of $a_{i,\sigma(j)}$ and $a_{i,\sigma(i)}$. However, this term now has the ***opposite sign*** of the original term. Thus, these two terms add up to ***zero***. Since all terms in $det(A)$ can be put into ***pairs*** that add up to zero, $det(A)$ must be zero.■

This proof also demonstrates the usefulness of the Theorem that we proved regarding the effect of switching ***any*** two coordinates in a permutation.

Let us now look at a non-trivial case where we only need to compute ***one*** term to compute the determinant:

Determinants of Triangular Matrices

We can easily compute the determinant of the triangular matrices introduced in Section 2.9:

> ***Theorem:*** Let A be an ***upper*** or a ***lower triangular*** matrix, that is, $a_{i,j} = 0$ for all $i > j$, ***or*** $a_{i,j} = 0$ for all $i < j$. Then:
>
> $$det(A) = a_{1,1} \cdot a_{2,2} \cdot \cdots \cdot a_{n-1,n-1} \cdot a_{n,n},$$
>
> that is, the product of the ***diagonal*** entries. In particular:
>
> $$\text{if } D = Diag(d_1, d_2, \ldots, d_n), \text{ then } det(D) = d_1 \cdot d_2 \cdot \cdots \cdot d_n.$$

Proof: Let us see the idea of the proof by looking at an upper triangular 4×4 matrix:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ 0 & a_{2,2} & a_{2,3} & a_{2,4} \\ 0 & 0 & a_{3,3} & a_{3,4} \\ 0 & 0 & 0 & a_{4,4} \end{bmatrix}.$$

The determinant formula says that we have to choose an entry from ***every row*** and from ***every column***, without repeating a row or column. But as soon as we pick one of the zeroes in column 1, that entire term will be zero. Thus, we must choose the term $a_{1,1}$ from column 1 in order to have any ***chance*** of producing a non-zero term. But now that we have chosen $a_{1,1}$, we cannot use any other term from ***row 1*** to go with this term. Thus, we cannot choose $a_{1,2}$ from column 2. Again, if we choose any of the zeroes appearing in column two, the term we produce is automatically zero. Thus, the only term we can choose from column 2 that has any ***chance*** of producing a non-zero term will be $a_{2,2}$. Continuing with this argument, we now cannot use any entry from row 1 or row 2 when we choose an entry from column 3. Thus, the only term we can choose from column 3 that has any chance of producing a non-zero term will be $a_{3,3}$. Similarly, we must choose $a_{4,4}$ in order to have any chance of producing a non-zero term.

We can easily see that this argument applies in general to an $n \times n$ upper triangular matrix. The only term that has any chance of being non-zero would be the term:

$$\pm a_{1,1} \cdot a_{2,2} \cdot \cdots \cdot a_{n-1,n-1} \cdot a_{n,n}.$$

The permutation that corresponds to this term is $\iota = (1, 2, \ldots, n-1, n)$, the identity permutation, with $sgn(\iota) = +1$. This completes our proof.■

Example: Consider the matrix:

$$A = \begin{bmatrix} 5 & 12 & 753 & 2^{12} \\ 0 & 3 & \sqrt{\pi} & 0 \\ 0 & 0 & -2 & 1/e \\ 0 & 0 & 0 & 1/10 \end{bmatrix}.$$

Then A is an upper triangular matrix, and so $det(A) = 5 \cdot 3 \cdot (-2) \cdot 1/10 = -3.$ □

Now that we appreciate how difficult it is to directly apply the definition in order to compute the determinant, in general, we want to develop techniques that will simplify the process. The first involves ***row operations***, so we proceed by reviewing elementary matrices.

Determinants of Elementary Matrices

Recall that an elementary matrix E is produced by performing a single elementary row operation on the identity matrix $\boldsymbol{I}_n$. We also saw in Chapter 2 that if A is any $n \times m$ matrix, then we can compute the product $E \cdot A$ by performing exactly the same row operation on A that was used to produce E from $\boldsymbol{I}_n$. The determinant properties let us compute the determinant of elementary matrices very easily:

Theorem: Suppose E is an ***elementary matrix***. If E is obtained from $\boldsymbol{I}_n$ by:
1. ***multiplying*** row i by $k \neq 0$, then $det(E) = k$.
2. ***exchanging*** row i and row j, then $det(E) = -1$.
3. ***adding*** k times row i to row j, then $det(E) = 1$.

Consequently, the determinant of every elementary matrix is ***non-zero***.

Proof: The Type 1 elementary matrix obtained from $\boldsymbol{I}_n$ by multiplying row i by $k \neq 0$ is already a ***diagonal*** matrix, with all 1's on the diagonal except for a single k. By our formula, $det(E) = k$. Similarly, a Type 3 elementary matrix obtained from $\boldsymbol{I}_n$ by adding k times row i to row j is either lower or upper ***triangular***, with all 1's on the diagonal. Thus $det(E) = 1$. Finally, if E is obtained from $\boldsymbol{I}_n$ by exchanging row i and row j, then the only non-zero term of $det(E)$ involves the entries which are all 1's. But the resulting permutation is $(1, 2, 3, \ldots, n)$, ***except*** i and j are exchanged. Since this permutation is obtained from the identity permutation ι by switching two components, the sign of this permutation is -1, and thus $det(E) = -1$. ■

Example: Shown below are examples of a Type 1, 2 and 3 elementary 4×4 matrix, respectively, and their determinants:

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -5 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = -5, \quad \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{vmatrix} = -1, \quad \text{and} \quad \begin{vmatrix} 1 & 0 & 7 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix} = 1. \square$$

The Effect of Row Operations

Now we are ready to study how elementary row operations affect the value of the determinant.

Theorem: Let A be an $n \times n$ matrix. Suppose B is obtained from A by:
1. ***multiplying*** row i of A by $k \neq 0$. Then: $det(B) = k \cdot det(A)$.
2. ***exchanging*** row i and row j of A. Then: $det(B) = -det(A)$.
3. ***adding*** k times row i of A to row j of A. Then: $det(B) = det(A)$.

Analogous statements can be made by replacing the word ***row*** with the word ***column***.

Consequently if E is the elementary matrix corresponding to the row operation performed, then $B = E \cdot A$, and so: $det(E \cdot A) = det(E) \cdot det(A)$.

In particular: $det(k \cdot A) = k^n \cdot det(A)$.

Proof: Let us see the ideas behind the proof by considering a general 4×4 matrix A:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}.$$

Again, we boxed that four entries that produce the term in $det(A)$ corresponding to the ***even*** permutation $\sigma = (3,2,4,1)$. Suppose now that we obtain B_1 from A by the Type 1 operation where we multiply row 3 of A by the non-zero number k:

$$B_1 = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \\ k \cdot a_{3,1} & k \cdot a_{3,2} & k \cdot a_{3,3} & \boxed{k \cdot a_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}.$$

We boxed the entries in B_1 which correspond to the four entries in A. The resulting term is exactly the same as the original, with an extra factor of k. Since every term of $det(B_1)$ will have an extra factor of k in the corresponding term from $det(A)$, we get the desired result that $det(B_1) = k \cdot det(A)$. We can see that this argument works in general for any $n \times n$ matrix: a typical term in $det(A)$ will look like:

$$sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{i,\sigma(i)} \cdot a_{i+1,\sigma(i+1)} \cdot \cdots \cdot a_{n-1,\sigma(n-1)} \cdot a_{n,\sigma(n)}$$

If B_1 is obtained from A by multiplying every entry in row i of A by a non-zero number k, then the corresponding term in $det(B_1)$ will be:

$$sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot (ka_{i,\sigma(i)}) \cdot a_{i+1,\sigma(i+1)} \cdot \cdots \cdot a_{n-1,\sigma(n-1)} \cdot a_{n,\sigma(n)}.$$

Since each term in $det(B_1)$ will have an extra factor of k, $det(B_1) = k \cdot det(A)$.

Now, suppose we obtain B_2 from A by exchanging rows 2 and 4 of A:

$$B_2 = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \end{bmatrix},$$

where we have boxed the ***same numbers*** that appear in B_2. Notice, though, that ***two*** of them are now found in ***different locations***. The corresponding permutation that will yield this term will be $\sigma' = (3,1,4,2)$, which is the permutation obtained from σ by exchanging the 2nd and 4th components.

Thus, $sgn(\sigma^{/}) = -sgn(\sigma)$. Since every term of $det(B_2)$ will have an extra factor of -1 compared to the corresponding term from $det(A)$, we get the desired result that $det(B_2) = -det(A)$. Again, this argument works in general, and we leave it as an Exercise to write a general proof for Type 2 row operations.

Finally, in order to see the idea for a Type 3 operation, we need the following more general result:

Lemma: Let A, B and C be $n \times n$ matrices that have all entries equal ***except*** for the entries in row i. However, row i of C is the ***sum*** of row i of A and row i of B. Then:

$$det(C) = det(A) + det(B).$$

Warning: This Theorem is not saying that $C = A + B$, nor is it saying that $det(A+B) = det(A) + det(B)$. In fact, in general, this equation is ***false***. Most of the time:

$$det(A+B) \neq det(A) + det(B).$$

Proof of the Lemma: Let us look at how A, B and C might look like for a 4×4 matrix. We have chosen all rows to be the same, except for the ***3rd row***, and so the three matrices are:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & \boxed{a_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}, \quad B = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \\ b_{3,1} & b_{3,2} & b_{3,3} & \boxed{b_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}, \quad \text{and}$$

$$C = \begin{bmatrix} a_{1,1} & a_{1,2} & \boxed{a_{1,3}} & a_{1,4} \\ a_{2,1} & \boxed{a_{2,2}} & a_{2,3} & a_{2,4} \\ a_{3,1}+b_{3,1} & a_{3,2}+b_{3,2} & a_{3,3}+b_{3,3} & \boxed{a_{3,4}+b_{3,4}} \\ \boxed{a_{4,1}} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}.$$

The term from $det(C)$ corresponding to the boxed entries above will be:

$$\begin{aligned} & sgn(\sigma) \cdot a_{1,3} \cdot a_{2,2} \cdot (a_{3,4} + b_{3,4}) \cdot a_{4,1} \\ = {} & sgn(\sigma) \cdot a_{1,3} \cdot a_{2,2} \cdot a_{3,4} \cdot a_{4,1} + \\ & sgn(\sigma) \cdot a_{1,3} \cdot a_{2,2} \cdot b_{3,4} \cdot a_{4,1}, \end{aligned}$$

where $\sigma = (3,2,4,1)$, as before. We can easily see that the term on the second line corresponds to a term in $det(A)$ and the term on the third line corresponds to a term in $det(B)$ Since every term in $det(C)$ can be ***distributed*** into the sum of a term in $det(A)$ and a term in $det(B)$, the conclusion of the Lemma follows for this example. More generally, if all three matrices are $n \times n$ and row k of C is the sum of row k of A and row k of B, with all other rows equal, a typical term in $det(C)$ will be:

$$sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{k-1,\sigma(k-1)} \cdot (a_{k,\sigma(k)} + b_{k,\sigma(k)}) \cdot a_{k+1,\sigma(k+1)} \cdot \cdots \cdot a_{n,\sigma(n)}$$

$$= sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{k-1,\sigma(k-1)} \cdot a_{k,\sigma(k)} \cdot a_{k+1,\sigma(k+1)} \cdot \cdots \cdot a_{n,\sigma(n)} +$$

$$sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{k-1,\sigma(k-1)} \cdot b_{k,\sigma(k)} \cdot a_{k+1,\sigma(k+1)} \cdot \cdots \cdot a_{n,\sigma(n)}.$$

Again, the term on the second line corresponds to a term in $det(A)$ and the term on the third line corresponds to a term in $det(B)$, and so the Lemma follows in general.■

Now we can go back to the Theorem to see the idea to prove the formula involving a Type 3 row operation. Suppose we obtain B_3 from A by adding k times row 1 to row 3 for our arbitrary 4×4 matrix A:

$$B_3 = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} + k \cdot a_{1,1} & a_{3,2} + k \cdot a_{1,2} & a_{3,3} + k \cdot a_{1,3} & a_{3,4} + k \cdot a_{1,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}$$

We can now apply the Lemma: B_3 has exactly the same entries as the matrices:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ k \cdot a_{1,1} & k \cdot a_{1,2} & k \cdot a_{1,3} & k \cdot a_{1,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix},$$

except that row 3 of B_3 is the sum of row 3 of A and row 3 of B. But notice that row 1 and row 3 of B are ***proportional***. Thus, $det(B) = 0$. We can now apply our Lemma to conclude:

$$det(B_3) = det(A) + det(B) = det(A) + 0 = det(A).$$

We leave it as an Exercise to generalize these ideas to a proof for any $n \times n$ elementary matrix E and any $n \times n$ matrix A.■

Finding det(A) Using Row and Column Operations

The previous Theorem allows us to perform row or column operations (keeping track only of Type 1 and Type 2 operations) until we obtain a ***triangular matrix***. If we ***factor out*** a non-zero number k from a row or column using a Type 1 operation, we need to ***multiply*** the final determinant by k to account for this row operation. Also, we count the total number t of ***row exchanges*** that we perform, and multiply our final determinant by $(-1)^t$ to account for all the Type 2 operations. Alternatively, you can also keep track of swaps by ***pairs***, switching signs back and forth. The Type 3 row operations do not change our determinant, so we need not keep track of them.

Example: Let us compute the determinant of:

$$A = \begin{bmatrix} 5 & 8 & 3 & -7 \\ -3 & 6 & 4 & 8 \\ 1 & 2 & -1 & -5 \\ 7 & -9 & -2 & 6 \end{bmatrix}$$

using row reduction. We see a "1" in row 3, column 1, so let us swap row 1 and row 3 first, but in so doing, we get a sign change:

$$det(A) = -\begin{vmatrix} 1 & 2 & -1 & -5 \\ -3 & 6 & 4 & 8 \\ 5 & 8 & 3 & -7 \\ 7 & -9 & -2 & 6 \end{vmatrix}.$$

Now, we clear the rest of column 1 using three Type 3 row operations, to get:

$$det(A) = -\begin{vmatrix} 1 & 2 & -1 & -5 \\ 0 & 12 & 1 & -7 \\ 0 & -2 & 8 & 18 \\ 0 & -23 & 5 & 41 \end{vmatrix}.$$

The easiest way to produce a leading 1 in row 2, column 2 is to factor out –2 from row 3, and switch row 3 with row 2. The beauty of this move is that we can forget about the first switch, since it now ***cancels*** with this second one. However, we still need to remember the "–2" factor. Clearing the rest of column 2 below row 2, we get:

$$det(A) = -2 \cdot \begin{vmatrix} 1 & 2 & -1 & -5 \\ 0 & 1 & -4 & -9 \\ 0 & 0 & 49 & 101 \\ 0 & 0 & -87 & -166 \end{vmatrix}.$$

This matrix is almost upper triangular, except for the –87. Now, we can leave the 49 alone and use a Type 3 operation by adding $\frac{87}{49}$ of row 3 to row 4, getting:

$$det(A) = -2 \cdot \begin{vmatrix} 1 & 2 & -1 & -5 \\ 0 & 1 & -4 & -9 \\ 0 & 0 & 49 & 101 \\ 0 & 0 & 0 & \frac{653}{49} \end{vmatrix}.$$

We get as our final answer:

$$det(A) = -2 \cdot 1 \cdot 1 \cdot 49 \cdot \frac{653}{49} = -1306. \square$$

5.2 Section Summary

Let A be an $n \times n$ matrix with entry $a_{i,j}$ in row i, column j. Then:

$$det(A) = \sum_{\substack{\text{all permutations} \\ \sigma \text{ of } \{1,2,\ldots,n\}}} sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{n-1,\sigma(n-1)} \cdot a_{n,\sigma(n)}$$

$$= \sum_{\substack{\text{all permutations} \\ \sigma \text{ of } \{1,2,\ldots,n\}}} sgn(\sigma) \cdot a_{\sigma(1),1} \cdot a_{\sigma(2),2} \cdot \cdots \cdot a_{\sigma(n-1),n-1} \cdot a_{\sigma(n),n}.$$

A and $A^\top$ have the same determinant.

If A has a ***row*** or a ***column*** of ***zeroes***, then $det(A) = 0$.

If A has two ***proportional*** rows or columns, then $det(A) = 0$.

If A is an upper or a lower ***triangular*** matrix, then $det(A) = a_{1,1} \cdot a_{2,2} \cdot \cdots \cdot a_{n,n}$.

In particular, if $D = Diag(d_1, d_2, \ldots, d_n)$, then $det(D) = d_1 \cdot d_2 \cdot \cdots \cdot d_n$.

Suppose E is an ***elementary matrix***. If E is obtained from $\boldsymbol{I}_n$ by:

1. ***multiplying*** row i by $k \neq 0$, then $det(E) = k$.
2. ***exchanging*** row i and row j, then $det(E) = -1$.
3. ***adding*** k times row i to row j, then $det(E) = 1$.

Consequently, the determinant of every elementary matrix is ***non-zero***.

Let A be an $n \times n$ matrix. Suppose B is obtained from A by:

1. ***multiplying*** row i of A by $k \neq 0$. Then: $det(B) = k \cdot det(A)$.
2. ***exchanging*** row i and row j of A. Then: $det(B) = -det(A)$.
3. ***adding*** k times row i of A to row j of A. Then: $det(B) = det(A)$.

Analogous statements can be made by replacing the word "row" with the word "column."

Consequently if E is the elementary matrix corresponding to the row operation performed, then $B = E \cdot A$, and so: $det(E \cdot A) = det(E) \cdot det(A)$.

In particular: $det(k \cdot A) = k^n \cdot det(A)$.

The determinant of A can be computed by performing row or column operations on A until we get a triangular matrix B. The determinant of A is obtained by multiplying the determinant of B by all the non-zero numbers k that were factored out of rows and columns using Type 1 operations, and by $(-1)^t$, where t is the total number of Type 2 operations. Type 3 operations do not affect the value of the determinant.

5.2 Exercises

For Exercises (1) to (6): Determine if the term has a positive or negative coefficient in the determinant formula for the corresponding matrix:

1. $\pm a_{1,3}a_{2,1}a_{3,4}a_{4,2}$
2. $\pm a_{1,2}a_{2,4}a_{3,3}a_{4,1}$
3. $\pm a_{1,4}a_{2,1}a_{3,5}a_{4,2}a_{5,3}$
4. $\pm a_{1,5}a_{2,3}a_{3,2}a_{4,4}a_{5,1}$
5. $\pm a_{1,4}a_{2,6}a_{3,1}a_{4,5}a_{5,2}a_{6,3}$
6. $\pm a_{1,6}a_{2,3}a_{3,4}a_{4,1}a_{5,2}a_{6,5}$

For Exercises (7) to (12): ***Pseudo-ku!*** Find the missing subscript/s in the determinant term, and decide if the resulting term has a positive or negative coefficient in the determinant formula:

7. $\pm a_{1,3}a_{2,4}a_{3,1}a_{4,?}$
8. $\pm a_{1,5}a_{2,3}a_{3,?}a_{4,2}a_{5,1}$
9. $\pm a_{5,4}a_{3,1}a_{1,5}a_{2,?}a_{4,2}$
10. $\pm a_{1,?}a_{2,3}a_{3,6}a_{4,2}a_{5,1}a_{6,4}$
11. $\pm a_{3,2}a_{5,4}a_{?,6}a_{1,3}a_{4,?}a_{6,1}$
12. $\pm a_{6,3}a_{3,5}a_{1,?}a_{8,6}a_{5,1}a_{2,8}a_{?,2}a_{7,4}$

For Exercises (13) to (20): Find the following determinants ***without*** using the original determinant formula. Explain your answers.

13. $\begin{vmatrix} 8 & \ln(1) & \sqrt{47} \\ 100 & \sin(5\pi) & 982 \\ e^{500} & \cos(\pi/2) & 537 \end{vmatrix}$

14. $\begin{vmatrix} 3 & -5 & 8 \\ 157 & 375 & -402 \\ 12 & -20 & 32 \end{vmatrix}$

15. $\begin{vmatrix} 3 & 61 & 3125 \\ 0 & -5 & -742 \\ 0 & 0 & 2 \end{vmatrix}$

16. $\begin{vmatrix} -2/3 & 1/5 & 3/5 \\ 0 & -6/5 & -3/8 \\ 0 & 0 & 7/4 \end{vmatrix}$

17. $\begin{vmatrix} 6 & 0 & 0 & 0 \\ 4 & 8 & 0 & 0 \\ -7 & 20 & -11 & 0 \\ 3 & -9 & 33 & 5 \end{vmatrix}$

18. $\begin{vmatrix} \sqrt{6} & 7 & 10^9 & -0.005 \\ 0 & \sqrt{3} & 4 & \pi \\ 0 & 0 & 5 & e^5 \\ 0 & 0 & 0 & \sqrt{8} \end{vmatrix}$

19. $\begin{vmatrix} -4 & 54 & 1 & 7 & 39 \\ 0 & 5 & -3 & 42 & 6 \\ 0 & 0 & 9 & 8 & 9 \\ 0 & 0 & 0 & -3 & 4 \\ 0 & 0 & 0 & 0 & 7 \end{vmatrix}$

20. $\begin{vmatrix} -2/3 & 0 & 0 & 0 & 0 \\ \sqrt{7} & 6/11 & 0 & 0 & 0 \\ e^{\pi} & 800 & 5/9 & 0 & 0 \\ 53 & \pi^{-e} & 0 & 33/4 & 0 \\ 4 & 40 & 400 & 99 & 3/20 \end{vmatrix}$

For Exercises (21) to (24): Explain why there is only one non-zero term in each of the following determinants. Compute that term, and hence the determinant. As part of your solution, find the only permutation that yields this term, and find the sign of this permutation.

21. $\begin{vmatrix} 0 & 7 & 0 & 0 \\ 0 & 0 & 0 & -8 \\ 5 & 0 & 0 & 0 \\ 0 & 0 & -2 & 0 \end{vmatrix}$

22. $\begin{vmatrix} 0 & 0 & 6 & 0 \\ 0 & 0 & 0 & -4 \\ 0 & 3 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{vmatrix}$

23. $\begin{vmatrix} 0 & 0 & 6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 \\ 0 & -2 & 0 & 0 & 0 \\ 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -5 & 0 \end{vmatrix}$

24. $\begin{vmatrix} 0 & 0 & 0 & -2/3 & 0 \\ 0 & -5/4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6/25 \\ 15 & 0 & 0 & 0 & 0 \\ 0 & 0 & -7/6 & 0 & 0 \end{vmatrix}$

25. Suppose A is a 3×3 matrix, and $\det(A) = 5$. Find the determinant of the following matrix products:

a. $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} A$

b. $\begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} A$

c. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & -4 & 0 \\ 0 & 0 & 1 \end{bmatrix} A$

d. $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{15} \end{bmatrix} A$

26. Suppose $A = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \vec{c}_3 & \vec{c}_4 \end{bmatrix}$ is a 4×4 matrix, partitioned into its four columns, and $\det(A) = -20$.

Find the determinant of the following matrices:

a. $\begin{bmatrix} 7\vec{c}_1 & 3\vec{c}_2 & \frac{1}{2}\vec{c}_3 & -\frac{1}{5}\vec{c}_4 \end{bmatrix}$

b. $\begin{bmatrix} \vec{c}_3 & \vec{c}_4 & \vec{c}_2 & \vec{c}_1 \end{bmatrix}$

c. $\begin{bmatrix} 5\vec{c}_4 & -3\vec{c}_1 & 4\vec{c}_3 & 9\vec{c}_2 \end{bmatrix}$

d. $\begin{bmatrix} \vec{c}_1 - 4\vec{c}_3 & \vec{c}_2 + 5\vec{c}_4 & 2\vec{c}_3 & -\vec{c}_4 \end{bmatrix}$

27. Suppose you are told that $\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = -9.$

Find the value of the following determinants, and provide an explanation. Hint: Think of the properties of determinants. Do not perform a brute-force computation.

a. $\begin{vmatrix} c_3 & b_3 & a_3 \\ c_1 & b_1 & a_1 \\ c_2 & b_2 & a_2 \end{vmatrix}$

b. $\begin{vmatrix} 5a_1 & 5a_2 & 5a_3 \\ -3b_1 & -3b_2 & -3b_3 \\ 2c_1 & 2c_2 & 2c_3 \end{vmatrix}$

c. $\begin{vmatrix} 2a_1 - 3c_1 & 7b_1 + 4c_1 & 2c_1 \\ 2a_2 - 3c_2 & 7b_2 + 4c_2 & 2c_2 \\ 2a_3 - 3c_3 & 7b_3 + 4c_3 & 2c_3 \end{vmatrix}$

d. $\begin{vmatrix} a_1 - b_1 & b_1 - c_1 & c_1 - a_1 \\ a_2 - b_2 & b_2 - c_2 & c_2 - a_2 \\ a_3 - b_3 & b_3 - c_3 & c_3 - a_3 \end{vmatrix}$

28. Suppose $A = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{bmatrix}$, and $\det(A) = 14$.

Find the determinants of the following matrices:

a. $\begin{bmatrix} -a_1 & -a_2 & -a_3 & -a_4 \\ b_1 - 3d_1 & b_2 - 3d_2 & b_3 - 3d_3 & b_4 - 3d_4 \\ c_1 + 4a_1 & c_2 + 4a_2 & c_3 + 4a_3 & c_4 + 4a_4 \\ 5d_1 & 5d_2 & 5d_3 & 5d_4 \end{bmatrix}$

b. $\begin{bmatrix} 10b_2 & 10d_2 & 10c_2 & 10a_2 \\ b_4/35 & d_4/35 & c_4/35 & a_4/35 \\ -3b_1 & -3d_1 & -3c_1 & -3a_1 \\ b_3/2 & d_3/2 & c_3/2 & a_3/2 \end{bmatrix}$

c. $\begin{bmatrix} 10a_2 & 5a_1 & 5a_3 & 20a_4 \\ -6b_2 & -3b_1 & -3b_3 & -12b_4 \\ 4c_2 & 2c_1 & 2c_3 & 8c_4 \\ \frac{2}{7}d_2 & \frac{1}{7}d_1 & \frac{1}{7}d_3 & \frac{4}{7}d_4 \end{bmatrix}$

d. $\begin{bmatrix} 6a_4 + 18b_4 & a_2 + 3b_2 & a_3 + 3b_3 - a_2 - 3b_2 & a_1 + 3b_1 \\ 42b_4 & 7b_2 & 7b_3 - 7b_2 & 7b_1 \\ 6c_4 - 12b_4 & c_2 - 2b_2 & c_3 - 2b_3 - c_2 + 2b_2 & c_1 - 2b_1 \\ 6d_4 + 30b_4 & d_2 + 5b_2 & d_3 + 5b_3 - d_2 - 5b_2 & d_1 + 5b_1 \end{bmatrix}$

For Exercises (29) to (40): Compute the following determinants by using row and/or column operations:

29. $\begin{vmatrix} 3 & -6 & 15 \\ -4 & -5 & -2 \\ 2 & 7 & 3 \end{vmatrix}$

30. $\begin{vmatrix} 4 & -1 & 5 \\ 8 & 2 & 7 \\ -9 & 3 & -6 \end{vmatrix}$

31. $\begin{vmatrix} 7 & -9 & 3 & 2 \\ 4 & 0 & 9 & -3 \\ 5 & 2 & -8 & 6 \\ -4 & 1 & 3 & -7 \end{vmatrix}$

32. $\begin{vmatrix} 2 & -1 & 5 & 1 \\ 0 & 4 & 1 & 2 \\ -5 & 2 & 3 & -6 \\ -1 & 3 & -2 & 4 \end{vmatrix}$

33. $\begin{vmatrix} 3 & 2 & -4 & 1 \\ -2 & 1 & 0 & 3 \\ 5 & 3 & 2 & -1 \\ 2 & -5 & 3 & 6 \end{vmatrix}$

34. $\begin{vmatrix} 2 & -1 & 7 & -1 \\ 5 & 0 & -2 & 2 \\ 7 & 2 & -3 & 8 \\ -4 & 6 & 2 & -9 \end{vmatrix}$

35. $\begin{vmatrix} 1 & 3 & -1 & 3 & -2 \\ 0 & -2 & -3 & 1 & 2 \\ -2 & 1 & 2 & -3 & 1 \\ 0 & 2 & 1 & -1 & 0 \\ -3 & 4 & 0 & 2 & 5 \end{vmatrix}$

36. $\begin{vmatrix} 2 & 5 & -1 & 3 & -2 \\ 3 & -2 & -3 & 1 & 2 \\ -2 & 1 & 2 & -3 & 1 \\ 1 & 2 & 1 & -1 & 0 \\ -3 & 4 & 0 & 2 & 5 \end{vmatrix}$

37. $\begin{vmatrix} 2 & 5 & -1 & 3 & -2 \\ -3 & -2 & 7 & 1 & 8 \\ 5 & 1 & 2 & -3 & 6 \\ 4 & 0 & 1 & -1 & 0 \\ 0 & 4 & 3 & 2 & 5 \end{vmatrix}$

38. $\begin{vmatrix} 7 & 5 & -1 & -3 & -2 \\ 2 & -2 & -3 & 1 & 9 \\ 3 & -1 & 2 & 3 & 1 \\ 6 & 0 & 1 & -5 & 1 \\ -3 & 1 & 4 & 2 & 2 \end{vmatrix}$

39. $\begin{vmatrix} 2 & 5 & -1 & -8 & 3 & -2 \\ 3 & -2 & -3 & 2 & 1 & 2 \\ 7 & -9 & 4 & 3 & 0 & 2 \\ -2 & 1 & 2 & -2 & -3 & 1 \\ 1 & 2 & 1 & 1 & -1 & 0 \\ -3 & 4 & 0 & 4 & 2 & 5 \end{vmatrix}$

40. $\begin{vmatrix} 7 & -3 & 4 & 2 & 0 & 9 \\ 3 & 8 & 1 & -4 & 1 & 3 \\ -2 & 4 & 5 & 2 & -3 & 2 \\ 1 & 0 & -4 & 3 & 2 & -7 \\ 4 & 6 & 3 & -1 & 6 & 8 \\ -5 & 3 & -2 & 5 & -4 & 2 \end{vmatrix}$

41. Let $A = \begin{bmatrix} 5 & -3 & 7 \\ 4 & 2 & -6 \\ 1 & 2 & -2 \end{bmatrix}$, $B = \begin{bmatrix} 5 & 9 & 7 \\ 4 & -9 & -6 \\ 1 & 3 & -2 \end{bmatrix}$, and $C = \begin{bmatrix} 5 & 6 & 7 \\ 4 & -7 & -6 \\ 1 & 5 & -2 \end{bmatrix}$.

a. Find $det(A)$, $det(B)$ and $det(C)$.

b. Look closely at the three matrices. How are A, B and C related to each other?

c. State the relationship among $det(A)$, $det(B)$ and $det(C)$.

d. Write a statement, analogous to the Lemma in the text, that is suggested by this Exercise. Prove this statement using the properties of the determinant.

42. Let A be an $n \times n$ matrix.

a. Suppose that B_2 is the matrix obtained from A by exchanging row i and row j of A. Prove that $det(B) = -det(A)$.

b. Suppose that B_3 is the matrix obtained from A by adding k times row i of A to row j of A. Prove that $det(B) = det(A)$.

Suggestion: Review the general proof in the text for Type 1 row operation. Mimic the proof and use the notation in that proof.

43. ***More Properties of The Cross Product:*** In the last Exercise of Section 5.1, we saw some properties of the cross product of two vectors in $\mathbb{R}^3$. Use the ***properties*** of the determinant function that we saw in this Section to prove the following statements (do not expand the determinants by brute force). Assume that $\vec{u}$, $\vec{v}$ and $\vec{w} \in \mathbb{R}^3$ and $k \in \mathbb{R}$.

a. $\vec{u} \times (k \cdot \vec{v}) = k \cdot (\vec{u} \times \vec{v}) = (k \cdot \vec{u}) \times \vec{v}$.

b. $\vec{u} \times (\vec{v} + \vec{w}) = \vec{u} \times \vec{v} + \vec{u} \times \vec{w}$.

c. $(\vec{u} + \vec{v}) \times \vec{w} = \vec{u} \times \vec{w} + \vec{v} \times \vec{w}$.

d. $\vec{u} \times \vec{v} = -(\vec{v} \times \vec{u})$.

e. $\vec{u} \circ (\vec{v} \times \vec{w}) = \begin{vmatrix} \vec{u} & \vec{v} & \vec{w} \end{vmatrix}$ (the matrix with $\vec{u}$, $\vec{v}$, and $\vec{w}$ in the columns).

f. Use (e) and the properties of the determinant to prove that:

$$\vec{u} \circ (\vec{v} \times \vec{w}) = \vec{v} \circ (\vec{w} \times \vec{u}) = \vec{w} \circ (\vec{u} \times \vec{v}).$$

(Recall that in the last Exercise of Section 5.1, we saw that the absolute value of all of these expressions is the volume of the parallelepiped determined by the three vectors.)

g. Suppose that $\vec{u}$ and $\vec{v}$ are ***non-parallel*** (therefore non-zero) vectors, and $\vec{w} = r\vec{u} + s\vec{v}$, where r and s are both ***non-zero*** scalars. Prove that $\vec{w}$ is also a non-zero vector which is not parallel to $\vec{u}$ or to $\vec{v}$, and furthermore: $\vec{u} \times \vec{v}$, $\vec{u} \times \vec{w}$, and $\vec{v} \times \vec{w}$ are all non-zero vectors which are ***parallel*** to each other.

h. Suppose that $A = (x_1, y_1, z_1)$, $B = (x_2, y_2, z_2)$, and $C = (x_3, y_3, z_3)$, are three ***non-collinear*** points in space. In other words, they form a ***triangle*** in space, and therefore, there is a unique plane Π containing these three points. Use the previous part to show that:

$$\overrightarrow{AB} \times \overrightarrow{AC}, \ \overrightarrow{BA} \times \overrightarrow{BC}, \ \text{and} \ \overrightarrow{CA} \times \overrightarrow{CB}$$

are all (non-zero) normal vectors to Π.

Thus, it does not matter which of the three points we choose as an anchor to find two vectors to cross in order to produce a normal vector to Π.

5.3 Properties of Determinants and Cofactor Expansion

We are now in a position to show that the determinant provides a litmus test for invertibility:

> ***Theorem:*** Let A be an $n \times n$ matrix. Then A is ***invertible*** ***if and only if*** $det(A)$ is ***non-zero***.

Proof: Suppose that A is an $n \times n$ matrix, and let R be the rref of A. Since A is square, we know from Chapter 1 that either R is $\boldsymbol{I}_n$ (which happens ***if and only if*** A is ***invertible***), or R contains a ***row of zeroes*** (which happens ***if and only if*** A is ***not invertible***). We also know from Chapter 2 that R can be obtained by a series of row operations, which we can simulate as a matrix product:

$$R = E_t \cdot \cdots \cdot E_2 \cdot E_1 \cdot A,$$

for some sequence of ***elementary matrices*** $E_1, E_2, \ldots, E_t$. In the previous Section, though, we proved that if E is an elementary $n \times n$ matrix, then $det(EA) = det(E) \cdot det(A)$. Thus, by repeated application of this principle, we get:

$$det(R) = det(E_t \cdot \cdots \cdot E_2 \cdot E_1 \cdot A) = det(E_t) \cdot \cdots \cdot det(E_2) \cdot det(E_1) \cdot det(A).$$

Notice that all of the elementary matrices in this sequence have a ***non-zero*** determinant. Thus $det(R)$ and $det(A)$ are either ***both zero*** or ***both non-zero***. Now, let us analyze the two cases for R:

Case 1. If $R = \boldsymbol{I}_n$, then $det(R) = det(\boldsymbol{I}_n) = 1$, and we get two conclusions: A is ***invertible***, ***and*** $det(A)$ is ***non-zero***.

Case 2. If R has a row of zeroes, then $det(R) = 0$, and our two conclusions are: A is ***not invertible***, ***and*** $det(A)$ is ***zero***.

Thus, by these two cases, A is ***invertible*** ***if and only if*** $det(A)$ is non-zero. ■

We said in the previous Section that the determinant function is not additive in general. However, the great miracle is that the determinant is in fact ***multiplicative:***

> ***Theorem:*** Let A and B be $n \times n$ matrices. Then: $det(A \cdot B) = det(A) \cdot det(B)$.

Proof: Let us divide our proof into two possibilities:

Case 1. A is invertible. Thus A is a product of elementary matrices, say $A = E_t \cdot \cdots \cdot E_2 \cdot E_1$. But by our Theorem on products with elementary matrices, we have:

$$det(A) = det(E_t \cdot \cdots \cdot E_2 \cdot E_1) = det(E_t) \cdot \cdots \cdot det(E_2) \cdot det(E_1).$$

But by the same token, we also have:

$$\begin{aligned} det(A \cdot B) &= det(E_t \cdot \cdots \cdot E_2 \cdot E_1 \cdot B) \\ &= det(E_t) \cdot \cdots \cdot det(E_2) \cdot det(E_1) \cdot det(B) = det(A) \cdot det(B). \end{aligned}$$

Case 2. A is not invertible. Then $det(A) = 0$, by our Theorem above. However, we saw in Chapter 2 that $A \cdot B$ is ***invertible*** ***if and only if*** both A ***and*** B are invertible. Thus, in this Case, $A \cdot B$ is also ***not*** invertible (otherwise both A and B would be invertible). Thus $det(A \cdot B) = 0$ also, and so we also get:

$$det(A \cdot B) = 0 = det(A) \cdot det(B). \blacksquare$$

As a bonus, we get the next Theorem, the proof of which is left as an Exercise:

Theorem: Let A be any $n \times n$ matrix. Then for any positive integer k:

$$det(A^k) = det(A)^k.$$

Furthermore, if A is ***invertible***, then:

$$det(A^{-1}) = \frac{1}{det(A)} = det(A)^{-1}.$$

Thus, if A is invertible, then for any ***integer*** power k :

$$det(A^k) = det(A)^k.$$

Minors and Cofactor Expansion

We saw that the determinant of the 3×3 matrix:

$$det(A) = \begin{vmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{vmatrix} = \begin{array}{c} a_{1,1}a_{2,2}a_{33} + a_{1,2}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}a_{32} \\ -a_{1,3}a_{2,2}a_{3,1} - a_{1,1}a_{2,3}a_{3,2} - a_{1,2}a_{2,1}a_{3,3}. \end{array}$$

If we collect the terms that contain $a_{1,1}$, $a_{1,2}$ and $a_{1,3}$, respectively, we get:

$$det(A) = a_{1,1}(a_{2,2}a_{33} - a_{2,3}a_{3,2}) + a_{1,2}(a_{2,3}a_{3,1} - a_{2,1}a_{3,3}) + a_{1,3}(a_{2,1}a_{32} - a_{2,2}a_{3,1}).$$

The terms in parentheses should look familiar, and that's because they represent determinants of 2×2 matrices. Notice that the ***first*** and ***third*** expressions in the parentheses can be written as:

$$a_{2,2}a_{33} - a_{2,3}a_{3,2} = \begin{vmatrix} a_{2,2} & a_{2,3} \\ a_{3,2} & a_{3,3} \end{vmatrix}, \text{ and}$$

$$a_{2,1}a_{32} - a_{2,2}a_{3,1} = \begin{vmatrix} a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \end{vmatrix}.$$

The first 2×2 determinant is associated with the factor $a_{1,1}$, and notice that we see this determinant if we ***erase row 1 and column 1*** of our original 3×3 matrix A. Similarly, the third 2×2 determinant is associated with the factor $a_{1,3}$, and we see this determinant if we ***erase row 1 and column 3*** of A. However, something is not quite right with the 2×2 determinant in the middle term. The two terms are backwards! We can correct this by factoring out -1:

$$a_{1,2}(a_{2,3}a_{3,1} - a_{2,1}a_{3,3}) = -a_{1,2}(a_{2,1}a_{3,3} - a_{2,3}a_{3,1}),$$

and we see the corresponding 2×2 determinant:

$$a_{2,1}a_{3,3} - a_{2,3}a_{3,1} = \begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix}$$

is obtained by ***erasing row 1 and column 2*** of A.

This observation is not exclusively associated to row 1. In fact, it is not even exclusive only to rows, but, as in any democratic society, applies also to the columns of a 3×3 matrix. For example, if we collect terms associated to the three entries in ***column 2***, namely, $a_{1,2}$, $a_{2,2}$ and $a_{3,2}$, in such a way that we get 2×2 determinants in the ***correct order***, we get:

$$\begin{aligned} det(A) &= a_{1,1}\boxed{a_{2,2}}a_{33} + \boxed{a_{1,2}}a_{2,3}a_{3,1} + a_{1,3}a_{2,1}\boxed{a_{3,2}} - a_{1,3}\boxed{a_{2,2}}a_{3,1} - a_{1,1}a_{2,3}\boxed{a_{3,2}} - \boxed{a_{1,2}}a_{2,1}a_{3,3} \\ &= \boxed{-a_{1,2}}(a_{2,1}a_{3,3} - a_{2,3}a_{3,1})\boxed{+a_{2,2}}(a_{1,1}a_{33} - a_{1,3}a_{3,1})\boxed{-a_{3,2}}(a_{1,1}a_{2,3} - a_{1,3}a_{2,1}). \end{aligned}$$

Notice that the three 2×2 determinants that we get are obtained from A by erasing the 2nd column entirely, and, one at a time, the 1st row, the 2nd row and the 3rd row respectively, getting us:

$$\begin{vmatrix} a_{2,1} & a_{2,3} \\ a_{3,1} & a_{3,3} \end{vmatrix}, \quad \begin{vmatrix} a_{1,1} & a_{1,3} \\ a_{3,1} & a_{3,3} \end{vmatrix} \text{ and } \begin{vmatrix} a_{1,1} & a_{1,3} \\ a_{2,1} & a_{2,3} \end{vmatrix}.$$

How do we know that we need to factor out -1? The answer lies in the row and column numbers of the ***coefficient*** beside the group. Notice that when they total an even number, such as for $a_{1,1}$ and $a_{1,3}$, we did not need to factor out -1. But when the row and column numbers total an odd number, such as for $a_{1,2}$ and $a_{3,2}$, we had to factor out -1. These observations lead us to the following:

Definition: Let A be an $n \times n$ matrix. The determinant of the submatrix obtained from A by erasing its *ith* row and *jth* column is called the ***i,j-minor*** of A, denoted:

$$M_{i,j}(A).$$

The ***i,j-cofactor*** of A is the number:

$$C_{i,j}(A) = (-1)^{i+j} \cdot M_{i,j}(A).$$

Note: One way to visually remember the factor $(-1)^{i+j}$ is to think of a chessboard:

+	−	+	−
−	+	−	+
+	−	+	−
−	+	−	+

The Signs of the Cofactors

The process that we saw above of rewriting the 3×3 determinant in terms of the entries of a chosen row or column can thus be written as:

$$\begin{aligned} det(A) &= a_{1,1}M_{1,1}(A) - a_{1,2}M_{1,2}(A) + a_{1,3}M_{1,3}(A) \\ &= a_{1,1}C_{1,1}(A) + a_{1,2}C_{1,2}(A) + a_{1,3}C_{1,3}(A), \end{aligned}$$

by using row 1. However, if we use column 2, we likewise have:

$$\begin{aligned} det(A) &= -a_{2,1}M_{2,1}(A) + a_{2,2}M_{2,2}(A) - a_{2,3}M_{2,3}(A) \\ &= a_{2,1}C_{2,1}(A) + a_{2,2}C_{2,2}(A) + a_{2,3}C_{2,3}(A). \end{aligned}$$

More generally, this "cofactor expansion" process works along any row or any column of a determinant of any size. However, the proof of this involves some heavy permutation theory, and so we omit it:

> ***Theorem:*** Let A be an $n \times n$ matrix. We can compute the determinant of A by a ***cofactor expansion along row i:***
>
> $$det(A) = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n},$$
>
> or a ***cofactor expansion along column j:***
>
> $$det(A) = a_{1,j}C_{1,j} + a_{2,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}.$$

The best choice for a row or column to expand along would obviously be one that contains the most zeroes, because in that case the corresponding cofactor is irrelevant.

Example: Consider the matrix:

$$A = \left[\begin{array}{cc|c|c} 7 & -4 & 0 & 2 \\ -2 & 5 & 3 & 8 \\ 0 & 9 & -5 & 7 \\ -3 & 4 & 0 & -6 \end{array}\right].$$

We have highlighted the 3rd column, because it contains two zeroes, and thus it is the best choice for a cofactor expansion. The formula tells us:

$$det(A) = 0 + (-1) \cdot 3 \cdot \begin{vmatrix} 7 & -4 & 2 \\ 0 & 9 & 7 \\ -3 & 4 & -6 \end{vmatrix} + (-5) \cdot \begin{vmatrix} 7 & -4 & 2 \\ -2 & 5 & 8 \\ -3 & 4 & -6 \end{vmatrix} + 0.$$

Notice that we need -1 for $a_{2,3}$ but not $a_{3,3}$. Now, we can compute the 3×3 determinants using the formula, but we can also expand the first 3×3 determinant along the first column or second row, thanks to the 0 entry. This first determinant becomes:

$$7 \cdot \begin{vmatrix} 9 & 7 \\ 4 & -6 \end{vmatrix} + (-3) \cdot \begin{vmatrix} -4 & 2 \\ 9 & 7 \end{vmatrix} = 7(-54 - 28) - 3(-28 - 18) = -436.$$

To compute the second determinant, notice that the third column has entries 2, 8 and -6, so let us use Type 3 row operations to produce zeroes in the 3rd column: add -4 times row 1 to row 2, and add 3 times row 1 to row 3. These do not change the determinant, and so:

$$\begin{vmatrix} 7 & -4 & 2 \\ -2 & 5 & 8 \\ -3 & 4 & -6 \end{vmatrix} = \begin{vmatrix} 7 & -4 & 2 \\ -30 & 21 & 0 \\ 18 & -8 & 0 \end{vmatrix} = 2\begin{vmatrix} -30 & 21 \\ 18 & -8 \end{vmatrix} = 2((-30) \cdot (-8) - 18 \cdot 21) = -276.$$

Plugging in these results into the first line, our final determinant is:

$$det(A) = -3(-436) - 5(-276) = 2688. \square$$

The Best of Both Worlds

The two techniques that we saw —— performing row or column reductions, and cofactor expansion along a row or column —— can be combined creatively to make short work of a determinant calculation.

Example: Consider the matrix:

$$A = \begin{bmatrix} -7 & -4 & 9 & 8 \\ -2 & 5 & 3 & 0 \\ 5 & 9 & -5 & -2 \\ -3 & 4 & 2 & 0 \end{bmatrix}.$$

A good beginning strategy would be to exploit the zeroes in column 4. However, we can first eliminate the "8" by adding 4 times row 3 to row 1. This is a Type 3 row operation, so it doesn't affect the determinant. We get:

$$det(A) = \begin{vmatrix} 13 & 32 & -11 & 0 \\ -2 & 5 & 3 & 0 \\ 5 & 9 & -5 & -2 \\ -3 & 4 & 2 & 0 \end{vmatrix}.$$

Now we can expand along column 4. That single row operation saved us from computing a 3×3 minor. We only need one cofactor:

$$det(A) = -(-2)\begin{vmatrix} 13 & 32 & -11 \\ -2 & 5 & 3 \\ -3 & 4 & 2 \end{vmatrix}.$$

We don't see any proportional rows or columns, but if we subtract row 3 from row 2 (again, a Type 3 row operation), we get:

$$det(A) = -(-2)\begin{vmatrix} 13 & 32 & -11 \\ 1 & 1 & 1 \\ -3 & 4 & 2 \end{vmatrix}.$$

To get a row with two zeroes, we can subtract column 1 from columns 2 and 3 (Type 3 column operations) and get:

$$det(A) = -(-2)\begin{vmatrix} 13 & 19 & -24 \\ 1 & 0 & 0 \\ -3 & 7 & 5 \end{vmatrix}.$$

Now we expand along row 2:

$$det(A) = -(-2)(-1)(1)\begin{vmatrix} 19 & -24 \\ 7 & 5 \end{vmatrix} = -2(19 \cdot 5 - (-24) \cdot 7) = -526. \square$$

In practice, of course, there can be more than one good way to compute the determinant, but since it is a ***function***, there will be only one correct final answer.

5.3 Section Summary

Let A and B be $n \times n$ matrices. Then:

- A is ***invertible*** ***if and only if*** $det(A)$ is ***non-zero***.
- $det(A \cdot B) = det(A) \cdot det(B)$.
- For any positive integer k: $det(A^k) = [det(A)]^k$.
- If A is ***invertible***, then: $det(A^{-1}) = 1/det(A) = [det(A)]^{-1}$.

  Thus, in this case, $det(A^k) = [det(A)]^k$ for all integers k.

The determinant of the submatrix obtained from A by erasing its *ith* row and *jth* column is called the ***i,j-minor*** of A, denoted $M_{i,j}(A)$. The ***i,j-cofactor*** of A is the number: $C_{i,j}(A) = (-1)^{i+j} \cdot M_{i,j}(A)$.

We can compute $det(A)$ by a ***cofactor expansion along row i:***

$$det(A) = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n},$$

or a ***cofactor expansion along column j:***

$$det(A) = a_{1,j}C_{1,j} + a_{2,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}.$$

Determinants can be computed by using a combination of two strategies — row and/or column operations and cofactor expansion along a row or column.

5.3 Exercises

1. Let $A = \begin{bmatrix} 5 & 8 \\ 3 & -2 \end{bmatrix}$ and $B = \begin{bmatrix} 6 & -4 \\ 1 & 7 \end{bmatrix}$.

   a. Compute $det(A)$ and $det(B)$.

   b. Compute AB and $det(A \cdot B)$.

   c. Check that $det(A \cdot B) = det(A) \cdot det(B)$

   d. Compute $A + B$ and $det(A + B)$.

   e. Show that $det(A + B) \neq det(A) + det(B)$.

   f. Compute $3B$ and $det(3B)$.

   g. What is the relationship between $det(3B)$ and $det(B)$?

2. The goal is to compute the determinant of the following matrix. Follow the instructions.

$$A = \begin{bmatrix} 2 & -3 & 7 & 2 \\ -1 & 3 & 0 & -4 \\ 2 & -8 & 0 & 3 \\ 6 & 5 & -2 & 7 \end{bmatrix}$$

a. Expand $det(A)$ using column 3. You should have two 3×3 determinants in your set-up.
b. Compute the two 3×3 determinants in your set-up. Check the answer key first.
c. Plug in your answers to (b) to complete the computation of $det(A)$.

3. The goal is to compute the determinant of the following matrix. Follow the instructions.

$$A = \begin{bmatrix} 5 & -3 & 7 & -2 \\ -1 & 3 & 6 & 4 \\ 2 & -8 & -2 & 3 \\ 6 & 0 & -4 & 0 \end{bmatrix}$$

a. Expand $det(A)$ using row 4. You should have two 3×3 determinants in your set-up.
b. Compute the two 3×3 determinants in your set-up. Check the answer key first.
c. Plug in your answers to (b) to complete the computation of $det(A)$.

4. The goal is to compute the determinant of the following matrix. Follow the instructions.

$$A = \begin{bmatrix} 4 & -2 & 3 & 8 \\ -7 & 8 & 5 & -4 \\ 2 & 3 & -2 & 3 \\ -3 & 0 & 9 & -5 \end{bmatrix}$$

a. Use row 1 and a Type 3 elementary row operation to turn the 8 in row 2 into a 0. Does this change the value of $det(A)$?
b. Use row 3 and a Type 3 elementary row operation to turn the –2 in row 1 into a 1. Does this change the value of $det(A)$?
c. Use (the new) row 1 and a Type 3 elementary row operation to turn the 3 in row 3, column 2, into a 0. Does this change the value of $det(A)$?
d. Expand $det(A)$ using column 2. You should have a single 3×3 determinant in your set-up.
e. Compute this 3×3 determinant and complete the computation of $det(A)$.

For Exercises (5) to (7): find the following determinants by cofactor expansion along a convenient row or column.

5. a. $\begin{vmatrix} 5 & 0 & -4 \\ 2 & -3 & 2 \\ 7 & 0 & 9 \end{vmatrix}$ b. $\begin{vmatrix} -7 & 4 & -2 \\ 2 & -3 & 9 \\ -6 & 0 & 0 \end{vmatrix}$

6. a. $\begin{vmatrix} 3 & 0 & 7 & 2 \\ -1 & 3 & 0 & -4 \\ 2 & -8 & 1 & 3 \\ 0 & 5 & -2 & 0 \end{vmatrix}$ b. $\begin{vmatrix} 5 & -2 & 0 & -3 \\ 3 & 0 & -8 & 5 \\ -7 & 9 & 1 & 0 \\ 6 & 0 & 3 & 2 \end{vmatrix}$

7. a. $\begin{vmatrix} -5 & -1 & 5 & 0 & -3 \\ 4 & 0 & 3 & -2 & 2 \\ 3 & 2 & 1 & 0 & 3 \\ 7 & 0 & -3 & 0 & 4 \\ 2 & -1 & 4 & 3 & -2 \end{vmatrix}$ b. $\begin{vmatrix} 7 & -8 & 3 & 2 & 4 \\ -5 & 1 & -1 & -4 & 0 \\ 3 & 0 & 7 & 0 & 5 \\ 0 & 6 & 0 & 0 & -7 \\ -2 & 4 & -1 & 3 & 0 \end{vmatrix}$

For Exercises (8) and (9): A matrix is called ***sparse*** when many of the entries are zeroes (you may be the judge as to what "many" is supposed to mean). Find the determinants of the following sparse matrices by cofactor expansions.

8. a. $\begin{vmatrix} 3 & -5 & 0 & 7 \\ 0 & 0 & 1 & 2 \\ -2 & 3 & 0 & 0 \\ 0 & 4 & -1 & -2 \end{vmatrix}$ b. $\begin{vmatrix} 0 & -5 & 0 & -3 \\ 7 & 0 & 1 & 0 \\ -2 & 3 & 0 & 0 \\ 0 & 0 & -1 & 2 \end{vmatrix}$

9. a. $\begin{vmatrix} 0 & 0 & -4 & 0 & 3 \\ 0 & -2 & 0 & 3 & 0 \\ 7 & 0 & 0 & 0 & -5 \\ 0 & 0 & 1 & 0 & 0 \\ -4 & 0 & 0 & -1 & 6 \end{vmatrix}$ b. $\begin{vmatrix} 4 & 9 & 0 & 0 & 5 \\ 0 & -2 & 0 & 3 & 0 \\ -3 & 0 & 3 & 0 & -5 \\ 0 & 0 & -1 & 0 & 2 \\ 6 & 0 & 0 & -1 & 6 \end{vmatrix}$

10. For any positive integer k, prove that $det(A^k) = det(A)^k$.

11. Prove that if A is invertible, then $det(A^{-1}) = 1/det(A) = det(A)^{-1}$.

12. We know in general that $A \bullet B \neq B \bullet A$ for two $n \times n$ matrices. However, prove that:

$$det(A \bullet B) = det(B \bullet A).$$

13. Suppose that X and Y are $n \times n$ matrices, where Y is ***invertible***, and we define Z by: $Z = Y \bullet X \bullet Y^{-1}$. Prove that $det(X) = det(Z)$.

14. Suppose that A is a 5×5 matrix with $det(A) = -60$, and E_1 through E_4 are elementary 5×5 matrices obtained from $\boldsymbol{I}_5$ in the following ways, respectively: E_1: exchange rows 2 and 4 of $\boldsymbol{I}_5$; E_2: multiply row 5 of $\boldsymbol{I}_5$ by 7; E_3: divide row 1 of $\boldsymbol{I}_5$ by 3; E_4: add 6 times row 3 of $\boldsymbol{I}_5$ to row 4. Find $det(B)$, where $B = E_1 \bullet E_2 \bullet E_3 \bullet E_4 \bullet A$.

15. ***Matrices in Block Diagonal Form:*** Suppose that $A_1, A_2, \ldots, A_k$ are all square matrices, not necessarily of the same size, with $k \geq 2$. We defined the direct sum of these matrices:

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_k$$

in the Exercises of Section 2.8. Prove that: $det(A) = det(A_1)det(A_2)\cdots det(A_k)$.

Hint: Use Induction. Start by showing that if $A = A_1 \oplus A_2$, then $det(A) = det(A_1)det(A_2)$.

16. ***The Vandermonde Determinant:*** We saw the matrix:

$$[E_{\vec{a}}]_{B,B'} = \begin{bmatrix} 1 & -4 & 16 & -64 \\ 1 & -1 & 1 & -1 \\ 1 & 2 & 4 & 8 \\ 1 & 3 & 9 & 27 \end{bmatrix}$$

in Section 3.8, which we used to find a polynomial $p(x)$ with specified y-coordinates at $x = -4, -1, 2$ and 3 (notice these are the entries in the ***2nd column***). We said that $E_{\vec{a}}$ is ***invertible*** because of the Fundamental Theorem of Algebra, so $det(E_{\vec{a}})$ cannot be zero. In this Exercise, we will prove an elegant formula for this determinant.

We define the ***Vandermonde Determinant***, denoted $V(a_1, a_2, \ldots, a_n)$, as:

$$V(a_1, a_2, \ldots, a_n) = \begin{vmatrix} 1 & a_1 & a_1^2 & \cdots & a_1^{n-1} \\ 1 & a_2 & a_2^2 & \cdots & a_2^{n-1} \\ 1 & a_3 & a_3^2 & \cdots & a_3^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_n & a_n^2 & \cdots & a_n^{n-1} \end{vmatrix}.$$

Notice that the 3rd and 4th columns of our first matrix $[E_{\vec{a}}]_{B,B'}$ indeed contain the squares and cubes of our four x coordinates, respectively, so the determinant of this matrix is an example of a Vandermonde Determinant.

We will guide you through a proof by ***Mathematical Induction*** to show that:

$$V(a_1, a_2, \ldots, a_n) = \prod_{1 \le i < j \le n} (a_j - a_i).$$

Notice this formula is a ***product***, not a sum. For example:

$$V(a_1, a_2, a_3) = (a_3 - a_1)(a_3 - a_2)(a_2 - a_1).$$

a. Warm-up: Use row reduction to compute $V(-4, -1, 2, 3)$, i.e., the determinant above.

b. Use the formula to compute $V(-4, -1, 2, 3)$, and check it with your answer to part (a).

c. State and prove the formula for $n = 2$.

d. Write down what the formula says for $n = k$, and assume that this formula is true.

e. Write down what the formula says for $n = k + 1$. This is the formula you need to prove to complete the induction. Go to the next step.

f. Show that if we perform k Type 3 column operations by adding a multiple α_i of column i, where $i = 1..k$, to the last column, then the Vandermonde determinant of size $(k+1) \times (k+1)$ can be written as:

$$\begin{vmatrix} 1 & a_1 & a_1^2 & \cdots & r(a_1) \\ 1 & a_2 & a_2^2 & \cdots & r(a_2) \\ 1 & a_3 & a_3^2 & \cdots & r(a_3) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & a_{k+1} & a_{k+1}^2 & \cdots & r(a_{k+1}) \end{vmatrix},$$

where $r(x)$ is a polynomial with largest term x^k. Hint: write the determinant for $V(a_1, a_2, a_3)$ and see what column 3 would look like if you add a multiple of column 1 and a multiple of column 2.

g. Now, the big mental leap: what ***polynomial*** $r(x)$, of degree k, can be used to make ***all*** the entries in the last column ***zero***, except for the bottom entry? What would this bottom entry be? Hint: what polynomial has ***roots*** $a_1,\ a_2,\ \ldots,\ a_k$?

h. Complete the induction process by performing a cofactor expansion along the new last column.

i. Explain why the formula proves that if $a_1,\ a_2,\ \ldots,\ a_n$ are ***distinct*** real numbers, then the matrix $[E_{\vec{a}}]_{B,B'}$ that we saw above, in general, is ***invertible*** as we claimed. This gives us another proof that we can always find a unique polynomial of degree at most n if we specify its y-coordinates at $n+1$ distinct x-coordinates.

17. ***The Group $SL_n(\mathbb{Z})$***: The set of all $n \times n$ matrices with determinant 1 whose entries consist of integers and whose ***inverses*** also consist of integer entries is called $\boldsymbol{SL_n}(\mathbb{Z})$, which stands for the ***Special Linear Group*** of $n \times n$ Matrices over $\mathbb{Z}$ (Reminder: the letter $\mathbb{Z}$ stands for "Zahlen," the German word for number, as mentioned in Chapter Zero).

A ***group*** is a non-empty set G, together with a binary operation $*$ on G, such that:

1. $*$ is ***associative:*** for all $a, b, c \in G$: $a * (b * c) = (a * b) * c$;
2. there exists an ***identity*** element e, such that $e * a = a = a * e$ for all $a \in G$; and
3. for every $a \in G$, there exists an ***inverse*** $a^{-1} \in G$, such that $a * a^{-1} = e = a^{-1} * a$.

For $\boldsymbol{SL_n}(\mathbb{Z})$, $*$ is matrix multiplication, which we know is associative, $e = \boldsymbol{I_n}$, and the inverse of A is obviously the matrix inverse A^{-1}.

a. Consider the two matrices: $S = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ and $T = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

Show that S and T are both members of $\boldsymbol{SL_2}(\mathbb{Z})$, that is, both have determinant 1, both are invertible and their inverses are also integer matrices. What are their inverses?

b. Compute S^2, S^3 and S^4. What do you notice?

c. Compute T^2, T^3, T^4 and T^5. What do you notice?

d. Guess a general formula for T^n and prove it by Induction.

e. Find the matrix product $U = ST$.

f. Compute U^2, U^3, U^4, U^5 and U^6. What do you notice?

g. Prove that $\boldsymbol{SL_n}(\mathbb{Z})$ is ***closed*** under matrix multiplication (for ***any*** positive integer n). In other words, prove that the product of two matrices from $\boldsymbol{SL_n}(\mathbb{Z})$ is also a member of $\boldsymbol{SL_n}(\mathbb{Z})$. Hint: You must check ***three*** properties.

h. Is it also true that $\boldsymbol{SL_n}(\mathbb{Z})$ is closed under matrix ***addition***? Either prove that it is true or provide a counterexample to show that it is false.

The purpose of the rest of this Exercise is to prove the following elegant statement:

Theorem: Let $A \in \boldsymbol{SL_2}(\mathbb{Z})$. Then A can be expressed as a ***finite product*** of $S,\ T,\ S^{-1}$ and T^{-1}, in some order.

This Theorem says that the two matrices S and T (and their inverses) ***generate*** the group $\boldsymbol{SL_2}(\mathbb{Z})$. For example, A might be written as:

$$A = STTS^{-1}TTTST^{-1}T^{-1}S^{-1}TST^{-1}.$$

To begin the proof of this Theorem, let us suppose that: $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \boldsymbol{SL}_2(\mathbb{Z})$.

i. Find a formula for SA.

j. Find a formula for TA and for $T^{-1}A$.

k. Let us go back to the matrix A. Suppose that $|a| \geq |c|$. Show that by repeatedly multiplying A by T or T^{-1}, we will eventually obtain a matrix:

$$A' = \begin{bmatrix} a' & b' \\ c' & d' \end{bmatrix},$$

where $|a'| < |c'|$. As a warm-up, you might want to apply this idea to the matrix:

$$A = \begin{bmatrix} -55 & -23 \\ 12 & 5 \end{bmatrix},$$

How many times will you have to multiply A by T or T^{-1}, and which should you use? What is your final matrix A'?

l. Now, use your formula for SA to show that $SA' = A''$, where:

$$A'' = \begin{bmatrix} a'' & b'' \\ c'' & d'' \end{bmatrix},$$

and $|a''| \geq |c''|$ and $|c''| < |c|$ (recall that c is the lower left entry of the original matrix A). In other words, A'' has exactly the same property as A, but with a smaller number on the lower left entry. What is A'' in your numerical example?

m. Recall that all the entries in these matrices, at every step, are ***integers***. Show that by repeatedly performing the last two steps, we finally obtain a matrix where the lower left entry is 0. In other words, we get an ***upper triangular*** matrix.

n. Show that an upper triangular matrix in $\boldsymbol{SL}_2(\mathbb{Z})$ has the form:

$$\text{either} \begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \text{or} \begin{bmatrix} -1 & k \\ 0 & -1 \end{bmatrix},$$

for some integer k. In other words, the matrix that we obtain at the end of Step (m) has one of these forms.

o. Show that either matrix above can be expressed as a product of one or more copies of S, T, S^{-1} and T^{-1}. As a warm-up, you might want to show how to express the matrices:

$$\begin{bmatrix} 1 & 5 \\ 0 & 1 \end{bmatrix} \text{and} \begin{bmatrix} -1 & 7 \\ 0 & -1 \end{bmatrix}$$

in this form. This completes the proof, since each step above is performed by multiplying by S, T, S^{-1} or T^{-1}.

p. Show how to express the numerical matrix in part (k) as a product of copies of S, T, S^{-1} and T^{-1}.

5.4 The Adjugate Matrix and Cramer's Rule

In this section, we will see how to assemble the cofactors of a matrix A into a new matrix called the adjugate matrix of A. We will see that this new matrix can be used to find the inverse of A, when A is in fact invertible. We will also see an alternative although largely impractical way to solve an invertible system.

The Cofactor and Adjugate Matrices

Definition: Let A be an $n \times n$ matrix. The ***cofactor matrix*** of A, denoted $cof(A)$, is the matrix whose entries are the corresponding cofactors of each entry of A:

$$cof(A) = [C_{i,j}(A)].$$

We recall that:

$$C_{i,j}(A) = (-1)^{i+j} \cdot M_{i,j}(A),$$

where $M_{i,j}(A)$, is the determinant of the submatrix obtained from A by erasing its *ith* row and *jth* column of A. The ***adjugate matrix*** of A is the ***transpose*** of the cofactor matrix, and is written as:

$$adj(A) = cof(A)^{\mathsf{T}} = [C_{j,i}(A)].$$

Note: Some older books refer to the adjugate matrix as the ***adjoint*** or ***classical adjoint*** of A. However, the word "adjoint" now usually refers to the ***Hermitian adjoint*** of A, which we will be defining in Chapter 8. In keeping with modern terminology, we will use the word "adjugate" in this book.

Example: Consider an arbitrary 2×2 matrix:

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

The cofactor matrix of A is:

$$cof(A) = \begin{bmatrix} +d & -c \\ -b & +a \end{bmatrix},$$

where we included the positive signs for emphasis. The adjugate matrix of A is:

$$adj(A) = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

Notice that this matrix suspiciously reminds us of the inverse of A :

$$A^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix},$$

assuming of course that $ad - bc \neq 0$. This is not a coincidence, as we shall soon see.$\square$

Example: Let us look at a 3×3 matrix, say:

$$A = \begin{bmatrix} 5 & -2 & 3 \\ -4 & 1 & 7 \\ -6 & 4 & -9 \end{bmatrix}.$$

Let us systematically construct $cof(A)$, again emphasizing the signs of the corresponding cofactors:

$$cof(A) = \begin{bmatrix} +\begin{vmatrix} 1 & 7 \\ 4 & -9 \end{vmatrix} & -\begin{vmatrix} -4 & 7 \\ -6 & -9 \end{vmatrix} & +\begin{vmatrix} -4 & 1 \\ -6 & 4 \end{vmatrix} \\ -\begin{vmatrix} -2 & 3 \\ 4 & -9 \end{vmatrix} & +\begin{vmatrix} 5 & 3 \\ -6 & -9 \end{vmatrix} & -\begin{vmatrix} 5 & -2 \\ -6 & 4 \end{vmatrix} \\ +\begin{vmatrix} -2 & 3 \\ 1 & 7 \end{vmatrix} & -\begin{vmatrix} 5 & 3 \\ -4 & 7 \end{vmatrix} & +\begin{vmatrix} 5 & -2 \\ -4 & 1 \end{vmatrix} \end{bmatrix} = \begin{bmatrix} -37 & -78 & -10 \\ -6 & -27 & -8 \\ -17 & -47 & -3 \end{bmatrix}.$$

Taking the transpose, we obtain the adjugate matrix:

$$adj(A) = \begin{bmatrix} -37 & -6 & -17 \\ -78 & -27 & -47 \\ -10 & -8 & -3 \end{bmatrix}.$$

This time, let us see what happens when we multiply our original matrix by its adjugate:

$$\begin{aligned} &A \cdot adj(A) \\ &= \begin{bmatrix} 5 & -2 & 3 \\ -4 & 1 & 7 \\ -6 & 4 & -9 \end{bmatrix}\begin{bmatrix} -37 & -6 & -17 \\ -78 & -27 & -47 \\ -10 & -8 & -3 \end{bmatrix} = \begin{bmatrix} -59 & 0 & 0 \\ 0 & -59 & 0 \\ 0 & 0 & -59 \end{bmatrix} = -59\boldsymbol{I}_3. \end{aligned}$$

A quick computation will tell us that $det(A) = -59$, so this equation again tells us that:

$$A \cdot adj(A) = det(A) \cdot \boldsymbol{I}_3.$$

In other words:

$$A \cdot \frac{1}{det(A)} adj(A) = \boldsymbol{I}_3,$$

and thus A is invertible, with:

$$A^{-1} = \frac{1}{det(A)} adj(A) = -\frac{1}{59}\begin{bmatrix} -37 & -6 & -17 \\ -78 & -27 & -47 \\ -10 & -8 & -3 \end{bmatrix}. \square$$

Our next goal, of course, is to prove this fact in general.

A New Formula for the Inverse of a Matrix

> ***Theorem:*** Let A be ***any*** $n \times n$ matrix. Then:
>
> $$A \cdot adj(A) = det(A) \cdot \boldsymbol{I_n}.$$
>
> Consequently, if A is ***invertible***, then: $A^{-1} = \dfrac{1}{det(A)} adj(A).$

Proof: We have to investigate every entry in the matrix product $A \cdot adj(A)$ and show that we obtain $det(A)$ along the main diagonal and 0 for the other entries.

Using the dot product formulation of the matrix product, the entry in row i, column j of the product is the dot product of row i of A with column j of $adj(A)$. However, since $adj(A)$ is the ***transpose*** of $cof(A)$, the entries in column j of $adj(A)$ are:

$$C_{j,1},\ C_{j,2},\ \ldots,\ C_{j,n},$$

that is, the cofactors of the entries from ***row*** j of A. Thus, the final entry is:

$$a_{i,1}C_{j,1} + a_{i,2}C_{j,2} + \cdots + C_{j,n}.$$

Case 1. If $i = j$, that is, we have a main diagonal entry, then we get:

$$a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + C_{i,n}.$$

However, this is exactly the formula for $det(A)$ by applying a ***cofactor expansion*** along row i of A. Thus the diagonal entries are indeed $det(A)$.

Case 2. If $i \neq j$, we must show that this entry is 0. We will do this by cleverly replacing row j of A with row i of A. Let A' be the resulting matrix:

$$A' = \begin{array}{cl} \begin{bmatrix} a_{1,1} & a_{1,2} & \ldots & a_{1,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i,1} & a_{i,2} & \ldots & a_{i,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i,1} & a_{i,2} & \ldots & a_{i,n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & a_{n,2} & \ldots & a_{n,n} \end{bmatrix} & \begin{array}{l} \\ \\ \leftarrow \textbf{row } \boldsymbol{i} \\ \\ \leftarrow \textbf{row } \boldsymbol{j} \\ \\ \\ \end{array} \end{array}$$

We boxed the entries of row j in order to distinguish row j from row i. Since A' has two rows with identical entries, we must have $det(A') = 0$.

Let us denote by $C'_{j,1}$ the $j,1$-cofactor of A'. Since we must delete row j and column 1 from A', we will get exactly the ***same minor*** as the original matrix A, and hence $C'_{j,1} = C_{j,1}$.

Similarly, $C'_{j,2}$ through $C'_{j,n}$ are correspondingly equal to $C_{j,2}$ through $C_{j,n}$. Now, if we perform a cofactor expansion for A' along row j, we obtain:

$$0 = det(A') = a_{i,1}C'_{j,1} + a_{i,2}C'_{j,2} + \cdots + a_{i,n}C'_{j,n} = a_{i,1}C_{j,1} + a_{i,2}C_{j,2} + \cdots + a_{i,n}C_{j,n},$$

completing our proof.■

A by-product of this formula, and the ideas behind its proof, is an alternative, albeit mostly impractical method to solve an invertible matrix equation.

Cramer's Rule

This method is usually seen in an Algebra or Precalculus class. It is named after the Swiss mathematician ***Gabriel Cramer*** (1704-1752). We will now be in a position not just to use it, but also to prove it:

Theorem — Cramer's Rule: Let A be an ***invertible*** matrix. Then: the ***unique solution*** to the matrix equation:

$$A\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

has entries:

$$x_1 = \frac{det(A^{(1)})}{det(A)},\ x_2 = \frac{det(A^{(2)})}{det(A)},\ \ldots,\ x_n = \frac{det(A^{(n)})}{det(A)},$$

where $A^{(i)}$ is the matrix obtained from A by ***replacing column i*** of A with $\vec{b}$.

Proof: The unique solution $\vec{x}$ is of course given by $\vec{x} = A^{-1} \cdot \vec{b}$, since A is invertible. However, if we use our new formula for A^{-1}, we get:

$$\vec{x} = \frac{1}{det(A)} adj(A) \cdot \vec{b}.$$

Since $\vec{b}$ is a column vector, we can again use the dot product interpretation to get:

$$x_i = \frac{1}{det(A)} \cdot row\ i \text{ of } adj(A) \circ \vec{b} = \frac{1}{det(A)} \cdot (b_1 C_{1,i} + b_2 C_{2,i} + \cdots + b_n C_{n,i}),$$

where again, we note that $adj(A)$ is the transpose of $cof(A)$, and thus the entries of row i of $adj(A)$ are the corresponding cofactors for ***column*** i of A. However, in the same way that we proved the previous theorem by cleverly replacing row j of A, we will finish the proof of Cramer's rule by replacing column i of A with the column vector $\vec{b}$, and call the resulting matrix $A^{(i)}$:

$$A^{(i)} = \begin{bmatrix} a_{1,1} & \ldots & a_{1,i-1} & \boxed{b_1} & a_{1,i+1} & \ldots & a_{1,n} \\ a_{2,1} & \ldots & a_{2,i-1} & b_2 & a_{2,i+1} & \ldots & a_{2,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{n,1} & \ldots & a_{n,i-1} & b_n & a_{n,i+1} & \ldots & a_{n,n} \end{bmatrix}.$$

This time, naturally, we will compute $det(A^{(i)})$ using a cofactor expansion along column i. We will denote the j,i–cofactor of $A^{(i)}$ by $C^{(i)}_{j,i}$, where $j = 1..n$. Since we will delete row j and column i of $A^{(i)}$, we will get exactly the ***same minor*** that we would obtain by deleting row j and column i of A

itself. Thus:

$$C_{j,i}^{(i)} = C_{j,i}.$$

From this, we get the determinant by expanding along column i:

$$det(A^{(i)}) = b_1 C_{1,i}^{(i)} + b_2 C_{2,i}^{(i)} + \cdots + b_n C_{n,i}^{(i)} = b_1 C_{1,i} + b_2 C_{2,i} + \cdots + b_n C_{n,i},$$

and thus we get:

$$x_i = \frac{1}{det(A)} \cdot (b_1 C_{1,i} + b_2 C_{2,i} + \cdots + b_n C_{n,i}) = \frac{1}{det(A)} \cdot det(A^{(i)}),$$

completing our Proof.■

Example: Let us bring back our earlier 3×3 matrix:

$$A = \begin{bmatrix} 5 & -2 & 3 \\ -4 & 1 & 7 \\ -6 & 4 & -9 \end{bmatrix}.$$

We saw that $det(A) = -59,$ so we know that A is invertible. Let us find the unique solution to:

$$A \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 7 \\ -4 \\ 2 \end{bmatrix}.$$

We will need the determinants:

$$det(A^{(1)}) = \begin{vmatrix} 7 & -2 & 3 \\ -4 & 1 & 7 \\ 2 & 4 & -9 \end{vmatrix} = -269,$$

$$det(A^{(2)}) = \begin{vmatrix} 5 & 7 & 3 \\ -4 & -4 & 7 \\ -6 & 2 & -9 \end{vmatrix} = -532, \text{ and}$$

$$det(A^{(3)}) = \begin{vmatrix} 5 & -2 & 7 \\ -4 & 1 & -4 \\ -6 & 4 & 2 \end{vmatrix} = -44.$$

Thus:

$$x_1 = \frac{-269}{-59}, \quad x_2 = \frac{-532}{-59}, \quad \text{and} \quad x_3 = \frac{-44}{-59},$$

so the unique solution is:

$$\vec{x} = \left\langle \frac{269}{59}, \frac{532}{59}, \frac{44}{59} \right\rangle. \square$$

5.4 Section Summary

Let A be an $n \times n$ matrix. The ***cofactor matrix*** of A, denoted $cof(A)$, is the matrix whose entries are the corresponding cofactors of each entry of A:

$$cof(A) = [C_{i,j}(A)],$$

where $C_{i,j}(A) = (-1)^{i+j} \cdot M_{i,j}(A)$ and the minor $M_{i,j}(A)$, is the determinant of the submatrix obtained from A by erasing its *ith* row and *jth* column of A.

The ***adjugate matrix*** of A is the ***transpose*** of the cofactor matrix, and is written as:

$$adj(A) = cof(A)^{\intercal} = [C_{j,i}(A)].$$

Let A be any $n \times n$ matrix. Then:

$$A \cdot adj(A) = det(A) \cdot \boldsymbol{I_n}.$$

Consequently, if A is ***invertible***, then:

$$A^{-1} = \frac{1}{det(A)} adj(A).$$

Cramer's Rule: Let A be an ***invertible*** matrix. Then: the ***unique solution*** to the matrix equation:

$$A \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

has entries:

$$x_1 = \frac{det(A^{(1)})}{det(A)}, \quad x_2 = \frac{det(A^{(2)})}{det(A)}, \quad \ldots, \quad x_n = \frac{det(A^{(n)})}{det(A)},$$

where $A^{(i)}$ is the matrix obtained from A by ***replacing column i*** with $\vec{b}$.

5.4 Exercises

For Exercises (1) to (3): Find the adjugate matrices of the following matrices, and use them to find the inverse of each matrix, when possible:

1. a. $A = \begin{bmatrix} 3 & 5 \\ 1 & 4 \end{bmatrix}$ b. $B = \begin{bmatrix} -3 & -5 \\ 12 & 20 \end{bmatrix}$

2. a. $A = \begin{bmatrix} 4 & 3 & -1 \\ 5 & -2 & 0 \\ -3 & 1 & 2 \end{bmatrix}$ b. $B = \begin{bmatrix} -2 & 3 & 4 \\ 1 & 7 & -1 \\ 5 & 0 & 2 \end{bmatrix}$

3. a. $A = \begin{bmatrix} -5 & 2 & 1 & -3 \\ 0 & -4 & 7 & 5 \\ 6 & 3 & -8 & 0 \\ 9 & 4 & 3 & -2 \end{bmatrix}$ b. $B = \begin{bmatrix} 3 & 7 & 7 & 1 \\ -2 & 3 & 6 & 2 \\ 5 & -2 & -7 & -1 \\ 4 & 6 & 5 & 2 \end{bmatrix}$

For Exercises (4) to (9): Use Cramer's Rule to solve the following systems of linear equations, if applicable:

4. a. $\begin{aligned} 3x - 5y &= 4 \\ 2x + y &= -7 \end{aligned}$ b. $\begin{aligned} 7x + 4y &= -6 \\ 6x - 5y &= 8 \end{aligned}$

5. a. $\begin{aligned} 15x - 5y &= 20 \\ -6x + 2y &= -14 \end{aligned}$ b. $\begin{aligned} \frac{3}{4}x + \frac{5}{2}y &= -\frac{7}{4} \\ \frac{2}{3}x - \frac{11}{6}y &= \frac{4}{3} \end{aligned}$

6. a. $\begin{aligned} 2x + 4y - 5z &= 6 \\ x - 3y + 7z &= -1 \\ 5x + y - 3z &= 4 \end{aligned}$ b. $\begin{aligned} 2x + 3y - 2z &= 9 \\ x - 2y - z &= 5 \\ 4x - y - 4z &= -3 \end{aligned}$

7. a. $\begin{aligned} 5x + 2y + z &= 8 \\ 7x + 3y - 4z &= 1 \\ -3x + 6y - 2z &= -5 \end{aligned}$ b. $\begin{aligned} -x + 2y - 4z &= 3 \\ x - 4y - 5z &= -2 \\ -3x + y + 7z &= 4 \end{aligned}$

8. a. $\begin{aligned} -5x + 2y + z + 2w &= -3 \\ -4y + 7z - 6w &= 5 \\ 6x + 3y - 8z + 5w &= 0 \\ 9x + 4y + 3z + w &= -2 \end{aligned}$ b. $\begin{aligned} 3x + 7y + 7z - 4w &= 1 \\ -2x + 3y + 6z + 9w &= 2 \\ 5x - 2y - 7z + 3w &= -1 \\ 4x + 6y + 5z - 5w &= 2 \end{aligned}$

9. a. $\begin{aligned} 7x - 5y + 2z - w &= 3 \\ -2x - 4y - 4z - 3w &= -7 \\ 4x + y - 6z + 5w &= 4 \\ 3x + 2y + 9z + 2w &= 0 \end{aligned}$ b. $\begin{aligned} x + 2y - 2z + 2w &= 3 \\ -4x - 8y + 6z + w &= -2 \\ 3x - 6y - 4z - 3w &= 5 \\ x + y + z - 4w &= -4 \end{aligned}$

10. Use Cramer's Rule to solve for the variables b and c ***only*** in the linear system:

$$\begin{aligned} 3a - 2b - 5c + d - 4e &= 6 \\ a + b + 3c - 2d + 7e &= -3 \\ -2a + 3b - c + 4d - 2e &= 2 \\ 5a - 4b - 2c + 3d + 3e &= 1 \\ 4a + 5b + 7c - d + e &= -4 \end{aligned}$$

11. Suppose that $\left\{\vec{a}, \vec{b}, \vec{c}, \vec{d}\right\}$ is a linearly independent set of vectors from $\mathbb{R}^4$. Use Cramer's Rule to show that the system represented by the augmented matrix:

$$\left[\begin{array}{cccc|c} a_1 & b_1 & c_1 & d_1 & 5a_1 - 2c_1 + 7d_1 \\ a_2 & b_2 & c_2 & d_2 & 5a_2 - 2c_2 + 7d_2 \\ a_3 & b_3 & c_3 & d_3 & 5a_3 - 2c_3 + 7d_3 \\ a_4 & b_4 & c_4 & d_4 & 5a_4 - 2c_4 + 7d_4 \end{array}\right]$$

has exactly one solution, and find that solution. As usual, $\vec{a} = \langle a_1, a_2, a_3, a_4 \rangle$, and similarly for the other vectors. Hint: think of the properties of determinants.

12. Let E be an $n \times n$ ***elementary*** matrix, where $n \geq 2$.

    a. Warm-up: compute the adjoint matrices of the following elementary matrices. Which of the adjoints is also an elementary matrix?

    $$\begin{bmatrix} 1 & 0 \\ 0 & -5 \end{bmatrix}; \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix}; \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}; \begin{bmatrix} 1 & 0 & -3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

    b. If $n = 2$ and E is of Type 1 (multiply row i of $\boldsymbol{I_n}$ by a non-zero number c to obtain E), then $adj(E)$ is also an elementary matrix of Type 1.

    c. Show that if $n > 2$ and E is of Type 1, then $adj(E)$ is ***not*** an elementary matrix (except for the trivial case when $E = \boldsymbol{I}_n$). What goes wrong?

    d. Show that if E is of Type 2 (exchange two rows of $\boldsymbol{I_n}$ to obtain E), then $adj(E)$ is ***never*** an elementary matrix.

    e. Show that if E is of Type 3 (add a multiple of row i of $\boldsymbol{I_n}$ to row j to obtain E), then $adj(E)$ is also an elementary matrix of Type 3.

13. Prove that if A and B are ***invertible*** $n \times n$ matrices, then:

    $$adj(A \cdot B) = adj(B) \cdot adj(A).$$

    Note: this formula is true even if A or B is not invertible, but the proof is much more difficult. Also notice the similarity between this formula and those for the ***inverse*** and ***transpose*** of the matrix product $A \cdot B$.

14. The objective of this exercise is to prove that an $n \times n$ matrix A is ***invertible*** ***if and only if*** $adj(A)$ is also ***invertible***. Note that the formula:

    $$A \cdot adj(A) = det(A) \cdot \boldsymbol{I_n}$$

    is ***always true***, whether or not A is invertible.

    a. The easy part: use this formula to prove that if A is invertible, then $adj(A)$ is also invertible.

    b. Now for the converse: Suppose instead that $adj(A)$ is invertible. Use Proof by Contradiction to prove that A is also invertible. Hint: Suppose A is not invertible. Use the formula above to solve for A. What happens? Be sure to actually explain ***what*** the contradiction is. This is not as obvious as it looks.

15. Use the previous Exercise to show that if A is an $n \times n$ matrix, then:

$$det(adj(A)) = det(A)^{n-1}.$$

16. The objective of this Exercise is to give an alternative proof that the inverse of an ***invertible upper-triangular*** matrix A is again upper triangular.

    a. Write down the ***rigorous*** definition of what an upper triangular matrix is, from Section 2.9. The definition should mention a certain ***inequality***.

    b. Consider the upper triangular matrix:

    $$A = \begin{bmatrix} 2 & 5 & -1 & 4 \\ 0 & -3 & 6 & 2 \\ 0 & 0 & 7 & -8 \\ 0 & 0 & 0 & -1 \end{bmatrix}.$$

    Find $adj(A)$. What kind of a matrix is it?

    c. Now, let A be ***any*** $n \times n$ upper-triangular matrix. Show that if none of the entries on the main diagonal of A is zero, then all of the cofactors $C_{i,i}$ for the main diagonal are also non-zero.

    d. Next, show that if $i < j$, then the matrix obtained by the deleting row i and column j from A is still upper triangular.

    e. Continuing part (d), show that additionally, a ***zero*** now appears on the main diagonal. Conclude that $C_{i,j} = 0$.

    f. Explain why the last two parts shows that $adj(A)$ is an upper triangular matrix also.

    g. Finally, explain why A^{-1} is also upper triangular.

17. Mimic the outline of the previous Exercise, parts (c) to (g), to prove that if A is an ***invertible lower triangular*** matrix, then A^{-1} is also lower triangular.

5.5 The Wronskian

In Section 3.2, we defined a finite set of functions $S = \{f_1(x), f_2(x), \ldots, f_n(x)\}$ from some function space $F(I)$ to be linearly ***independent*** if the only solution to the dependence test equation:

$$c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) = z(x)$$

is the trivial solution: $c_1 = c_2 = \cdots = c_n = 0$. In other words, the only way for the linear combination on the left side of this equation to be zero at ***all*** points $x \in I$ is to have all coefficients zero. We saw a variety of ideas to determine whether or not S were linearly independent or dependent, such as plugging in several values of x and attempting to solve the resulting systems of equations, taking a limit as x approaches a or some infinite limit, applying the Fundamental Theorem of Algebra (if the functions happened to be polynomials, or can be converted to polynomials), and using known identities from trigonometry, again if applicable. In other words, we had no clear or obvious strategy on how to attack this question. In this Section, we present a method which uses the derivatives and the determinant concept in order to decide if S is independent or dependent.

To see how this method works, let us first assume that S is linearly ***dependent***. This means that we can find a non-trivial solution to the dependence test equation, that is, where at least one c_i is ***non-zero***. However, if $S \subset C^1(I)$, and we apply the derivative operator to both sides of the equation, we get:

$$c_1 f_1'(x) + c_2 f_2'(x) + \cdots + c_n f_n'(x) = z'(x) = z(x).$$

Let us keep applying this idea by taking a 2nd derivative, 3rd derivative, all the way to the $(n-1)$-st derivative, thus assuming that $S \subset C^{n-1}(I)$. We end up with the system:

$$\begin{aligned} c_1 f_1(x) + c_2 f_2(x) + \cdots + c_n f_n(x) &= z(x) \\ c_1 f_1'(x) + c_2 f_2'(x) + \cdots + c_n f_n'(x) &= z(x) \\ c_1 f_1''(x) + c_2 f_2''(x) + \cdots + c_n f_n''(x) &= z(x) \\ &\vdots \\ c_1 f_1^{(n-1)}(x) + c_2 f_2^{(n-1)}(x) + \cdots + c_n f_n^{(n-1)}(x) &= z(x). \end{aligned}$$

Since we only took $n-1$ derivatives, this system is ***square***, and we can write it in the form of a matrix product:

$$\begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & & f_n^{(n-1)}(x) \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} z(x) \\ z(x) \\ \vdots \\ \vdots \\ z(x) \end{bmatrix}.$$

Now, recall that the non-trivial solution $c_1, c_2, \ldots, c_n$ must be valid for ***all*** points $x \in I$. But this means that the square matrix ***cannot be invertible*** for any value of $x \in I$. Thus, its ***determinant*** must be zero for ***all*** $x \in I$. This determinant is called the ***Wronskian*** of S, denoted $W(S)$, and is named after ***Józef Maria Hoene-Wroński*** (Poland, 1776-1853). The ***contrapositive*** of this statement tells us that if $W(S)$ is non-zero for at least ***one*** $x \in I$, then S is an ***independent*** set. We summarize our conclusions in the following:

Definition/Theorem: Let $S = \{f_1(x), f_2(x), \ldots, f_n(x)\} \subset C^{n-1}(I)$ for some interval I. We define the ***Wronskian*** of S, $W_S(x)$ as the ***function:***

$$W_S(x) = W(\{f_1(x), f_2(x), \ldots, f_n(x)\}) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}.$$

If S is a ***dependent*** set, then $W_S(x) = z(x)$ for ***all*** values $x \in I$.
Thus, if $W_S(x)$ is ***non-zero*** for at least ***one*** $x \in I$, then S is an ***independent*** set.

Example: Let us consider the set $S = \{\sin(x), \sin(2x), \sin(3x)\}$.
Using the Chain Rule, the Wronskian of S is:

$$W_S(x) = \begin{vmatrix} \sin(x) & \sin(2x) & \sin(3x) \\ \cos(x) & 2\cos(2x) & 3\cos(3x) \\ -\sin(x) & -4\sin(2x) & -9\sin(3x) \end{vmatrix} = \begin{vmatrix} \sin(x) & \sin(2x) & \sin(3x) \\ \cos(x) & 2\cos(2x) & 3\cos(3x) \\ 0 & -3\sin(2x) & -8\sin(3x) \end{vmatrix}$$
$$\begin{aligned} &= \sin(x)(-16\cos(2x)\sin(3x) + 9\cos(3x)\sin(2x)) \\ &\quad - \cos(x)(-8\sin(2x)\sin(3x) + 3\sin(2x)\sin(3x)) \\ &= 9\sin(x)\cos(3x)\sin(2x) - 16\sin(x)\cos(2x)\sin(3x) + 5\cos(x)\sin(2x)\sin(3x). \end{aligned}$$

Note that we added row 1 to row 3 to produce the zero in column 1 before performing a cofactor expansion along column 1 to compute the determinant. We can attempt to simplify $W(S)$ further using trigonometric identities, but there is really no point in doing so in this case. Let us evaluate $W_S(x)$ at some convenient value for x to see if we obtain at least one non-zero result. Notice that $\sin(3x)$ appears in two terms. If we let $x = \pi/3$, then $\sin(3x) = 0$, and so:

$$W_S(\pi/3) = 9\sin(\pi/3)\cos(\pi)\sin((2\pi/3)) = -\frac{27}{4}.$$

Since $W_S(x)$ is non-zero at $x = \pi/3$, we can safely conclude that S is linearly ***independent***. □

What if we had evaluated $W_S(x)$ at several values for x and kept getting zero? Would this mean that S were dependent? We will answer this question in more depth in the next subsection, but for now, let us say that when this happens, you can either keep trying with more values for x until you get a non-zero result, or you can try to think of identities (trigonometric or otherwise) that will show that $W_S(x) = z(x)$ for all $x \in I$. A valid conclusion in one case will be presented in the next subsection.

What About the Converse?

Suppose that $W_S(x) = z(x)$ for ***all*** $x \in I$. Does this also mean that S is a ***dependent*** set? Unfortunately, the answer in general is ***no***. Giuseppe Peano, in 1889, found a counterexample. Consider the two functions:

$$f(x) = x^2, \text{ and } g(x) = x|x|,$$

and consider the set $S = \{f(x), g(x)\} \subset F(\mathbb{R})$. Note that we can write the second function as:

$$g(x) = \begin{cases} x^2 & \text{if } x \geq 0 \\ -x^2 & \text{if } x < 0 \end{cases}, \quad \text{and so} \quad g'(x) = \begin{cases} 2x & \text{if } x \geq 0 \\ -2x & \text{if } x < 0 \end{cases} = 2|x|.$$

Thus, both $f(x)$ and $g(x)$ are ***differentiable*** for ***all*** x, and:

$$W_S(x) = \begin{vmatrix} x^2 & x|x| \\ 2x & 2|x| \end{vmatrix} = 2x^2|x| - 2x^2|x| = z(x),$$

so $W_S(x) = z(x)$ for ***all*** $x \in \mathbb{R}$. However, S contains only two (non-zero) functions, and so S is dependent ***if and only if*** these functions are ***parallel*** as vectors, that is:

$$f(x) = k \cdot g(x), \text{ for some non-zero } k \in \mathbb{R}, \text{ for } \textbf{\textit{all}}\ x \in \mathbb{R}.$$

But since $f(x) = x^2 = g(x)$ when $x \geq 0$, the only possible solution for this equation would be $k = 1$. But $f(x) = -g(x)$ when $x < 0$, and so $k = 1$ does not work for ***all*** $x \in \mathbb{R}$. Thus, S is actually an ***independent*** set, even though $W_S(x) = z(x)$ for all $x \in \mathbb{R}$.

What went wrong with this set? Notice that $g'(x) = 2|x|$, and so $g'(x)$ is ***not*** differentiable at $x = 0$. Peano himself also discovered that the converse will be true if we require that the functions in S are ***real analytic***, that is they are members of $C^\infty(\mathbb{R})$, the set of all functions whose higher derivatives ***all*** exist at all $x \in \mathbb{R}$. Since $g(x)$ is not in $C^\infty(\mathbb{R})$, this Example also shows that the following converse could fail if our functions are ***not*** in $C^\infty(\mathbb{R})$.

Theorem: Let $S = \{f_1(x), f_2(x), \ldots, f_n(x)\} \subset C^\infty(\mathbb{R})$.
If $W(S) = z(x)$ for ***all*** $x \in \mathbb{R}$, then S is a linearly ***dependent*** set.

Example: Let us consider the set $S = \{\cos^2(x), \sin^2(x), 1\} \subset C^\infty(\mathbb{R})$. We know that:

$$\cos^2(x) + \sin^2(x) = 1,$$

and so S is certainly a ***dependent*** set. To prepare ourselves to compute $W(S)$, we have:

$$\frac{d}{dx}\cos^2(x) = 2\cos(x)(-\sin(x)) = -\sin(2x), \text{ and so:}$$
$$\frac{d^2}{dx^2}\cos^2(x) = -2\cos(2x). \quad \text{Similarly:}$$
$$\frac{d}{dx}\sin^2(x) = 2\sin(x)\cos(x) = \sin(2x), \text{ and so:}$$
$$\frac{d^2}{dx^2}\sin^2(x) = 2\cos(2x).$$

Thus, the Wronskian of this set is:

$$W_S(x) = \begin{vmatrix} \cos^2(x) & \sin^2(x) & 1 \\ -\sin(2x) & \sin(2x) & 0 \\ -2\cos(2x) & 2\cos(2x) & 0 \end{vmatrix} = -2\sin(2x)\cos(2x) + 2\sin(2x)\cos(2x) = z(x),$$

where we computed the determinant using a cofactor expansion along the 3rd column. Thus, we verify that S is indeed a dependent set.$\square$

5.5 Section Summary

Let $S = \{f_1(x), f_2(x), \ldots, f_n(x)\} \subset C^{n-1}(I)$ for some interval I. The ***Wronskian*** of S, $W_S(x)$, is:

$$W_S(x) = W(\{f_1(x), f_2(x), \ldots, f_n(x)\}) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}.$$

If S is a ***dependent*** set, then $W_S(x) = z(x)$ for all values $x \in I$.
Thus, if $W(S)$ is ***non-zero*** for at least ***one*** $x \in I$, then S is an ***independent*** set.
As a partial converse, suppose $S = \{f_1(x), f_2(x), \ldots, f_n(x)\} \subset C^\infty(I)$, for some interval I.
If $W(S) = z(x)$ for ***all*** $x \in I$, then S is a linearly ***dependent*** set.

5.5 Exercises

For Exercises (1) to (12): (a) Find the Wronskian $W_S(x)$ of the set of functions S, and (b) decide whether S is independent or dependent. If S is independent, give at least one value for x so that the Wronskian is non-zero at x. Approximations are acceptable.

1. $S = \{\cos(x),\ \cos(2x),\ \cos(3x)\}$
2. $S = \{e^x \sin(x),\ e^x \cos(x)\}$
3. $S = \{e^{kx} \sin(nx),\ e^{kx} \cos(nx)\}$, where k and n are fixed non-zero real numbers.
4. $S = \{\tan^2(x),\ \sec^2(x),\ 1\}$
5. $S = \{\cot^2(x),\ \csc^2(x),\ 1\}$
6. $S = \{\cos(x),\ \sin(x),\ \cos(2x),\ \sin(2x)\}$
7. $S = \{\tan(x),\ \tan(2x),\ \tan(3x)\}$
8. $S = \{x^{1/2}, x^{3/5}, x^{7/4}\}$
9. $S = \{x^{1/2}, x^{1/3}, x^{1/4}, x^{1/5}\}$
10. $S = \left\{\sqrt{x-1},\ \sqrt{x-2},\ \sqrt{x-3},\ \sqrt{x-4}\right\}$
11. $S = \{3^x, 4^x, 5^x\}$
12. $S = \{\log_3(x),\ \log_4(x),\ \log_5(x)\}$

For Exercises (13) to (16): In Section 3.3, we said that if $S = \{f_i(x) \mid i \in I\}$ is an ***infinite*** set of functions, then S is a linearly ***independent*** set if every ***finite*** subset of S is linearly ***independent***. Thus, we can apply the idea of the Wronskian on an arbitrary finite subset of S to determine if S is dependent or independent. For the following infinite sets S: (a) Write down what an arbitrary finite subset S' of S would look like, where S' contains n functions; (b) Find the Wronskian $W_{S'}(x)$ of the set S' from (a); (c) Determine if S is linearly independent or dependent using $W_{S'}(x)$. Hint: in all four problems, one factor for $W_{S'}(x)$ is a Vandermonde determinant. You may need to perform some row and/or column operations before the Vandermonde determinant reveals itself.

13. $S = \{e^{kx} \mid k \in \mathbb{R}\} \subset C^\infty(\mathbb{R})$.
14. $S = \{b^x \mid b \in (0,\infty)\} \subset C^\infty(\mathbb{R})$.
15. $S = \{x^k \mid k \in (0,\infty)\} \subset C^\infty((0,\infty))$.
16. $S = \{(x-k)^m \mid k \in \mathbb{R}\}$, where m is some fixed real number.
    Hint: the answer to (c) depends on whether or not m is an integer. Think of Exercise 10.

A Summary of Chapter 5

An ordered list consisting of the numbers $1, 2, \ldots, n$, with each number appearing exactly once, is called a ***permutation*** of the set $\{1, 2, 3, .., n\}$. There are $n!$ permutations of $\{1, 2, 3, .., n\}$.

An ***inversion*** occurs in a permutation σ if a number on the left is bigger than a number to its right.

A permutation σ is ***even*** if it has an even number of inversions, and σ is ***odd*** if it has an odd number of inversions. The ***sign*** of σ, denoted $sgn(\sigma)$, is $+1$ if σ is even, and -1 if σ is odd.

A permutation σ is also a ***bijection*** of the set $\{1, 2, 3, .., n\}$, and as such possesses an ***inverse***, σ^{-1}.

Permutations can be represented by a directed bipartite graph. We will use two copies of the set $\{1, 2, 3, .., n\}$, one on top of the other, to serve as our vertices. If $\sigma(i) = j$, we will have a directed edge (an arrow) from i on the top row to j on the bottom row. An inversion occurs when two of these edges intersect.

The graph representing σ^{-1} is the same as that of σ, only with the arrows reversed. Thus, the number of ***inversions*** in σ and σ^{-1} are the ***same***, so $sgn(\sigma) = sgn(\sigma^{-1})$.

If σ' is obtained from σ by ***exchanging*** any two components, then $sgn(\sigma') = -sgn(\sigma)$.

Consequently, ***half*** of the $n!$ permutations of $\{1, 2, 3, .., n\}$ are ***even***, and ***half*** are ***odd***.

In this Chapter, all matrices are $n \times n$ or ***square***. The ***determinant*** of an $n \times n$ matrix A is:

$$det(A) = |A| = \sum_{\substack{\text{all permutations} \\ \sigma \text{ of } \{1,2,\ldots,n\}}} sgn(\sigma) \cdot a_{1,\sigma(1)} \cdot a_{2,\sigma(2)} \cdot \cdots \cdot a_{n-1,\sigma(n-1)} \cdot a_{n,\sigma(n)}.$$

$det(A) = det(A^\top)$.

If A has a ***row*** or a ***column*** of ***zeroes***, then $det(A) = 0$.

If A has two ***proportional*** rows or columns, then $det(A) = 0$.

If A is an upper or a lower ***triangular*** matrix, then $det(A) = a_{1,1} \cdot a_{2,2} \cdot \cdots \cdot a_{n,n}$.

In particular, if $D = Diag(d_1, d_2, \ldots, d_n)$, then $det(D) = d_1 \cdot d_2 \cdot \cdots \cdot d_n$.

Suppose E is an ***elementary matrix***. If E is obtained from $\boldsymbol{I}_n$ by:

1. ***multiplying*** row i by $k \neq 0$, then $det(E) = k$.
2. ***exchanging*** row i and row j, then $det(E) = -1$.
3. ***adding*** k times row i to row j, then $det(E) = 1$.

Consequently, the determinant of every elementary matrix is ***non-zero***.

Let A be an $n \times n$ matrix. Suppose B is obtained from A by:

1. ***multiplying*** row i of A by $k \neq 0$. Then: $det(B) = k \cdot det(A)$.
2. ***exchanging*** row i and row j of A. Then: $det(B) = -det(A)$.
3. ***adding*** k times row i of A to row j of A. Then: $det(B) = det(A)$.

Analogous statements can be made by replacing the word "row" with the word "column."

Consequently if E is the elementary matrix corresponding to the row operation performed, then $B = E \cdot A$, and so: $det(E \cdot A) = det(E) \cdot det(A)$.

In particular: $det(k \cdot A) = k^n \cdot det(A)$.

The determinant of A can be computed by performing row or column operations on A until we get a triangular matrix B. The determinant of A is obtained by multiplying the determinant of B by all the non-zero numbers k that were factored out of rows and columns using Type 1 operations, and by $(-1)^t$, where t is the total number of Type 2 operations. Type 3 operations do not affect the value of the determinant.

Let A and B be $n \times n$ matrices. Then:

- A is ***invertible*** ***if and only if*** $det(A) \neq 0$.
- $det(A \cdot B) = det(A) \cdot det(B) = det(B \cdot A)$.
- For any positive integer k: $det(A^k) = [det(A)]^k$.
- If A is ***invertible***, then: $det(A^{-1}) = 1/det(A) = [det(A)]^{-1}$.
  Thus, in this case, $det(A^k) = [det(A)]^k$ for all integers k.

The determinant of the submatrix obtained from A by erasing its ith row and jth column is called the ***i,j-minor*** of A, denoted $M_{i,j}(A)$. The ***i,j-cofactor*** of A is the number: $C_{i,j}(A) = (-1)^{i+j} \cdot M_{i,j}(A)$.

We can compute $det(A)$ by a ***cofactor expansion along row i:***

$$det(A) = a_{i,1}C_{i,1} + a_{i,2}C_{i,2} + \cdots + a_{i,n}C_{i,n},$$

or a ***cofactor expansion along column j:*** $det(A) = a_{1,j}C_{1,j} + a_{2,j}C_{2,j} + \cdots + a_{n,j}C_{n,j}$.
Determinants can be computed by using a combination of two strategies — row and/or column operations and cofactor expansion along a row or column.

The ***cofactor matrix*** of A, denoted $cof(A)$, is the matrix whose entries are the corresponding cofactors of each entry of A: $cof(A) = [C_{i,j}(A)]$.

The ***adjugate matrix*** of A is the ***transpose*** of the cofactor matrix, and is written as:

$$adj(A) = cof(A)^{\mathsf{T}} = [C_{j,i}(A)].$$

Let A be any $n \times n$ matrix. Then: $A \cdot adj(A) = det(A) \cdot \boldsymbol{I_n}$.

Consequently, if A is ***invertible***, then: $A^{-1} = \dfrac{1}{det(A)} adj(A)$.

Cramer's Rule: Let A be an ***invertible*** matrix. Then: the unique solution to the matrix equation $A\vec{x} = \vec{b}$ has entries: $x_1 = det(A^{(1)})/det(A)$, $x_2 = det(A^{(2)})/det(A)$, … , $x_n = det(A^{(n)})/det(A)$, where $A^{(i)}$ is obtained from A by ***replacing column*** i with $\vec{b}$.

Let $S = \{f_1(x), f_2(x), \ldots, f_n(x)\} \subset C^{n-1}(I)$ for some interval I. The ***Wronskian*** of S, $W_S(x)$, is:

$$W_S(x) = W(\{f_1(x), f_2(x), \ldots, f_n(x)\}) = \begin{vmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ f_1''(x) & f_2''(x) & \cdots & f_n''(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{vmatrix}.$$

If S is a ***dependent*** set, then $W_S(x) = z(x)$ for all values $x \in I$.
Thus, if $W(S)$ is ***non-zero*** for at least ***one*** $x \in I$, then S is an ***independent*** set.

As a partial converse, suppose $S = \{f_1(x), f_2(x), \ldots, f_n(x)\} \subset C^\infty(\mathbb{R})$.

If $W(S) = z(x)$ for ***all*** $x \in I$, then S is a linearly ***dependent*** set.

Chapter 6

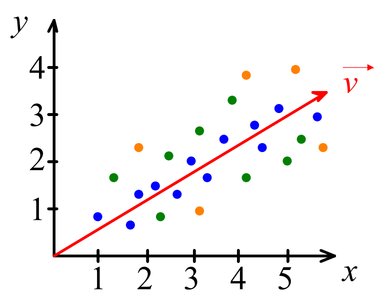
Painting the Lines:

Eigentheory, Diagonalization and Similarity

We continue to focus on $n \times n$ or ***square*** matrices A, and analogously, on ***operators*** $T : \mathbb{R}^n \to \mathbb{R}^n$. More generally, we will extend our ideas to operators on finite dimensional vector spaces.

We will be using the determinant function to find a polynomial, called the ***characteristic polynomial***, which is associated to A. The roots of this polynomial, which is of degree n, are special numbers called ***eigenvalues***, denoted by the Greek letter $\boldsymbol{\lambda}$. Each eigenvalue corresponds to an infinite set of non-zero vectors called ***eigenvectors***. These eigenvectors have special geometric properties: If $\vec{v}$ is an eigenvector for A corresponding to λ, then $A\vec{v} = \lambda\vec{v}$.

Alternatively, this means that if T is the operator with standard matrix A, then $T(\vec{v}) = \lambda\vec{v}$. ***Eigentheory***, the science of studying matrices, their eigenvalues and ***eigenspaces***, has important applications. We will see in Chapter 9, for example, that we can use eigentheory to determine if a set of data points is strongly linearly correlated via a constant of proportionality:



Eigenvectors have the major advantage in that if we knew that $\vec{v}$ is an eigenvector with associated eigenvalue λ, then we can compute the matrix product $A\vec{v}$ using the much easier scalar product $\lambda\vec{v}$.

Arguably the most important application of eigenvalues and eigenvectors is the process of ***diagonalizing*** a square matrix, when possible. This means that if T is the linear operator of $\mathbb{R}^n$ corresponding to A, then we can find a basis S for $\mathbb{R}^n$ so that $[T]_S$ is diagonal. We will see that this is possible if and only if we can find a set S of n ***linearly independent*** eigenvectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for our matrix A. In this case: $[T]_S = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where each λ_i is the eigenvalue corresponding to $\vec{v}_i$.

We know that there are an infinite number of bases for a vector space. We will explore what happens if we perform computations using two different bases for the same space. Our observations will lead us to the concept of ***similarity:*** two matrices that are similar, as the word implies basically represent the ***same operator***. They share certain properties that are intrinsic to these matrices, such as invertibility, rank and eigenvalues. The concept of similarity will enable us to find eigenvalues and eigenvectors for a linear operator $T : V \to V$ on an abstract vector space V.

6.1 The Eigentheory of Square Matrices

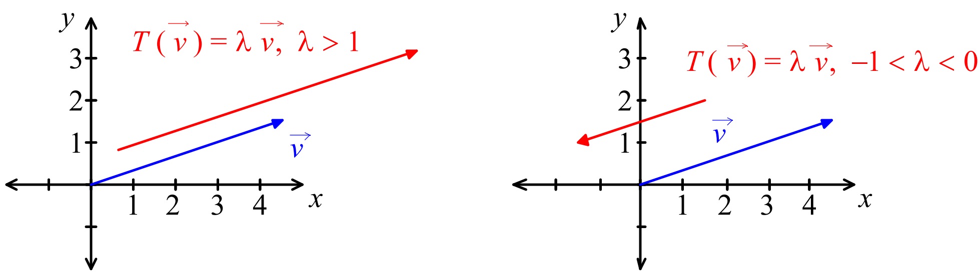
We know that it is time-consuming and often tedious to compute the product of an $n \times n$ matrix with an $n \times 1$ matrix: each row requires n multiplications, so there are n^2 multiplications in all. Then there are $n - 1$ additions to perform for each row, for a total of $n^2 - n$ additions. This is a total of $2n^2 - n$ arithmetic operations, and as such we expect that the time to perform this matrix product varies directly with n^2. Thus, our computation time approximately ***quadruples*** when we ***double*** the dimension of the matrices. This becomes a significant issue when the matrices are large, even with a modern computer.

On the other hand, multiplying an $n \times 1$ matrix by a ***scalar*** requires only n multiplications, and no additions. Clearly this is a much faster operation, and we should try to perform a scalar product instead of a matrix product whenever possible. This naturally leads us to the following:

Definition: Let A be an $n \times n$ matrix. We say that $\lambda \in \mathbb{R}$ (the Greek letter ***lambda***) is an ***eigenvalue*** of A, and a ***non-zero*** vector $\vec{v} \in \mathbb{R}^n$ is an ***eigenvector*** for A associated to λ, or simply an eigenvector for λ, if:

$$A\vec{v} = \lambda\vec{v}.$$

In other words, if $T : \mathbb{R}^n \to \mathbb{R}^n$, with $[T] = A$, then $T(\vec{v})$ is ***parallel*** to $\vec{v}$:



Examples of Possible Eigenvectors and Eigenvalues of an Operator on $\mathbb{R}^2$

We note that the eigenvalue λ depends on the matrix A, and the eigenvector $\vec{v}$ depends on the matrix A as well as the eigenvalue λ. Thus, as we shall see below, a matrix could have several different eigenvalues, and each eigenvalue could have different eigenvectors. We will refer to the study of the eigenvalues and eigenvectors of a matrix as ***Eigentheory***, for the sake of brevity.

Let us see how we would go about finding such a λ and $\vec{v}$. Suppose we ***assume*** that such a λ and such a $\vec{v}$ exist. Then we must have:

$$A\vec{v} = \lambda\vec{v} = (\lambda \boldsymbol{I}_n)(\vec{v}),$$

where the right side of the equation is now also a ***matrix product***. We can now put both sides of the equation together as:

$$(\lambda \boldsymbol{I}_n)(\vec{v}) - A\vec{v} = \vec{\mathbf{0}}_n, \quad \text{or}$$

$$(\lambda \boldsymbol{I}_n - A)\vec{v} = \vec{\mathbf{0}}_n.$$

Since we require that $\vec{v} \neq \vec{\mathbf{0}}_n$ this equation says that we can find a ***non-trivial*** solution $\vec{v}$ to the ***homogeneous system*** $(\lambda \boldsymbol{I}_n - A)\vec{v} = \vec{\mathbf{0}}_n$. By our Really Big Theorem on Invertible Matrices in Chapter 2, this is possible ***if and only if*** $\lambda \boldsymbol{I}_n - A$ is a ***non-invertible*** matrix (otherwise, we would have the ***unique*** trivial solution $\vec{v} = \vec{\mathbf{0}}_n$). But we saw in the previous Chapter that a matrix is ***non-invertible*** ***if and only if*** its ***determinant*** is ***zero***. Thus, we have:

Definition/Theorem: Let A be an $n \times n$ matrix. Then we can find a real number λ and a non-zero vector $\vec{v} \in \mathbb{R}^n$ such that:

$$A\vec{v} = \lambda\vec{v}$$

if and only if $det(\lambda \boldsymbol{I}_n - A) = 0$.
The equation above is called the ***characteristic equation*** of the matrix A.
The determinant in this equation is a polynomial whose highest term is λ^n, and it is called the ***characteristic polynomial*** of A**,** denoted $\boldsymbol{p_A(\lambda)}$, or $\boldsymbol{p(\lambda)}$:

$$p_A(\lambda) = p(\lambda) = det(\lambda \boldsymbol{I}_n - A).$$

Thus, in order to find the eigenvalues of a matrix, we need to find the ***real roots*** of the characteristic polynomial $p(\lambda)$. In order to find the eigenvectors associated to λ, we must find all the non-trivial solutions to the system of equations:

$$(\lambda \boldsymbol{I}_n - A)\vec{v} = \vec{\mathbf{0}}_n.$$

This means we find the non-zero vectors in the ***nullspace*** of $\lambda \boldsymbol{I}_n - A$. However, we can also use the system:

$$(A - \lambda \boldsymbol{I}_n)\vec{v} = \vec{\mathbf{0}}_n$$

instead, because all that we have to do to compute this safely is to ***subtract*** λ along the ***diagonal*** of A. The rest of the entries of A are unchanged.

The only part of this Definition/Theorem that needs to be proven is that λ^n is the ***highest term*** of this polynomial. Recall from the definition of the determinant that each of the $n!$ terms in the determinant is the product of a factor from each row and column, multiplied by the sign of the permutation associated with this product. In particular, $\lambda - a_{i,i}$ is on the main diagonal, for $i = 1..n$, and thus the term corresponding to the main diagonal is $sgn(\sigma)(\lambda - a_{1,1}) \cdot (\lambda - a_{2,2}) \cdot \cdots \cdot (\lambda - a_{n,n})$.

However, the permutation associated to this product is $(1, 2, \ldots, n)$, which has no inversions, and thus $sgn(\sigma) = 1$. Moreover, any other term will have fewer than n of such factors $\lambda - a_{i,i}$, and thus will contribute a term of degree strictly ***less*** than n. Thus, λ^n is the highest term appearing in the characteristic polynomial.■

Let us demonstrate the Eigentheory of a 2×2 matrix:

Example: Let $A = \begin{bmatrix} -37 & 21 \\ -70 & 40 \end{bmatrix}$. The characteristic polynomial is:

$$\begin{aligned} p(\lambda) &= det(\lambda \boldsymbol{I}_2 - A) \\ &= det\left(\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} -37 & 21 \\ -70 & 40 \end{bmatrix}\right) \\ &= \begin{vmatrix} \lambda+37 & -21 \\ 70 & \lambda-40 \end{vmatrix} \\ &= (\lambda+37)(\lambda-40) - (-21)(70) = \lambda^2 - 3\lambda - 10 = (\lambda+2)(\lambda-5). \end{aligned}$$

Thus, the eigenvalues are $\lambda = -2$ and 5. Let us find the eigenvectors, one at a time, for each of these two eigenvalues. For $\lambda = -2$, we need to find the ***nullspace*** of the matrix:

$$A - (-2) \cdot \boldsymbol{I}_2 = \begin{bmatrix} -37+2 & 21 \\ -70 & 40+2 \end{bmatrix} = \begin{bmatrix} -35 & 21 \\ -70 & 42 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & -\frac{3}{5} \\ 0 & 0 \end{bmatrix}.$$

Thus, y is free, and we get:

$$\vec{v} = \left\langle \frac{3}{5}t, t \right\rangle = \frac{t}{5}\langle 3, 5\rangle, \quad \text{for some } t \in \mathbb{R},\ t \neq 0.$$

as our eigenvectors. Recall that we are only interested in the ***non-zero*** members of the nullspace. We can check that these are indeed eigenvectors by directly multiplying A by the column vector:

$$A\vec{v} = \begin{bmatrix} -37 & 21 \\ -70 & 40 \end{bmatrix}\begin{bmatrix} \frac{3}{5}t \\ t \end{bmatrix} = \begin{bmatrix} -\frac{6}{5}t \\ -2t \end{bmatrix} = -2\begin{bmatrix} \frac{3}{5}t \\ t \end{bmatrix},$$

and thus $\vec{v}$ is an eigenvector for $\lambda = -2$. Similarly, for $\lambda = 5$, we have to find the ***nullspace*** of:

$$A - 5 \cdot \boldsymbol{I}_2 = \begin{bmatrix} -37-5 & 21 \\ -70 & 40-5 \end{bmatrix} = \begin{bmatrix} -42 & 21 \\ -70 & 35 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & 0 \end{bmatrix}.$$

Again, y is free, and the eigenvectors are thus of the form:

$$\vec{v} = \left\langle \frac{1}{2}t, t \right\rangle = \frac{1}{2}t\langle 1, 2\rangle, \quad \text{for some } t \in \mathbb{R},\ t \neq 0.$$

We can check that these are indeed eigenvectors as before:

$$A\vec{v} = \begin{bmatrix} -37 & 21 \\ -70 & 40 \end{bmatrix}\begin{bmatrix} \frac{1}{2}t \\ t \end{bmatrix} = \begin{bmatrix} \frac{5}{2}t \\ 5t \end{bmatrix} = 5\begin{bmatrix} \frac{1}{2}t \\ t \end{bmatrix},$$

and thus $\left\langle \frac{1}{2}t, t \right\rangle$ is indeed an eigenvector for $\lambda = 5$. □

The 2×2 matrix that we looked at had integer roots. Suppose for simplicity that A is any 2×2 matrix with rational (possibly integer) entries. The characteristic polynomial $p(\lambda) = \lambda^2 + b\lambda + c$ will have rational coefficients. Let $\Delta = b^2 - 4c$ be its discriminant. If $\Delta < 0$, the eigenvalues will be ***imaginary***. We will not deal with imaginary eigenvalues for now, but we will deal with them in Chapter 8. If $\Delta = 0$, there will be a ***unique*** (double) root, and this eigenvalue will be rational. If $\Delta > 0$, we will get two distinct ***real*** eigenvalues, and they will be rational ***if and only if*** Δ is a perfect square.

Eigenspaces

We said in the definition that an eigenvector $\vec{v}$ ***cannot*** be the zero vector. However, we also know that:

$$A\vec{\mathbf{0}}_n = \vec{\mathbf{0}}_n = \lambda \cdot \vec{\mathbf{0}}_n,$$

for ***any*** $n \times n$ matrix A and ***any*** scalar λ. In other words, $\vec{\mathbf{0}}_n$ behaves somewhat like a "trivial eigenvector." For this reason, we give it an honorary membership in this set, but we will continue to say that the zero vector is ***not*** an eigenvector of A :

Definition/Theorem — Eigenspaces:
Let A be an $n \times n$ matrix and $\lambda \in \mathbb{R}$. We define the ***eigenspace*** of A associated to λ, denoted $\boldsymbol{Eig}(\boldsymbol{A},\boldsymbol{\lambda})$, to be:

$$\boldsymbol{Eig}(\boldsymbol{A},\boldsymbol{\lambda}) = \{\vec{v} \in \mathbb{R}^n \,|\, A\vec{v} = \lambda\vec{v}\}.$$

Notice that $A\vec{\mathbf{0}}_n = \vec{\mathbf{0}}_n = \lambda\vec{\mathbf{0}}_n$, so $\vec{\mathbf{0}}_n \in \boldsymbol{Eig}(\boldsymbol{A},\boldsymbol{\lambda})$.
If λ is an actual ***eigenvalue*** for A, then $Eig(A,\lambda) = nullspace(A - \lambda \boldsymbol{I}_n)$, which is a ***non-zero subspace*** of $\mathbb{R}^n$ containing all the ***eigenvectors*** of A associated to λ, and thus its dimension is strictly ***positive***.
If λ is ***not*** an eigenvalue of A, then $Eig(A,\lambda)$ consists only of $\vec{\mathbf{0}}_n$. In this case, we can refer to $Eig(A,\lambda)$ as a ***trivial eigenspace***.
Thus, we can say that λ is an ***eigenvalue*** of A ***if and only if*** the eigenspace $\boldsymbol{Eig}(\boldsymbol{A},\boldsymbol{\lambda})$ is at least ***1-dimensional***.

Proof: First of all, $Eig(A,\lambda)$ is a ***non-empty*** set, because $\vec{\mathbf{0}}_n \in Eig(A,\lambda)$, as seen above. If λ is an actual eigenvalue for A, then A has at least one (non-zero!) eigenvector $\vec{v}$ associated to λ. Thus $Eig(A,\lambda)$ contains at least one non-zero vector, so its dimension is at least 1.

Now, we have to show that $Eig(A,\lambda)$ is closed under ***vector addition*** and ***scalar multiplication:***

Let $\vec{v}_1$ and $\vec{v}_2$ be two eigenvectors for A associated to λ. This means that:

$$A\vec{v}_1 = \lambda\vec{v}_1 \text{ and } A\vec{v}_2 = \lambda\vec{v}_2.$$

Thus we have:

$$A(\vec{v}_1 + \vec{v}_2) = A\vec{v}_1 + A\vec{v}_2 = \lambda\vec{v}_1 + \lambda\vec{v}_2 = \lambda(\vec{v}_1 + \vec{v}_2),$$

and thus $\vec{v}_1 + \vec{v}_2$ is also an eigenvector for A associated to λ (or possibly the zero vector, but this is still a member of $Eig(A,\lambda)$). Thus $Eig(A,\lambda)$ is closed under vector addition. Similarly:

$$A(k \cdot \vec{v}_1) = k \cdot A(\vec{v}_1) = k \cdot \lambda\vec{v}_1 = \lambda \cdot (k\vec{v}_1).$$

Thus, $k \cdot \vec{v}_1$ is an eigenvector for A associated to λ (or it is the zero vector). Thus $Eig(A,\lambda)$ is a subspace of $\mathbb{R}^n$.$\blacksquare$

Example: We saw in the previous Example that the eigenvalues of:

$$A = \begin{bmatrix} -37 & 21 \\ -70 & 40 \end{bmatrix}$$

are $\lambda = -2$ and 5. We can write the associated eigenspaces in terms of the eigenvectors that we computed above:

$$Eig(A,-2) = Span(\{\langle 3, 5\rangle\}), \quad \text{and}$$
$$Eig(A,5) = Span(\{\langle 1, 2\rangle\}),$$

and thus each eigenspace is 1-dimensional. □

Eigentheory for Triangular Matrices

We saw in Section 5.2 that the easiest non-trivial matrices for which we can find the determinant are the triangular matrices. Similarly, we can find the characteristic polynomial, as well as the eigenvalues of these special matrices, with hardly any effort:

> ***Theorem:*** Let A be an upper or lower ***triangular*** $n \times n$ matrix, and suppose the entries along the main diagonal are $c_1, c_2, \ldots, c_n$. Then: the ***characteristic polynomial*** of A is:
>
> $$p(\lambda) = (\lambda - c_1)(\lambda - c_2)\cdots(\lambda - c_n),$$
>
> and therefore the ***eigenvalues*** are precisely $c_1, c_2, \ldots, c_n$. Moreover, if:
>
> $$D = Diag(d_1, d_2, \ldots, d_n)$$
>
> is a ***diagonal*** matrix, then for every $i = 1\ldots n$: $\vec{e}_i$ is an ***eigenvector*** for d_i.

The proof is left as an Exercise.

Examples: Consider the lower triangular matrix $A = \begin{bmatrix} 4/3 & 0 & 0 \\ 22/3 & -7/3 & 0 \\ -10/3 & 5/3 & 4/3 \end{bmatrix}$

The characteristic polynomial of A is:

$$p(\lambda) = \left(\lambda - \tfrac{4}{3}\right)\left(\lambda + \tfrac{7}{3}\right)\left(\lambda - \tfrac{4}{3}\right) = \left(\lambda - \tfrac{4}{3}\right)^2\left(\lambda + \tfrac{7}{3}\right),$$

and the eigenvalues are indeed $\lambda = 4/3$ and $-7/3$.

Let us find the eigenvectors associated to each eigenvalue. We will need the ***nullspaces*** of:

$$A - \tfrac{4}{3}\boldsymbol{I}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 22/3 & -11/3 & 0 \\ -10/3 & 5/3 & 0 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and}$$

$$A + \tfrac{7}{3}\boldsymbol{I}_3 = \begin{bmatrix} 11/3 & 0 & 0 \\ 22/3 & 0 & 0 \\ -10/3 & 5/3 & 11/3 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{11}{5} \\ 0 & 0 & 0 \end{bmatrix}.$$

We can see from the rrefs that $Eig(A,4/3) = Span(\{\langle 1, 2, 0\rangle, \langle 0, 0, 1\rangle\})$, which is a 2-dimensional subspace, and $Eig(A,-7/3) = Span(\{\langle 0,-11, 5\rangle\})$, which is a 1-dimensional subspace.

Now, suppose we change A by changing 22/3 to 23/3:

$$A_1 = \begin{bmatrix} 4/3 & 0 & 0 \\ 23/3 & -7/3 & 0 \\ -10/3 & 5/3 & 4/3 \end{bmatrix}$$

Since we did not change any of the diagonal entries, the characteristic polynomial and the eigenvalues are exactly the same as before. Let us study the eigenspaces:

$$A_1 - \frac{4}{3}\boldsymbol{I}_3 = \begin{bmatrix} 0 & 0 & 0 \\ 23/3 & -11/3 & 0 \\ -10/3 & 5/3 & 0 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and}$$

$$A_1 + \frac{7}{3}\boldsymbol{I}_3 = \begin{bmatrix} 11/3 & 0 & 0 \\ 23/3 & 0 & 0 \\ -10/3 & 5/3 & 11/3 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \frac{11}{5} \\ 0 & 0 & 0 \end{bmatrix}.$$

Notice that the rref is the same for $\lambda = -7/3$, and so that eigenspace is also the same as before. However, the rref for $\lambda = 4/3$ is now ***different***, with only one free variable x_3. Thus, $Eig(A_1, 4/3) = Span(\{\langle 0, 0, 1\rangle\})$ is only 1-dimensional.$\square$

Let us conclude this Section with a Theorem that is almost obvious, but it still deserves to be stated. Its proof will be left as an Exercise.

Theorem: Suppose that λ_1 and λ_2 are two ***distinct*** eigenvalues for A. Then:

$$Eig(A, \lambda_1) \cap Eig(A, \lambda_2) = \left\{\vec{\mathbf{0}}_n\right\}.$$

In other words, an eigenvector for A belongs to exactly ***one*** eigenspace.

6.1 Section Summary

Let A be an $n \times n$ matrix. We say that λ is an ***eigenvalue*** of A, and $\vec{v}$ is an ***eigenvector*** for A associated to λ, or simply an eigenvector for λ, if $A\vec{v} = \lambda\vec{v}$, where $\vec{v}$ is a ***non-zero*** vector of $\mathbb{R}^n$.

Let A be an $n \times n$ matrix. Then we can find a real number λ and a non-zero vector $\vec{v} \in \mathbb{R}^n$ such that $A\vec{v} = \lambda\vec{v}$ ***if and only if***: $det(\lambda \boldsymbol{I}_n - A) = 0$.

This equation is the ***characteristic equation*** of the matrix A. The determinant above is a polynomial whose highest term is λ^n, called the ***characteristic polynomial*** of A, and denoted $p_A(\lambda)$ or just $p(\lambda)$.

If $\lambda \in \mathbb{R}$, we define the ***eigenspace*** of A associated to λ, denoted $\boldsymbol{Eig(A,\lambda)}$, to be:

$$\boldsymbol{Eig(A,\lambda)} = \{\vec{v} \in \mathbb{R}^n \mid A\vec{v} = \lambda\vec{v}\} = nullspace(A - \lambda\boldsymbol{I}_n).$$

This set includes $\vec{\mathbf{0}}_n$ for any real number λ. If λ is an actual eigenvalue for A, then $Eig(A, \lambda)$ is a ***non-zero subspace*** of $\mathbb{R}^n$. It contains all the eigenvectors of A associated to λ, and thus, its dimension is strictly ***positive***. If λ is ***not*** an eigenvalue of A, then $Eig(A, \lambda)$ consists only of the zero vector. In this case, we can refer to $Eig(A,\lambda)$ as a ***trivial*** eigenspace.

Thus, λ is an ***eigenvalue*** of A ***if and only if*** the eigenspace $\boldsymbol{Eig(A,\lambda)}$ is at least ***1-dimensional***.

Let A be an upper or lower ***triangular*** $n \times n$ matrix, and suppose the entries along the main diagonal are $c_1, c_2, \ldots, c_n$. Then: the characteristic polynomial of A is: $p(\lambda) = (\lambda - c_1)(\lambda - c_2)\cdots(\lambda - c_n)$.

Thus, the eigenvalues are precisely $c_1, c_2, \ldots, c_n$. Moreover, if $D = Diag(d_1, d_2, \ldots, d_n)$ is a diagonal matrix, then for every $i = 1\ldots n$: $\vec{e}_i$ is an eigenvector for d_i.

Suppose that λ_1 and λ_2 are two ***distinct*** eigenvalues for A. Then: $Eig(A,\lambda_1) \cap Eig(A,\lambda_2) = \left\{\vec{\mathbf{0}}_n\right\}$.

6.1 Exercises

For Exercises (1) to (15): For each of the following matrices: (a) find the characteristic polynomial, (b) find the eigenvalues, (c) find a basis for each eigenspace consisting of vectors with integer coefficients, and (d) find the dimension of each eigenspace:

1. $\begin{bmatrix} -8 & -10 \\ 5 & 7 \end{bmatrix}$

2. $\begin{bmatrix} -5 & 4 \\ -20 & 13 \end{bmatrix}$

3. $\begin{bmatrix} 8 & 6 \\ 6 & 3 \end{bmatrix}$

4. $\begin{bmatrix} -61 & -84 \\ 42 & 58 \end{bmatrix}$

5. $\begin{bmatrix} 3 & 5 \\ -9 & -3 \end{bmatrix}$

6. $\begin{bmatrix} -94 & 245 \\ -42 & 109 \end{bmatrix}$

7. $\begin{bmatrix} 5 & 4 & -7 \\ 0 & -2 & -3 \\ 0 & 0 & -4 \end{bmatrix}$

8. $\begin{bmatrix} 4 & 0 & 0 \\ -1 & 7 & 0 \\ 3 & -6 & -2 \end{bmatrix}$

9. $\begin{bmatrix} -5 & -3 & 6 \\ 0 & 0 & -7 \\ 0 & 0 & 8 \end{bmatrix}$

10. $\begin{bmatrix} 3 & 3 & -15 & 12 \\ 0 & 2 & 5 & -4 \\ 0 & 0 & 3 & -1 \\ 0 & 0 & 0 & 4 \end{bmatrix}$

11. $\begin{bmatrix} -2 & 0 & 0 & 0 \\ -4 & 3 & 0 & 0 \\ 8 & -10 & -2 & 0 \\ -24 & 30 & 15 & 3 \end{bmatrix}$

12. $\begin{bmatrix} 5 & 0 & 0 & 0 \\ 7 & -3 & 0 & 0 \\ -21 & 24 & 5 & 0 \\ 14 & -16 & 0 & 5 \end{bmatrix}$

Note: for Exercises (13) to (15), follow the same instructions as above, but we warn that you should follow exactly the same algorithm that you have been using. ***DO NOT*** clear the denominators before performing the algorithm. Deal with the fractions properly.

13. $\begin{bmatrix} \frac{43}{9} & \frac{49}{9} \\ -\frac{28}{9} & -\frac{34}{9} \end{bmatrix}$

14. $\begin{bmatrix} -1/3 & 5/3 & 2/3 \\ 0 & 4/3 & -1/3 \\ 0 & 0 & 2/3 \end{bmatrix}$

15. $\begin{bmatrix} 5/2 & -7/2 & 1/2 & 0 \\ 0 & 0 & 9/2 & 7/2 \\ 0 & 0 & -3/2 & -5/2 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}$

For Exercises (16) to (23): In each item, you are given two or three matrices that are almost equal to each other. For each matrix, follow the same instructions as above, and write down an observation as to how the eigenvalues and/or eigenvectors differ among the matrices in each item. Use technology for Exercise 19 onwards, if allowed by your instructor.

16. a. $\begin{bmatrix} 0 & 9 \\ 4 & 0 \end{bmatrix}$ b. $\begin{bmatrix} 0 & 9 \\ -4 & 0 \end{bmatrix}$

17. a. $\begin{bmatrix} 3 & 15 & -6 \\ 0 & -2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$ b. $\begin{bmatrix} 3 & 14 & -6 \\ 0 & -2 & 2 \\ 0 & 0 & 3 \end{bmatrix}$

18. a. $\begin{bmatrix} -7 & 0 & 0 \\ -3 & 2 & 0 \\ 6 & -18 & -7 \end{bmatrix}$ b. $\begin{bmatrix} -7 & 0 & 0 \\ -3 & 2 & 0 \\ 6 & 18 & -7 \end{bmatrix}$

19. a. $\begin{bmatrix} -2 & -10 & 15 & 4 \\ 0 & 3 & -7 & 21 \\ 0 & 0 & -2 & 15 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ b. $\begin{bmatrix} -2 & -10 & 14 & 4 \\ 0 & 3 & -7 & 21 \\ 0 & 0 & -2 & 15 \\ 0 & 0 & 0 & 3 \end{bmatrix}$ c. $\begin{bmatrix} -2 & -10 & 14 & 4 \\ 0 & 3 & -7 & 20 \\ 0 & 0 & -2 & 15 \\ 0 & 0 & 0 & 3 \end{bmatrix}$

20. a. $\begin{bmatrix} -2 & -10 & 8 & -16 \\ 0 & 3 & -2 & 4 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & -2 \end{bmatrix}$ b. $\begin{bmatrix} -2 & -20 & 8 & -16 \\ 0 & 3 & -2 & 4 \\ 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & -2 \end{bmatrix}$ c. $\begin{bmatrix} -2 & -20 & 8 & -16 \\ 0 & 3 & -2 & 4 \\ 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & -2 \end{bmatrix}$

21. a. $\begin{bmatrix} 1 & 4 & -12 & 36 & 80 \\ 0 & 3 & -6 & 18 & 45 \\ 0 & 0 & 1 & 6 & 15 \\ 0 & 0 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ b. $\begin{bmatrix} 1 & 4 & -12 & 36 & 90 \\ 0 & 3 & -6 & 18 & 45 \\ 0 & 0 & 1 & 6 & 15 \\ 0 & 0 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ c. $\begin{bmatrix} 1 & 4 & -12 & 40 & 90 \\ 0 & 3 & -6 & 20 & 45 \\ 0 & 0 & 1 & 6 & 15 \\ 0 & 0 & 0 & 3 & 5 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$

22. a. $\begin{bmatrix} \sqrt{3} & 0 & 0 \\ \sqrt{3}-\sqrt{2} & \sqrt{2} & 0 \\ 5 & -5 & \sqrt{3} \end{bmatrix}$ b. $\begin{bmatrix} \sqrt{3} & 0 & 0 \\ \sqrt{3}-\sqrt{2} & \sqrt{2} & 0 \\ 4 & -5 & \sqrt{3} \end{bmatrix}$

23. a. $\begin{bmatrix} 3\pi^2 & 0 & 0 \\ 6\pi-4 & 2\pi & 0 \\ 1 & -\pi & 3\pi^2 \end{bmatrix}$ b. $\begin{bmatrix} 3\pi^2 & 0 & 0 \\ 6\pi-4 & 2\pi & 0 \\ 2 & -\pi & 3\pi^2 \end{bmatrix}$

24. Let $A = \begin{bmatrix} -8 & -10 \\ 5 & 7 \end{bmatrix}$, be the matrix from Exercise 1.

   a. Find $A^\top$.
   b. Find the characteristic polynomial for $A^\top$.
   c. Is this the same characteristic polynomial as that of A?
   d. Describe the eigenvectors for each of the eigenvalues for $A^\top$.
   e. Compare your answers to (d) with your eigenvectors for A. Are they the same or different?

25. Prove in general that A and $A^\top$ have the same characteristic polynomial: $p_A(\lambda) = p_{A^\top}(\lambda)$. Thus, they have the same eigenvalues. Hint: what word best describes the matrix $\lambda \boldsymbol{I}_n$, as it relates to this Exercise? The deeper relationship between the corresponding eigenspaces will be explained in Chapter 9.

26. Let A be an upper or lower triangular $n \times n$ matrix, and suppose the entries along the main diagonal are $c_1, c_2, \ldots, c_n$. Prove that the characteristic polynomial of A is:

$$p(\lambda) = (\lambda - c_1)(\lambda - c_2)\cdots(\lambda - c_n),$$

and therefore the eigenvalues are $c_1, c_2, \ldots, c_n$.

27. Prove that if $D = Diag(d_1, d_2, \ldots, d_n)$ is a diagonal matrix with ***distinct*** diagonal entries d_i, then for every $i = 1 \ldots n$: $Eig(D, d_i) = Span(\{\vec{e}_i\})$.

28. Prove that $\vec{e}_1$ is always an eigenvector of an upper triangular matrix A. Describe the eigenvalue it is associated to (where do you find it?).

29. State and prove an analogous theorem for lower triangular matrices similar to the previous Exercise. Hint: look at the last Example in this Section.

30. Suppose that $B = kA$, for some scalar k, and A is a square matrix. Prove that if $\vec{v}$ is an eigenvector for A with corresponding eigenvalue λ, then $\vec{v}$ is also an eigenvector for B. What is the corresponding eigenvalue?

31. The objective of this Exercise is to show that the ***rotation matrix:***

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

does not have any real eigenvalues, for any value of θ except when $\theta = \pi n$, where n is an integer.

   a. Find the characteristic polynomial of this matrix.
   b. Show that the discriminant of the characteristic equation is ***negative***, unless $\theta = \pi n$ for some integer n. If so, show that there are exactly two different matrices R_θ, and find their eigenvalues.
   c. Now, argue ***geometrically*** that the rotation matrix can only have real eigenvalues if $\theta = \pi n$ for some integer n. Hint: draw the effect of R_θ on an eigenvector.

32. In contrast, the objective of this Exercise is to show that the matrix:

$$A = \begin{bmatrix} -\cos(\theta) & \sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

always has real eigenvalues.

a. Show that $A = DR_\theta$, where R_θ is the rotation matrix from the previous Exercise, and D is a diagonal matrix. What is D?

b. Based on your answer to (a), explain in words the geometric action of A. Hint: a matrix product is equivalent to a composition of two transformations.

c. Find the characteristic polynomial of A.

d. Show that the eigenvalues of A do not depend on θ.

e. Find the general eigenvector for each eigenvalue (these do depend on θ).

f. Use the half angle formulas:

$$\tan\left(\frac{\theta}{2}\right) = \frac{1-\cos(\theta)}{\sin(\theta)} = \frac{\sin(\theta)}{1+\cos(\theta)}$$

to show that the eigenvectors can be expressed in terms of $\cos(\theta/2)$ and $\sin(\theta/2)$.

g. Use your answers to (b) and (f) to explain geometrically why we get the eigenvectors in (e) for each eigenvalue. Draw a picture for each eigenvalue.

h. Do you notice something special about the two sets of eigenvectors?

i. Assemble the matrix A for $\theta = \cos^{-1}(-5/13)$ and find the corresponding eigenvectors, expressed with integer coefficients.

j. Repeat steps (a) to (i) for the matrix:

$$B = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}.$$

Warning: Part (a) should be slightly modified.

33. Suppose that λ_1 and λ_2 are two ***distinct*** eigenvalues for A. Prove that:

$$Eig(A,\lambda_1) \cap Eig(A,\lambda_2) = \left\{\vec{\mathbf{0}}_n\right\}.$$

34. ***Matrices in Block Diagonal Form:*** Suppose that A_1, A_2, …, A_k are all square matrices, not necessarily of the same size, with $k \geq 2$. We defined the direct sum of these matrices:

$$A = A_1 \oplus A_2 \oplus \cdots \oplus A_k$$

in the Exercises of Section 2.8.

a. Prove that: $p_A(\lambda) = p_{A_1}(\lambda) \cdot p_{A_2}(\lambda) \cdot \cdots \cdot p_{A_k}(\lambda)$. Hint: use Exercise 15 in Section 5.3 regarding the determinant of matrices in block diagonal form.

b. Now, suppose that:

$$A_1 = \begin{bmatrix} 23 & 40 \\ -8 & -13 \end{bmatrix}, \text{ and } A_2 = \begin{bmatrix} -13 & -16 & 16 \\ -8 & -5 & 8 \\ -16 & -16 & 19 \end{bmatrix}, \text{ with}$$

$$\begin{aligned} Eig(A_1,3) &= Span(\{\langle -2, 1\rangle\}), \\ Eig(A_1,7) &= Span(\{\langle -5, 2\rangle\}), \\ Eig(A_2,3) &= Span(\{\langle -1, 1, 0\rangle, \langle 1, 0, 1\rangle\}), \text{ and} \\ Eig(A_2,-5) &= Span(\{\langle 1, 2, 1\rangle\}). \end{aligned}$$

Find the eigenvalues of $A = A_1 \oplus A_2$ and find a basis for each eigenspace. Hint/Warning: A_1 and A_2 have a ***common*** eigenvalue.

6.2 The Geometry of Eigentheory and Computational Techniques

The concepts of an eigenvalue and an eigenvector obviously have geometric implications: $\vec{v}$ is an eigenvector for an operator T ***if and only if*** $T(\vec{v})$ is ***parallel*** to $\vec{v}$. We begin this section by seeing instances when eigentheory naturally appears from the geometric description of an operator.

The Kernel as an Eigenspace

Although $\vec{\mathbf{0}}_n$ is ***not*** allowed to be an eigenvector, the scalar 0 is allowed to be an eigenvalue. Notice, though, that $\lambda = 0$ is an eigenvalue for A if and only if there exists a non-zero vector $\vec{v}$ such that:

$$A\vec{v} = \lambda \cdot \vec{v} = 0 \cdot \vec{v} = \vec{\mathbf{0}}_n,$$

or in other words, ***if and only if*** A has a non-zero ***kernel***. We know from The Really Big Theorem on Invertibility in Chapter 2 that for a square matrix, this is equivalent to A being ***not*** invertible. The contrapositive of this statement thus tells us that $\lambda = 0$ is ***not*** an eigenvalue for A ***if and only if*** A is ***invertible***. Together with one of our main results in Chapter 5 that A is invertible ***if and only if*** $det(A)$ is non-zero, we can now formally add two more conditions to our Really Big Theorem:

Theorem — Addenda to the Really Big Theorem on Invertibility:
Let A be an $n \times n$ matrix. Then, the condition that A is ***invertible*** is equivalent to the following:
23. $det(A)$ is ***not*** 0.
24. $\lambda = 0$ is ***not*** an ***eigenvalue*** for A.

Example: Let $A = \begin{bmatrix} -1 & -2 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$.

The third row is the sum of the first two rows, and therefore this matrix is not invertible. Let us find its characteristic polynomial and verify:

$$p(\lambda) = det(\lambda \boldsymbol{I}_3 - A) = det\left(\begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} - \begin{bmatrix} -1 & -2 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}\right)$$

$$= \begin{vmatrix} \lambda+1 & 2 & -2 \\ -2 & \lambda-2 & -1 \\ -1 & 0 & \lambda-3 \end{vmatrix}$$

$$= (\lambda+1)(\lambda-2)(\lambda-3) + 2 + 0 - 2(\lambda-2) + 4(\lambda-3) - 0$$

$$= \lambda^3 - 4\lambda^2 + \lambda + 6 + 2 - 2\lambda + 4 + 4\lambda - 12$$

$$= \lambda^3 - 4\lambda^2 + 3\lambda = \lambda(\lambda-1)(\lambda-3).$$

Thus, $\lambda = 0$ is indeed an eigenvalue, along with 1 and 3. The rref of A is:

$$\left[\begin{array}{ccc} 1 & 0 & 3 \\ 0 & 1 & -\frac{5}{2} \\ 0 & 0 & 0 \end{array}\right],$$

and from this we can see that:

$$Eig(A,0) = nullspace(A) = Span(\{\langle -6, 5, 2\rangle\})$$

A basis for the other two eigenspaces can be found, as usual, by looking at the nullspace of $A - \lambda \boldsymbol{I}_3$.□

We repeat that $\lambda = 0$ ***is*** allowable as an ***eigenvalue***, but $\vec{v} = \vec{\mathbf{0}}_n$ is ***not*** allowable as an ***eigenvector***, even though it is a member of every ***eigenspace***.

Eigentheory for Geometric Operators

The ***scaling operators*** S_k have matrices $k \cdot \boldsymbol{I}_n$. The characteristic polynomial of these matrices are thus:

$$p(\lambda) = det(\lambda \cdot \boldsymbol{I}_n - k \cdot \boldsymbol{I}_n) = (\lambda - k)^n,$$

so $\lambda = k$ is the only eigenvalue. But then, for all $\vec{v} \in \mathbb{R}^n$:

$$S_k(\vec{v}) = k\vec{v},$$

and so all $\vec{v} \in \mathbb{R}^n$ are eigenvectors for S_k (except the zero vector), that is:

$$Eig(k \cdot \boldsymbol{I}_n, k) = \mathbb{R}^n.$$

The ***shear operators*** in $\mathbb{R}^2$ have matrices:

$$A = \left[\begin{array}{cc} 1 & a \\ 0 & 1 \end{array}\right] \text{ or } B = \left[\begin{array}{cc} 1 & 0 \\ b & 1 \end{array}\right].$$

Recall that the first kind are ***horizontal*** shears and the second kind are ***vertical*** shears. We assume that a and b are not zero, otherwise we get the identity matrix (which is an example of a scaling operator with $\lambda = 1$). Both of these matrices have characteristic polynomial:

$$p(\lambda) = (\lambda - 1)^2,$$

and so $\lambda = 1$ is the only eigenvalue for either kind of shear operator. But then:

$$A - 1 \cdot \boldsymbol{I}_2 = \left[\begin{array}{cc} 0 & a \\ 0 & 0 \end{array}\right] \text{ or } \left[\begin{array}{cc} 0 & 0 \\ b & 0 \end{array}\right].$$

The rref of these matrices are, respectively:

$$\left[\begin{array}{cc} 0 & 1 \\ 0 & 0 \end{array}\right] \text{ and } \left[\begin{array}{cc} 1 & 0 \\ 0 & 0 \end{array}\right], \quad \text{thus}$$

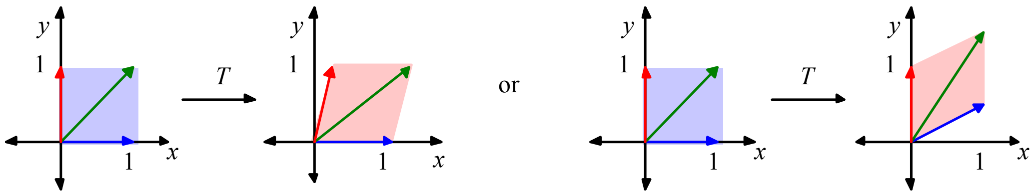
$$Eig(A,1) = Span(\{\langle 0, 1\rangle\}) = Span\left(\left\{\vec{j}\right\}\right), \text{ and}$$

$$Eig(B,1) = Span(\{\langle 1, 0\rangle\}) = Span\left(\left\{\vec{i}\right\}\right).$$

We can check that:

$$\begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}\begin{bmatrix} x \\ 0 \end{bmatrix} = \begin{bmatrix} x \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ b & 1 \end{bmatrix}\begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix}.$$

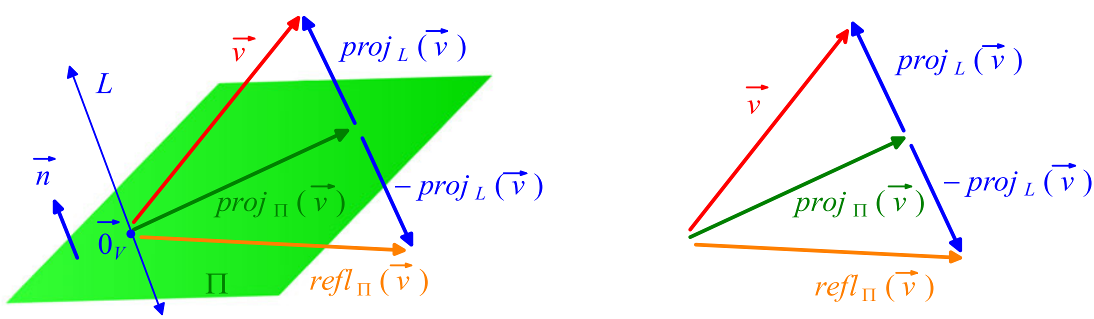
For a horizontal shear, we indeed see from the action on the basic box that a vector on the x-axis does not change, whereas for a vertical shear, a vector on the y-axis does not change:



The Effect of a Shear Operator on the Basic Box

Our computations above show that there are ***no other*** eigenvectors for shear operators.

Now, let us think of a plane Π and its normal line L, and the associated ***projection*** and ***reflection operators:***



Projections Onto and Reflections Across a Plane Π

If a vector $\vec{v}$ is on Π, then its projection onto as well as its reflection across Π are already ***itself***:

$$proj_\Pi(\vec{v}) = \vec{v} \text{ and } refl_\Pi(\vec{v}) = \vec{v}.$$

In other words, $\vec{v}$ is an eigenvector for both operators, with eigenvalue 1 in both cases. Conversely, any vector $\vec{u}$ which is ***not*** on Π will be projected to a vector on Π, and its reflection will certainly not be the same vector $\vec{u}$ either. Thus:

$$Eig([proj_\Pi], 1) = \Pi = Eig([refl_\Pi], 1).$$

On the other hand, if $\vec{w}$ is on L, then its projection onto Π will be the ***zero vector***, and its reflection across Π will be its ***negative:***

$$proj_\Pi(\vec{w}) = \vec{\mathbf{0}}_3 \text{ and } refl_\Pi(\vec{w}) = -\vec{w}.$$

Thus, $\vec{w}$ is an eigenvector for $proj_\Pi$ with eigenvalue 0 (in other words, $\vec{w}$ is in the ***nullspace*** of $proj_\Pi$) and $\vec{w}$ is an eigenvector for $refl_\Pi$ with eigenvalue -1. Conversely, a vector $\vec{u}$ which is not on L will have a non-zero projection onto Π, and its reflection across Π will not be in the exact opposite direction as $\vec{u}$. Thus:

$$Eig([proj_\Pi], 0) = L = Eig([refl_\Pi], -1).$$

We leave it as an Exercise to see the effect of $proj_L$ and $refl_L$ on these two kinds of vectors. Similar ideas can be applied to projections onto a line L in $\mathbb{R}^2$ as well as to reflections across L. Along with these operators, you will also investigate in the Exercises the eigentheory of rotation matrices in $\mathbb{R}^2$, and some of their related matrices.

More generally, the effect of a 2×2 matrix with two distinct eigenspaces can easily be visualized using a vector from each eigenspace.

Example: Let $A = \begin{bmatrix} \frac{9}{26} & \frac{5}{13} \\ -\frac{6}{13} & \frac{43}{26} \end{bmatrix}$. Its characteristic polynomial is:

$$p(\lambda) = \begin{vmatrix} \lambda - \frac{9}{26} & -\frac{5}{13} \\ \frac{6}{13} & \lambda - \frac{43}{26} \end{vmatrix} = \left(\lambda - \tfrac{9}{26}\right)\left(\lambda - \tfrac{43}{26}\right) + \left(\tfrac{5}{13}\right)\left(\tfrac{6}{13}\right)$$
$$= \lambda^2 - 2\lambda + \frac{3}{4} = \frac{1}{4}(2\lambda - 1)(2\lambda - 3).$$

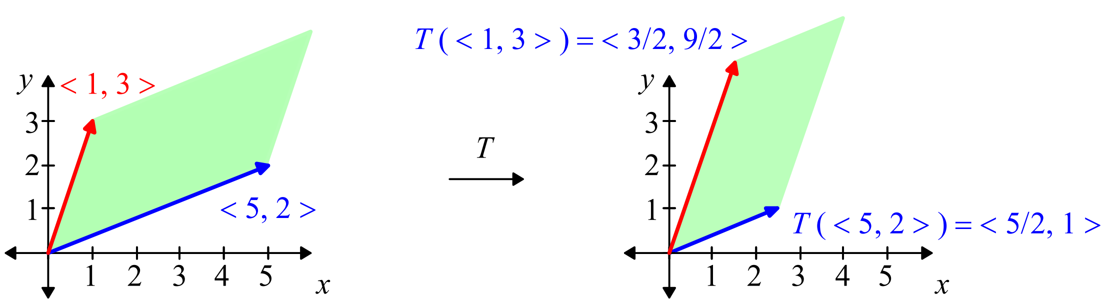
Thus, the eigenvalues are $1/2$ and $3/2$. For $\lambda = 1/2$:

$$A - \frac{1}{2}\boldsymbol{I}_2 = \begin{bmatrix} \frac{-2}{13} & \frac{5}{13} \\ -\frac{6}{13} & \frac{15}{13} \end{bmatrix}, \quad \text{with rref} \quad R_1 = \begin{bmatrix} 1 & -\frac{5}{2} \\ 0 & 0 \end{bmatrix}.$$

Thus, $Eig(A, 1/2) = Span(\langle 5, 2 \rangle)$. For $\lambda = 3/2$:

$$A - \frac{3}{2}\boldsymbol{I}_2 = \begin{bmatrix} -\frac{15}{13} & \frac{5}{13} \\ -\frac{6}{13} & \frac{2}{13} \end{bmatrix}, \quad \text{with rref} \quad R_2 = \begin{bmatrix} 1 & -\frac{1}{3} \\ 0 & 0 \end{bmatrix}.$$

Thus, $Eig(A, 3/2) = Span(\langle 1, 3 \rangle)$. We can therefore visualize the effect of the operator T corresponding to A using the parallelogram formed by the two basis vectors $\langle 5, 2 \rangle$ and $\langle 1, 3 \rangle$:



Visualizing the Effect of an Operator Through its Eigenvectors

Notice that the corresponding sides of the parallelogram are parallel to each other, although the proportions are different.□

Finding the eigenvalues of a non-triangular 3×3 matrix or bigger can be a daunting task, unless one can use mathematical software or a graphing calculator. We will need to find the roots of a cubic or higher-degree polynomial, which may be irrational or imaginary. However, if the entries of the matrix are integers, hopefully there are enough integer and rational roots so that any irrational root will involve only a square root. Let us now look at Theorems and techniques that will help us to find eigenvalues:

The Integer and Rational Roots Theorems

Our basic tools are two Theorems that we usually see in Precalculus:

> ***Theorem — The Integer Roots Theorem:***
> Let $p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ be a polynomial with ***integer*** coefficients, and $c_0 \neq 0$. Then, all the rational roots of $p(x)$ are in fact integers, and if $x = c$ is an integer root of $p(x)$, then c is a ***factor*** of the constant coefficient c_0.

Note: If $c_0 = 0$, then we can factor $p(x)$ in the form $x^k \cdot q(x)$ for some positive power k, and the constant coefficient of $q(x)$ is now non-zero. Thus we can apply the Integer Roots Theorem to $q(x)$.

If we have the misfortune of having a matrix with entries that are non-integer rational numbers (i.e. fractions), the characteristic polynomial will still have highest term λ^n, but some of the coefficients of the lower powers of λ may be fractions. To find the roots of such a polynomial, we would normally "clear denominators" by multiplying the polynomial by the least common denominator of the coefficients. This will give us a polynomial that will now have integer coefficients. The roots of this polynomial will be the same as the roots of original polynomial $p(\lambda)$. The highest term, though, will now have a coefficient that is ***not*** 1. To find the roots of such a polynomial, we can use:

> ***Theorem — The Rational Roots Theorem:***
> Let $q(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ be a polynomial with ***integer*** coefficients, with $c_0 \neq 0$. Then, all the rational roots of $q(x)$ are of the form $x = c/d$, where c is a factor of the constant coefficient c_0 and d is a factor of the leading coefficient c_n.

These two Theorems give us possible ***candidates*** for the integer and rational roots of the characteristic polynomial. It is still up to us to perform the tedious task of finding the actual roots, either by plugging them directly into $p(\lambda)$ or using ***synthetic division***, a technique that we also see in precalculus. We can likewise use the following Theorem for some additional assistance:

> ***Theorem — Descartes' Rule of Signs:***
> Let $p(x)$ be a polynomial with ***real*** coefficients. Then: the number of ***positive roots*** of $p(x)$ is equal to the number of sign changes in consecutive coefficients of $p(x)$, or it is less than this number by an ***even*** integer. Similarly, the number of ***negative roots*** of $p(x)$ is the number of sign changes in consecutive coefficients of $p(-x)$, or it is less than this number by an even integer.

In any case, if we are fortunate enough to find a root quickly, say $\lambda = c$, then we can factor out $\lambda - c$ from the characteristic polynomial, resulting in a polynomial of degree $n - 1$. We repeat this process of finding roots until we are down to a quadratic factor, at which point we use the quadratic formula or factoring techniques to find the remaining roots.

Example: Let $A = \begin{bmatrix} -13 & 8 & 8 \\ -20 & 15 & 8 \\ 4 & -4 & 3 \end{bmatrix}$. First we find the characteristic polynomial:

$$det(\lambda \boldsymbol{I_3} - A) = \begin{vmatrix} \lambda+13 & -8 & -8 \\ 20 & \lambda-15 & -8 \\ -4 & 4 & \lambda-3 \end{vmatrix}$$

$$\begin{aligned} &= (\lambda+13)(\lambda-15)(\lambda-3) + (-8)(-8)(-4) + (-8)(20)(4) \\ &\quad - (-8)(\lambda-15)(-4) - (\lambda+13)(-8)(4) - (-8)(20)(\lambda-3) \\ &= \lambda^3 - 5\lambda^2 - 189\lambda + 585 - 256 - 640 - 32\lambda + 480 + 32\lambda + 416 + 160\lambda - 480 \\ &= \lambda^3 - 5\lambda^2 - 29\lambda + 105, \end{aligned}$$

with a little work on a calculator. Let us hope that there are integer roots for $p(\lambda)$, and try the factors of $105 = 3 \times 5 \times 7$. Thus, even though 105 is a big number, it has a small number of factors, namely:

$$\pm\{1, 3, 5, 7, 15, 21, 35, 105\}.$$

Notice also that $p(\lambda)$ has ***two*** sign changes, and $p(-\lambda) = -\lambda^3 - 5\lambda^2 + 29\lambda + 105$ has only ***one*** sign change, and therefore we are guaranteed a ***unique negative real root***. We keep our fingers crossed that this negative root is an integer, as we sequentially try the negative factors of 105 :

$$\begin{aligned} p(-1) &= -1 - 5 + 29 + 105 = 128 && \text{No.} \\ p(-3) &= -27 - 45 + 87 + 105 = 120 && \text{Try again.} \\ p(-5) &= -125 - 125 + 145 + 105 = 0 && \text{Success!} \end{aligned}$$

Now, we divide out $\lambda + 5$ from $p(\lambda)$, either via long division or synthetic division, and get a quadratic, thus:

$$\lambda^3 - 5\lambda^2 - 29\lambda + 105 = (\lambda+5)(\lambda^2 - 10\lambda + 21).$$

The quadratic above now easily factors as $(\lambda-3)(\lambda-7)$, and so:

$$p(\lambda) = (\lambda+5)(\lambda-3)(\lambda-7).$$

Thus, the eigenvalues are: $\lambda = -5,\ 3$ and 7.□

Once the eigenvalues are found, it is now a straightforward task to find the corresponding eigenvectors using the Gauss-Jordan algorithm, as seen in the previous Section. We leave it as an Exercise to find the eigenvectors of the matrix above.

We have some further remarks about this Example. You may gamble and try to look for the ***positive*** roots instead, but unfortunately Descartes' Rule of Sign says that we have either 2 or 0 positive roots, and thus we are not guaranteed in advance that we will find positive roots, much less positive integer roots. The gamble would have paid off in this case because $\lambda = 3$ is an eigenvalue, and we would have discovered it next after failing with $\lambda = 1$.

Notice also that we were not guaranteed either that the unique negative root would be an integer. Since $p(-1)$ and $p(-3)$ are both positive, if we reach a point where $p(\lambda)$ is negative for some integer candidate $\lambda < -3$, we would have a sign change without hitting an integer root. Thus, this unique negative root would have been ***irrational***. If this happens, we should try to find the positive roots, and hope that they exist, and hope at least one is an integer. When all else fail, most graphing calculators and mathematical software can easily approximate the roots of polynomials. Some can also find an approximate basis for eigenspaces.

Here is one thing we learned in Calculus, though, that might give us some comfort:

Theorem: Let $p(x)$ be a polynomial with ***odd degree***. Then $p(x)$ has at least one ***real*** root.

Thus, all 3×3 matrices, all 5×5 matrices, etc., are guaranteed to have at least ***one real eigenvalue***. Unfortunately, this still does not guarantee us that the eigenvalue is an integer or a rational number.

Example: Consider the matrix:

$$A = \begin{bmatrix} 3 & 8 & -1 \\ 8 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}.$$

With a little bit of work, we will find that the characteristic polynomial of A is:

$$p(\lambda) = \lambda^3 - 10\lambda^2 - 34\lambda + 103.$$

Descartes' Rule of Signs tells us that there are two ***or*** zero positive roots for $p(\lambda)$, and thus we are ***not*** guaranteed a positive root. However:

$$p(-\lambda) = -\lambda^3 - 10\lambda^2 + 34\lambda + 103,$$

and so Descartes' Rule now tells us that there is ***exactly one*** negative root, since $p(-\lambda)$ has only one sign change. Unfortunately, 103 is a prime number, and so the only possible rational roots for $p(\lambda)$ are the integer roots ± 1 and ± 103. Directly plugging in -1 and -103 yield 126 and -1195212, respectively, so neither one of them is a root. This tells us that the negative root must be ***irrational***. However, since there is a sign change between the values of $p(-1)$ and $p(-103)$, the Intermediate Value Theorem of Calculus tells us that there must be a zero for $p(\lambda)$ somewhere between -103 and -1, and our instincts should tell us that this root should be much closer to -1 than -103, judging by their values under p. Let us try to narrow the gap a bit:

$$\begin{aligned} p(-2) &= 123, \\ p(-3) &= 88, \\ p(-4) &= 15, \quad \text{and at last:} \\ p(-5) &= -102. \end{aligned}$$

Thus, our irrational root is in the interval $[-5, -4]$. To refine our root further, we can apply ***Newton's Method***, which needs the derivative, $p'(\lambda) = 3\lambda^2 - 20\lambda - 34$. Recall that this method begins with an initial ***guess*** which we will call x_0, and the next guess is inductively defined to be:

$$x_{k+1} = x_k - \frac{p(x_k)}{p'(x_k)}.$$

Let us use $x_0 = -4$ as our initial guess, since 15 is closer to zero than -102. We will stop Newton's Method when the first four digits after the decimal point do not change between x_k and x_{k+1}. We get:

$$\begin{aligned}
x_1 &= -4 - \frac{p(-4)}{p^{/}(-4)} = -4 - \frac{15}{94} \\
&\approx -4.15957, \\
x_2 &= -4.15957 - \frac{p(-4.15957)}{p^{/}(-4.15957)} \approx -4.15957 - \frac{-0.5638199325}{101.0974678} \\
&\approx -4.153993, \quad \text{and} \\
x_3 &= -4.153993 - \frac{p(-4.153993)}{p^{/}(-4.153993)} \approx -4.153993 - \frac{-0.0006983350646}{100.8468335} \\
&\approx -4.153986075.
\end{aligned}$$

Rounding this off to -4.153986, we verify that $p(-4.153986) \approx 7.59167 \times 10^{-6}$, so it would appear that we have an excellent approximation for our negative root. We can now use ***synthetic division*** to factor out $\lambda + 4.153986$ from $p(\lambda)$:

$$\begin{array}{rr|rrrr}
-4.153986 &) & 1 & -10 & -34 & 103 \\
 & & & -4.153986 & 58.79546 & -102.9999937 \\
\hline
 & & 1 & -14.153986 & 24.79546 & \approx 0
\end{array}$$

The bottom line tells us that other factor of $p(\lambda)$ is the approximate quadratic:

$$\lambda^2 - 14.153986\lambda + 24.79546.$$

Applying the Quadratic Formula, we get our two other approximate roots:

$$\lambda \approx 2.048238689 \quad \text{and} \quad \lambda \approx 12.10574731.$$

Thus, all three roots are in fact irrational. Of course, a graphing calculator would also find these three approximate roots for us, with much less effort.

We should point out that since these are not exact values for λ, extra care must be given in applying the Gauss Jordan Algorithm to find the eigenspaces. Recall that each matrix $A - \lambda \cdot \boldsymbol{I}_3$ must ***not*** be invertible, that is, the correct rref should yield at least one row of ***zeroes***. Let us illustrate for $\lambda \approx 2.048239$. As usual, we first find $A - (2.048239)\boldsymbol{I}_3$:

$$\begin{bmatrix} 3 - 2.048239 & 8 & -1 \\ 8 & 5 - 2.048239 & 0 \\ -1 & 0 & 2 - 2.048239 \end{bmatrix} \approx \begin{bmatrix} 0.951761 & 8 & -1 \\ 8 & 2.951761 & 0 \\ -1 & 0 & -0.048239 \end{bmatrix}.$$

If we were to directly use technology at this point to find the rref of this matrix, we would be disappointed to get the identity matrix $\boldsymbol{I}_3$: this is both ***wrong*** and ***useless***. Let us find the correct approximate rref intelligently, step by step. Let us swap the 1st and 3rd row to get a leading 1 in the first column (after dividing row 1 by -1), and clear out the other two entries of column 1, as usual:

$$\to \begin{bmatrix} 1 & 0 & 0.048239 \\ 8 & 2.951761 & 0 \\ 0.951761 & 8 & -1 \end{bmatrix} \to \begin{bmatrix} 1 & 0 & 0.048239 \\ 0 & 2.951761 & -0.385912 \\ 0 & 8 & -1.045912 \end{bmatrix}.$$

To get a row of zeroes, the 2nd and 3rd row should be ***parallel*** to each other. We can easily verify this by dividing each row by the first non-zero entry (in other words, the corresponding entry in column 2):

$$\to \begin{bmatrix} 1 & 0 & 0.048239 \\ 0 & 1 & -0.1307395822 \\ 0 & 1 & -0.130739 \end{bmatrix}.$$

The 2nd and 3rd row are now approximately equal. If the λ we obtained were exact, the two rows should be exactly equal. This gives us the more useful approximate rref:

$$R \approx \begin{bmatrix} 1 & 0 & 0.048239 \\ 0 & 1 & -0.130739 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, our eigenspace is approximately:

$$Eig(A, 2.048239) \approx Span(\{\langle -0.048239, 0.130739, 1\rangle\}).$$

We can check that:

$$\begin{bmatrix} 3 & 8 & -1 \\ 8 & 5 & 0 \\ -1 & 0 & 2 \end{bmatrix}\begin{bmatrix} -0.048239 \\ 0.130739 \\ 1 \end{bmatrix} \approx \begin{bmatrix} -0.098805 \\ 0.267783 \\ 2.048239 \end{bmatrix}, \text{ whereas}$$

$$2.048239\begin{bmatrix} -0.048239 \\ 0.130739 \\ 1 \end{bmatrix} \approx \begin{bmatrix} -0.098805 \\ 0.2677847 \\ 2.048239 \end{bmatrix}.$$

We can apply these ideas to find an approximate basis for the other two eigenspaces:

$$Eig(A, -4.153986) \approx Span(\{\langle 6.153986, -5.3781913, 1\rangle\}) \text{ and}$$
$$Eig(A, 12.105747) \approx Span(\{\langle -10.105747, -11.37755, 1\rangle\}). \square$$

This last Example illustrates an important point: technology can be very useful in performing messy computations for us, but their precision is ***limited***. Their use could yield ***misleading*** or outright ***false*** answers. We need to intelligently interpret whatever results we obtain from them, and re-do our computations if necessary to account for the lack of precision.

We also remark that the matrix in the previous Example is ***symmetric***, and it will be stated in Chapter 7 and proven in Chapter 8 that all the eigenvalues of a symmetric matrix are ***real***, and thus our efforts to find their eigenspaces will not be in vain.

6.2 Section Summary

Let A be an $n \times n$ matrix. Then A is ***invertible*** ***if and only if*** $\lambda = 0$ is ***not*** an ***eigenvalue*** for A.

We can try to find integer or rational eigenvalues by using the Integer Roots Theorem or the Rational Roots Theorem, and Descartes' Rule of Signs. Technology can help us approximate irrational eigenvalues and a basis for the corresponding eigenspaces.

The Integer Roots Theorem: Let $p(x) = x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ be a polynomial with ***integer*** coefficients, with c_0 non-zero. Then, all the ***rational*** roots of $p(x)$ are in fact ***integers***, and if $x = c$ is an integer root of $p(x)$, then c is a ***factor*** of the constant coefficient c_0.

The Rational Roots Theorem: Let $q(x) = c_n x^n + c_{n-1}x^{n-1} + \cdots + c_1 x + c_0$ be a polynomial with ***integer*** coefficients, with $c_0 \neq 0$. Then, all the ***rational*** roots of $q(x)$ are of the form $x = c/d$, where c is a ***factor*** of the constant coefficient c_0 and d is a ***factor*** of the leading coefficient c_n.

Descartes' Rule of Signs: Let $p(x)$ be a polynomial with ***real*** coefficients. Then: the number of ***positive roots*** of $p(x)$ is equal to the number of sign changes in consecutive coefficients of $p(x)$, or less than this number by an ***even*** integer; the number of ***negative roots*** of $p(x)$ is the number of sign changes in consecutive coefficients of $p(-x)$, or less than this number by an ***even*** integer.

6.2 Exercises

1. Find a basis for the eigenspaces of each of the eigenvalues $\lambda = -5$, 3 and 7 for the matrix that is in the 3rd Example of this Section.

2. We saw that if A only has integer entries, then its characteristic polynomial $p(\lambda)$ has only integer coefficients. According to the Integer Roots Theorem, the factors of the constant coefficient c_0 (if $c_0 \neq 0$) are the only possible candidates for the integer roots of $p(\lambda)$.

   Prove that if c_0 factors as:

$$c_0 = p_1^{n_1} \bullet p_2^{n_2} \bullet \cdots \bullet p_k^{n_k}$$

   where $p_1, p_2, \ldots, p_k$ are the distinct prime factors of c_0, then there are exactly:

$$2(n_1 + 1)(n_2 + 1)\ldots(n_k + 1)$$

   distinct factors of c_0. Hint: What are the possible choices for the power of p_i that appears in a factor? Don't forget that you can have positive and negative factors.

   For Exercises (3) to (5): Use the previous Exercise to get a ***count*** of the number of candidates for the integer roots of the following polynomials, list ***all*** of them, then factor the polynomials completely and find their roots:

3. $p(\lambda) = \lambda^3 - 8\lambda^2 - 3\lambda + 90$
4. $p(\lambda) = \lambda^3 + \lambda^2 - 30\lambda - 72$
5. $p(\lambda) = \lambda^3 - 11\lambda^2 + 33\lambda - 15$

For Exercises (6) to (29): Find the characteristic polynomial, eigenvalues, and a basis for each eigenspace (for the real eigenvalues only), and the dimension of each eigenspace, of the following matrices. Warning: irrational eigenvalues will appear in Exercise 9, and you will need the Rational Roots Theorem in Exercise 11, 18, 19 and 20.

6. $\begin{bmatrix} 2 & 0 & 1 \\ 0 & 5 & 0 \\ 1 & 0 & 2 \end{bmatrix}$

7. $\begin{bmatrix} -4 & -2 & 0 \\ 0 & 3 & 0 \\ 1 & 5 & 3 \end{bmatrix}$

8. $\begin{bmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}$

9. $\begin{bmatrix} 3 & 1 & 0 \\ -2 & -3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

10. $\begin{bmatrix} -6 & -2 & 12 \\ 3 & -5 & 6 \\ -8 & -2 & 14 \end{bmatrix}$

11. $\begin{bmatrix} \frac{19}{4} & -2 & -2 \\ 5 & -\frac{9}{4} & -2 \\ \frac{7}{2} & -2 & -\frac{3}{4} \end{bmatrix}$

12. $\begin{bmatrix} 3 & 4 & 4 \\ -21 & -17 & -21 \\ 15 & 10 & 14 \end{bmatrix}$

13. $\begin{bmatrix} -11 & -6 & 12 \\ 15 & 10 & -12 \\ -15 & -6 & 16 \end{bmatrix}$

14. $\begin{bmatrix} -5 & -6 & 6 \\ 12 & 10 & -9 \\ -9 & -6 & 10 \end{bmatrix}$

15. $\begin{bmatrix} 29 & 64 & -32 \\ -16 & -35 & 16 \\ -8 & -16 & 5 \end{bmatrix}$

16. $\begin{bmatrix} -13 & 11 & 11 \\ -44 & 42 & 44 \\ 22 & -22 & -24 \end{bmatrix}$

17. $\begin{bmatrix} -19 & 16 & 16 \\ -72 & 53 & 48 \\ 48 & -32 & -27 \end{bmatrix}$

18. $\begin{bmatrix} 17 & \frac{22}{3} & -11 \\ \frac{55}{3} & 6 & -11 \\ \frac{110}{3} & \frac{44}{3} & -\frac{70}{3} \end{bmatrix}$

19. $\begin{bmatrix} -\frac{97}{4} & 75 & 15 \\ -\frac{25}{2} & \frac{153}{4} & \frac{15}{2} \\ 25 & -75 & -\frac{57}{4} \end{bmatrix}$

20. $\begin{bmatrix} -\frac{32}{5} & 14 & \frac{21}{5} \\ -7 & \frac{73}{5} & \frac{21}{5} \\ 14 & -28 & -\frac{39}{5} \end{bmatrix}$

21. $\begin{bmatrix} 0 & 0 & 3 & 0 \\ 0 & 3 & 0 & 0 \\ 3 & 0 & 0 & 4 \\ 0 & 0 & 4 & 0 \end{bmatrix}$

22. $\begin{bmatrix} 0 & 0 & 0 & 7 \\ 0 & 0 & 7 & 0 \\ 0 & 7 & 0 & 0 \\ 7 & 0 & 0 & 0 \end{bmatrix}$

23. $\begin{bmatrix} 0 & 3 & 0 & 7 \\ 3 & 0 & 7 & 0 \\ 0 & 7 & 0 & 3 \\ 7 & 0 & 3 & 0 \end{bmatrix}$

24. $\begin{bmatrix} 5 & -9 & -6 & -9 \\ 8 & -8 & -24 & -5 \\ 0 & -3 & 3 & -3 \\ -8 & 10 & 24 & 7 \end{bmatrix}$

25. $\begin{bmatrix} 68 & 80 & -80 & -30 \\ -32 & -39 & 37 & 15 \\ 32 & 37 & -39 & -15 \\ -20 & -25 & 25 & 13 \end{bmatrix}$

26. $\begin{bmatrix} -25 & 30 & 14 & -2 \\ 14 & -12 & -7 & 1 \\ -70 & 75 & 38 & -5 \\ 14 & -8 & -7 & -3 \end{bmatrix}$

27. $\begin{bmatrix} -26 & -42 & 14 & 42 \\ 14 & 23 & -7 & -21 \\ -42 & -63 & 23 & 63 \\ 14 & 21 & -7 & -19 \end{bmatrix}$

28. $\begin{bmatrix} -12 & -11 & 27 & 10 \\ 6 & 7 & -13 & -5 \\ -22 & -17 & 43 & 15 \\ 50 & 39 & -95 & -33 \end{bmatrix}$

29. $\begin{bmatrix} 2 & 0 & 0 & 0 & 2 \\ 0 & 2 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & 2 & 0 \\ 2 & 0 & 0 & 0 & 2 \end{bmatrix}$

For Exercises (30) to (32): The following symmetric matrices have irrational eigenvalues, but each eigenspace is only 1-dimensional. Find $p(\lambda)$ and the eigenvalues, correct to 4 decimal places. Find an approximate basis for each eigenspace, also correct to 4 decimal places. You may use technology, but heed the warnings in the final Example of this Section with regards to finding the approximate rref of $A - \lambda \boldsymbol{I}_n$.

30. $\begin{bmatrix} 1 & -3 & 2 \\ -3 & 0 & 4 \\ 2 & 4 & -2 \end{bmatrix}$ 31. $\begin{bmatrix} -1 & 1 & 3 \\ 1 & -2 & 0 \\ 3 & 0 & -5 \end{bmatrix}$ 32. $\begin{bmatrix} 2 & -1 & 0 & 1 \\ -1 & 3 & 2 & -1 \\ 0 & 2 & 1 & 0 \\ 1 & -1 & 0 & -3 \end{bmatrix}$

33. Suppose that A is an $n \times n$ matrix. Prove that if $\vec{v}$ is an eigenvector of A associated to the eigenvalue λ, and $k \neq 0$, then $k \cdot \vec{v}$ is again an eigenvector of A associated to λ.

34. Suppose that A is an $n \times n$ matrix. Prove that if λ is an eigenvalue for A with associated eigenvector $\vec{v}$, then λ^k is an eigenvalue for A^k for any positive integer k, with associated eigenvector $\vec{v}$ as well. Hint: use Induction.

35. Suppose that A is an ***invertible*** $n \times n$ matrix.

    a. Prove that if λ is an eigenvalue for A with associated eigenvector $\vec{v}$, then $1/\lambda$ is an eigenvalue for A^{-1}, with associated eigenvector $\vec{v}$ as well. As part of your proof, explain why the expression $1/\lambda$ makes sense if A is invertible.

    b. Use (a) to show that for every eigenvalue λ: $Eig(A, \lambda) = Eig(A^{-1}, 1/\lambda)$.

36. ***The Cayley-Hamilton Theorem*** states that: If A is an $n \times n$ matrix with characteristic polynomial $p(\lambda)$, then $p(A) = \mathbf{0}_{n\times n}$. We can think of this as saying that A is a ***root*** of its characteristic polynomial. This Theorem is very deep, and its proof is complicated.

    Demonstrate that the Cayley-Hamilton Theorem is true when applied to the matrices in Exercise 10, 15 and 24.

37. ***True or False:*** Determine if the statement is true or false. If the statement is true, cite a definition or Theorem that supports your conclusion, or give a convincing argument why the statement is true. If the statement is false, cite a definition or Theorem that supports your conclusion, or provide a counterexample, or give a convincing argument why the statement is false.

    a. If A is an $n \times n$ matrix and λ is one of its eigenvalues, then $\vec{\mathbf{0}}_n$ is an eigenvector for λ.

    b. If A is an $n \times n$ matrix and λ is one of its eigenvalues, then $\vec{\mathbf{0}}_n$ is a member of the eigenspace $Eig(A, \lambda)$.

    c. The diagonal entries $a_{i,i}$ are the eigenvalues of any $n \times n$ matrix A.

    d. Every 3×3 matrix has at least one rational eigenvalue.

    e. Every 3×3 matrix has at least one real eigenvalue.

    f. Every 4×4 matrix has at least one real eigenvalue.

    g. Every 5×5 matrix has at least one real eigenvalue.

6.3 Diagonalization of Square Matrices

One of the most elegant applications of Eigentheory is the process of diagonalizing a square matrix or a linear operator. In this Section, we will study the process for square matrices, and see the process for operators in Section 6.6:

> ***Definition:*** Let A be an $n \times n$ matrix. We say that A is ***diagonalizable*** if we can find an ***invertible*** matrix C such that:
> $$C^{-1}AC = D,$$
> where $D = Diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a ***diagonal*** matrix, or equivalently:
> $$AC = CD \quad \text{or} \quad A = CDC^{-1}.$$
> We also say that C ***diagonalizes*** A. The matrix product $C^{-1}AC$ is also referred to as ***conjugating*** A by C. A matrix which is ***not*** diagonalizable is also called ***defective***.

The key to understanding the connection between diagonalization and Eigentheory is actually the second equation: $AC = CD$.

If we partition C into its ***column*** vectors as:

$$C = \left[\ \vec{v}_1 \,|\, \vec{v}_2 \,|\, \cdots \,|\, \vec{v}_n\ \right],$$

we can think of both sides of the equation in terms of matrix multiplication. Recall that AC is the matrix whose columns are $A\vec{v}_i$. Similarly, in Section 2.9, we saw that multiplying a matrix C on the ***right*** by a diagonal matrix D has the effect of multiplying the ***columns*** of C by the corresponding diagonal entry in D. Thus, we get:

$$AC = CD \qquad \text{is equivalent to:}$$
$$\left[\ A\vec{v}_1 \,|\, A\vec{v}_2 \,|\, \cdots \,|\, A\vec{v}_n\ \right] = \left[\alpha_1\vec{v}_1 \,|\, \alpha_2\vec{v}_2 \,|\, \cdots \,|\, \alpha_n\vec{v}_n\right].$$

By comparing columns, we see that we must satisfy:

$$A\vec{v}_i = \alpha_i\vec{v}_i,$$

for each column $\vec{v}_i$. Thus, the columns $\vec{v}_i$ of C are ***eigenvectors*** of A, and the corresponding entry α_i in D is its ***eigenvalue***. Since C is ***invertible***, these columns must be linearly ***independent***, and consequently the set $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a ***basis*** for $\mathbb{R}^n$. We may summarize our discovery in the following:

> ***Theorem — The Basis Test for Diagonalizability:***
> Let A be an $n \times n$ matrix. Then, A is ***diagonalizable*** ***if and only if*** we can find a ***basis*** for $\mathbb{R}^n$ consisting of n linearly ***independent eigenvectors*** for A, say $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. If this is the case, then the diagonalizing matrix C is the matrix whose ***columns*** are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and the diagonal matrix D contains the corresponding ***eigenvalues*** along the main ***diagonal***.

Although there could be as many as $n!$ ways to arrange D, let us agree for the sake of uniformity to arrange the eigenvalues in ***increasing*** order.

Example: We saw the matrix:

$$A = \begin{bmatrix} -37 & 21 \\ -70 & 40 \end{bmatrix}$$

in Section 6.1, with eigenvalues $\lambda = -2$ and 5, and we found that:

$$Eig(A,-2) = Span(\{\langle 3, 5\rangle\}), \quad \text{and}$$

$$Eig(A,5) = Span(\{\langle 1, 2\rangle\}),$$

The two basis vectors are definitely not parallel, and therefore they are linearly ***independent***. Thus, we can assemble them into the columns of:

$$C = \begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix}, \text{ whose inverse is: } C^{-1} = \begin{bmatrix} 2 & -1 \\ -5 & 3 \end{bmatrix}.$$

We assemble the associated eigenvalues in the ***same order***, thus:

$$D = \begin{bmatrix} -2 & 0 \\ 0 & 5 \end{bmatrix}.$$

We can check that:

$$\begin{aligned} CDC^{-1} &= \begin{bmatrix} 3 & 1 \\ 5 & 2 \end{bmatrix}\begin{bmatrix} -2 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ -5 & 3 \end{bmatrix} \\ &= \begin{bmatrix} -6 & 5 \\ -10 & 10 \end{bmatrix}\begin{bmatrix} 2 & -1 \\ -5 & 3 \end{bmatrix} = \begin{bmatrix} -37 & 21 \\ -70 & 40 \end{bmatrix} = A. \end{aligned}$$

and thus A is indeed diagonalizable.□

Example: Let $A = \begin{bmatrix} 0 & -2 \\ 8 & 0 \end{bmatrix}$. Then:

$$p(\lambda) = det\left(\begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 8 & 0 \end{bmatrix}\right) = \begin{vmatrix} \lambda & 2 \\ -8 & \lambda \end{vmatrix} = \lambda^2 + 16.$$

This matrix does not have real eigenvalues, and thus does not have any eigenvectors. Thus, A is not diagonalizable, even though it ***almost*** looks like a diagonal matrix.□

In general, any matrix which has imaginary eigenvalues cannot be diagonalized. However, we will see in Chapter 8 that it will be possible to diagonalize such matrices over the set of matrices with ***complex*** entries, so to be more precise, we say:

> ***Theorem:*** Let A be an $n \times n$ matrix with ***imaginary*** eigenvalues. Then A is ***not*** diagonalizable over the set of ***real*** invertible matrices.

On the other hand, let us look at the following:

Example: Let $A = \begin{bmatrix} 2 & -3 & 7 \\ 0 & 2 & -1 \\ 0 & 0 & 5 \end{bmatrix}$.

This matrix is upper triangular, so the characteristic polynomial is $p(\lambda) = (\lambda - 2)^2(\lambda - 5)$, and the eigenvalues are 2 and 5. To find the eigenvectors, we must compute the matrices $\lambda \boldsymbol{I}_3 - A$ and their rrefs, resulting in:

$$\text{For } \lambda = 2\text{:} \quad \begin{bmatrix} 0 & -3 & 7 \\ 0 & 0 & -1 \\ 0 & 0 & 3 \end{bmatrix} \quad \text{with rref} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and}$$

$$\text{For } \lambda = 5\text{:} \quad \begin{bmatrix} -3 & -3 & 7 \\ 0 & -3 & -1 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{with rref} \quad \begin{bmatrix} 1 & 0 & -8/3 \\ 0 & 1 & 1/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus we can conclude that:

$$Eig(A, 2) = Span(\{\langle 1, 0, 0\rangle\}), \text{ and}$$
$$Eig(A, 5) = Span(\{\langle 8, -1, 3\rangle\}).$$

Thus, we only get a total of only ***two*** linearly independent eigenvectors for A instead of ***three***, so A is ***not*** diagonalizable. This shows that it is not always possible to diagonalize a matrix, even if all of its eigenvalues are real.□

Example: Let $B = \begin{bmatrix} 2 & -9 & 6 \\ 0 & 5 & -2 \\ 0 & 0 & 2 \end{bmatrix}$.

This matrix is again upper triangular, with characteristic polynomial $p(\lambda) = (\lambda - 2)^2(\lambda - 5)$, which is exactly the same polynomial as in the last Example. The eigenvalues are again $\lambda = 2$ and 5. To find the eigenvectors, we must compute the matrices $\lambda \boldsymbol{I}_3 - B$ and their rrefs, resulting in:

$$\text{For } \lambda = 2\text{:} \quad \begin{bmatrix} 0 & -9 & 6 \\ 0 & 3 & -2 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{with rref} \quad \begin{bmatrix} 0 & 1 & -\frac{2}{3} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad \text{and}$$

$$\text{For } \lambda = 5\text{:} \quad \begin{bmatrix} -3 & -9 & 6 \\ 0 & 0 & -2 \\ 0 & 0 & -3 \end{bmatrix} \quad \text{with rref} \quad \begin{bmatrix} 1 & 3 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus we can conclude that:

$$Eig(B, 2) = Span(\{\langle 1, 0, 0\rangle, \langle 0, 2, 3\rangle\}), \text{ and}$$
$$Eig(B, 5) = Span(\{\langle -3, 1, 0\rangle\}).$$

This time, we have ***three*** linearly independent eigenvectors, and so B ***is*** diagonalizable.□

Hopefully, you noticed that the suspicious issue is the fact that the factor $\lambda - 2$ was ***squared*** in both characteristic polynomials. This will be further explained below.

Independence of Distinct Eigenspaces

We saw in the previous example that it may not always be possible to get n linearly independent eigenvectors for an $n \times n$ matrix. However, since we had two eigenvalues, we were still fortunate enough to find two linearly independent eigenvectors. More generally, the number of eigenvalues gives us a ***lower bound*** for the number of linearly independent eigenvectors we can find:

> ***Theorem — Independence of Distinct Eigenspaces:***
> Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_k$ are ***distinct*** eigenvalues for an $n \times n$ matrix A, and suppose that $\vec{v}_i$ is an eigenvector of A corresponding to λ_i, for $i = 1..k$.
> Then: the set $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is ***linearly independent***. Thus, if A has a total of m distinct eigenvalues, we can find ***at least m*** linearly independent eigenvectors for A.

Proof: This is another proof that deserves to be called ***magical***. Since $\vec{v}_1 \neq \vec{0}_n$, the set $\{\vec{v}_1\}$ is linearly independent. Now let's see what happens when we include the 2nd vector. Suppose $\{\vec{v}_1, \vec{v}_2\}$ is a set of two eigenvectors with corresponding distinct eigenvalues λ_1 and λ_2. If this set were dependent, then $\vec{v}_2$ would parallel to $\vec{v}_1$, that is, $\vec{v}_2 = c\vec{v}_1$. Since subspaces are closed under scalar multiplication, this implies that $\vec{v}_2$ belongs to the same eigenspace as $\vec{v}_1$. This contradicts our condition that $\lambda_1 \neq \lambda_2$.

The rest of the proof proceeds by Mathematical Induction: Let us assume that the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_j\}$ is linearly independent for our Induction Hypothesis. We must show that the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_j, \vec{v}_{j+1}\}$ is still linearly independent as our Inductive Step.

Let us consider the ***dependence test equation*** for this set:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_j\vec{v}_j + c_{j+1}\vec{v}_{j+1} = \vec{\mathbf{0}}_{\boldsymbol{n}}.$$

By multiplying both sides of this equation by the matrix A on the left, we get:

$$\begin{aligned} A(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_j\vec{v}_j + c_{j+1}\vec{v}_{j+1}) &= A\vec{\mathbf{0}}_{\boldsymbol{n}} \quad \Rightarrow \\ c_1A\vec{v}_1 + c_2A\vec{v}_2 + \cdots + c_jA\vec{v}_j + c_{j+1}A\vec{v}_{j+1} &= \vec{\mathbf{0}}_{\boldsymbol{n}} \quad \Rightarrow \\ c_1\lambda_1\vec{v}_1 + c_2\lambda_2\vec{v}_2 + \cdots + c_j\lambda_j\vec{v}_j + c_{j+1}\lambda_{j+1}\vec{v}_{j+1} &= \vec{\mathbf{0}}_{\boldsymbol{n}}, \end{aligned}$$

where we were able to perform the last step because each $\vec{v}_i$ is an eigenvector with eigenvalue λ_i. Now, starting with the ***original*** highlighted dependence equation above, we can also multiply both sides by λ_{j+1}, thus getting:

$$\begin{aligned} \lambda_{j+1}(c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_j\vec{v}_j + c_{j+1}\vec{v}_{j+1}) &= \lambda_{j+1}\vec{\mathbf{0}}_{\boldsymbol{n}} \quad \Rightarrow \\ c_1\lambda_{j+1}\vec{v}_1 + c_2\lambda_{j+1}\vec{v}_2 + \cdots + c_j\lambda_{j+1}\vec{v}_j + c_{j+1}\lambda_{j+1}\vec{v}_{j+1} &= \vec{\mathbf{0}}_{\boldsymbol{n}}. \end{aligned}$$

Let us put our two resulting equations on top of each other:

$$\begin{aligned} c_1\lambda_1\vec{v}_1 \;+\; c_2\lambda_2\vec{v}_2 \;+\; \cdots + c_j\lambda_j\vec{v}_j \;+\; c_{j+1}\lambda_{j+1}\vec{v}_{j+1} &= \vec{\mathbf{0}}_{\boldsymbol{n}}, \quad \text{and} \\ c_1\lambda_{j+1}\vec{v}_1 + c_2\lambda_{j+1}\vec{v}_2 + \cdots + c_j\lambda_{j+1}\vec{v}_j + c_{j+1}\lambda_{j+1}\vec{v}_{j+1} &= \vec{\mathbf{0}}_{\boldsymbol{n}}. \end{aligned}$$

Notice that the last terms in both equations are ***identical***. Now, here comes the magic: if we ***subtract*** the corresponding sides of these equations from each other, we get:

$$c_1(\lambda_1 - \lambda_{j+1})\vec{v}_1 + c_2(\lambda_2 - \lambda_{j+1})\vec{v}_2 + \cdots + c_j(\lambda_j - \lambda_{j+1})\vec{v}_j = \vec{\mathbf{0}}_{\boldsymbol{n}}.$$

This now looks like a ***dependence test equation*** for the set of vectors in $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_j\}$. This set is assumed to be ***independent*** in the Induction Hypothesis, and so each $c_i(\lambda_i - \lambda_{j+1})$ must be zero! Since the λ_i are distinct, each $\lambda_i - \lambda_{j+1}$ is a ***non-zero*** scalar, and so this means that $c_i = 0$ for $i = 1 \ldots j$. But going back to our original dependence equation, we get: $c_{j+1}\vec{v}_{j+1} = \vec{\mathbf{0}}_{\boldsymbol{n}}$.

Since $\vec{v}_{j+1} \neq \vec{\mathbf{0}}_{\boldsymbol{n}}$, the Zero-Factors Theorem tells us that $c_{j+1} = 0$. Thus, the bigger set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_j, \vec{v}_{j+1}\}$ is still linearly independent. This argument shows that we can keep adding eigenvectors to this set, as long as we are adding an eigenvector from a ***new*** eigenspace. Thus, $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is linearly independent.■

Example: Consider the matrix from one of our previous Examples:

$$A = \begin{bmatrix} 2 & -3 & 7 \\ 0 & 2 & -1 \\ 0 & 0 & 5 \end{bmatrix}.$$

We saw that A is ***not*** diagonalizable because we could not find three linearly independent eigenvectors for A. But the two vectors that we did find, $\langle 1,0,0 \rangle$ and $\langle 8,-1,3 \rangle$, are still linearly ***independent***, i.e., not parallel.□

Geometric and Algebraic Multiplicities

We saw earlier that if the matrix A does not have n linearly independent eigenvectors, then we cannot diagonalize A. However, notice also that the eigenspace which made this impossible came from a ***double*** (i.e. repeated) eigenvalue for $p(\lambda)$. Clearly this is what produces the complication, and therefore it requires further investigation. First, let us introduce some new terminology:

Definitions — Algebraic and Geometric Multiplicities:
Let A be an $n \times n$ matrix with ***distinct*** (possibly imaginary) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Suppose $p(\lambda)$ factors as:

$$p(\lambda) = (\lambda - \lambda_1)^{n_1} \bullet (\lambda - \lambda_2)^{n_2} \bullet \cdots \bullet (\lambda - \lambda_k)^{n_k},$$

where $n_1 + n_2 + \cdots + n_k = n$.
We call the exponent n_i the ***algebraic multiplicity*** of λ_i.
We call $dim(Eig(A, \lambda_i))$ the ***geometric multiplicity*** of λ_i.
We agree that $dim(Eig(A, \lambda_i)) = 0$ if λ_i is an ***imaginary*** eigenvalue (at least for now).

Note: The Fundamental Theorem of Algebra states that the sum of the algebraic multiplicities of the λ_i must be the degree n of $p(\lambda)$.

A very deep result from a field of mathematics called (just by coincidence) ***Algebraic Geometry*** gives the connection between these two multiplicities:

Theorem — The Geometric vs. Algebraic Multiplicity Theorem:

For any eigenvalue λ_i of an $n \times n$ matrix. A, the ***geometric multiplicity*** of λ_i is ***at most equal*** to the ***algebraic multiplicity*** of λ_i. Thus, following our notation in the previous definitions:

$$dim(Eig(A, \lambda_i)) \leq n_i \text{ for every } i = 1..k.$$

Example: Consider again the matrix:

$$A = \begin{bmatrix} 2 & -3 & 7 \\ 0 & 2 & -1 \\ 0 & 0 & 5 \end{bmatrix}$$

from our previous Example. The characteristic polynomial of A is $p(\lambda) = (\lambda - 2)^2(\lambda - 5)$, and thus $\lambda = 2$ has algebraic multiplicity 2, and $\lambda = 5$ has algebraic multiplicity 1. However, we saw that $Eig(A, 2)$ and $Eig(A, 5)$ are both only 1-dimensional, thus verifying the Theorem. On the other hand, the matrix:

$$B = \begin{bmatrix} 2 & -9 & 6 \\ 0 & 5 & -2 \\ 0 & 0 & 2 \end{bmatrix}$$

has exactly the same characteristic polynomial, but $Eig(B, 2)$ is 2-dimensional and $Eig(B, 5)$ still 1-dimensional. Thus, the Geometric vs. Algebraic Multiplicities Theorem is also satisfied.□

Example: Let $p(\lambda) = (\lambda + 5)^3 \cdot (\lambda - 2)^4 \cdot (\lambda^2 + 4)$ be the characteristic polynomial of a matrix A. The degree of $p(\lambda)$ is $3 + 4 + 2 = 9$, and thus A must be a 9×9 matrix. The Geometric vs. Algebraic Multiplicity Theorem tells us that:

$$dim(Eig(A, -5)) \leq 3,$$
$$dim(Eig(A, 2)) \leq 4, \text{ and}$$
$$dim(Eig(A, 2i)) = 0 = dim(Eig(A, -2i)).$$

The last two eigenspaces have dimension 0, by convention, because $2i$ and $-2i$ are imaginary numbers. Thus, A is definitely not diagonalizable.□

Thanks to the Geometric vs. Algebraic Multiplicity Theorem, together with our Theorem on the linear independence of eigenvectors from different eigenspaces, we have the following Theorem:

Theorem — The Multiplicity Test for Diagonalizability:

Let A be an $n \times n$ matrix. Then A is ***diagonalizable*** ***if and only if*** for all of its eigenvalues λ_i, the ***geometric multiplicity*** of λ_i is ***exactly equal*** to its ***algebraic multiplicity***.

Idea of the Proof: We will demonstrate the ideas behind the Proof using an 8×8 matrix A. The ideas can be applied in general, and will be outlined in the Exercises. Let us suppose that:

$$p(\lambda) = (\lambda + 3)^2(\lambda - 2)^3(\lambda + 5)(\lambda - 7)^2.$$

Thus, A has exactly four distinct eigenvalues, and four distinct eigenspaces. We also know from ***The Geometric vs. Algebraic Multiplicity Theorem*** that:

$$\begin{aligned} dim(Eig(A,-3)) &\leq 2, \\ dim(Eig(A,2)) &\leq 3, \\ dim(Eig(A,-5)) &= 1, \text{ and} \\ dim(Eig(A,7)) &\leq 2. \end{aligned}$$

Notice that we were able to definitely say that $Eig(A,-5)$ is ***exactly*** 1-dimensional, since we are guaranteed at least one (non-zero!) eigenvector. Now, let us demonstrate both directions of the Theorem:

($\Rightarrow$) Suppose we know that A is diagonalizable. Then there exists a basis $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_8\}$ of 8 eigenvectors for A. Each of these eigenvectors belongs to exactly one of the eigenspaces above. However, since B is linearly independent, ***every subset*** of B is also independent. In order to make all 8 eigenvectors fit into the four subspaces under the restrictions above, exactly 2 have to go to $Eig(A,-3)$, exactly 3 have to go to $Eig(A,2)$, exactly 1 has to go to $Eig(A,-5)$, and exactly 2 have to go to $Eig(A,7)$. Thus, each dimension above is exactly equal to its maximum allowable value, which happens to be the corresponding algebraic multiplicity.

($\Leftarrow$) Suppose that the dimensions of the eigenspaces above are all ***exactly equal*** to the corresponding algebraic multiplicity. Thus, we can construct a basis for each eigenspace:

$$\begin{aligned} \{\vec{v}_1, \vec{v}_2\} &: \text{ a basis for } Eig(A,-3) \\ \{\vec{v}_3, \vec{v}_4, \vec{v}_5\} &: \text{ a basis for } Eig(A,2) \\ \{\vec{v}_6\} &: \text{ a basis for } Eig(A,-5) \\ \{\vec{v}_7, \vec{v}_8\} &: \text{ a basis for } Eig(A,7) \end{aligned}$$

We note that, separately, ***each*** of the bases above must be linearly independent. However, this does not mean that ***taken together***, all 8 vectors are still linearly independent. Unfortunately, ***The Independence of Distinct Eigenspaces Theorem*** only guarantees that a set consisting of ***one*** basis vector from ***each*** eigenspace is ***independent***. For example, the sets:

$$\{\vec{v}_1, \vec{v}_4, \vec{v}_6, \vec{v}_8\} \text{ and } \{\vec{v}_2, \vec{v}_5, \vec{v}_6, \vec{v}_7\}$$

are both independent. Thus, as usual, let us begin the process of proving that the aggregate set of all 8 vectors is independent by writing down the ***dependence test equation:***

$$c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + c_4\vec{v}_4 + c_5\vec{v}_5 + c_6\vec{v}_6 + c_7\vec{v}_7 + c_8\vec{v}_8 = \vec{\mathbf{0}}_8.$$

We want to show that all the coefficients c_i must be zero. To do this, let us separate the terms above into four natural groups:

$$\begin{aligned} \text{Let } \vec{w}_1 &= c_1\vec{v}_1 + c_2\vec{v}_2, \\ \vec{w}_2 &= c_3\vec{v}_3 + c_4\vec{v}_4 + c_5\vec{v}_5, \\ \vec{w}_3 &= c_6\vec{v}_6, \text{ and} \\ \vec{w}_4 &= c_7\vec{v}_7 + c_8\vec{v}_8. \end{aligned}$$

The dependence test equation can now be rewritten into the shorter equation:

$$\vec{w}_1 + \vec{w}_2 + \vec{w}_3 + \vec{w}_4 = \vec{\mathbf{0}}_8,$$

where each of the vectors above is in a ***distinct eigenspace***. We are now in a position to use the power of the Independence of Distinct Eigenspaces Theorem.

We will now use the fact that each of the (separate) bases is linearly ***independent***. Thus, if either c_1 or c_2 is non-zero, then $\vec{w}_1 \neq \vec{0}_8$. Similarly, if either c_3 or c_4 or c_5 is non-zero, then $\vec{w}_2 \neq \vec{0}_8$. If $c_6 \neq 0$, then $\vec{w}_3 \neq \vec{0}_8$. If either c_7 or c_8 is non-zero, then $\vec{w}_4 \neq \vec{0}_8$. In other words, the corresponding $\vec{w}_i$ becomes an actual ***eigenvector***, and not just a member of the eigenspace.

If exactly ***one*** of the $\vec{w}_i$ is non-zero, say $\vec{w}_2 \neq \vec{0}_8$, the test equation becomes $\vec{w}_2 = \vec{0}_8$, which yields a contradiction. If ***two or more*** of the $\vec{w}_i$ are non-zero, then the test equation becomes, for example:

$$\vec{w}_1 + \vec{w}_2 + \vec{w}_4 = \vec{\mathbf{0}}_8, \quad \text{or perhaps}$$
$$\vec{w}_2 + \vec{w}_3 = \vec{\mathbf{0}}_8,$$

where we assume that none of the $\vec{w}_i$ in each equation is $\vec{\mathbf{0}}_8$. But now, the test equation says that the vectors involved are ***dependent***. This is impossible, because they come from ***distinct*** eigenspaces.■

As we saw in one of our Examples above, it is time consuming to have to find a basis for every eigenspace of A, before being able to determine whether or not A is diagonalizable. However, one important and easy consequence of this Theorem is the following:

Theorem: Let A be an $n \times n$ matrix with n ***real, distinct, eigenvalues***. Then A is ***diagonalizable***.

Proof: Suppose that $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the n distinct eigenvalues of A. We know that each eigenvalue λ_i has at least one (non-zero) eigenvector, say $\vec{v}_i$. Thus, $dim(Eig(A, \lambda_i)) \geq 1$ for all λ_i. But since there are n distinct real eigenvalues, the algebraic multiplicity of each eigenvalue is 1. Since the geometric multiplicity is ***at most equal*** to the algebraic multiplicity, we have:

$$1 \leq dim(Eig(A, \lambda_i)) \leq 1,$$

and thus each geometric multiplicity must be ***exactly*** 1. By the previous Theorem, A is diagonalizable.■

Example: Let $A = \begin{bmatrix} -1 & -2 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$.

We saw in the previous Section that the characteristic polynomial of this matrix is:

$$p(\lambda) = \lambda(\lambda - 1)(\lambda - 3),$$

and the eigenvalues are $\lambda = 0,\ 1$ and 3. Even though this matrix is ***not invertible*** because 0 is an eigenvalue, it is ***diagonalizable*** because we have three distinct eigenvalues for this 3×3 matrix. To find the eigenvectors, we must compute the matrices $A - \lambda \boldsymbol{I}_3$ and their rrefs, resulting in:

$$\text{For } \lambda = 0: \begin{bmatrix} -1 & -2 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & -5/2 \\ 0 & 0 & 0 \end{bmatrix};$$

$$\text{For } \lambda = 1: \begin{bmatrix} -2 & -2 & 2 \\ 2 & 1 & 1 \\ 1 & 0 & 2 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & -3 \\ 0 & 0 & 0 \end{bmatrix};$$

$$\text{For } \lambda = 3: \begin{bmatrix} -4 & -2 & 2 \\ 2 & -1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

The eigenspaces are thus:

$$Eig(A, 0) = Span(\{\langle -6, 5, 2 \rangle\}),$$
$$Eig(A, 1) = Span(\{\langle -2, 3, 1 \rangle\}), \quad \text{and}$$
$$Eig(A, 3) = Span(\{\langle 0, 1, 1 \rangle\}).$$

We assemble the basis vectors into the columns of C:

$$C = \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \text{ with inverse } C^{-1} = \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix}.$$

The diagonalizing matrix is $D = Diag(0, 1, 3)$. We verify that:

$$CDC^{-1} = \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -2 & 0 \\ 0 & 3 & 3 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix} = \begin{bmatrix} -1 & -2 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} = A.$$

as expected.□

Powers of Diagonalizable Matrices

One useful application of diagonalization is the ability to find powers of a matrix without a lot of effort. Suppose we have the factorization $A = CDC^{-1}$, then:

$$\begin{aligned} A^2 &= (CDC^{-1})(CDC^{-1}) \\ &= CD(C^{-1}C)DC^{-1} \\ &= CD^2C^{-1}, \end{aligned}$$

and proceeding by induction, we get:

Theorem: Let A be a ***diagonalizable*** $n \times n$ matrix with $A = CDC^{-1}$. Then: for all positive integers k:

$$A^k = CD^kC^{-1}.$$

Furthermore, if A is ***invertible***, then:

$$A^{-1} = CD^{-1}C^{-1}.$$

Since D^k is easy to compute for a diagonal matrix D, this gives us an easy way to compute A^k indirectly. The formula for A^{-1} will be proven in the Exercises.

Example: We saw in the previous Example where A can be diagonalized as:

$$A = \begin{bmatrix} -1 & -2 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix} = \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix}.$$

Thus, if we want the 8th power of this matrix, we get:

$$\begin{aligned} A^8 &= \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}^8 \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix} \\ &= \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0^8 & 0 & 0 \\ 0 & 1^8 & 0 \\ 0 & 0 & 3^8 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix} \\ &= \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6561 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix} \\ &= \begin{bmatrix} 0 & -2 & 0 \\ 0 & 3 & 6561 \\ 0 & 1 & 6561 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix} = \begin{bmatrix} -1 & -2 & 2 \\ 1095 & -2184 & 8745 \\ 1094 & -2186 & 8747 \end{bmatrix}. \square \end{aligned}$$

6.3 Section Summary

Let A be an $n \times n$ matrix. We say that A is ***diagonalizable*** if we can find an ***invertible*** matrix C such that $C^{-1}AC = D$, where $D = Diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$ is a diagonal matrix, or equivalently:

$$A = CDC^{-1} \quad \text{or} \quad AC = CD$$

We also say that C ***diagonalizes*** A.

The Basis Test for Diagonalizability: A is diagonalizable ***if and only if*** we can find a set of n ***linearly independent eigenvectors*** for A, say $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ If this is the case, then the diagonalizing matrix C is the matrix whose ***columns*** are $\vec{v}_1$, $\vec{v}_2$, $\ldots$, $\vec{v}_n$, and the diagonal matrix D contains the corresponding ***eigenvalues*** along the main ***diagonal***.

Let A have ***imaginary*** eigenvalues. Then A is ***not*** diagonalizable over the set of ***real*** invertible matrices.

Independence of Distinct Eigenspaces: Suppose that λ_1, λ_2, $\ldots$, λ_k are ***distinct*** eigenvalues for an $n \times n$ matrix A, and suppose that $\vec{v}_i$ is an eigenvector of A corresponding to λ_i, for $i = 1..k$. Then: the set $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is ***linearly independent***. Thus, if A has a total of m distinct eigenvalues, we can find ***at least m*** linearly independent eigenvectors for A.

Let A have ***distinct*** (possibly imaginary) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$.

Algebraic and Geometric Multiplicities: Suppose $p(\lambda)$ factors as:

$$p(\lambda) = (\lambda - \lambda_1)^{n_1} \cdot (\lambda - \lambda_2)^{n_2} \cdot \cdots \cdot (\lambda - \lambda_k)^{n_k}.$$

We call the exponent n_i the ***algebraic multiplicity*** of λ_i.

We call $dim(Eig(A, \lambda_i))$ the ***geometric multiplicity*** of λ_i, where we agree that $dim(Eig(A, \lambda_i)) = 0$ if λ_i is an imaginary eigenvalue.

The Geometric vs. Algebraic Multiplicity Theorem: For any eigenvalue λ_i of A, the ***geometric multiplicity*** of λ_i is ***at most*** equal to the ***algebraic multiplicity*** of λ_i.

The Multiplicity Test for Diagonalizability: A is ***diagonalizable*** ***if and only if*** for all of its eigenvalues λ_i, the geometric multiplicity of λ_i is ***exactly equal*** to its algebraic multiplicity.

If A has n ***real, distinct eigenvalues***, then A is ***diagonalizable***.

If $A = CDC^{-1}$, then $A^k = CD^kC^{-1}$, for any positive integer k.

6.3 Exercises

For Exercises (1) to (24): Diagonalize the matrix in the indicated Exercise, that is, find an invertible matrix C and a diagonal matrix D such that $A = CDC^{-1}$, or explain why it is impossible to do so. You need not find C^{-1}, at least for now. You can use the answers to Sections 6.1 and 6.2 in the Key.

1. Section 6.1, Exercise 1.
2. Section 6.1, Exercise 4.
3. Section 6.1, Exercise 5.
4. Section 6.1, Exercise 13.
5. Section 6.1, Exercise 7.
6. Section 6.1, Exercise 14.

7. Section 6.1, Exercise 10.
8. Section 6.2, Exercise 1.
9. Section 6.2, Exercise 6.
10. Section 6.2, Exercise 7.
11. Section 6.2, Exercise 8.
12. Section 6.2, Exercise 10.
13. Section 6.2, Exercise 11.
14. Section 6.2, Exercise 12.
15. Section 6.2, Exercise 13.
16. Section 6.2, Exercise 14.
17. Section 6.2, Exercise 15.
18. Section 6.2, Exercise 16.
19. Section 6.2, Exercise 22.
20. Section 6.2, Exercise 24.
21. Section 6.2, Exercise 25.
22. Section 6.2, Exercise 26.
23. Section 6.2, Exercise 27.
24. Section 6.2, Exercise 29.

For Exercises (25) to (32): Decide which of the matrices in each Exercise, if any, is diagonalizable. You may use the answers in the Key to Section 6.1. There are no further computations necessary.

25. The two matrices in Exercise 16, Section 6.1.
26. The two matrices in Exercise 17, Section 6.1.
27. The two matrices in Exercise 18, Section 6.1.
28. The three matrices in Exercise 19, Section 6.1.
29. The three matrices in Exercise 20, Section 6.1.
30. The three matrices in Exercise 21, Section 6.1.
31. The two matrices in Exercise 22, Section 6.1.
32. The two matrices in Exercise 23, Section 6.1.

For Exercises (33) to (52): Use the diagonal matrix D and diagonalizing matrix C to find the 5th power of the following matrices. You will need to find C^{-1} for these matrices. Use technology, if allowed by your instructor, to find C^{-1}.

33. The matrix in Exercise 2 above.
34. The matrix in Exercise 4 above.
35. The matrix in Exercise 5 above.
36. The matrix in Exercise 7 above.
37. The matrix in Exercise 8 above.
38. The matrix in Exercise 12 above.
39. The matrix in Exercise 14 above.
40. The matrix in Exercise 15 above.

41. The matrix in Exercise 18 above.
42. The matrix in Exercise 19 above.
43. The matrix in Exercise 20 above.
44. The matrix in Exercise 21 above.
45. The matrix in Exercise 23 above.
46. The matrix in Exercise 24 above.
47. The matrix in Exercise 16 (a), Section 6.1.
48. The matrix in Exercise 17 (a), Section 6.1.
49. The matrix in Exercise 18 (a), Section 6.1.
50. The matrix in Exercise 19 (b), Section 6.1.
51. The matrix in Exercise 20 (c), Section 6.1.
52. The matrix in Exercise 21 (b), Section 6.1.
53. Suppose that A is a 9×9 matrix, and:

$$\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} \text{ is a basis for } Eig(A,-2), \quad \{\vec{v}_4, \vec{v}_5\} \text{ is a basis for } Eig(A,3), \text{ and}$$
$$\{\vec{v}_6, \vec{v}_7, \vec{v}_8, \vec{v}_9\} \text{ is a basis for } Eig(A,7).$$

Prove that the set $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6, \vec{v}_7, \vec{v}_8, \vec{v}_9\}$ is linearly independent, and therefore, A is diagonalizable.

54. Suppose that A is a 10×10 matrix, and:

$$\{\vec{v}_1, \vec{v}_2\} \text{ is a basis for } Eig(A,-4), \qquad \{\vec{v}_3, \vec{v}_4, \vec{v}_5\} \text{ is a basis for } Eig(A,-1),$$
$$\{\vec{v}_6, \vec{v}_7, \vec{v}_8\} \text{ is a basis for } Eig(A,2), \text{ and } \quad \{\vec{v}_9, \vec{v}_{10}\} \text{ is a basis for } Eig(A,5).$$

Prove that the set $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{v}_6, \vec{v}_7, \vec{v}_8, \vec{v}_9, \vec{v}_{10}\}$ is linearly independent, and therefore, A is diagonalizable.

55. We will generalize the two previous Exercises: let A be an $n \times n$ matrix with distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Our goal in this Exercise is to prove that:

$$A \text{ is } \textbf{\textit{diagonalizable}} \textbf{\textit{ if and only if }} \sum_{i=1}^{k} dim(Eig(A, \lambda_i)) = n.$$

In other words, the dimensions of the eigenspaces of A must add up to n.

a. Suppose that A is ***diagonalizable***. Use the linearly independent set of n eigenvectors guaranteed by the main Theorem on diagonalizability to show that the dimensions of the eigenspaces of A must add up to n. The fact that distinct eigenspaces intersect only in the zero vector is also useful.

b. Now, suppose that the dimensions of the eigenspaces of A add up to n. Let us construct a ***basis*** for each eigenspace (using double-subscript notation):

$$\{\vec{v}_{1,1}, \vec{v}_{1,2}, \ldots, \vec{v}_{1,n_1}\}, \text{ a basis for } Eig(A, \lambda_1);$$
$$\{\vec{v}_{2,1}, \vec{v}_{2,2}, \ldots, \vec{v}_{2,n_2}\}, \text{ a basis for } Eig(A, \lambda_2); \ldots$$
$$\{\vec{v}_{k,1}, \vec{v}_{k,2}, \ldots, \vec{v}_{k,n_k}\}, \text{ a basis for } Eig(A, \lambda_k),$$

where $n_1 + n_2 + \cdots + n_k = n$. Prove that the aggregate set of basis vectors:

$$S = \{\vec{v}_{1,1}, \vec{v}_{1,2}, \ldots, \vec{v}_{1,n_1}, \vec{v}_{2,1}, \vec{v}_{2,2}, \ldots, \vec{v}_{2,n_2}, \ldots, \vec{v}_{k,1}, \vec{v}_{k,2}, \ldots, \vec{v}_{k,n_k}\}$$

is itself linearly ***independent***, and consequently A is diagonalizable by the Basis Test for Diagonalizability. Hint: write down a dependence equation for S and use the fact that a set of eigenvectors from distinct eigenspaces is linearly independent.

56. Let A be an $n \times n$ matrix. Prove that A is diagonalizable ***if and only if*** for all of its eigenvalues λ_i, the geometric multiplicity of λ_i is ***exactly equal*** to its algebraic multiplicity. Hint: use the Theorem from Algebraic Geometry that the geometric multiplicity is always at most equal to the algebraic multiplicity, and use the previous Exercise.

57. Let A be an $n \times n$ matrix, and $T : \mathbb{R}^n \to \mathbb{R}^n$ the corresponding operator such that $[T] = A$. Prove that A is diagonalizable ***if and only if*** there exists a diagonal matrix D and a basis B for $\mathbb{R}^n$ such that $[T]_B = D$. Note that $[T]$ is the standard matrix for T, whereas $[T]_B$ is the matrix of T with respect to the basis B.

58. Prove that if A is ***invertible*** and ***diagonalizable***, with $A = CDC^{-1}$, then $A^{-1} = CD^{-1}C^{-1}$. As part of your proof, explain why D must also be invertible.

59. Let A be an ***invertible*** and ***diagonalizable*** $n \times n$ matrix. Prove that if the characteristic polynomial of A factors as:
$$p_A(\lambda) = (\lambda - \lambda_1)^{n_1} \cdot (\lambda - \lambda_2)^{n_2} \cdot \cdots \cdot (\lambda - \lambda_k)^{n_k},$$
where the λ_i are unique, then the characteristic polynomial of A^{-1} factors as:
$$p_{A^{-1}}(\lambda) = \left(\lambda - \frac{1}{\lambda_1}\right)^{n_1} \cdot \left(\lambda - \frac{1}{\lambda_2}\right)^{n_2} \cdot \cdots \cdot \left(\lambda - \frac{1}{\lambda_k}\right)^{n_k}.$$
Hint: use the Exercise in Section 6.2 on the eigentheory of an invertible matrix. As part of your proof, show that $Eig(A, \lambda_i) = Eig(A^{-1}, \lambda_i^{-1})$ for all i.

60. Let $D = Diag(c_1, c_2, \ldots, c_n)$ be a diagonal matrix. We know that the eigenvalues of D are c_1 through c_n.
    a. Show that the standard basis vector $\vec{e}_i$ is an eigenvector for $\lambda = c_i$ for every $i = 1 \ldots n$.
    b. Show that if the c_i are all ***distinct***, then every eigenvector of c_i is a non-zero multiple of $\vec{e}_i$.
    c. Let $D = Diag(3, -2, 3)$. Give an example of an eigenvector for $\lambda = 3$ which is ***not*** a multiple of a standard basis vector.

61. ***True or False:*** Determine if the statement is true or false. If the statement is true, cite a definition or Theorem that supports your conclusion, or give a convincing argument why the statement is true. If the statement is false, cite a definition or Theorem that supports your conclusion, provide a counterexample, or give a convincing argument why the statement is false.
    a. An invertible matrix is automatically diagonalizable.
    b. A diagonalizable matrix is automatically invertible.
    c. If all of the eigenvalues of a matrix are real, then the matrix is diagonalizable.
    d. If a matrix A has characteristic polynomial $p(\lambda) = \lambda^4 - 29\lambda^2 + 100$, then A has to be diagonalizable.
    e. If a matrix A has characteristic polynomial $p(\lambda) = (\lambda + 9)^2(\lambda - 5)$, then A is ***not*** diagonalizable.
    f. If a matrix A has characteristic polynomial $p(\lambda) = (\lambda^2 + 9)(\lambda - 5)$, then A is ***not*** diagonalizable. Why is this different from (e)?
    g. A triangular matrix is automatically diagonalizable.
    h. A diagonal matrix is automatically diagonalizable. (This isn't as obvious as it looks!)
    i. We can diagonalize a triangular matrix whose entries on the main diagonal are all distinct.
    j. If a matrix is diagonalizable, then all the eigenspaces have dimension 1.

6.4 The Exponential of a Matrix

In Pre-Calculus, we encounter the natural exponential function, e^x, where Euler's number e is:

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n.$$

Although we don't fully understand the concept of a limit at this level, we can see by experiment that as n gets bigger and bigger, the expression above ***converges*** or gets closer and closer to a familiar number:

$$\left(1+\frac{1}{100}\right)^{100} = 1.01^{100} \approx 2.7048,$$

$$\left(1+\frac{1}{10000}\right)^{100000} = 1.00001^{100000} \approx 2.7183,$$

and so on. The graph of e^x is of course strictly ***increasing*** over the real numbers, so we can define its ***inverse***, the natural logarithmic function or $\ln(x)$. Later, we reverse this chronology by using the Fundamental Theorem of Calculus to first define $\ln(x)$ as the definite integral:

$$\ln(x) = \int_1^x \frac{1}{t}dt, \quad \text{where } x > 0,$$

and define e^x as the inverse of this one-to-one function. However, it is not until we get to the study of ***Taylor*** and ***Maclaurin Series*** that we are able to compute as many digits to e as we want, using the famous formula:

$$e^x = \sum_{n=0}^{\infty} \frac{1}{n!}x^n, \quad \text{so in particular,}$$

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 1+1+\frac{1}{2}+\frac{1}{6}+\frac{1}{24}+\frac{1}{120}+\cdots \approx 2.718281828\ldots$$

We will use this Maclaurin series to extend the operation of exponentiation to square matrices:

Definition — The Exponential of a Square Matrix:
Let A be an $m \times m$ matrix. We define the ***exponential*** of A, denoted e^A, by the infinite series:

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!}A^n = \boldsymbol{I}_m + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \frac{1}{24}A^4 + \cdots$$

This definition seems easy and straightforward enough, but there are certainly some issues that we have to deal with. First, we know how to compute a linear combination of matrices, but what is the meaning of an ***infinite series*** of matrices? We can answer this, as usual, by taking a limit. We can compute the sequence of ***partial sums:***

$$\{S_k(A)\}_{k=1}^{\infty}, \quad \text{where} \quad S_k(A) = \sum_{n=0}^{k} \frac{1}{n!}A^n.$$

These are just polynomial evaluations of A, so they certainly exist. The next issue is now: what does it mean for a sequence of matrices to converge to a ***limit matrix*** B? We can define it in the natural way: we will say that:

$$\lim_{k\to\infty} S_k(A) = B$$

if the entry in row i, column j of $S_k(A)$ converges to the corresponding entry $B_{i,j}$.

Although these definitions are of course precise and quantifiable, they do not give us an easy way to compute e^A for an arbitrary matrix A, or even allow us to know if we have a reasonable approximation for e^A. Fortunately, we will be sidestepping these technical issues by focusing on the case when A is a ***diagonalizable*** matrix. As our next step, let us see what happens when A is itself a diagonal matrix:

Theorem: Suppose that $D = Diag(d_1, d_2, \ldots, d_m)$ is a ***diagonal*** matrix. Then:

$$e^D = Diag(e^{d_1}, e^{d_2}, \ldots, e^{d_m}).$$

Proof: We know from Section 2.9 that for a diagonal matrix in the notation above:

$$\begin{aligned} D^n &= Diag(d_1^n, d_2^n, \ldots, d_m^n), \text{ and so} \\ \frac{1}{n!}D^n &= Diag\left(\frac{1}{n!}d_1^n, \frac{1}{n!}d_2^n, \ldots, \frac{1}{n!}d_m^n\right), \text{ and} \\ \sum_{n=0}^{\infty}\frac{1}{n!}D^n &= Diag\left(\sum_{n=0}^{\infty}\frac{1}{n!}d_1^n, \sum_{n=0}^{\infty}\frac{1}{n!}d_2^n, \ldots, \sum_{n=0}^{\infty}\frac{1}{n!}d_m^n\right) \\ &= Diag(e^{d_1}, e^{d_2}, \ldots, e^{d_m}). \blacksquare \end{aligned}$$

This now gives us an easy way to compute the exponential of a diagonalizable matrix:

Theorem: Suppose that A is an $m \times m$ ***diagonalizable*** matrix, with $A = CDC^{-1}$, for some invertible matrix C and diagonal matrix $D = Diag(d_1, d_2, \ldots, d_m)$. Then:

$$e^A = C \cdot e^D \cdot C^{-1} = C \cdot Diag(e^{d_1}, e^{d_2}, \ldots, e^{d_m}) \cdot C^{-1}.$$

In particular, if t is a real variable, we have:

$$e^{tA} = C \cdot e^{Dt} \cdot C^{-1} = C \cdot Diag(e^{d_1 t}, e^{d_2 t}, \ldots, e^{d_m t}) \cdot C^{-1}.$$

Proof: We know from Section 6.3 that in the notation above:

$$\begin{aligned} A^n &= CD^nC^{-1}, \text{ and so} \\ \frac{1}{n!}A^n &= \frac{1}{n!}CD^nC^{-1}, \text{ and} \\ \sum_{n=1}^{\infty}\frac{1}{n!}A^n &= \sum_{n=1}^{\infty}\frac{1}{n!}CD^nC^{-1} = C \cdot \left(\sum_{n=1}^{\infty}\frac{1}{n!}D^n\right) \cdot C^{-1} = C \cdot e^D \cdot C^{-1}. \blacksquare \end{aligned}$$

The expression e^{tA} is important in a course in ***Differential Equations***. In this case, the matrix A usually represents the coefficients of a ***system*** of linear ordinary differential equations in the variable t, with a given initial or boundary condition. The expression e^{tA} appears in the ***solution*** to this system.

Example: Consider the matrix from Section 6.3: $A = \begin{bmatrix} -1 & -2 & 2 \\ 2 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$.

We saw that A is diagonalizable, with:

$$D = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad C = \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix}, \quad \text{and} \quad C^{-1} = \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix}.$$

Thus, we can find the matrix exponential:

$$e^A = C \cdot e^D \cdot C^{-1} = \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} e^0 & 0 & 0 \\ 0 & e^1 & 0 \\ 0 & 0 & e^3 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix}$$

$$= \begin{bmatrix} -e+2 & -2e+2 & 2e-2 \\ \frac{3}{2}e + \frac{1}{6}e^3 - \frac{5}{3} & 3e - \frac{1}{3}e^3 - \frac{5}{3} & -3e + \frac{4}{3}e^3 + \frac{5}{3} \\ \frac{1}{2}e + \frac{1}{6}e^3 - \frac{2}{3} & e - \frac{1}{3}e^3 - \frac{2}{3} & -e + \frac{4}{3}e^3 + \frac{2}{3} \end{bmatrix}.$$

If we wanted e^{tA}, we would get:

$$e^{tA} = C \cdot e^{Dt} \cdot C^{-1} = \begin{bmatrix} -6 & -2 & 0 \\ 5 & 3 & 1 \\ 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} e^0 & 0 & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{3t} \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ 1/2 & 1 & -1 \\ 1/6 & -1/3 & 4/3 \end{bmatrix}$$

$$= \begin{bmatrix} -e^t+2 & -2e^t+2 & 2e^t-2 \\ \frac{3}{2}e^t + \frac{1}{6}e^{3t} - \frac{5}{3} & 3e^t - \frac{1}{3}e^{3t} - \frac{5}{3} & -3e^t + \frac{4}{3}e^{3t} + \frac{5}{3} \\ \frac{1}{2}e^t + \frac{1}{6}e^{3t} - \frac{2}{3} & e^t - \frac{1}{3}e^{3t} - \frac{2}{3} & -e^t + \frac{4}{3}e^{3t} + \frac{2}{3} \end{bmatrix}. \square$$

Notice that we get e^A by substituting $t = 1$ into our answer for e^{tA}.

6.4 Section Summary

Let A be an $m \times m$ matrix. We define the ***exponential*** of A, denoted e^A, by the infinite series:

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I_m + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \frac{1}{24}A^4 + \cdots$$

If $D = Diag(d_1, d_2, \ldots, d_m)$ is a ***diagonal*** matrix, then:

$$e^D = Diag(e^{d_1}, e^{d_2}, \ldots, e^{d_m}).$$

Suppose that A is an $m \times m$ ***diagonalizable*** matrix, with $A = CDC^{-1}$, for some invertible matrix C and diagonal matrix $D = Diag(d_1, d_2, \ldots, d_m)$. Then:

$$e^A = Ce^D C^{-1} = C \cdot Diag(e^{d_1}, e^{d_2}, \ldots, e^{d_m}) \cdot C^{-1}.$$

In particular, if t is a real variable, we have:

$$e^{tA} = Ce^{Dt} C^{-1} = C \cdot Diag(e^{d_1 t}, e^{d_2 t}, \ldots, e^{d_m t}) \cdot C^{-1}.$$

6.4 Exercises

For Exercises (1) to (20): Find e^A and e^{tA} for the following matrices A, which were diagonalized in Exercises (33) to (52) of Section 6.3. You may use the matrices C and D which are found in the Answer Key. For your reference, the item also indicates the Exercise where the matrix originally appears (in either Section 6.1 or 6.2).

1. Exercise 33. (see Exercise 4, Section 6.1)
2. Exercise 34. (see Exercise 13, Section 6.1)
3. Exercise 35. (see Exercise 7, Section 6.1)
4. Exercise 36. (see Exercise 10, Section 6.1)
5. Exercise 37. (see Exercise 1, Section 6.2)
6. Exercise 38. (see Exercise 10, Section 6.2)
7. Exercise 39. (see Exercise 12, Section 6.2)
8. Exercise 40. (see Exercise 13, Section 6.2)
9. Exercise 41. (see Exercise 16, Section 6.2)
10. Exercise 42. (see Exercise 22, Section 6.2)
11. Exercise 43. (see Exercise 24, Section 6.2)
12. Exercise 44. (see Exercise 25, Section 6.2)
13. Exercise 45. (see Exercise 27, Section 6.2)
14. Exercise 46. (see Exercise 29, Section 6.2)
15. Exercise 47. (see Exercise 16 (a), Section 6.1)
16. Exercise 48. (see Exercise 17 (a), Section 6.1)
17. Exercise 49. (see Exercise 18 (a), Section 6.1)
18. Exercise 50. (see Exercise 19 (b), Section 6.1)
19. Exercise 51. (see Exercise 20 (c), Section 6.1)
20. Exercise 52. (see Exercise 21 (b), Section 6.1)

6.5 Change of Basis and Linear Transformations on Euclidean Spaces

In this Section, we will find out how to relate the ***coordinates*** of the same vector $\vec{v}$ from $\mathbb{R}^n$, with respect to two different bases for $\mathbb{R}^n$. We will use this relationship in order to construct different matrices for the same linear transformations $T : \mathbb{R}^n \to \mathbb{R}^m$. These ideas will be generalized in Section 6.6 for a general linear transformation $T : V \to W$.

The Change of Basis Matrix

If B and B' are two bases for a vector space V, let us investigate the relationship between the two coordinate matrices $[\vec{v}]_B$ and $[\vec{v}]_{B'}$. We will focus on $\mathbb{R}^n$ for now, but we will consider the general case in the next Section. The relationship for coordinates in $\mathbb{R}^n$ is given as follows:

Theorem/Definition: For any bases B and B' for $\mathbb{R}^n$, there exists an ***invertible*** $n \times n$ matrix $C_{B,B'}$, which depends only on B and B', such that for all vectors $\vec{v}$ of $\mathbb{R}^n$:

$$[\vec{v}]_{B'} = C_{B,B'}[\vec{v}]_B.$$

The matrix $C_{B,B'}$ is called the ***change of basis matrix*** from B to B'. We can explicitly find $C_{B,B'}$ by performing the Gauss-Jordan algorithm on the augmented matrix:

$$\left[B' \mid B\right],$$

where this notation means that we assemble as the columns of this matrix the vectors in B', followed by the vectors of B. At the end of the process, we will obtain:

$$\left[\boldsymbol{I_n} \mid C_{B,B'}\right].$$

Consequently, $C_{B,B'} = [B']^{-1}[B]$, and the columns of $C_{B,B'}$ are the ***coordinate matrices*** of the members of B with respect to B'. In other words, if $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then:

$$C_{B,B'} = [B']^{-1}[B] = \left[\, [\vec{v}_1]_{B'}\ [\vec{v}_2]_{B'} \cdots [\vec{v}_n]_{B'} \,\right].$$

Furthermore, $C_{B,B'}$ is ***invertible***, and:

$$C_{B,B'}^{-1} = C_{B',B} = [B]^{-1}[B'].$$

Proof: In keeping with the notation above, we denote by $[B]$ (pronounced ***the matrix of B***) the $n \times n$ matrix whose columns are the vectors of $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$:

$$[B] = [\vec{w}_1\, \vec{w}_2 \cdots \vec{w}_n\,],$$

and similarly for $[B']$. Both $[B]$ and $[B']$ are ***invertible*** matrices since B and B' are both ***bases*** for $\mathbb{R}^n$. Now, for any vector $\vec{v} \in \mathbb{R}^n$, we have:

$$\vec{v} = [B][\vec{v}]_B, \quad \text{and} \quad \vec{v} = [B'][\vec{v}]_{B'},$$

which follow directly from the definition of a matrix product $A\vec{x}$ as a linear combination of the ***columns*** of A with coefficients from $\vec{x}$. Thus:

$$[B'][\vec{v}]_{B'} = [B][\vec{v}]_B,$$

and therefore:

$$[\vec{v}]_{B'} = [B']^{-1}[B][\vec{v}]_B.$$

Notice that this equation does ***not*** depend on $\vec{v}$. Thus, we get:

$$C_{B,B'} = [B']^{-1}[B],$$

which is an ***invertible*** matrix, since it is the product of two invertible matrices. The recipe to compute $C_{B,B'}$ by performing the Gauss-Jordan Algorithm on:

$$\left[B' \mid B\right],$$

now follows from the Proof in Section 2.7 regarding the computation of the ***inverse*** of a matrix, except this time $[B]$ is on the right side of the augmented matrix instead of $\boldsymbol{I_n}$. Recall that the idea is that every elementary row operation corresponds to an elementary matrix E. Thus, if:

$$E_n \bullet \cdots \bullet E_2 \bullet E_1 \bullet [B'] = \boldsymbol{I_n},$$

then:

$$[B']^{-1} = E_n \bullet \cdots \bullet E_2 \bullet E_1.$$

Therefore, at the end of the Gauss-Jordan Algorithm applied to $\left[B' \mid B\right]$, we obtain:

$$\left[\boldsymbol{I_n} \mid E_n \bullet \cdots \bullet E_2 \bullet E_1 \bullet [B]\right],$$

and the right side is thus $[B']^{-1}[B]$, which is exactly what we want. The formula for $C_{B,B'}$ now follows because the solution to the augmented system $\left[B' \mid B\right]$ is the coordinate matrix of each member of B with respect to B'. Finally, we get $C_{B,B'}^{-1} = C_{B',B}$ by reversing the roles of B and B' in this proof.

Example: Consider the two bases for $\mathbb{R}^3$:

$$B = \{\langle 1, 0, 1\rangle, \langle -1, 1, 0\rangle, \langle 0, -1, 1\rangle\}, \quad \text{and}$$
$$B' = \{\langle 1, -1, -2\rangle, \langle 0, -1, -3\rangle, \langle 0, 0, 2\rangle\}.$$

Let us find $\langle \vec{v}\rangle_B$ and $\langle \vec{v}\rangle_{B'}$ for $\vec{v} = \langle -4, 8, -2\rangle$.

We perform the Gauss-Jordan Algorithm on the augmented matrices:

$$\left[\begin{array}{ccc|c} 1 & -1 & 0 & -4 \\ 0 & 1 & -1 & 8 \\ 1 & 0 & 1 & -2 \end{array}\right] \text{ and } \left[\begin{array}{ccc|c} 1 & 0 & 0 & -4 \\ -1 & -1 & 0 & 8 \\ -2 & -3 & 2 & -2 \end{array}\right].$$

The rrefs of these two matrices are:

$$\left[\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & -3 \end{array}\right] \text{ and } \left[\begin{array}{ccc|c} 1 & 0 & 0 & -4 \\ 0 & 1 & 0 & -4 \\ 0 & 0 & 1 & -11 \end{array}\right].$$

Thus $\langle \vec{v}\rangle_B = \langle 1, 5, -3\rangle$ and $\langle \vec{v}\rangle_{B'} = \langle -4, -4, -11\rangle$. Notice that one desirable by-product of this computation is a verification that the sets B and B' are indeed bases, because the rrefs have $\boldsymbol{I_3}$ in the first three columns.

Let us find the change of basis matrix from B to B'. We form the augmented matrix:

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & -1 & 0 \\ -1 & -1 & 0 & 0 & 1 & -1 \\ -2 & -3 & 2 & 1 & 0 & 1 \end{array}\right].$$

The rref of this augmented matrix is:

$$\left[\begin{array}{rrr|rrr} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 & 2 \end{array}\right].$$

Thus $C_{B,B'} = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 2 \end{bmatrix}$. We can verify the formula:

$$\begin{aligned} [\vec{v}]_{B'} &= C_{B,B'}[\vec{v}]_B \\ &= \begin{bmatrix} 1 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 5 \\ -3 \end{bmatrix} = \begin{bmatrix} -4 \\ -4 \\ -11 \end{bmatrix}. \end{aligned}$$

Therefore, $\langle \vec{v} \rangle_{B'} = \langle -4, -4, -11 \rangle$, just as we computed.□

Change of Basis for Span(S)

We saw in Section 1.9 that if $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\} \subset \mathbb{R}^{n}$, and $W = Span(S)$, we can find a basis for W in two different ways. If $A = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_k \end{bmatrix}$, with the vectors in S assembled into the ***columns*** of A, then, ***The Minimizing Theorem*** tells us that the columns of A corresponding to the leading 1's in the rref R of A form a basis B for W. However, if R' is the rref of $A^{\top}$, whose ***rows*** are the vectors in S, then the non-zero rows in the rref R' of C also form a basis B' for W. Let us see how to construct the change of basis matrices in this situation.

Example: Consider the set:

$$S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5\} = \{\langle 1, 2, -3, 1, -6, -5 \rangle, \langle -3, -4, -1, -2, 9, 1 \rangle, \langle 3, 2, 11, 1, 0, 13 \rangle,$$
$$\langle 4, 7, -7, 3, -17, -12 \rangle, \langle 4, 18, -62, 7, -59, -86 \rangle\}$$

and consider $W = Span(S) \trianglelefteq \mathbb{R}^6$. Assembling the vectors in S into ***columns***, we get:

$$A = \begin{bmatrix} 1 & -3 & 3 & 4 & 4 \\ 2 & -4 & 2 & 7 & 18 \\ -3 & -1 & 11 & -7 & -62 \\ 1 & -2 & 1 & 3 & 7 \\ -6 & 9 & 0 & -17 & -59 \\ -5 & 1 & 13 & -12 & -86 \end{bmatrix}, \quad \text{with rref} \quad R = \begin{bmatrix} 1 & 0 & -3 & 0 & 9 \\ 0 & 1 & -2 & 0 & 7 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, W is 3-dimensional, and columns 1, 2 and 4 form a basis B for W:

$$B = \{\vec{v}_1, \vec{v}_2, \vec{v}_4\} = \{\langle 1, 2, -3, 1, -6, -5 \rangle, \langle -3, -4, -1, -2, 9, 1 \rangle, \langle 4, 7, -7, 3, -17, -12 \rangle\}.$$

We can also see from R the coordinates of the rejected vectors with respect to B:

$$\begin{aligned} \vec{v}_3 &= -3\vec{v}_1 - 2\vec{v}_2, \quad \text{and} \\ \vec{v}_5 &= 9\vec{v}_1 + 7\vec{v}_2 + 4\vec{v}_4, \end{aligned}$$

so $\langle \vec{v}_3 \rangle_B = \langle -3, -2, 0 \rangle$, and $\langle \vec{v}_5 \rangle_B = \langle 9, 7, 4 \rangle$.

Now, transposing A, so that the $\vec{v}_1$ through $\vec{v}_5$ are now in ***rows***, we get:

$$A^\top = \begin{bmatrix} 1 & 2 & -3 & 1 & -6 & -5 \\ -3 & -4 & -1 & -2 & 9 & 1 \\ 3 & 2 & 11 & 1 & 0 & 13 \\ 4 & 7 & -7 & 3 & -17 & -12 \\ 4 & 18 & -62 & 7 & -59 & -86 \end{bmatrix}, \quad \text{with rref } R' = \begin{bmatrix} 1 & 0 & 7 & 0 & 3 & 9 \\ 0 & 1 & -5 & 0 & -2 & -6 \\ 0 & 0 & 0 & 1 & -5 & -2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Thus, our second basis for W is made of the three non-zero rows of R':

$$B' = \{\langle 1, 0, 7, 0, 3, 9 \rangle, \langle 0, 1, -5, 0, -2, -6 \rangle, \langle 0, 0, 0, 1, -5, -2 \rangle\} = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}.$$

This time, to construct $C_{B,B'}$, we will use the alternative formula where the columns of $C_{B,B'}$ are the coordinates of the vectors in $B = \{\vec{v}_1, \vec{v}_2, \vec{v}_4\}$ with respect to our second basis $B' = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$. We can easily do this by looking at the 1st, 2nd and 4th components of these vectors:

$$\begin{aligned} \vec{v}_1 &= \left\langle \boxed{1}, \boxed{2}, -3, \boxed{1}, -6, -5 \right\rangle = \vec{w}_1 + 2\vec{w}_2 + \vec{w}_3, \\ \vec{v}_2 &= \left\langle \boxed{-3}, \boxed{-4}, -1, \boxed{-2}, 9, 1 \right\rangle = -3\vec{w}_1 - 4\vec{w}_2 - 2\vec{w}_3, \quad \text{and} \\ \vec{v}_4 &= \left\langle \boxed{4}, \boxed{7}, -7, \boxed{3}, -17, -12 \right\rangle = 4\vec{w}_1 + 7\vec{w}_2 + 3\vec{w}_3. \end{aligned}$$

Notice that it is easier to find the coordinates of the vectors in B with respect to B', instead of the other way around.

Now, we assemble the three coordinate vectors into columns to form:

$$C_{B,B'} = \begin{bmatrix} 1 & -3 & 4 \\ 2 & -4 & 7 \\ 1 & -2 & 3 \end{bmatrix}, \quad \text{with inverse } C_{B',B} = \begin{bmatrix} -2 & -1 & 5 \\ -1 & 1 & -1 \\ 0 & 1 & -2 \end{bmatrix}.$$

Earlier, we found out that $\langle \vec{v}_3 \rangle_B = \langle -3, -2, 0 \rangle$, and $\langle \vec{v}_5 \rangle_B = \langle 9, 7, 4 \rangle$, where $\vec{v}_3 = \langle 3, 2, 11, 1, 0, 13 \rangle$, and $\vec{v}_5 = \langle 4, 18, -62, 7, -59, -86 \rangle$. We can check that:

$$\begin{bmatrix} 1 & -3 & 4 \\ 2 & -4 & 7 \\ 1 & -2 & 3 \end{bmatrix} \begin{bmatrix} -3 \\ -2 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix} = [\vec{v}_3]_{B'}, \text{ and}$$

$$\begin{bmatrix} 1 & -3 & 4 \\ 2 & -4 & 7 \\ 1 & -2 & 3 \end{bmatrix}\begin{bmatrix} 9 \\ 7 \\ 4 \end{bmatrix} = \begin{bmatrix} 4 \\ 18 \\ 7 \end{bmatrix} = [\vec{v}_5]_{B'}.$$

Again, we can check using the 1st, 2nd and 4th coordinates of $\vec{v}_3$ and $\vec{v}_5$ that these coordinate are correct.

Going in the other direction, the columns of $C_{B',B}$ give us the coordinates of the vectors in B' with respect to B. For example, the 3rd column of $C_{B',B}$ tells us that:

$$\begin{aligned} \langle \vec{w}_3 \rangle_B &= \langle 5, -1, -2 \rangle, \quad \text{and we verify that:} \\ \vec{w}_3 &= 5\vec{v}_1 - \vec{v}_2 - 2\vec{v}_4 \\ &= 5\langle 1, 2, -3, 1, -6, -5 \rangle - \langle -3, -4, -1, -2, 9, 1 \rangle - 2\langle 4, 7, -7, 3, -17, -12 \rangle \\ &= \langle 0, 0, 0, 1, -5, -2 \rangle. \square \end{aligned}$$

Matrices for Linear Transformations of Euclidean Spaces

You might be wondering: why are we making things more complicated than they have to be? Wouldn't it be simpler if we computed everything in terms of the standard bases? Of course, the answer is usually ***yes***, but sometimes, we are given linear transformations that are best described using vectors that are ***not*** from the standard bases. We would therefore like to be able to reconstruct the standard matrix $[T]$ if we are given the less convenient matrix $[T]_{B,B'}$:

Theorem: Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, B a ***basis*** for $\mathbb{R}^n$, and B' a ***basis*** for $\mathbb{R}^m$. Then: the ***standard matrix*** $[T]$ is related to $[T]_{B,B'}$ via:

$$[T] = [B'][T]_{B,B'}[B]^{-1},$$

where again, $[B]$ is the ***invertible*** matrix whose ***columns*** are the vectors of B, and similarly for $[B']$. In particular, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is an ***operator***, and B is a basis for $\mathbb{R}^n$ (used to encode both the domain and the codomain), then:

$$[T] = [B][T]_B[B]^{-1}.$$

Proof: We will prove that $[T] = [B'][T]_{B,B'}[B]^{-1}$ by proving instead that:

$$[T][B] = [B'][T]_{B,B'}.$$

Notice, we got this equation by multiplying both sides of the previous equation on the ***right*** by $[B]$. Now, all we have to do is compare the ***columns*** of each matrix product.

Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$. Column i of $[T][B]$ is thus $[T][\vec{w}_i] = T(\vec{w}_i)$.

But column i of $[T]_{B,B'}$, by definition, is $[T(\vec{w}_i)]_{B'}$, that is, the coordinate matrix of $T(\vec{w}_i)$ with respect to B'. Thus, column i of $[B'][T]_{B,B'}$ is a linear combination of the vectors of B' with coefficients from $[T(\vec{w}_i)]_{B'}$. But this is exactly the ***decoding*** process for $T(\vec{w}_i)$. Thus the two matrix products are equal.■

Example: Let $B = \{\langle 1, 0, -1\rangle, \langle 1, 1, 1\rangle, \langle 0, -1, 1\rangle\}$. Let us assemble the matrix with B in the first three columns and $\boldsymbol{I_3}$ in the last three. We obtain:

$$\left[\begin{array}{ccc|ccc} 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 & 0 \\ -1 & 1 & 1 & 0 & 0 & 1 \end{array}\right] \quad \text{with rref} \quad \left[\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ 0 & 1 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{array}\right].$$

Thus, B is a basis and we find $[B]^{-1}$ in the last three columns:

$$[B]^{-1} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$

Let us put aside $[B]^{-1}$ for a minute. Suppose we are given that $T : \mathbb{R}^3 \to \mathbb{R}^4$, with:

$$[T]_{B,B'} = \begin{bmatrix} 2 & 5 & -4 \\ -1 & 6 & 3 \\ 4 & -2 & -1 \\ 7 & 3 & 9 \end{bmatrix},$$

where B is the basis above, and B' is the basis:

$$\{\langle 1, -1, 0, 1\rangle, \langle 0, 1, -1, 1\rangle, \langle 1, 1, -1, 1\rangle, \langle 1, 0, 0, -1\rangle\}.$$

Let us compute $T(\vec{v})$ for $\vec{v} = \langle 8, -1, 5\rangle$. First, we need to ***encode*** $\langle 8, -1, 5\rangle$ using B, so we must solve:

$$\left[\begin{array}{ccc|c} 1 & 1 & 0 & 8 \\ 0 & 1 & -1 & -1 \\ -1 & 1 & 1 & 5 \end{array}\right], \quad \text{with rref:} \quad \left[\begin{array}{ccc|c} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 4 \\ 0 & 0 & 1 & 5 \end{array}\right].$$

Thus $\langle \vec{v}\rangle_B = \langle 4, 4, 5\rangle$. However, since we already found $[B]^{-1}$ above, we could also have found this vector using a simpler matrix product:

$$\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 8 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 5 \end{bmatrix}.$$

Next step, we now ***multiply:***

$$[T]_B[\vec{v}]_B = \begin{bmatrix} 2 & 5 & -4 \\ -1 & 6 & 3 \\ 4 & -2 & -1 \\ 7 & 3 & 9 \end{bmatrix} \begin{bmatrix} 4 \\ 4 \\ 5 \end{bmatrix} = \begin{bmatrix} 8 \\ 35 \\ 3 \\ 85 \end{bmatrix}.$$

Now, we can ***decode***, this time using B':

$$T(\vec{v}) = 8\langle 1, -1, 0, 1\rangle + 35\langle 0, 1, -1, 1\rangle + 3\langle 1, 1, -1, 1\rangle + 85\langle 1, 0, 0, -1\rangle = \langle 96, 30, -38, -39\rangle.$$

This was a lot of work just to find $T(\vec{v})$. Let us find $[T]$ to simplify the process:

$$[T] = [B][T]_B[B]^{-1}$$

$$= \begin{bmatrix} 1 & 0 & 1 & 1 \\ -1 & 1 & 1 & 0 \\ 0 & -1 & -1 & 0 \\ 1 & 1 & 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 5 & -4 \\ -1 & 6 & 3 \\ 4 & -2 & -1 \\ 7 & 3 & 9 \end{bmatrix}\begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & -\frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} 12 & -5 & -1 \\ \frac{7}{3} & -\frac{14}{3} & \frac{4}{3} \\ -4 & 1 & -1 \\ -3 & 10 & -1 \end{bmatrix}.$$

Now, we can compute $T(\vec{v})$ by ***direct multiplication***:

$$\begin{bmatrix} 12 & -5 & -1 \\ \frac{7}{3} & -\frac{14}{3} & \frac{4}{3} \\ -4 & 1 & -1 \\ -3 & 10 & -1 \end{bmatrix}\begin{bmatrix} 8 \\ -1 \\ 5 \end{bmatrix} = \begin{bmatrix} 96 \\ 30 \\ -38 \\ -39 \end{bmatrix},$$

thus giving us $\langle 96, 30, -38, -39 \rangle$, as before. Of course, this "easier" computation came with a price: we have to compute $[B]^{-1}$ and the matrix product $[B][T]_B[B]^{-1}$. □

Incidentally, the idea behind finding $[T]_{B,B'}$ also gives us a different perspective on $C_{B,B'}$. We leave the proof of the following Theorem as an Exercise.

Theorem: Let $\boldsymbol{I}_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n$ be the identity operator, and let B and B' be bases for $\mathbb{R}^n$. Then: $[\boldsymbol{I}_{\mathbb{R}^n}]_{B,B'} = C_{B,B'}$.

Revisiting Projections and Reflections

Let us present a third method to find the matrix for the projection and reflection operators onto and across a plane through the origin in $\mathbb{R}^3$. Recall that in our second method in Section 3.6, we saw that one of the steps felt a lot like finding the ***inverse*** of a matrix. Now that we have some understanding of the change of basis matrix, we shall see why this is not a coincidence.

Example: Suppose that Π is the plane with equation: $5x + 2y - 6z = 0$.

This is exactly the same plane that we studied in the last Example in Section 3.6. Recall that we chose $B = \{\langle 2, -5, 0\rangle, \langle 0, 3, 1\rangle, \langle 5, 2, -6\rangle\}$ for our basis for $\mathbb{R}^3$, where the first two vectors are on Π and the third is a normal vector $\vec{n}$ to Π.

This time, let us kill two birds with one stone: let us find $[proj_\Pi]_B$, $[proj_\Pi]$, $[refl_\Pi]_B$, and $[refl_\Pi]$ at the same time. Again, because $\vec{v}_1$ and $\vec{v}_2$ are both on Π, their projections onto Π and their reflections across Π are themselves:

$$proj_\Pi(\langle 2, -5, 0\rangle) = \langle 2, -5, 0\rangle,\ proj_\Pi(\langle 0, 3, 1\rangle) = \langle 0, 3, 1\rangle, \text{ and}$$

$$refl_\Pi(\langle 2, -5, 0\rangle) = \langle 2, -5, 0\rangle,\ refl_\Pi(\langle 0, 3, 1\rangle) = \langle 0, 3, 1\rangle.$$

On the other hand, since $\vec{n} = \langle 5, 2, -6\rangle$ is already orthogonal to Π, we have:

$$proj_{\Pi}(\langle 5, 2, -6\rangle) = \langle 0, 0, 0\rangle \text{ and } refl_{\Pi}(\langle 5, 2, -6\rangle) = -\langle 5, 2, -6\rangle.$$

Thus, we can see that:

$$[proj_{\Pi}]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and } [refl_{\Pi}]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

To apply the formula in the previous Theorem, we must assemble $[B]$ and compute $[B]^{-1}$:

$$[B] = \begin{bmatrix} 2 & 0 & 5 \\ -5 & 3 & 2 \\ 0 & 1 & -6 \end{bmatrix}, \text{ and } [B]^{-1} = \frac{1}{65}\begin{bmatrix} 20 & -5 & 15 \\ 30 & 12 & 29 \\ 5 & 2 & -6 \end{bmatrix}.$$

Note that this is exactly the same matrix that we saw in Section 3.6 that contained the coordinates of $\vec{e}_1, \vec{e}_2$ and $\vec{e}_3$ with respect to B in its columns. Now, to get the standard matrices for our projection and reflection operators, we apply the formula:

$$[T] = [B][T]_B[B]^{-1},$$

one at a time, to $[proj_{\Pi}]_B$ and $[refl_{\Pi}]_B$:

$$[proj_{\Pi}] = \begin{bmatrix} 2 & 0 & 5 \\ -5 & 3 & 2 \\ 0 & 1 & -6 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \cdot \frac{1}{65}\begin{bmatrix} 20 & -5 & 15 \\ 30 & 12 & 29 \\ 5 & 2 & -6 \end{bmatrix} = \frac{1}{65}\begin{bmatrix} 40 & -10 & 30 \\ -10 & 61 & 12 \\ 30 & 12 & 29 \end{bmatrix};$$

$$[refl_{\Pi}] = \begin{bmatrix} 2 & 0 & 5 \\ -5 & 3 & 2 \\ 0 & 1 & -6 \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \cdot \frac{1}{65}\begin{bmatrix} 20 & -5 & 15 \\ 30 & 12 & 29 \\ 5 & 2 & -6 \end{bmatrix} = \frac{1}{65}\begin{bmatrix} 15 & -20 & 60 \\ -20 & 57 & 24 \\ 60 & 24 & -7 \end{bmatrix}.$$

We can check from the ready-made formulas in Section 2.2 that these are correct. A similar idea can be applied to find $[proj_L]$.□

6.5 Section Summary

For any bases B and B' for $\mathbb{R}^n$, there exists an ***invertible*** $n \times n$ matrix $C_{B,B'}$ such that for all vectors $\vec{v}$ of $\mathbb{R}^n$, $[\vec{v}]_{B'} = C_{B,B'}[\vec{v}]_B$.

The matrix $C_{B,B'}$ is called the ***change of basis matrix from B to B'***. We can explicitly find $C_{B,B'}$ by performing the Gauss-Jordan algorithm on the augmented matrix $\left[B' \mid B\right]$, where this notation means that we assemble as the columns of this matrix the vectors in B', followed by the vectors of B.

At the end of the process, we obtain $\left[\boldsymbol{I_n} | C_{B,B'}\right]$.

Consequently, the columns of $C_{B,B'}$ are the ***coordinate matrices*** of the members of B with respect to B', that is, if $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then: $C_{B,B'} = [[\vec{v}_1]_{B'}[\vec{v}_2]_{B'} \cdots [\vec{v}_n]_{B'}]$.

Moreover, $C_{B,B'}^{-1} = C_{B',B}$.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, B a **basis** for $\mathbb{R}^n$, and B' a **basis** for $\mathbb{R}^m$. Then:

$$[T] = [B'][T]_{B,B'}[B]^{-1},$$

where $[B]$ is the matrix whose ***columns*** are the vectors of B, and similarly for $[B']$.

In particular, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is an ***operator***, and B is a **basis** for $\mathbb{R}^n$ (used to encode both the domain and the codomain), then $[T] = [B][T]_B[B]^{-1}$.

6.5 Exercises

1. Let $\vec{v} = \langle 4, -3, 7 \rangle$, and consider the following bases for $\mathbb{R}^3$:

$$B = \{\langle 1,0,-1\rangle, \langle 1,1,2\rangle, \langle 0,1,1\rangle\}, \text{ and}$$

$$B' = \{\langle 0,-1,1\rangle, \langle 1,-1,1\rangle, \langle 1,2,1\rangle\}.$$

   a. Find $\langle \vec{v} \rangle_B$ and $\langle \vec{v} \rangle_{B'}$ using the Gauss-Jordan algorithm.
   b. Explain why your computations in (a) prove that B and B' are indeed bases for $\mathbb{R}^3$.
   c. Find $C_{B,B'}$.
   d. Verify by direct matrix multiplication that $[\vec{v}]_{B'} = C_{B,B'}[\vec{v}]_B$.

2. Repeat Exercise 1 with $\vec{v} = \langle 5, 3, 1, -4 \rangle$ and the following bases for $\mathbb{R}^4$:

$$B = \{\langle 1,0,-1,1\rangle, \langle 2,1,1,0\rangle, \langle -3,1,0,0\rangle, \langle 2,0,0,0\rangle\},$$

$$B' = \{\langle 1,0,1,2\rangle, \langle 0,1,1,-1\rangle, \langle 0,0,2,1\rangle, \langle 0,0,0,-1\rangle\}.$$

3. Let:

$$B = \{\langle 1,0,-1,1\rangle, \langle 2,1,1,0\rangle, \langle -3,1,0,0\rangle, \langle 2,0,0,0\rangle\}$$

be the first basis for $\mathbb{R}^4$ in Exercise 2, and let:

$$B' = \{\langle 0,-1,1\rangle, \langle 1,-1,1\rangle, \langle 1,2,1\rangle\}$$

be the second basis for $\mathbb{R}^3$ in Exercise 1. Suppose that a linear transformation $T : \mathbb{R}^4 \to \mathbb{R}^3$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 5 & 3 & -2 & 1 \\ 3 & -1 & 4 & 0 \\ 2 & 0 & 3 & -2 \end{bmatrix}$$

   a. Compute $T(\vec{v})$, where $\vec{v} = \langle 5, 3, 1, -4 \rangle$. Express your final answer as an ordinary vector in $\mathbb{R}^3$ (which means that you should not forget to decode). Use part of your answer in Exercise 2.
   b. Find the standard matrix of T.
   c. Recompute $T(\vec{v})$ using your standard matrix.

4. Now, let:

$$B = \{\langle 1,0,-1\rangle, \langle 1,1,2\rangle, \langle 0,1,1\rangle\},$$

be the first basis for $\mathbb{R}^3$ in Exercise 1, and let:

$$B' = \{\langle 1,0,1,2\rangle, \langle 0,1,1,-1\rangle, \langle 0,0,2,1\rangle, \langle 0,0,0,-1\rangle\}$$

be the second basis for $\mathbb{R}^4$ in Exercise 2. Suppose that a linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^4$ is given by:

$$[T]_{B,B'} = \begin{bmatrix} 4 & 3 & 1 \\ -3 & 1 & 0 \\ -5 & -2 & 4 \\ 0 & -1 & -2 \end{bmatrix}.$$

a. Compute $T(\vec{v})$, where $\vec{v} = \langle 4, -3, 7 \rangle$. Express your final answer as an ordinary vector in $\mathbb{R}^4$. Use part of your answer in Exercise 1.

b. Find the standard matrix of T.

c. Recompute $T(\vec{v})$ using your standard matrix.

5. Let:

$$B = \{\langle 1, 0, -1 \rangle, \langle 1, 1, 2 \rangle, \langle 0, 1, 1 \rangle\}$$

be the first basis for $\mathbb{R}^3$ in Exercise 1. Suppose that an operator $T : \mathbb{R}^3 \to \mathbb{R}^3$ is given by:

$$[T]_B = \begin{bmatrix} 6 & -3 & -1 \\ -2 & 1 & 0 \\ -7 & 2 & 4 \end{bmatrix}.$$

a. Compute $T(\vec{v})$, where $\vec{v} = \langle 4, -3, 7 \rangle$. Express your final answer as an ordinary vector in $\mathbb{R}^3$. Use part of your answer in Exercise 1.

b. Find the standard matrix of T.

c. Recompute $T(\vec{v})$ using your standard matrix.

6. Suppose that an operator $T : \mathbb{R}^3 \to \mathbb{R}^3$ is given by its standard matrix:

$$[T] = \begin{bmatrix} 7 & 3 & 1 \\ -1 & -4 & 0 \\ 3 & 5 & -2 \end{bmatrix}.$$

a. Compute $T(\vec{v})$, where $\vec{v} = \langle 4, -3, 7 \rangle$.

b. Suppose that:

$$B = \{\langle 1, 0, -1 \rangle, \langle 1, 1, 2 \rangle, \langle 0, 1, 1 \rangle\},$$

is the first basis for $\mathbb{R}^3$ in Exercise 1. Find $[T]_B$.

c. Recompute $T(\vec{v})$ using $[T]_B$.

For Exercises (7) to (14): You are given the set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$, and suppose that $W = Span(S)$.

(a) Find a basis B for W using the Minimizing Theorem;

(b) Find a basis B' for W by assembling S into the rows of a matrix and finding its rref;

(c) Express the non-basis vectors in S in terms of the basis B from part (a);

(d) Express the non-basis vectors in S in terms of the basis B' from part (c);

(e) Find the change of basis matrix $C_{B,B'}$;

(f) Verify that $[\vec{v}\,]_{B'} = C_{B,B'}[\vec{v}\,]_B$, for every non-basis vector $\vec{v}$.

7. $S = \{\langle -3, 1, 6, -5\rangle, \langle 4, 2, -4, -4\rangle, \langle 18, 4, -24, -2\rangle, \langle 1, 4, 7, 3\rangle\}$

8. $S = \{\langle -3, 1, 6, -5\rangle, \langle 4, 2, -4, -4\rangle, \langle 1, 4, 7, 3\rangle, \langle -10, -3, 1, -42\rangle\}$

9. $S = \{\langle -3, 12, 5, 2, -2\rangle, \langle 1, -4, 4, 3, -4\rangle, \langle 4, -16, -6, -4, 18\rangle\}$

10. $S = \{\langle -3, -4, -2, 9, 1, 1\rangle, \langle 1, 2, 4, 9, 11, -11\rangle, \langle 4, 3, 5, 16, 1, 8\rangle,$
$\langle -21, -36, -26, 59, -45, 79\rangle, \langle -20, -37, -23, 84, -43, 88\rangle\}$

11. $S = \{\langle -5, 3, -3, 2, -14, -4\rangle, \langle 3, -4, -7, -5, -21, 7\rangle, \langle -21, 17, 5, 16, 0, -26\rangle,$
$\langle 2, -1, 2, 0, 11, 2\rangle, \langle -1, 2, 5, 3, 17, -8\rangle\}$

12. $S = \{\langle -4, -5, -1, 3, 7, 1\rangle, \langle 2, 3, -1, -1, -8, 9\rangle, \langle -2, -1, -5, 3, -10, 29\rangle,$
$\langle -1, 0, -4, 2, 2, 1\rangle, \langle 3, 2, 6, -5, -4, -12\rangle\}$

13. $S = \{\langle -3, 1, 4, -21, -20\rangle, \langle -4, 2, 3, -36, -37\rangle, \langle -2, 4, 5, -26, -23\rangle,$
$\langle 9, 9, 16, 59, 84\rangle, \langle 1, 11, 1, -45, -43\rangle, \langle 1, -11, 8, 79, 88\rangle\}$

14. $S = \{\langle -4, -5, 3, 19, 2, -8\rangle, \langle -8, -1, 2, -28, 3, -26\rangle, \langle 2, 2, -1, -5, 0, 15\rangle,$
$\langle 7, 3, -4, 5, -5, -8\rangle, \langle 8, 7, -10, -32, -8, -57\rangle\}$

For Exercises (15) to (18): Your goal is to find the matrices $[proj_\Pi]$, $[refl_\Pi]$ and $[proj_L]$ for the following planes Π (with normal line $L = Span(\{\vec{n}\})$) that appear in Exercises 28 to 31 in Section 3.6. The answers can be found in the Key to Section 3.6. Review the last Example in this Section, and follow this outline:

(a) Find two non-parallel vectors $\vec{v}_1$ and $\vec{v}_2$ on Π.
Keep it simple by choosing vectors with integer coordinates, where one coordinate is 0.

(b) Form the matrix $[B]$, where $B = \{\vec{v}_1, \vec{v}_2, \vec{n}\}$ is a basis for $\mathbb{R}^3$.

(c) Find $[B]^{-1}$.

(d) Write $[proj_\Pi]_B$, $[refl_\Pi]_B$ and $[proj_L]_B$ as diagonal matrices.

(e) Use the formula $[T] = [B][T]_B[B]^{-1}$ three times to find $[proj_\Pi]$, $[refl_\Pi]$ and $[proj_L]$.

15. $\Pi : 3x + 7y - 8z = 0$.

16. $\Pi : 5x - 3y + 7z = 0$.

17. $\Pi : 2x - y + 5z = 0$.

18. $\Pi : x = \frac{2}{3}z$. As in Section 3.6, think very carefully about part (a). Hint: which coordinate axis is contained in Π?

19. Prove that if $I_{\mathbb{R}^n} : \mathbb{R}^n \to \mathbb{R}^n$ is the ***identity operator***, then:

$$[I_{\mathbb{R}^n}]_{B,B'} = C_{B,B'}$$

for ***any bases*** B and B' of $\mathbb{R}^n$. Hint: start by writing down the ***definitions*** of both sides of this equation, and then unwind these definitions.

6.6 Change of Basis for Abstract Spaces and Determinants for Operators

We will now use the ideas from the previous Section in order to extend the concept of a change of basis matrix for coordinate vectors in abstract vector spaces and the linear transformations that act on them. This will also enable us to find the determinant of a linear operator, and in the next Section, a characteristic equation, eigenvalues and eigenspaces for them.

Change of Basis for Abstract Vector Spaces

The computations in Section 6.4 can usually be adapted to abstract vectors spaces when we have a nice "standard basis" that we can use. For the polynomial spaces $\mathbb{P}^n$, for example, we have the monomials $\{1, x, x^2, \ldots, x^n\}$. Similarly, we defined other finite-dimensional function spaces as $Span(\{f_1(x), \ldots, f_n(x)\})$, so we can use this indicated set as a standard basis. Let us begin with the following generalization:

Definition/Theorem: For any ***bases*** B and $B^/$ for a vector space V, with $\dim(V) = n$, there exists an ***invertible*** $n \times n$ matrix $C_{B,B^/}$ such that for all $\vec{v} \in V$:

$$[\vec{v}]_{B^/} = C_{B,B^/}[\vec{v}]_B.$$

The matrix $C_{B,B^/}$ is called the ***change of basis matrix*** from B to $B^/$.

The columns of $C_{B,B^/}$ are the ***coordinate matrices*** of the members of B with respect to $B^/$, that is, if $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then:

$$C_{B,B^/} = [[\vec{v}_1]_{B^/}[\vec{v}_2]_{B^/}\ldots[\vec{v}_n]_{B^/}].$$

Notice that we left out the part where we derive how to efficiently compute $C_{B,B^/}$, but this is simply because the members of B and $B^/$ are not necessarily from $\mathbb{R}^n$. Thus, it doesn't make sense to enter the members of B and $B^/$ as the columns of a matrix. But this is easily fixed by using ***any*** arbitrary basis, let's call it S, for our vector space V. The most convenient choice for S would most often be a "standard basis" for this space. Let us make this more explicit:

Suppose $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ and $B^/ = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ are two arbitrary bases for V, and S some kind of a "standard basis" for V. Let us denote by $[B]_S$ and $[B^/]_S$ the $n \times n$ matrices:

$$[B]_S = [[\vec{v}_1]_S[\vec{v}_2]_S\ldots[\vec{v}_n]_S], \text{ and}$$

$$[B^/]_S = [[\vec{w}_1]_S[\vec{w}_2]_S\ldots[\vec{w}_n]_S].$$

With this notation, we can explicitly find $C_{B,B^/}$ by performing the Gauss-Jordan algorithm on the augmented matrix:

$$\left[[B^/]_S | [B]_S\right],$$

At the end of the process, we obtain:

$$\left[\boldsymbol{I}_n | C_{B,B^/}\right].$$

Analogous to our Theorem in Section 3.8, we have the formula:

$$C_{B,B'} = [B']_S^{-1}[B]_S.$$

Example: Let $V = Span(S)$, where $S = \{\sin(x), \cos(x)\}$. First, we saw in Chapter 3 that S is linearly independent, so V is 2-dimensional. Consider:

$$B = \{3\sin(x) + 2\cos(x), 2\sin(x) + \cos(x)\}, \text{ and}$$
$$B' = \{\sin(x) + 3\cos(x), \sin(x) + 4\cos(x)\}.$$

We easily see that the two vectors in B and B' are not parallel, so they also form bases for V. Now, let us find the associated coordinate matrices:

$$[B]_S = \begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix}, \text{ and } [B']_S = \begin{bmatrix} 1 & 1 \\ 3 & 4 \end{bmatrix}.$$

Let us find $C_{B,B'}$ in two different ways.

First Solution: we perform the Gauss Jordan Algorithm on the augmented matrix:

$$\left[[B']_S | [B]_S \right] = \left[\begin{array}{cc|cc} 1 & 1 & 3 & 2 \\ 3 & 4 & 2 & 1 \end{array}\right],$$

whose rref is:

$$\left[\begin{array}{cc|cc} 1 & 0 & 10 & 7 \\ 0 & 1 & -7 & -5 \end{array}\right].$$

Thus, the change of basis matrix is $C_{B,B'} = \begin{bmatrix} 10 & 7 \\ -7 & -5 \end{bmatrix}$.

Second Solution: We find $[B']_S^{-1}$ using our formula for 2×2 matrices:

$$[B']_S^{-1} = \frac{1}{4-3}\begin{bmatrix} 4 & -1 \\ -3 & 1 \end{bmatrix} = \begin{bmatrix} 4 & -1 \\ -3 & 1 \end{bmatrix}.$$

Next, we perform a matrix product to get, as before:

$$C_{B,B'} = [B']_S^{-1}[B]_S = \begin{bmatrix} 4 & -1 \\ -3 & 1 \end{bmatrix}\begin{bmatrix} 3 & 2 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 10 & 7 \\ -7 & -5 \end{bmatrix}. \square$$

Matrices for Linear Transformations of Abstract Vector Spaces

We can use the generalization above to find different matrices for a linear transformation from one abstract vector space to another. With the notation that we established in the previous theorem, we have the following:

Theorem: Let $T : V \to W$ be a linear transformation of finite dimensional vector spaces, B and S ***bases*** for V, and B' and S' ***bases*** for W. Then:

$$[T]_{S,S'} = [B']_{S'}[T]_{B,B'}[B]_S^{-1},$$

where, again, $[B]_S$ is the matrix whose columns are the ***coordinate vectors*** of the members of B with respect to S, and similarly for $[B']_{S'}$.

In particular, if $T : V \to V$ is an ***operator***, then:

$$[T]_S = [B]_S[T]_B[B]_S^{-1}.$$

Example: Consider the bases:

$$B = \{2, 3 - x, 5 + 7x + x^2\}, \quad \text{and } B' = \{-3, 2 + x, 1 - x^2, x + x^3\}.$$

for $\mathbb{P}^2$ and $\mathbb{P}^3$, respectively. Now, let $T : \mathbb{P}^2 \to \mathbb{P}^3$, whose matrix with respect to B and B' is:

$$[T]_{B,B'} = \begin{bmatrix} 7 & -1 & 5 \\ 2 & 6 & -2 \\ -3 & 8 & 0 \\ 4 & 2 & 3 \end{bmatrix}.$$

As a warm-up, let us remember how to ***compute*** $T(\vec{v})$, where $\vec{v} = 6 - 4x + 3x^2$.

First, we need to ***encode*** this polynomial using the basis B. Fortunately, the degrees of the members of B are all ***distinct***, so starting with the quadratic member, we find the coefficients by inspection, as:

$$6 - 4x + 3x^2 = -42(2) + 25(3 - x) + 3(5 + 7x + x^2),$$

thus $\langle \vec{v} \rangle_B = \langle -42, 25, 3 \rangle$. Now, we ***multiply***:

$$[T(\vec{v})]_{B'} = [T]_{B,B'}[\vec{v}]_B = \begin{bmatrix} 7 & -1 & 5 \\ 2 & 6 & -2 \\ -3 & 8 & 0 \\ 4 & 2 & 3 \end{bmatrix} \begin{bmatrix} -42 \\ 25 \\ 3 \end{bmatrix} = \begin{bmatrix} -304 \\ 60 \\ 326 \\ -109 \end{bmatrix}.$$

Finally, we ***decode*** these coefficients using B' to get:

$$\begin{aligned} T(\vec{v}) &= -304(-3) + 60(2 + x) + 326(1 - x^2) - 109(x + x^3) \\ &= 1358 - 49x - 326x^2 - 109x^3. \end{aligned}$$

Clearly $[T]_{B,B'}$ is not a very convenient matrix to use. Let us therefore find the matrix of T with respect to the ***standard bases*** $S = \{1, x, x^2\}$ for $\mathbb{P}^2$ and $S' = \{1, x, x^2, x^3\}$ for $\mathbb{P}^3$. We have:

$$[B]_S = \begin{bmatrix} 2 & 3 & 5 \\ 0 & -1 & 7 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad [B']_{S'} = \begin{bmatrix} -3 & 2 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We use the Gauss-Jordan algorithm or technology to find the ***inverse*** of the first matrix:

$$[B]_S^{-1} = \begin{bmatrix} 1/2 & 3/2 & -13 \\ 0 & -1 & 7 \\ 0 & 0 & 1 \end{bmatrix}.$$

Now, we are ready to apply the formula to obtain our ***standard matrix:***

$$[T]_{S,S'} = [B']_{S'}[T]_{B,B'}[B]_S^{-1}$$

$$= \begin{bmatrix} -3 & 2 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 7 & -1 & 5 \\ 2 & 6 & -2 \\ -3 & 8 & 0 \\ 4 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1/2 & 3/2 & -13 \\ 0 & -1 & 7 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -10 & -53 & 402 \\ 3 & 1 & -21 \\ 3/2 & 25/2 & -95 \\ 2 & 4 & -35 \end{bmatrix}.$$

To check that this matrix is correct, we recompute $T(6 - 4x + 3x^2)$. This time, we ***encode*** $\vec{v}$ using the ***standard basis*** S, thus producing $[\vec{v}]_S$, and then we multiply $[T]_{S,S'}$ by this matrix:

$$[T(\vec{v})]_{S'} = [T]_{S,S'}[\vec{v}]_S = \begin{bmatrix} -10 & -53 & 402 \\ 3 & 1 & -21 \\ 3/2 & 25/2 & -95 \\ 2 & 4 & -35 \end{bmatrix} \begin{bmatrix} 6 \\ -4 \\ 3 \end{bmatrix} = \begin{bmatrix} 1358 \\ -49 \\ -326 \\ -109 \end{bmatrix}.$$

Finally, we ***decode*** this result with respect to the ***standard basis*** S' as:

$$T(\vec{v}) = 1358 - 49x - 326x^2 - 109x^3.$$

As expected, we get the same answer, so we can be fairly confident that our standard matrix is correct.□

The Determinant of an Operator

We are now in a position to extend the concepts of the determinant and eigentheory to linear ***operators*** $T : V \to V$. This would almost seem like a natural thing to do because any matrix for T with respect to some basis S would be be a ***square*** matrix $[T]_S$. It would thus be a simple matter of computing the determinant, characteristic polynomial, and eigenvectors of this matrix. The million dollar question, though, is: Will we get the same answers regardless of the ***choice*** of the basis S? The answer, of course, is ***yes***. The key is the equation:

$$[T]_S = [B]_S[T]_B[B]_S^{-1},$$

where B and S are any two bases for V, and $[B]_S$ is the matrix whose columns are the coordinate vectors of the members of B with respect to S. Notice that all of the matrices involved in this equation are $n \times n$ matrices, where $n = \dim(V)$. This equation immediately leads us to our first goal:

Definition/Theorem: Let S and B be ***bases*** for a finite dimensional vector space V, and let $T : V \to V$ be a linear ***operator*** acting on V. Then: $det([T]_S) = det([T]_B)$.

Thus, we can define: $det(T) = |T| = det([T]_B)$, where B is ***any*** basis for V.

This number does ***not*** depend on the choice of basis B.

Proof: By the multiplicative property of determinants, and the determinant of the inverse of a matrix, we have:

$$\begin{aligned} det([T]_S) &= det\left([B]_S[T]_B[B]_S^{-1}\right) = det([B]_S) \bullet det([T]_B) \bullet det\left([B]_S^{-1}\right) \\ &= det([B]_S) \bullet det([T]_B) \bullet det([B]_S)^{-1} = det([T]_B) \bullet det([B]_S) \bullet det([B]_S)^{-1} = det([T]_B). \end{aligned}$$

Notice that we are free to rearrange the three factors above since a determinant is just a ***number*** and not a matrix.■

Example: Let us bring back our old friends, the projection and reflection operators onto and across a plane Π. In Section 2.2, we saw that for the plane Π given by the equation $3x - 5y + 2z = 0$, the standard matrices of these two operators are:

$$[proj_\Pi] = \frac{1}{38}\begin{bmatrix} 29 & 15 & -6 \\ 15 & 13 & 10 \\ -6 & 10 & 34 \end{bmatrix} \text{ and } [refl_\Pi] = \frac{1}{19}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix}.$$

It would not be a terrible chore to find the determinants of these two matrices, but it wouldn't exactly be a pleasant one either. It is also quite likely that you will make one or two arithmetic mistakes along the way. Fortunately, we can find these determinants by looking at two very different but simpler matrices. Recall that in Section 3.8, Exercise 7, we chose two linearly independent vectors $\vec{v}_1 = \langle 5,3,0\rangle$ and $\vec{v}_2 = \langle 0,2,5\rangle$ that are on the plane Π, and $\vec{n} = \langle 3,-5,2\rangle$ the obvious normal for the plane. The set $B = \{\vec{n}, \vec{v}_1, \vec{v}_2\}$ is a basis for $\mathbb{R}^3$. With respect to this basis, we get:

$$\begin{aligned} [proj_\Pi]_B &= Diag(0, 1, 1), \quad \text{and} \\ [refl_\Pi]_B &= Diag(-1, 1, 1). \end{aligned}$$

Obviously, it is much easier to find the determinant of a diagonal matrix. We immediately get:

$$\begin{aligned} det([proj_\Pi]_B) &= 0 \bullet 1 \bullet 1 = 0, \text{ and} \\ det([refl_\Pi]_B) &= -1 \bullet 1 \bullet 1 = -1. \square \end{aligned}$$

Example: Let $T : \mathbb{P}^2 \to \mathbb{P}^2$ be the linear operator given by:

$$T(p(x)) = 2p(x) + (x+5)p'(x) + (x^2 - 3x + 7)p''(x).$$

We leave the reader to check that T is indeed additive and homogeneous. Let us find its matrix with respect to the standard basis $S = \{1, x, x^2\}$:

$$\begin{aligned} T(1) &= 2 + 0 + 0 = 2, \\ T(x) &= 2x + (x+5) \bullet 1 + 0 = 5 + 3x, \text{ and} \\ T(x^2) &= 2x^2 + (x+5) \bullet 2x + (x^2 - 3x + 7) \bullet 2 = 14 + 4x + 6x^2. \end{aligned}$$

We encode their coefficients into the columns of:

$$[T]_S = \begin{bmatrix} 2 & 5 & 14 \\ 0 & 3 & 4 \\ 0 & 0 & 6 \end{bmatrix}.$$

This is an upper triangular matrix, and thus we easily find: $det(T) = 2 \bullet 3 \bullet 6 = 36.$ □

As a bonus, we leave it as an Exercise for you to prove:

Theorem: Let $T : V \to V$ be a linear ***operator*** acting on a finite dimensional vector space V. Then T is ***invertible*** ***if and only if*** $det(T) \neq 0$.

Example: Let $T : \mathbb{P}^2 \to \mathbb{P}^2$ be the linear operator from the previous Example. We saw that $det(T) = 36$, so this operator is invertible. This means that if $q(x)$ is ***any*** polynomial from $\mathbb{P}^2$, then we can find another polynomial $p(x)$ from $\mathbb{P}^2$ so that:

$$2p(x) + (x+5)p'(x) + (x^2 - 3x + 7)p''(x) = q(x),$$

and therefore this differential equation is ***solvable*** for any choice of $q(x)$.□

6.6 Section Summary

For any ***bases*** B and B' for a vector space V, with $n = \dim(V)$, there exists an ***invertible*** $n \times n$ matrix $C_{B,B'}$ such that for all vectors $\vec{v} \in V$, $[\vec{v}]_{B'} = C_{B,B'}[\vec{v}]_B$.

The matrix $C_{B,B'}$ is called the ***change of basis matrix from B to B'***. The columns of $C_{B,B'}$ are the ***coordinate matrices of the members of B with respect to B'***, that is, if $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then: $C_{B,B'} = [[\vec{v}_1]_{B'}\,[\vec{v}_2]_{B'} \ldots [\vec{v}_n]_{B'}] = [B]_{B'}$, the matrix whose ***columns*** are the coordinate matrices of the $\vec{v}_i$ with respect to B'.

Let $T : V \to W$ be a linear transformation of finite dimensional vector spaces, B and S ***bases*** for V, and B' and S' ***bases*** for W. Then: $[T]_{S,S'} = [B']_{S'}[T]_{B,B'}[B]_S^{-1}$.

In particular, if $T : V \to V$ is an ***operator***, then: $[T]_S = [B]_S[T]_B[B]_S^{-1}$.

Let S and B be ***bases*** for a finite dimensional vector space V, and let $T : V \to V$ be a linear ***operator*** acting on V. Then: $det([T]_S) = det([T]_B)$. Thus, we can define: $det(T) = |T| = det([T]_B)$, where B is ***any*** basis for V. This number does ***not*** depend on the choice of basis B.

Let $T : V \to V$ be a linear ***operator*** acting on a finite dimensional vector space V. Then T is ***invertible*** ***if and only if*** $det(T) \neq 0$.

6.6 Exercises

1. Let $B = \{1 + x^2, 1 - x + x^2, 1 + x\}$, $B' = \{x - x^2, 1 + x, 2 - x^2\}$ and $\vec{v} = 4 + 3x - 5x^2$.
   a. Find $\langle \vec{v} \rangle_B$ and $\langle \vec{v} \rangle_{B'}$. using the Gauss-Jordan algorithm.
   b. Explain why your computations in (a) prove that B and B' are bases for $\mathbb{P}^2$.
   c. Find $C_{B,B'}$.
   d. Verify by direct matrix multiplication that $[\vec{v}]_{B'} = C_{B,B'}[\vec{v}]_B$.
2. Repeat Exercise 1, with $B = \{1 + x - x^3, 2 - x + x^2, 5 - x, 2\}$, $B' = \{1 - 3x^2 + x^3, x - x^2 - x^3, x^2 - 5x^3, -x^3\}$, and $\vec{v} = 5 - 3x + 4x^2 - 2x^3$, with $\mathbb{P}^2$ replaced with $\mathbb{P}^3$ in part (b).
3. Let $B = \{1 + x - x^3, 2 - x + x^2, 5 - x, 2\}$ be the first basis for $\mathbb{P}^3$ in Exercise 2, and $B' = \{x - x^2, 1 + x, 2 - x^2\}$ the second basis for $\mathbb{P}^2$ in Exercise 1.

   Suppose that a linear transformation $T : \mathbb{P}^3 \to \mathbb{P}^2$ is given by:

$$[T]_{B,B^{/}} = \begin{bmatrix} 4 & 3 & 2 & -1 \\ -3 & -1 & 1 & 8 \\ 2 & 7 & -4 & -2 \end{bmatrix}.$$

a. Compute $T(\vec{v})$, where $\vec{v} = 5 - 3x + 4x^2 - 2x^3$. Express your final answer as an ordinary vector in $\mathbb{P}^2$ (which means that you should not forget to decode). Use part of your answer in Exercise 2.

b. Let $S = \{1, x, x^2, x^3\}$ and $S^{/} = \{1, x, x^2\}$. Find $[T]_{S,S^{/}}$.

c. Recompute $T(\vec{v})$, where $\vec{v} = 5 - 3x + 4x^2 - 2x^3$, using $[T]_{S,S^{/}}$.

4. Let $B = \{1 + x^2, 1 - x + x^2, 1 + x\}$ be the first basis for $\mathbb{P}^2$ in Exercise 1, and $B^{/} = \{1 - 3x^2 + x^3, x - x^2 - x^3, x^2 - 5x^3, -x^3\}$ the second basis for $\mathbb{P}^3$ in Exercise 2.

Suppose that a linear transformation $T : \mathbb{P}^2 \to \mathbb{P}^3$ is given by:

$$[T]_{B,B^{/}} = \begin{bmatrix} 0 & -3 & 5 \\ -3 & -1 & 1 \\ 2 & 1 & -4 \\ -1 & 7 & 6 \end{bmatrix}.$$

a. Compute $T(\vec{v})$, where $\vec{v} = 4 + 3x - 5x^2$. Express your final answer as an ordinary vector in $\mathbb{P}^3$. Use part of your answer in Exercise 1.

b. Let $S = \{1, x, x^2\}$ and $S^{/} = \{1, x, x^2, x^3\}$. Find $[T]_{S,S^{/}}$.

c. Recompute $T(\vec{v})$, where $\vec{v} = 4 + 3x - 5x^2$, using $[T]_{S,S^{/}}$.

5. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^2$ is an operator, with:

$$[T]_B = \begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & -3 \\ -1 & -2 & 1 \end{bmatrix},$$

where $B = \{1 + x^2, 1 - x + x^2, 1 + x\}$ is the first basis from Exercise 1.

a. Compute $T(\vec{v})$ where $\vec{v} = 4 + 3x - 5x^2$ is the vector from Exercise 1.

b. Let $S = \{1, x, x^2\}$. Find the change of basis matrix $[B]_S$ and its inverse $[B]_S^{-1}$.

c. Use the formula $[T]_S = [B]_S [T]_B [B]_S^{-1}$ to find $[T]_S$.

d. Recompute $T(\vec{v})$, where $\vec{v} = 4 + 3x - 5x^2$, using $[T]_S$.

e. Compute $det(T)$.

f. Is T invertible? If so, find $[T^{-1}]_B$.

6. Suppose that $T : \mathbb{P}^3 \to \mathbb{P}^3$ is an operator, with:

$$[T]_B = \begin{bmatrix} -1 & 0 & -2 & 3 \\ -1 & 2 & 1 & -4 \\ 0 & 1 & 3 & -2 \\ -1 & 1 & 1 & 1 \end{bmatrix},$$

where $B = \{1 + x - x^3, 2 - x + x^2, 5 - x, 2\}$ is the first basis from Exercise 2.

a. Compute $T(\vec{v})$ where $\vec{v} = 5 - 3x + 4x^2 - 2x^3$ is the vector from Exercise 2.

b. Let $S = \{1, x, x^2, x^3\}$. Find the change of basis matrix $[B]_S$ and its inverse $[B]_S^{-1}$.

c. Use the formula $[T]_S = [B]_S[T]_B[B]_S^{-1}$ to find $[T]_S$.

d. Recompute $T(\vec{v})$, where $\vec{v} = 5 - 3x + 4x^2 - 2x^3$, using $[T]_S$.

e. Compute $det(T)$.

f. Is T invertible? If so, find $[T^{-1}]_B$.

7. Let $D : \mathbb{P}^2 \to \mathbb{P}^2$ be the differentiation operator, and let $B = \{1 + x^2, 1 - x + x^2, 1 + x\}$ be the first basis from Exercise 1.

a. Find $[D]_B$.

b. Let $S = \{1, x, x^2\}$. Use the formula $[T]_S = [B]_S[T]_B[B]_S^{-1}$ to find $[D]_S$. You may use your work from Exercise 5.

c. Compute $det(T)$.

d. Is T invertible? If so, find $[T^{-1}]_B$.

8. Repeat the previous Exercise for the differentiation operator $D : \mathbb{P}^3 \to \mathbb{P}^3$, where:

$$B = \{1 + x - x^3, 2 - x + x^2, 5 - x, 2\}$$

is the first basis for $\mathbb{P}^3$ from Exercise 2 and $S = \{1, x, x^2, x^3\}$. You may use your work from Exercise 6.

9. Let $V = Span(B)$, where $B = \{\sin(x), \cos(x)\}$, and $B' = \{\sin(x + \pi/6), \sin(x + \pi/3)\}$.

a. Show that B' is also a basis for V. Reminder: you must show that B' is a subset of V to begin with.

b. Find the change of basis matrix from B to B'.

c. Suppose that $T : V \to V$ is given by: $[T]_{B'} = \begin{bmatrix} 1 & -3 \\ 3 & -7 \end{bmatrix}$. Find $[T]_B$.

d. Compute $det(T)$.

e. Is T invertible? If so, find $[T^{-1}]_B$.

f. Suppose that $D : V \to V$ is the differentiation operator. Find $[D]_B$.

g. Compute $det(D)$.

h. Is D invertible? If so, find $[D^{-1}]_B$.

10. Let $V = Span(B)$, where $B = \{e^{-2x}, x \cdot e^{-2x}, x^2 \cdot e^{-2x}\}$, and D the differentiation operator $D : V \to V$.

a. Find $[D]_B$.

b. Compute $det(D)$.

c. Is D invertible? If so, find $[D^{-1}]_B$.

11. Let $T : V \to V$ be a linear operator acting on a finite dimensional vector space V. Prove that T is ***invertible*** ***if and only if*** $det(T) \neq 0$.

12. Prove that if $\boldsymbol{I}_V : V \to V$ is the ***identity*** operator of V, then $[\boldsymbol{I}_V]_{B,B'} = C_{B,B'}$, for any ***bases*** B and B' of V. Note: this is analogous to the last Exercise from the previous Section, and the proof is exactly the same idea there.

6.7 Similarity and The Eigentheory of Operators

You may have noticed that in the last few Sections, we saw two equations that bear some resemblance. We said that a square matrix A is ***diagonalizable*** if we can find an ***invertible*** matrix C of the same dimension such that:

$$D = C^{-1}AC,$$

where D is a diagonal matrix. We also saw the equation:

$$[T]_S = [B]_S[T]_B[B]_S^{-1},$$

relating two different matrices for the same linear transformation T. Both of these equations involve a ***product of three matrices***, where the middle matrix is essentially "sandwiched" by ***a matrix*** and ***its inverse***. We called this a ***conjugation*** process. This kind of matrix expression is of such importance in Linear Algebra that we generalize its role in the following:

> ***Definition:*** Let P and Q be $n \times n$ matrices. We say that P is ***similar*** to Q if we can find an ***invertible*** $n \times n$ matrix R such that:
>
> $$P = R^{-1}QR.$$
>
> We also say that Q is ***conjugated*** by R to ***produce*** P, and vice versa.

This definition says that a square matrix A is ***diagonalizable*** if we can find a ***diagonal*** matrix D that is ***similar*** to A. It also says that the matrix of a linear transformation T with respect to one basis S is ***similar*** to the matrix of T with respect to ***any*** other basis B.

Example: Let $R = \begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix}$, an invertible matrix with $R^{-1} = \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix}$.

Let $Q = \begin{bmatrix} 5 & 9 \\ -3 & 7 \end{bmatrix}$. Then:

$$P = R^{-1}QR = \begin{bmatrix} -2 & 3 \\ 3 & -4 \end{bmatrix}\begin{bmatrix} 5 & 9 \\ -3 & 7 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} -67 & -51 \\ 105 & 79 \end{bmatrix}.$$

Thus, we can say that:

$$Q = \begin{bmatrix} 5 & 9 \\ -3 & 7 \end{bmatrix} \text{ is similar to } P = \begin{bmatrix} -67 & -51 \\ 105 & 79 \end{bmatrix},$$

even though you might insist that they don't "look" at all "similar" in the everyday sense. Clearly, there are deeper issues involved here, and we will explore them further. □

Equivalence Relations and Similarity

The relationship of "similarity" is an example of an important class of relationships in mathematics that are called ***equivalence relations***. First, let us make a general definition for the term "relation":

Definitions: A ***relation*** $\sim$ on the members of a set X is a ***function*** that gives a value of either ***true*** or ***false*** given an ordered pair (x,y) of members of X.
If the value of $\sim$ is ***true*** for the pair (x,y), we write this symbolically as $\boldsymbol{x \sim y}$. We also say that "x is ***related*** to y."
If the value of $\sim$ is ***false*** for the pair (x,y), we write $\boldsymbol{x \nsim y}$ and say that "x is ***not related*** to y."

Example: We can construct a relation among the set of integers $\mathbb{Z}$ by saying that:

$$x \sim y \textit{ if and only if } x - y \text{ is an } \textit{even} \text{ integer.}$$

Thus, $6 \sim 2$, and $11 \sim 5$, because both $6 - 2 = 4$ and $11 - 5 = 6$ are even numbers. However, $9 \nsim 6$, because $9 - 6 = 3$, which is odd. We could also have said that $x \sim y$ ***if and only if*** both x and y are even, or both x and y are odd (keeping in mind that 0 is an even number).□

This particular Example of a relation has three important qualities that elevate it to a special status:

Definitions: We say that a relation $\sim$ is an ***equivalence relation*** if it satisfies three properties: for any x, y and $z \in X$:

1. $\sim$ is ***reflexive:*** $x \sim x$, that is, every x is related to ***itself***.
2. $\sim$ is ***symmetric:*** if $x \sim y$, then $y \sim x$.
3. $\sim$ is ***transitive:*** if $x \sim y$ ***and*** $y \sim z$, then $x \sim z$.

Example: The simplest equivalence relation among the members of ***any*** set X is the relation "***is equal to***" : (1) any member x of X is equal to itself, thus $x = x$, which we know as the ***reflexive*** property of equality; (2) if $x = y$, then $y = x$, which we know as the ***symmetric*** property of equality; and (3) if $x = y$ and $y = z$, then $x = z$, which we know as the ***transitive*** property of equality.□

Example: In science, ***"has the same blood type"*** is an equivalence relation among the set of all people: a person has the same blood type as himself or herself; if Ed has the same blood type as Sue, then Sue has the same blood type as Ed. If Ed has the same blood type as Sue, and Sue has the same blood type as Pat, then Ed has the same blood type as Pat. Notice we don't say that "two people are equal," but we say that "two people have the same blood type."□

Example: The relationship ***"is a sibling of"*** among human beings (i.e., is a brother or sister of, or more precisely, have the same mother and father) is ***not*** an equivalence relation because it is ***not reflexive:*** we do not say that we are our own sibling.□

Example: In mathematics, the relationship ***"less than or equal to"*** or ***"≤"*** is not an equivalence relation on the set of real numbers because it is ***not symmetric:*** $3 \le 8$, but $8 \not\le 3$.□

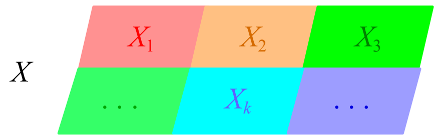
Equivalence Classes

Equivalence relations can give us a way to separate X into subsets where the members of each subset are related to each other:

Definition/Theorem: Any equivalence relation $\sim$ acting on a set X partitions X into ***equivalence classes:***

$$X = X_1 \cup X_2 \cup \cdots \cup X_k \cup \cdots,$$

where $x,\ y \in X$ belong to the same equivalence class X_i ***if and only if*** $x \sim y$. Furthermore, every element $x \in X$ belongs to ***exactly one*** equivalence class X_i. We can thus visualize X partitioned into these equivalence classes:



where two distinct equivalence classes ***do not*** intersect.
Note: it is possible to have an ***infinite*** number of equivalence classes.

Proof: The only statement that needs to be proved is that every element $x \in X$ belongs to ***exactly one*** equivalence class X_i. Since $x \sim x$, x belongs to its own equivalence class, according to the definition. Suppose that x belongs to ***two*** equivalence classes:

$$x \in X_i \quad \text{and} \quad x \in X_j.$$

We must show that $X_i = X_j$. So suppose $y \in X_i$. Then, by definition, $x \sim y$. But since $x \in X_j$ and $x \sim y$, then again, by definition, $y \in X_j$. Thus $X_i \subseteq X_j$. Repeating the argument with $z \in X_j$, we also see that $X_j \subseteq X_i$. Thus, $X_i = X_j$.■

Examples: In our first Example, we defined a relation on the set of integers $\mathbb{Z}$ by saying that:

$$x \sim y \ \textit{\textbf{if and only if}}\ x - y \text{ is an } \textit{\textbf{even}} \text{ integer.}$$

Let us check that this is an equivalence relation. For any $x \in \mathbb{Z}$, $x - x = 0$, which is an even integer, so $x \sim x$ and the relation is ***reflexive***. Now, if $x \sim y$, then $x - y = 2n$ for some integer n. But then $y - x = -2n = 2(-n)$, which is also an even integer, and so the relation is ***symmetric***. Finally, if $x \sim y$ and $y \sim z$, then:

$$x - y = 2n, \ \text{and} \ y - z = 2m, \ \text{thus:}$$

$$x - z = 2n + 2m = 2(n + m),$$

for some integers n and m. Since $2(n+m)$ is also an even integer, $x \sim z$ and the relation is ***transitive***.

Let us think of the equivalence classes of $\mathbb{Z}$ under $\sim$. Consider the integer 0. Under our relation, $x \sim 0$ ***if and only if*** $x - 0 = 2n$, so in other words, x is an ***even*** integer. Thus, one equivalence class is:

$$X_1 = \{\text{even integers}\}.$$

This leads us to suspect that the odd numbers must form another equivalence class. Indeed, the number 1 is not a member of X_1. Under our relation, $x \sim 1$ ***if and only if*** $x - 1 = 2n$, so in other words, x is an ***odd*** integer. Since every integer is either even or odd, we can therefore conclude that there is only one other equivalence class, and that is:

$$X_2 = \{\text{odd integers}\}.$$

Thus, we have the partitioning of X into two equivalence classes:

$$\mathbb{Z} = X_1 \cup X_2 = \{\text{even integers}\} \cup \{\text{odd integers}\}. \square$$

Example: The relation "has the same blood type" is an equivalence relation, and this partitions the set of all human beings into those of type A, type B, type AB and type O, thus creating ***four*** equivalence classes (or eight, if you want to monkey around with Rhesus positives and negatives).□

Similarity as an Equivalence Relation

Now let us go back to similarity of matrices:

> ***Theorem:*** The relationship "***is similar to***," symbolized by ~, is an ***equivalence relation*** on the set of all $n \times n$ matrices.
> In other words, for all $n \times n$ matrices P, Q and S:
>
> 1. ***Similarity is Reflexive:*** $P \sim P$.
> 2. ***Similarity is Symmetric:*** If $P \sim Q$, then $Q \sim P$.
> 3. ***Similarity is Transitive:*** If $P \sim Q$ and $Q \sim S$, then $P \sim S$.

Because of the symmetric property, we can say that P and Q are similar to ***each other***.

Proof:

The Reflexive Property: The identity matrix $\boldsymbol{I}_n$ is invertible, and $P = \boldsymbol{I_n} P \boldsymbol{I_n^{-1}}$.
Thus, P is similar to itself.

The Symmetric Property: If P is similar to Q, then we can find an invertible matrix R such that:

$$P = R^{-1}QR.$$

But we can also move R^{-1} and R to the left side of the equation, yielding: $RPR^{-1} = Q$.
This equation now says that Q is similar to P.

The Transitive Property: If P is similar to Q, and Q is similar to S, then we can find ***two*** invertible matrices R_1 and R_2 such that:

$$P = R_1 Q R_1^{-1}, \quad \text{and} \quad Q = R_2 S R_2^{-1}.$$

Now, we substitute the second equation into the first, and get:

$$P = R_1(R_2 S R_2^{-1})R_1^{-1} = (R_1R_2)S(R_2^{-1}R_1^{-1}) = (R_1R_2)S(R_1R_2)^{-1},$$

and thus P is similar to S. Notice that we used the formula for the inverse of a product, and the fact that the product of two invertible matrices is also invertible.■

Invariant Properties under Similarity

As expected, the word "similar" is loaded with meaning, and has many consequences. We list below properties that are shared by two similar matrices, starting with one we have already proven. The proofs of the rest of the properties are left as Exercises:

Theorem: Let P and Q be similar $n \times n$ matrices, that is, $P \sim Q$. Then, the following properties are true:

1. ***Equality of Determinants*** — $det(P) = det(Q)$
2. ***Equality of Nullities*** — $nullity(P) = nullity(Q)$
3. ***Equality of Ranks*** — $rank(P) = rank(Q)$
4. ***Equality of Traces*** — $tr(P) = tr(Q)$
5. ***Equality of Characteristic Polynomials*** — $det(\lambda \boldsymbol{I_n} - P) = det(\lambda \boldsymbol{I_n} - Q)$
6. ***Common Eigenvalues*** — The eigenvalues of P are identical to the eigenvalues of Q.
7. ***Equality of Algebraic Multiplicities . . .*** — If λ is a common eigenvalue of P and Q, then $AlgMult(P,\lambda) = AlgMult(Q,\lambda)$
8. ***. . . and Geometric Multiplicities*** — and $dim(Eig(P,\lambda)) = dim(Eig(Q,\lambda))$.
9. ***Invariance of Invertibility*** — P is invertible ***if and only if*** Q is invertible.
10. ***Invariance of Diagonalizability*** — P is diagonalizable ***if and only if*** Q is diagonalizable.

Notice that the properties above are ***necessary*** if two matrices are similar to each other, but they are ***not sufficient*** to prove that two matrices are indeed similar to each other. Thus, if two matrices have ***different*** ranks, then they ***cannot*** possibly be similar. However, two matrices with the ***same*** rank ***may or may not*** be similar.

Also, we should point out that ***eigenvectors*** are not preserved by similarity. Even though P and Q have identical eigenvalues, an eigenvector for P with respect to λ may not be an eigenvector for Q with respect to λ. We will see in the Exercises, though, that there is an elegant connection between the associated ***eigenspaces***.

Let us bring back the definition of the trace function, which we saw in Exercise 28 of Section 3.5, where you were asked to prove its linearity properties:

Definition: The ***trace*** of an $n \times n$ matrix A is the linear transformation:

$$tr \;:\; Mat(n,n) \to \mathbb{R}, \text{ given by:}$$

$$tr(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}.$$

Example: Let $A = \begin{bmatrix} 5 & 3 & -2 \\ 7 & 4 & 6 \\ 2 & 0 & -3 \end{bmatrix}$. Then $tr(A) = 5 + 4 - 3 = 6$. □

Example: Consider the matrices $P = \begin{bmatrix} 3 & 2 \\ 0 & 3 \end{bmatrix}$, and $Q = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$.

Both P and Q have determinant 9, trace 6, and characteristic polynomial $p(\lambda) = (\lambda - 3)^2$. So it looks like P and Q should be similar to each other. However, $Q = 3\boldsymbol{I_2}$, so if R is invertible, then:

$$R^{-1}QR = R^{-1}(3\boldsymbol{I_2})R = 3R^{-1}R = 3\boldsymbol{I_2} = Q,$$

so it is impossible to get P no matter what R is. Thus, P and Q are ***not*** similar. In fact, the above computation shows that Q is the ***only*** matrix that is similar to Q! We can also rephrase this by saying that Q is the only ***only member*** of its ***equivalence class*** under similarity. □

Eigentheory for Operators

We are now ready to define the characteristic polynomial, eigenvalues, eigenvectors and eigenspaces for linear operators:

Definition/Theorem: If $T : V \to V$ is a linear operator acting on a ***finite dimensional*** vector space V, then we can define the ***characteristic polynomial*** of T to be the characteristic polynomial of any matrix $[T]_S$ for T, that is, with respect to any basis S of V.

We say that λ is an ***eigenvalue*** of T if λ is a root of the characteristic polynomial of T. We say that a non-zero vector $\vec{v} \in V$ is an ***eigenvector*** for T associated to λ, if:

$$[T]_S[\vec{v}]_S = \lambda[\vec{v}]_S,$$

again, for any choice of basis S.

More generally, if V is infinite dimensional, we say that λ is an eigenvalue for T and a non-zero vector $\vec{v} \in V$ is an eigenvector for T associated to λ if:

$$T(\vec{v}) = \lambda\vec{v}.$$

We will denote the corresponding ***eigenspace*** by:

$$Eig(T,\lambda) = \{\vec{v} \in V | T(\vec{v}) = \lambda\vec{v}\},$$

which is a subspace of V. Again, $\vec{\mathbf{0}}_V \in Eig(T,\lambda)$, even though $\vec{\mathbf{0}}_V$ is ***never*** an eigenvector.

Proof: We already know that any two matrices for T (with respect to two different bases), are similar, thus we have the same characteristic polynomials and the same eigenvalues. Now, if a non-zero vector $\vec{v} \in V$ is an eigenvector for T associated to λ, and S and B are two different bases for V, we have to show that:

$$[T]_S[\vec{v}]_S = \lambda[\vec{v}]_S \textit{ if and only if } [T]_B[\vec{v}]_B = \lambda[\vec{v}]_B,$$

that is, it does not matter how we choose ***coordinates*** for V and T. But from the previous Sections, we have the relationships:

$$[\vec{v}]_S = C_{B,S}[\vec{v}]_B, \quad \text{and} \quad [T]_S = [B]_S[T]_B[B]_S^{-1},$$

where $C_{B,S}$ is the change of basis matrix from B to S. In general, we saw that if B, B' and S are any three bases for V, then:

$$C_{B,B'} = [B']_S^{-1}[B]_S.$$

Thus, if we let $B' = S$, we get:

$$C_{B,S} = [S]_S^{-1}[B]_S = \boldsymbol{I}_n^{-1}[B]_S = [B]_S,$$

since the coordinates of the members of S with respect to S form the identity matrix. Thus, we are ready to substitute:

$$\begin{aligned}[T]_S[\vec{v}]_S &= [B]_S[T]_B[B]_S^{-1} \cdot C_{B,S}[\vec{v}]_B \\ &= [B]_S[T]_B[B]_S^{-1} \cdot [B]_S[\vec{v}]_B = [B]_S[T]_B[\vec{v}]_B.\end{aligned}$$

Thus, we always get $[T]_S[\vec{v}]_S = [B]_S[T]_B[\vec{v}]_B$.

Now, suppose $[T]_B[\vec{v}]_B = \lambda[\vec{v}]_B$. We must show that $[T]_S[\vec{v}]_S = \lambda[\vec{v}]_S$. But:

$$[T]_S[\vec{v}]_S = [B]_S[T]_B[\vec{v}]_B = [B]_S \cdot \lambda[\vec{v}]_B = \lambda \cdot [B]_S[\vec{v}]_B = \lambda \cdot [\vec{v}]_S,$$

using the Change of Basis Formula in the last step.

Similarly, we can reverse this process because $[B]_S$ is an ***invertible*** matrix, and $[B]_S^{-1} = [S]_B$. Thus, the eigenvectors of T with respect to λ do not depend on the choice of basis for V. The proof that $Eig(V, \lambda)$ is a subspace of V is identical to our proof for eigenspaces of a matrix A as subspaces of $\mathbb{R}^n$.■

Example: Consider the differentiation operator D on the infinite dimensional function space $C^1(\mathbb{R})$.

The constant functions $f(x) = c$ are the only functions with zero derivatives, and thus they are eigenvectors for $\lambda = 0$ (in other words, they make up the ***nullspace*** of this operator). Similarly:

$$\frac{d}{dx}e^{5x} = 5e^{5x},$$

and therefore $f(x) = e^{5x}$ is an eigenvector for $\lambda = 5$. More generally, $f(x) = e^{\lambda x}$ is an eigenvector for any real number λ. Notice that we do not have a characteristic polynomial to help us find these eigenvalues and eigenvectors. We will need a course in ***Differential Equations*** to prove that in fact the ***only*** eigenvectors of D from $C^1(\mathbb{R})$ associated to λ are of the form $f(x) = Ce^{\lambda x}$.□

Example: Let us bring back $T : \mathbb{P}^2 \to \mathbb{P}^2$, the linear operator given by:

$$T(p(x)) = 2p(x) + (x+5)p'(x) + (x^2 - 3x + 7)p''(x).$$

We found that its matrix with respect to the standard basis $S = \{1, x, x^2\}$ is:

$$[T]_S = \begin{bmatrix} 2 & 5 & 14 \\ 0 & 3 & 4 \\ 0 & 0 & 6 \end{bmatrix}.$$

Since this is ***upper triangular***, the characteristic polynomial is: $p(\lambda) = (\lambda - 2)(\lambda - 3)(\lambda - 6)$, with ***distinct*** eigenvalues $\lambda = 2$, 3 and 6. Thus, each eigenspace must be 1-dimensional, as we saw in Section 6.3. As before, let us simultaneously find the eigenvectors by finding the reduced row echelon forms of the matrices $[T]_S - \lambda\boldsymbol{I}_3$ for each of the values of λ:

$$[T]_S - 2\boldsymbol{I}_3 = \begin{bmatrix} 0 & 5 & 14 \\ 0 & 1 & 4 \\ 0 & 0 & 4 \end{bmatrix}, \quad \text{with rref} \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix};$$

$$[T]_S - 3\boldsymbol{I}_3 = \begin{bmatrix} -1 & 5 & 14 \\ 0 & 0 & 4 \\ 0 & 0 & 3 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & -5 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and}$$

$$[T]_S - 6\boldsymbol{I}_3 = \begin{bmatrix} -4 & 5 & 14 \\ 0 & -3 & 4 \\ 0 & 0 & 0 \end{bmatrix}, \text{ with rref } \begin{bmatrix} 1 & 0 & -31/6 \\ 0 & 1 & -4/3 \\ 0 & 0 & 0 \end{bmatrix}.$$

The ***coordinates*** with respect to S of the single basis vector for each of the respective eigenspaces are therefore:

$$\langle 1,0,0\rangle, \quad \langle 5,1,0\rangle \quad \text{and} \quad \langle 31,8,6\rangle.$$

Lastly, we ***decode*** these coordinates using the standard basis $S = \{1, x, x^2\}$, and get:

$$\begin{aligned} Eig(T,2) &= Span\{1\}, \\ Eig(T,3) &= Span\{5+x\}, \text{ and} \\ Eig(T,6) &= Span\{31+8x+6x^2\}. \end{aligned}$$

We can compute their images under T:

$$\begin{aligned} T(1) &= 2 = 2 \cdot 1, \\ T(5+x) &= 2(5+x) + (x+5)\cdot 1 = 3\cdot(5+x), \text{ and} \\ T(31+8x+6x^2) &= 2(31+8x+6x^2) + (x+5)(8+12x) + (x^2-3x+7)(12) \\ &= 186 + 48x + 36x^2 = 6\cdot(31+8x+6x^2). \end{aligned}$$

and see that they are indeed eigenvectors, respectively, for $\lambda = 2,\ 3$ and 6. □

Diagonalization of Operators

Now that we know how to find eigenvalues and eigenspaces for linear operators, we can generalize the diagonalization process to operators:

> ***Definition/Theorem:*** Let $T : V \to V$ be a linear operator acting on a finite dimensional vector space V.
> We say that T is ***diagonalizable*** if we can find a ***basis*** B for V such that $[T]_B$ is a ***diagonal*** matrix.
> Thus, T is ***diagonalizable*** ***if and only if*** $[T]_S$ is a ***diagonalizable*** matrix for ***any*** choice of basis S of V.

Example: Our previous operator $T : \mathbb{P}^2 \to \mathbb{P}^2$ is diagonalizable because $[T]_S$ has 3 distinct eigenvalues. Thus, with respect to the basis:

$$B = \{1,\ 5+x,\ 31+8x+6x^2\}$$

consisting of eigenvectors, the matrix of T is diagonal:

$$[T]_B = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$

Notice also that if we construct $[B]_S$, where $S = \{1, x, x^2\}$, then:

$$[B]_S = \begin{bmatrix} 1 & 5 & 31 \\ 0 & 1 & 8 \\ 0 & 0 & 6 \end{bmatrix}, \text{ with inverse: } [B]_S^{-1} = \begin{bmatrix} 1 & -5 & 3/2 \\ 0 & 1 & -4/3 \\ 0 & 0 & 1/6 \end{bmatrix}, \text{ and so:}$$

$$\begin{aligned} [B]_S^{-1}[T]_S[B]_S &= \begin{bmatrix} 1 & -5 & 3/2 \\ 0 & 1 & -4/3 \\ 0 & 0 & 1/6 \end{bmatrix}\begin{bmatrix} 2 & 5 & 14 \\ 0 & 3 & 4 \\ 0 & 0 & 6 \end{bmatrix}\begin{bmatrix} 1 & 5 & 31 \\ 0 & 1 & 8 \\ 0 & 0 & 6 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{bmatrix} = [T]_B, \end{aligned}$$

as it should be.□

6.7 Section Summary

Let P and Q be $n \times n$ matrices. We say that P is ***similar*** to Q, and write $P \sim Q$ if we can find an ***invertible*** $n \times n$ matrix R such that $P = R^{-1}QR$.

The relationship $\sim$ is an ***equivalence relation*** on the set of $n \times n$ matrices. This means that for all $n \times n$ matrices P, Q and S:

1. ***Similarity is Reflexive:*** $P \sim P$.
2. ***Similarity is Symmetric:*** If $P \sim Q$, then $Q \sim P$.
3. ***Similarity is Transitive:*** If $P \sim Q$, and $Q \sim S$, then $P \sim S$.

Similarity preserves many properties: ***If P ~ Q, then:***

- $det(P) = det(Q)$.
- P is invertible ***if and only if*** Q is invertible.
- $nullity(P) = nullity(Q)$.
- $rank(P) = rank(Q)$.
- the characteristic polynomial of P and Q are exactly the same.
- the eigenvalues of P are exactly the same as the eigenvalues of Q.
- if λ is a common eigenvalue, then the algebraic and geometric multiplicities of λ are the same for P and Q.
- P is diagonalizable ***if and only if*** Q is diagonalizable.
- $tr(P) = tr(Q)$, where the ***trace*** of A is: $tr(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}$.

The converses of these implications can be ***false:*** for example, two matrices can have exactly the same characteristic polynomial, but they may ***not*** be similar.

If $T : V \to V$ is a linear operator acting on a finite dimensional vector space V, then we can define the characteristic polynomial of T to be the characteristic polynomial of any matrix $[T]_S$ for T, that is, with respect to any basis S of V.

We say that λ is an ***eigenvalue*** of T if λ is a root of the characteristic polynomial of T. We say that a non-zero vector $\vec{v} \in V$ is an ***eigenvector*** for T associated to λ, if $[T]_S[\vec{v}]_S = \lambda[\vec{v}]_S$, again, for any choice of basis S. More generally, even if V is infinite dimensional, we say that λ is an eigenvalue for T and a non-zero vector $\vec{v} \in V$ is an eigenvector for T associated to λ, if $T(\vec{v}) = \lambda\vec{v}$.

We denote the corresponding ***eigenspace*** by $Eig(T, \lambda) = \{ \vec{v} \in V \mid T(\vec{v}) = \lambda\vec{v} \} \trianglelefteq V$.

If V is finite dimensional, we say that T is ***diagonalizable*** if we can find a basis B for V such that $[T]_B$ is a ***diagonal*** matrix. Thus, T is ***diagonalizable*** ***if and only if*** $[T]_S$ is a ***diagonalizable*** matrix for ***any*** choice of basis S of V.

6.7 Exercises

For Exercises (1) to (5): For each of the operators below: (a) find $[T]_S$, (b) find $det(T)$, (c) find the characteristic polynomial of T, (d) find the eigenvalues of T, (e) find a basis for each eigenspace of T, properly decoded as vectors of V, (f) diagonalize T, if possible, that is, find a basis B for which $[T]_B$ is diagonal, and find $[T]_B$ itself. You may assume (or convince yourself mentally) that T is indeed linear. If V is given as Span(S), you may safely assume that S is linearly independent and use S as the standard basis for V.

1. $T : \mathbb{P}^2 \to \mathbb{P}^2$, given by: $T(p(x)) = (3x - 5)p'(x) + (4x^2 - 7)p''(x)$; $S = \{1, x, x^2\}$.
2. $T : \mathbb{P}^3 \to \mathbb{P}^3$, given by: $T(p(x)) = 4p(x) + (2x + 5)p'(x) + (3x^2 + 2x - 4)p''(x)$; $S = \{1, x, x^2, x^3\}$.
3. $D : V \to V$, where D is the differentiation operator and: $V = Span(\{\sin(5x), \cos(5x)\})$.
4. $D : V \to V$, where D is the differentiation operator and: $V = Span(\{e^{-x}, e^{2x}, e^{5x}\})$.
5. $D : V \to V$, where D is the differentiation operator and: $V = Span(\{e^{3x}, xe^{3x}, x^2e^{3x}\})$. Warning: don't forget the product rule and chain rule for this problem.
6. Let $D^2 = D \circ D$ be the ***second derivative*** operator: $D^2 : C^2(\mathbb{R}) \to C^0(\mathbb{R})$, acting on the vector space of all twice differentiable functions with continuous first and second derivatives defined on all real numbers. Note that both of these spaces are infinite dimensional, so we cannot construct a matrix for D^2, and thus we do not have a characteristic polynomial to work with either.
   a. Show that $f(x) = \sin(x)$ and $g(x) = \cos(x)$ are both eigenvectors for D^2. What are the corresponding eigenvalues?
   b. Show that $h(x) = e^{kx}$ is an eigenvector for D^2 for all real numbers k. What is the corresponding eigenvalue?
   c. Show that every ***positive*** number μ is an eigenvalue for D^2, and find at least one eigenvector.
   d. Show that $p(x) = \sin(kx)$ is an eigenvector for D^2 for all real numbers k. What is the corresponding eigenvalue?
   e. What can we conclude for $q(x) = \cos(kx)$?

f. Show that $f(x) = \sin(x)$ and $g(x) = \cos(x)$ are both eigenvectors for $D^4 = D^2 \circ D^2$, and they have the same eigenvalue.

7. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^2$ is an operator whose matrix with respect to $S = \{1, x, x^2\}$ is:

$$[T]_S = \begin{bmatrix} 5 & 2 & -3 \\ 0 & -1 & 6 \\ 0 & 0 & 4 \end{bmatrix}.$$

Show that T is diagonalizable, find a basis B for $\mathbb{P}^2$ such that $[T]_B$ is diagonal, and find $[T]_B$.

8. Suppose that $T : \mathbb{P}^2 \to \mathbb{P}^2$ is an operator whose matrix is: $[T]_B = Diag(4, -7, 3)$, with respect to the basis $B = \{3 - 5x, 2 + x^2, 1 - x - x^2\}$.

Find $[T]_S$, where S is the standard basis $S = \{1, x, x^2\}$.

9. Show that the relationship $x \sim y$ among human beings, where $x \sim y$ if x and y have the ***same birthday*** is an equivalence relation (not necessarily on the same year).

How many equivalence classes are there?

10. Fix an integer $n > 1$. Define a relation on $\mathbb{Z}$ via: $x \sim y$ ***if and only if*** $x - y$ is a ***multiple*** of n, in other words, $x - y = kn$ for some integer k.

a. Show that this is an equivalence relation.

b. If $n = 2$, show that the equivalence classes of $\mathbb{Z}$ under this relation are the sets of ***even*** and ***odd*** integers, as seen in one of the Examples.

c. If $n = 3$, show that there are 3 equivalence classes of $\mathbb{Z}$ under this relation.

d. In general, show that there are n equivalence classes of $\mathbb{Z}$ under this relation. What is the smallest positive member in each of these equivalence classes?

11. Suppose that S is the set of ***all*** vector spaces (both finite and infinite dimensional spaces). Define a relation on S via: $V \sim W$ ***if and only if*** V is ***isomorphic*** to W.

Recall this means that there exists a linear transformation $T : V \to W$ that is both one-to-one and onto. Prove that $\sim$ is an equivalence relation.

Follow up: suppose that V has dimension n. Describe the other members of the equivalence class of V.

12. Prove that for every fixed scalar k, the only matrix that is similar to $k \cdot \boldsymbol{I_n}$ is $k \cdot \boldsymbol{I_n}$ itself.

13. Suppose that $D = Diag(d_1, d_2, \ldots, d_n)$ is a diagonal matrix, and E is another $n \times n$ diagonal matrix that contains the same entries on the main diagonal as D, except possibly in a different order. Prove that E is similar to D.

Note: the entries do not have to be distinct. Hint: Think of permutations and row operations.

14. Prove that if $S = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is a set of n ***distinct*** real numbers, and A and B are $n \times n$ matrices whose eigenvalues are exactly the members of S, then A and B are similar.

Warning: you are not allowed to use any of the properties that are preserved by similarity, since you may only use these properties if you ***already know*** that A and B are similar. Instead, recall what we know about matrices with distinct eigenvalues from Section 6.3.

15. Prove that for any two $n \times n$ matrices A and B: $tr(AB) = tr(BA)$. Hint: all you need to do is look at ***each*** diagonal entry of both AB and BA.

16. ***Properties Preserved by Similarity:*** We will complete the proofs of the properties stated in the main theorem for similar matrices. Suppose that P and Q are $n \times n$ matrices and $P \sim Q$, that is, P is ***similar*** to Q.

   a. Write down the ***definition*** of $P \sim Q$.

   b. Use (a) directly to show that $det(P) = det(Q)$.

   c. Use (a) directly to show that P is ***invertible*** ***if and only if*** Q is ***invertible***.

   d. Show that the ***characteristic polynomials*** of P and Q are exactly the ***same***.

      Hint: First show that $\lambda \boldsymbol{I_n} = R^{-1}(\lambda \boldsymbol{I_n})R$, which is basically the idea behind Exercise 13 above.

   e. Show that the ***eigenvalues*** of P are exactly the ***same*** as the ***eigenvalues*** of Q.

   f. Show that if λ is a common eigenvalue of P and Q, then the ***algebraic multiplicity*** of λ with respect to P is the ***same*** as its ***algebraic multiplicity*** with respect to Q.

   g. Preliminary to the next part: Suppose that $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a ***linearly independent*** set of vectors from $Eig(Q, \lambda)$. Prove that the set of vectors: $\{R^{-1}\vec{v}_1, R^{-1}\vec{v}_2, \ldots, R^{-1}\vec{v}_k\}$ is a linearly independent set of vectors from $Eig(P, \lambda)$, where $P = R^{-1}QR$.

      Hint: slowly compute $(R^{-1}QR)(R^{-1}\vec{v}_i)$, and use the fact that R is ***invertible***.

      State and prove an analogous statement regarding a set of vectors from $Eig(P, \lambda)$.

   h. Show that if λ is a common eigenvalue of P and Q, then the ***geometric multiplicity*** of λ with respect to P is the ***same*** as its ***geometric multiplicity*** with respect to Q.

      Hint: use the previous Exercise to convert a ***basis*** for $Eig(P, \lambda)$ to a set of ***linearly independent vectors*** from $Eig(Q, \lambda)$, and vice versa, starting with a basis for $Eig(Q, \lambda)$. What does each construction imply about the relative ***dimensions*** of $Eig(Q, \lambda)$ and $Eig(P, \lambda)$?

   i. Show that $nullity(P) = nullity(Q)$. Hint: consider two cases: both P and Q are invertible and both P and Q are not invertible.

   j. Show that $rank(P) = rank(Q)$.

   k. Show that P is ***diagonalizable*** ***if and only if*** Q is also ***diagonalizable***.

      Reminder: you must show both implications.

   l. Show that $tr(P) = tr(Q)$. Hint: Use Exercise 15.

17. Suppose that A is a ***diagonalizable*** $n \times n$ matrix, with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (possibly with repetitions). Prove that:

$$tr(A) = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

   Thus, the trace of a diagonalizable matrix is the sum of all its eigenvalues, taking into account multiplicities.

18. Let $T : V \to V$ be an operator on a finite-dimensional vector space V. Prove that T is ***diagonalizable*** ***if and only if*** there exists a basis B for V consisting of eigenvectors for T.

A Summary of Chapter 6

In this Chapter, all matrices are $n \times n$ or ***square***.

We say that λ is an ***eigenvalue*** of A, and $\vec{v}$ is an ***eigenvector*** for A associated to λ, or simply an eigenvector for λ, if $A\vec{v} = \lambda\vec{v}$, where $\vec{v}$ is a ***non-zero*** vector of $\mathbb{R}^n$.

We can find an eigenvalue λ and an eigenvector $\vec{v} \in \mathbb{R}^n$ such that $A\vec{v} = \lambda\vec{v}$ ***if and only if*** $det(\lambda I_n - A) = 0$. This equation is called the ***characteristic equation*** of A. The determinant in this equation is a polynomial whose highest term is λ^n. It is called the ***characteristic polynomial*** of A, denoted $p(A,\lambda)$ or simply $p(\lambda)$.

Let A be a ***triangular*** $n \times n$ matrix, with main diagonal entries $c_1, c_2, \ldots, c_n$. Then: $p(\lambda) = (\lambda - c_1)(\lambda - c_2)\cdots(\lambda - c_n)$, and the eigenvalues are $c_1, c_2, \ldots, c_n$.

A is ***invertible*** ***if and only if*** $\lambda = 0$ is ***not*** an ***eigenvalue*** for A.

If λ is a fixed eigenvalue of A, we define the ***eigenspace*** of A associated to λ, to be:

$$\boldsymbol{Eig}(A,\lambda) = \{\vec{v} \in \mathbb{R}^n \,|\, A\vec{v} = \lambda\vec{v}\} \trianglelefteq \mathbb{R}^n.$$

We say that A is ***diagonalizable*** if we can find an ***invertible*** matrix C such that $C^{-1}AC = Diag(\alpha_1, \alpha_2, \ldots, \alpha_n)$. We also say that C ***diagonalizes*** A.

A is ***diagonalizable*** ***if and only if*** we can find a set of n ***linearly independent eigenvectors*** for A, say $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. If this is the case, then the diagonalizing matrix C is the matrix whose ***columns*** are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and the diagonal matrix D contains the corresponding ***eigenvalues*** along the main ***diagonal***.

Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be an ordered set of eigenvectors for A, and suppose that the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ for these eigenvectors are all ***distinct***. Then, S is ***linearly independent***.

Let A have ***distinct*** (possibly imaginary) eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$. Suppose $p(\lambda)$ factors as $p(\lambda) = (\lambda - \lambda_1)^{n_1} \bullet (\lambda - \lambda_2)^{n_2} \bullet \cdots \bullet (\lambda - \lambda_k)^{n_k}$.

We call n_i the ***algebraic multiplicity*** of λ_i, and $dim(\boldsymbol{Eig}(A,\lambda_i))$ the ***geometric multiplicity*** of λ_i. We agree that $dim(\boldsymbol{Eig}(A,\lambda_i)) = 0$ if λ_i is an imaginary eigenvalue (this will be upgraded in Chapter 8).

For any eigenvalue λ_i of A, the ***geometric multiplicity*** of λ_i is ***at most*** equal to the ***algebraic multiplicity*** of λ_i. Thus, A is ***diagonalizable*** ***if and only if*** the ***geometric multiplicity*** of λ_i is ***exactly equal*** to its ***algebraic multiplicity***, for ***all*** λ_i.

If A has n ***distinct*** (real) eigenvalues, then A is ***diagonalizable***.

For any ***bases*** B and B' for $\mathbb{R}^n$, there exists an ***invertible*** $n \times n$ matrix $C_{B,B'}$ such that for all vectors $\vec{v}$ of $\mathbb{R}^n$, $[\vec{v}]_{B'} = C_{B,B'}[\vec{v}]_B$.

The matrix $C_{B,B'}$ is called the ***change of basis matrix*** from B to B'. We can explicitly find $C_{B,B'}$ by performing the Gauss-Jordan algorithm on the augmented matrix $\left[B' \,|\, B\right]$, where this notation means that we assemble as the columns of this matrix the vectors in B', followed by the vectors of B. At the end of the process, we obtain $\left[\boldsymbol{I_n} \,|\, C_{B,B'}\right]$.

Consequently, the columns of $C_{B,B'}$ are simply the ***coordinate matrices*** of the members of B with respect to B', that is, if $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, then:

$$C_{B,B'} = \left[\ [\vec{v}_1]_{B'}\ [\vec{v}_2]_{B'} \cdots [\vec{v}_n]_{B'}\ \right].$$

Moreover, $C_{B,B'}^{-1} = C_{B',B}$.

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a ***linear transformation***, B a ***basis*** for $\mathbb{R}^n$, and B' a ***basis*** for $\mathbb{R}^m$. Then: $[T] = [B'][T]_{B,B'}[B]^{-1}$, where $[B]$ is the matrix whose columns are the vectors of B, and similarly for $[B']$. In particular, if $T : \mathbb{R}^n \to \mathbb{R}^n$ is an ***operator***, and B is a basis for $\mathbb{R}^n$ (used to encode both the domain and the codomain), then $[T] = [B][T]_B[B]^{-1}$.

More generally, for any two ***bases*** B and B' for a vector space V, with $n = \dim(V)$, there exists an ***invertible*** $n \times n$ matrix $C_{B,B'}$ such that for all vectors $\vec{v} \in V$, $[\vec{v}]_{B'} = C_{B,B'}[\vec{v}]_B$. The matrix $C_{B,B'}$ is called the ***change of basis matrix*** from B to B'.

Let $T : V \to W$ be a linear transformation of finite dimensional vector spaces, B and S bases for V, and B' and S' bases for W. Then: $[T]_{S,S'} = [B']_{S'}[T]_{B,B'}[B]_S^{-1}$.

In particular, if $T : V \to V$ is an ***operator***, then $[T]_S = [B]_S[T]_B[B]_S^{-1}$.

Let P and Q be $n \times n$ matrices. We say that P is ***similar*** to Q, and write $P \sim Q$ if we can find an ***invertible*** $n \times n$ matrix R such that $P = R^{-1}QR$.

Similarity is an ***equivalence relation*** on the set of $n \times n$ matrices, that is, for all $n \times n$ matrices P, Q and S: (a) $P \sim P$, (b) If $P \sim Q$, then $Q \sim P$, (c) If $P \sim Q$, and $Q \sim S$, then $P \sim S$.

Similarity preserves many properties: ***If P ~ Q, then:***

- $det(P) = det(Q)$.
- P is invertible ***if and only if*** Q is invertible.
- $nullity(P) = nullity(Q)$.
- $rank(P) = rank(Q)$.
- the characteristic polynomial of P and Q are exactly the same.
- the eigenvalues of P are exactly the same as the eigenvalues of Q.
- if λ is a common eigenvalue, then the algebraic and geometric multiplicities of λ are the same with respect to P or Q.
- P is diagonalizable ***if and only if*** Q is diagonalizable.
- $tr(P) = tr(Q)$, where $tr(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}$.

Let $T : V \to V$ be a linear operator acting on a finite dimensional vector space V. If S and B are bases for V, then $[T]_S$ is ***similar*** to $[T]_B$. Thus, we can define the ***determinant*** of T as: $det(T) = |T| = det([T]_B)$, where B is any basis for V, and this number does not depend on the choice of B. The ***characteristic polynomial*** of T is the characteristic polynomial of any matrix $[T]_B$ for T.

We say that λ is an ***eigenvalue*** of T if λ is a root of the characteristic polynomial of T. We say that a ***non-zero*** vector $\vec{v} \in V$ is an ***eigenvector*** for T associated to λ, if $[T]_B[\vec{v}]_B = \lambda[\vec{v}]_B$, again, for any choice of basis B.

More generally, even if V is infinite dimensional, we say that λ is an ***eigenvalue*** for T and a ***non-zero*** vector $\vec{v} \in V$ is an ***eigenvector*** for T associated to λ, if $T(\vec{v}) = \lambda\vec{v}$. We denote the corresponding ***eigenspace*** by $Eig(T, \lambda) = \{\vec{v} \in V | T(\vec{v}) = \lambda\vec{v}\} \trianglelefteq V$.

If V is finite dimensional, we say that T is ***diagonalizable*** if we can find a basis B for V such that $[T]_B$ is a ***diagonal*** matrix. Thus, T is ***diagonalizable*** ***if and only if*** $[T]_S$ is a ***diagonalizable*** matrix for ***any*** choice of basis S of V.

Chapter 7

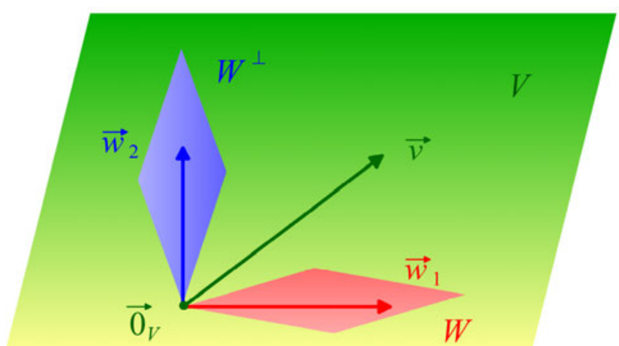
Geometry in the Abstract:

Inner Product Spaces

In this Chapter, we will look for generalizations of the dot product operation in abstract vector spaces, which are called ***inner products***. We will require that these inner products possess four of the properties that dot products possess, and from these properties, derive other properties that are shared with the dot product. Because of these properties, we will see that the ***Cauchy-Schwarz Inequality*** from Chapter 1 is still true in such an inner product space, and thus we will be able to generalize the concepts of the ***length*** of a vector, and the ***angle*** and ***distance*** between two vectors. In particular, we will be able to decide when two vectors are ***orthogonal*** or perpendicular to each other.

We will show that for any subspace W of an inner product space, we can construct the ***orthogonal complement***, $W^{\perp}$, such than any member of W is orthogonal to any member of $W^{\perp}$. Recall that we did this in Chapter 1 for subspaces of Euclidean space, and to find a basis for $W^{\perp}$, we find the nullspace of a matrix whose rows form a basis for W. Unfortunately, this does not generalize well in an abstract inner product space, but ***The Gram-Schmidt Algorithm*** will do this for us.

When we constructed the ***projection*** and ***reflection*** operators across lines and planes in 2- and 3-dimensional Euclidean space back in Chapter 2, we first showed that we can always decompose a vector in these spaces as the sum of a vector on the given line or plane, and a vector orthogonal to this line or plane. Similarly, we will generalize this ***orthogonal decomposition*** in terms of pairs of subspaces W and $W^{\perp}$ of an inner product space: any vector $\vec{v} \in V$ can be expressed as a sum: $\vec{v} = \vec{w}_1 + \vec{w}_2$, where $\vec{w}_1 \in W$ and $\vec{w}_2 \in W^{\perp}$. Likewise we will generalize the construction of a projection operator onto a subspace W of V.



We will see a special family of invertible matrices, called ***orthogonal matrices***, that have the special property that their ***inverse*** is simply their ***transpose***. This will require an orthogonality condition among the rows and columns of the matrix. We will demonstrate that ***symmetric matrices*** can always be ***diagonalized*** by an orthogonal matrix, but this property will be proven in Chapter 8 in greater generality. We will find approximate solutions to inconsistent linear systems using The ***Method of Least Squares*** and factor a matrix with independent columns into an orthogonal and upper triangular matrix (called the ***QR-decomposition***).

7.1 Axioms for an Inner Product Space

Way back in Chapter 1, at the beginning of our voyage, we saw the dot product in $\mathbb{R}^n$:

$$\vec{u} \circ \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n,$$

where $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$ and $\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$ as usual. The dot product takes two vectors $\vec{u}$ and $\vec{v}$ and gives a ***scalar*** as its value. We derived several desirable properties of the dot product, and we will use four of them in order to generalize its construction in abstract vector spaces:

Definition — The Axioms of an Inner Product Space:
Let V be a vector space. A ***bilinear form*** $\langle | \rangle$ on V is a ***function*** that takes ***two vectors*** $\vec{u}$, $\vec{v} \in V$, and produces a ***scalar***, denoted $\langle \vec{u} | \vec{v} \rangle$.
An ***inner product*** on V is a bilinear form on V, such that the following properties are satisfied by all vectors $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$:

1. ***The Symmetric Property*** $\quad \langle \vec{u} | \vec{v} \rangle = \langle \vec{v} | \vec{u} \rangle.$
2. ***The Homogeneity Property*** $\quad \langle k \cdot \vec{u} | \vec{v} \rangle = k \cdot \langle \vec{u} | \vec{v} \rangle.$
3. ***The Additivity Property*** $\quad \langle \vec{u} + \vec{v} | \vec{w} \rangle = \langle \vec{u} | \vec{w} \rangle + \langle \vec{v} | \vec{w} \rangle.$
4. ***The Positivity Property*** $\quad$ If $\vec{v} \neq \vec{\mathbf{0}}_V$, then $\langle \vec{v} | \vec{v} \rangle > 0.$

We also say that V is an ***inner product space*** under the inner product $\langle | \rangle$.

Notice that the Positivity Property deals only with ***non-zero*** vectors. It turns out that the Additivity and Symmetric Properties are enough to show that the inner product of the zero vector with any vector, as expected, is zero:

Theorem: Let V be an inner product space. Then, for any $\vec{v} \in V$:

$$\left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle = \left\langle \vec{\mathbf{0}}_V | \vec{v} \right\rangle = 0.$$

In particular:

$$\left\langle \vec{\mathbf{0}}_V | \vec{\mathbf{0}}_V \right\rangle = 0.$$

Proof: Since we know that $\vec{\mathbf{0}}_V = \vec{\mathbf{0}}_V + \vec{\mathbf{0}}_V$, we get:

$$\left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle = \left\langle \vec{v} | \vec{\mathbf{0}}_V + \vec{\mathbf{0}}_V \right\rangle = \left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle + \left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle,$$

by the Additivity Property. Since $\left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle$ is ***some*** real number, it has a ***negative.*** We can add this negative to both sides of the equation and get:

$$\left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle + \left(-\left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle \right) = \left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle + \left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle + \left(-\left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle \right), \quad \text{thus}$$

$$0 = \left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle + 0 = \left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle.$$

By the Symmetric Property, $\left\langle \vec{\mathbf{0}}_V | \vec{v} \right\rangle = \left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle = 0$ as well. ■

Note that the Symmetric Property gives us Homogeneity and Additivity properties in the ***right*** vector as well, that is:

$$\langle \vec{u} | k \cdot \vec{v} \rangle = k \cdot \langle \vec{u} | \vec{v} \rangle, \quad \text{and}$$
$$\langle \vec{u} | \vec{v} + \vec{w} \rangle = \langle \vec{u} | \vec{v} \rangle + \langle \vec{u} | \vec{w} \rangle.$$

If $k = -1$, we also have:

$$\begin{aligned} \langle \vec{u} - \vec{v} | \vec{w} \rangle &= \langle \vec{u} + (-1 \cdot \vec{v}) | \vec{w} \rangle = \langle \vec{u} | \vec{w} \rangle + \langle -1 \cdot \vec{v} | \vec{w} \rangle \\ &= \langle \vec{u} | \vec{w} \rangle - \langle \vec{v} | \vec{w} \rangle \quad \text{and similarly,} \\ \langle \vec{u} | \vec{v} - \vec{w} \rangle &= \langle \vec{u} | \vec{v} \rangle - \langle \vec{u} | \vec{w} \rangle. \blacksquare \end{aligned}$$

Euclidean spaces under the ordinary dot product are inner product spaces according to our definition above, where we write $\vec{u} \circ \vec{v}$ instead of $\langle \vec{u} | \vec{v} \rangle$. Obviously, there are many other kinds of inner products, and we will now see several types and examples of them.

Weighted Dot Products

The easiest way to change the dot product is to incorporate a list of ***weights*** for each term. For this purpose, let $\gamma_1, \gamma_2, \ldots, \gamma_n$ be n ***positive*** numbers. We define a new inner product on $\mathbb{R}^n$ by:

$$\langle \vec{u} | \vec{v} \rangle = \gamma_1 u_1 v_1 + \gamma_2 u_2 v_2 + \cdots + \gamma_n u_n v_n.$$

Example: Let us consider $\mathbb{R}^3$ under the bilinear form:

$$\langle \vec{u} | \vec{v} \rangle = 3u_1v_1 + 5u_2v_2 + 2u_3v_3,$$

where $\vec{u}$ and $\vec{v}$ are written as usual. Here, our weights are 3, 5 and 2. For example, if $\vec{u} = \langle 4, -1, 6 \rangle$ and $\vec{v} = \langle 2, 2, -3 \rangle$, then:

$$\langle \vec{u} | \vec{v} \rangle = 3 \cdot 4 \cdot 2 + 5 \cdot (-1) \cdot 2 + 2 \cdot 6 \cdot (-3) = -22.$$

Note that in contrast, $\vec{u} \circ \vec{v} = 8 - 2 - 18 = -12$.

The bilinear form is obviously ***symmetric*** (we can reverse u_i and v_i) and ***homogeneous*** (we can factor out k). Let us verify that it is ***additive:***

$$\begin{aligned} \langle \vec{u} + \vec{v} | \vec{w} \rangle &= 3(u_1 + v_1)w_1 + 5(u_2 + v_2)w_2 + 2(u_3 + v_3)w_3 \\ &= 3u_1w_1 + 3v_1w_1 + 5u_2w_2 + 5v_2w_2 + 2u_3w_3 + 2v_3w_3 \\ &= \langle \vec{u} | \vec{w} \rangle + \langle \vec{v} | \vec{w} \rangle, \end{aligned}$$

after some rearrangements. Now, suppose $\vec{v} \in \mathbb{R}^3$ is a ***non-zero*** vector, and consider:

$$\langle \vec{v} | \vec{v} \rangle = 3v_1^2 + 5v_2^2 + 2v_3^2.$$

Since at least one coordinate v_1, v_2 or v_3 is ***not zero***, one of these squares is ***strictly positive***, and thus $\langle \vec{v} | \vec{v} \rangle > 0$. Therefore our inner product is ***positive***.$\square$

The ideas behind these calculations can of course be generalized to show that this bilinear form is an inner product for all positive weights $\gamma_1, \gamma_2, \ldots, \gamma_n$.

Inner Products Generated by Isomorphisms

We can generalize the dot product in $\mathbb{R}^n$ further by considering any ***isomorphism:***

$$T : \mathbb{R}^n \to \mathbb{R}^n,$$

(that is, a one-to-one and onto operator) and define a new inner product on $\mathbb{R}^n$ by:

$$\langle \vec{u} | \vec{v} \rangle = T(\vec{u}) \circ T(\vec{v}).$$

This looks fairly abstract, but recall that $[T]$ is an ***invertible*** $n \times n$ matrix, and we will use this invertibility to show that all the properties of an inner product are satisfied.

Example: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ be given by:

$$[T] = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix}.$$

Since $\det([T]) = 1$, T is invertible. Now, suppose $\vec{u} = \langle 4, -7 \rangle$ and $\vec{v} = \langle -1, 6 \rangle$. Then:

$$T(\vec{u}) = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} 4 \\ -7 \end{bmatrix} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}, \text{ and}$$

$$T(\vec{v}) = \begin{bmatrix} 5 & 3 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 6 \end{bmatrix} = \begin{bmatrix} 13 \\ 9 \end{bmatrix}.$$

Thus, we have:

$$\langle \vec{u} | \vec{v} \rangle = T(\vec{u}) \circ T(\vec{v}) = \langle -1, -2 \rangle \circ \langle 13, 9 \rangle = -1 \bullet 13 + (-2) \bullet 9 = -31.$$

Again, note that $\vec{u} \circ \vec{v} = 4(-1) + (-7) \bullet 6 = -46$ is different from $\langle \vec{u} | \vec{v} \rangle = -31$.

Now let us verify the four properties, where $T : \mathbb{R}^n \to \mathbb{R}^n$ is ***any*** isomorphism. The bilinear form is obviously ***symmetric*** because the dot product is symmetric. It is also ***homogenous*** because T is homogenous:

$$\langle k \bullet \vec{u} | \vec{v} \rangle = T(k \bullet \vec{u}) \circ T(\vec{v}) = k \bullet T(\vec{u}) \circ T(\vec{v}) = k \bullet \langle \vec{u} | \vec{v} \rangle.$$

Likewise, we will leave it to the reader to show that the bilinear form is ***additive*** because T is additive. Finally, suppose $\vec{v} \in \mathbb{R}^n$. Then:

$$\langle \vec{v} | \vec{v} \rangle = T(\vec{v}) \circ T(\vec{v}) = \| T(\vec{v}) \|^2,$$

where we used the property that $\|\vec{x}\|^2 = \vec{x} \circ \vec{x}$ for any vector $\vec{x} \in \mathbb{R}^n$, and $\|\vec{x}\|$ is the length of $\vec{x}$. But we know that if $\vec{v} \neq \vec{\mathbf{0}}_n$, then $T(\vec{v}) \neq \vec{\mathbf{0}}_n$ because T is an isomorphism and thus $\ker(T) = \left\{ \vec{\mathbf{0}}_n \right\}$ only. Thus $\| T(\vec{v}) \|^2 > 0$ and our inner product is ***positive***. $\square$

In the Exercises, you will prove that this can be further generalized to abstract vector spaces, and not just Euclidean spaces.

Polynomial Evaluations

Let us now look at an example of an inner product in our polynomial spaces $\mathbb{P}^n$. Let us randomly choose $n+1$ ***distinct*** numbers $x_1, x_2, \ldots, x_{n+1}$. Now, define:

$$\langle p(x) | q(x) \rangle = p(x_1)q(x_1) + p(x_2)q(x_2) + \cdots + p(x_n)q(x_n) + p(x_{n+1})q(x_{n+1}).$$

Example: Let us consider $\mathbb{P}^2$, $x_1 = -2$, $x_2 = 0$, $x_3 = 1$, and the bilinear form:

$$\langle p(x) | q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$

For example, let $p(x) = 3x^2 - 5x + 2$, and $q(x) = 7x - 6$. In order to compute the inner product, let us make a table of values:

x_i	$p(x_i)$	$q(x_i)$
-2	24	-20
0	2	-6
1	0	1

Thus we get:

$$\langle 3x^2 - 5x + 2 | 7x - 6 \rangle = 24(-20) + 2(-6) + 0 \cdot 1 = -492.$$

Let us proceed with checking the four properties:

The bilinear form is clearly ***symmetric*** and ***homogeneous*** from the definition (just change order and factor out the k, respectively). Let us check the ***additivity*** property:

$$\begin{aligned}
&\langle p(x) + q(x) | r(x) \rangle \\
&= [p(-2) + q(-2)]r(-2) + [p(0) + q(0)]r(0) + [p(1) + q(1)]r(1) \\
&= p(-2)r(-2) + q(-2)r(-2) + p(0)r(0) + q(0)r(0) + p(1)r(1) + q(1)r(1) \\
&= p(-2)r(-2) + p(0)r(0) + p(1)r(1) + q(-2)r(-2) + q(0)r(0) + q(1)r(1) \\
&= \langle p(x) | r(x) \rangle + \langle q(x) | r(x) \rangle,
\end{aligned}$$

so the inner product is additive. Lastly, if $p(x)$ is any non-zero polynomial, then:

$$\langle p(x) | p(x) \rangle = [p(-2)]^2 + [p(0)]^2 + [p(1)]^2.$$

Since $p(x)$ has degree at most 2, it has ***at most 2 real roots***. Thus, at least one of the three values $p(-2)$, $p(0)$ or $p(1)$ is ***non-zero***, and therefore the corresponding square is strictly positive. Thus $\langle p(x) | p(x) \rangle > 0$, and the inner product is ***positive***. $\square$

This proof also shows that we need to evaluate our polynomials at $n+1$ ***distinct*** x-coordinates to define our inner product, in order to guarantee the positivity property for a polynomial of degree at most n. Recall that the dimension of $\mathbb{P}^n$ is $n+1$, and it is not a coincidence that we need this many scalars.

Inner Products Induced by Integrals

Let us bring Calculus into the picture. We will consider all ***continuous*** functions on a closed interval $I = [a, b]$, that is, the members of the vector space $C(I)$. We will define an inner product on this space as:

$$\langle f(x) | g(x) \rangle = \int_a^b f(x) \cdot g(x)\, dx.$$

We know that the product of two continuous functions is again continuous, so this integral definitely exists.

Example: Let $I = [0, \pi/2]$, so we have:

$$\langle f(x) | g(x) \rangle = \int_0^{\pi/2} f(x) \cdot g(x)\, dx.$$

Suppose $f(x) = \sin(x)$ and $g(x) = \cos(x)$. Then:

$$\langle \sin(x) | \cos(x) \rangle = \int_0^{\pi/2} \sin(x) \cdot \cos(x)\, dx = \frac{1}{2} \sin^2(x) \Big|_0^{\pi/2} = \frac{1}{2}.$$

where we used the substitution $u = \sin(x)$ in order to find the antiderivative.□

Now let us check that the four properties of an inner product space are valid for ***any*** interval $I = [a, b]$. Again, the ***symmetric*** and ***homogenous*** properties are easily verified. As for the ***additive*** property:

$$\begin{aligned} \langle f(x) + g(x) | h(x) \rangle &= \int_a^b [(f(x) + g(x)) \cdot h(x)]\, dx \\ &= \int_a^b f(x) \cdot h(x)\, dx + \int_a^b g(x) \cdot h(x)\, dx \\ &= \langle f(x) | h(x) \rangle + \langle g(x) | h(x) \rangle, \end{aligned}$$

and thus additivity is ***inherited*** from the additivity of the definite integral that we know is true from Calculus. However, it is again the ***positivity*** property that is most difficult to prove. Let $f(x)$ be any continuous function on I. Then:

$$\langle f(x) | f(x) \rangle = \int_a^b [f(x)]^2\, dx.$$

If $f(x)$ is ***not*** the zero function, then we must have $f(c) \neq 0$ for ***at least one*** point $c \in [a, b]$. Thus, $[f(c)]^2 > 0$. To simplify our notation for the rest of the proof, let us write:

$g(x) = [f(x)]^2$, a continuous function on all of $[a, b]$, with

$g(x) \geq 0$ for ***all*** $x \in [a, b]$, and $g(c) > 0$.

Notice that we used the property that the product of two continuous functions is also continuous, and so $[f(x)]^2$ is also continuous.

In order to prove the positivity property, we will need to ***rigorously define*** two terms from Calculus. Since we are dealing with a continuous function, we will first recall the definition of ***continuity:***

Definition: We say that $g(x)$ is ***continuous*** at $x = c$ if $\lim_{x \to c} g(x) = g(c)$.

Since it appears above, the second concept we must recall is the definition of a ***limit:***

Definition: We say that $\lim_{x \to c} g(x) = L$ if for any positive number ε, we can find another positive number δ, such that: ***if*** $0 < |x - c| < \delta$, ***then*** $|g(x) - L| < \varepsilon$.

We include the inequality $0 < |x - c|$ in the definition above to indicate that $|g(x) - L| < \varepsilon$ ***except possibly*** at $x = c$, because it is not necessary for a function to be defined at $x = c$ in order for it to have a limit at this point. Notice, however, that if $g(x)$ is also ***continuous*** at $x = c$, then $L = g(c)$, so $g(x) - L = 0$, and thus $|g(x) - L| < \varepsilon$ is automatically true. Thus, the concluding implication above can be slightly simplified to:

if $|x - c| < \delta$, ***then*** $|g(x) - L| < \varepsilon$.

Now we are ready for our Proof. For simplicity, let us assume that c is not an endpoint of $[a, b]$. By the definition of continuity, we must have:

$$\lim_{x \to c} g(x) = g(c).$$

Now, the definition of "limit" says that we can ***choose*** any positive number ε so that another positive number δ ***exists*** such that the displayed implication above is true. For reasons we will see below, we will choose the positive number $\varepsilon = \frac{1}{2} g(c)$. The definition guarantees that we can find δ such that:

$$\textbf{\textit{if }} |x - c| < \delta, \textbf{\textit{ then }} |g(x) - g(c)| < \tfrac{1}{2} g(c).$$

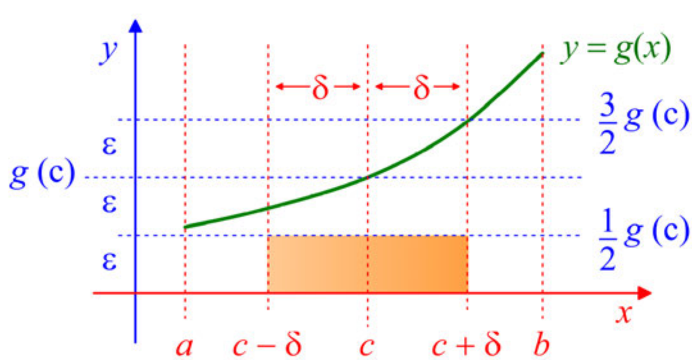
We can simplify the two inequalities above into compound inequalities:

$$\textbf{\textit{if }} c - \delta < x < c + \delta, \textbf{\textit{ then }} g(c) - \tfrac{1}{2} g(c) < g(x) < g(c) + \tfrac{1}{2} g(c).$$

Finally, simplifying this, we get:

$$\textbf{\textit{if }} c - \delta < x < c + \delta, \textbf{\textit{ then }} \tfrac{1}{2} g(c) < g(x) < \tfrac{3}{2} g(c).$$

We may also assume that $[c - \delta, c + \delta]$ is safely within $[a, b]$. These two inequalities are illustrated in the diagram below:



The Definition of ***Limit*** Applied at $(c, g(c))$

The crucial part is that $g(x) > \frac{1}{2}g(c) > 0$ on the interval $[c-\delta, c+\delta]$. Within the intervals $[a, c-\delta]$ and $[c+\delta, b]$, all we know is that $g(x) \geq 0$, and so the integral of $g(x)$ over these two intervals is also at least zero.

Now, let us break up the integral that we started with into three parts and get:

$$\begin{aligned}\int_a^b [f(x)]^2\,dx &= \int_a^b g(x)\,dx \\ &= \int_a^{c-\delta} g(x)\,dx + \int_{c-\delta}^{c+\delta} g(x)\,dx + \int_{c+\delta}^{b} g(x)\,dx \\ &\geq 0 + \int_{c-\delta}^{c+\delta} \frac{1}{2}g(c)\,dx + 0 \\ &= \frac{1}{2}g(c)\,(2\delta) \\ &= \delta[f(c)]^2,\end{aligned}$$

which is a strictly positive number. Thus the inner product is ***positive***.

This argument can be modified if the point c such that $f(c) \neq 0$ happens to be an ***endpoint*** of the interval $[a, b]$, i.e. if we are only told that $f(a) \neq 0$ or $f(b) \neq 0$, by using the definitions of ***right*** and ***left continuity***. This is found in the Exercises.

A Non-Example

Recall that when we created abstract vector spaces, we could define different and strange kinds of "vector addition" and "scalar multiplication," but we had to make sure that all Ten Axioms of a vector space are fulfilled. Similarly, just because we specify how to compute a bilinear form, it doesn't mean that all four properties of an inner product space are fulfilled by this bilinear form.

(Non-)Example: Let $\mathbb{R}^2$ be given the bilinear form:

$$\langle \vec{u} | \vec{v} \rangle = u_1 v_2 + u_2 v_1,$$

where $\vec{u} = \langle u_1, u_2 \rangle$ and $\vec{v} = \langle v_1, v_2 \rangle$ as usual. Notice that the subscripts of $\vec{v}$ are now ***switched***. However, this bilinear form is still ***symmetric*** because:

$$\langle \vec{v} | \vec{u} \rangle = v_1 u_2 + v_2 u_1 = u_1 v_2 + u_2 v_1 = \langle \vec{u} | \vec{v} \rangle.$$

It is still ***homogeneous*** because:

$$\langle k \cdot \vec{u} | \vec{v} \rangle = (ku_1)v_2 + (ku_2)v_1 = k \cdot \langle \vec{u} | \vec{v} \rangle.$$

Is it still ***additive***?

$$\begin{aligned}\langle \vec{u} + \vec{v} | \vec{w} \rangle &= (u_1 + v_1)w_2 + (u_2 + v_2)w_1 \\ &= u_1 w_2 + u_2 w_1 + v_1 w_2 + v_2 w_1 \\ &= \langle \vec{u} | \vec{w} \rangle + \langle \vec{v} | \vec{w} \rangle.\end{aligned}$$

Yes! So far so good. Now, consider:

$$\langle \vec{v} | \vec{v} \rangle = v_1 v_2 + v_2 v_1 = 2 v_1 v_2.$$

Thus, if v_1 and v_2 are of ***opposite signs***, then $\langle \vec{v} | \vec{v} \rangle < 0$. Thus, we violate the ***positive*** property, and this bilinear form is ***not*** an inner product on $\mathbb{R}^2$. $\square$

7.1 Section Summary

Let V be a vector space. An ***inner product*** on V is a ***bilinear form*** $\langle | \rangle$ on V, that is, a function that takes two vectors $\vec{u}$, $\vec{v} \in V$, and produces a ***scalar***, denoted $\langle \vec{u} | \vec{v} \rangle$, such that the following properties are satisfied by all vectors $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$:

1. ***The Symmetric Property*** $\langle \vec{u} | \vec{v} \rangle = \langle \vec{v} | \vec{u} \rangle$.
2. ***The Homogenous Property*** $\langle k \cdot \vec{u} | \vec{v} \rangle = k \cdot \langle \vec{u} | \vec{v} \rangle$.
3. ***The Additive Property*** $\langle \vec{u} + \vec{v} | \vec{w} \rangle = \langle \vec{u} | \vec{w} \rangle + \langle \vec{v} | \vec{w} \rangle$.
4. ***The Positive Property*** If $\vec{v} \neq \vec{\mathbf{0}}_V$, then $\langle \vec{v} | \vec{v} \rangle > 0$.

We say that V is an ***inner product space*** under the inner product $\langle | \rangle$.

Let V be an inner product space. Then, for any $\vec{v} \in V$, we have: $\left\langle \vec{v} | \vec{\mathbf{0}}_V \right\rangle = \left\langle \vec{\mathbf{0}}_V | \vec{v} \right\rangle = 0$. In particular, $\left\langle \vec{\mathbf{0}}_V | \vec{\mathbf{0}}_V \right\rangle = 0$.

Let $\gamma_1, \gamma_2, \ldots, \gamma_n$ be a list of n positive numbers. We can define a ***weighted inner product*** on $\mathbb{R}^n$ by: $\langle \vec{u} | \vec{v} \rangle = \gamma_1 u_1 v_1 + \gamma_2 u_2 v_2 + \cdots + \gamma_n u_n v_n$.

Any ***isomorphism*** $T : \mathbb{R}^n \to \mathbb{R}^n$ defines a new inner product on $\mathbb{R}^n$ by:

$$\langle \vec{u} | \vec{v} \rangle = T(\vec{u}) \circ T(\vec{v}).$$

If $c_1, c_2, \ldots, c_{n+1}$ are ***distinct*** real numbers, we can define an inner product on $\mathbb{P}^n$ by:

$$\langle p(x) | q(x) \rangle = p(c_1)q(c_1) + p(c_2)q(c_2) + \cdots + p(c_n)q(c_n) + p(c_{n+1})q(c_{n+1}).$$

We can define an inner product on $C([a, b])$ by: $\langle f(x) | g(x) \rangle = \int_a^b f(x) \cdot g(x)\, dx$.

7.1 Exercises

For Exercises (1) to (8): Prove that the following bilinear forms are inner products in the indicated vector spaces V:

1. $V = \mathbb{R}^3$, and for $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$:
$$\langle \vec{u} | \vec{v} \rangle = 2u_1 v_1 + u_2 v_2 + 5u_3 v_3.$$
2. $V = \mathbb{R}^3$, and for $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$:
$$\langle \vec{u} | \vec{v} \rangle = \frac{2}{5}u_1 v_1 + \frac{1}{5}u_2 v_2 + \frac{2}{5}u_3 v_3.$$
3. $V = \mathbb{R}^4$, and for $\vec{u} = \langle u_1, u_2, u_3, u_4 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3, v_4 \rangle$:
$$\langle \vec{u} | \vec{v} \rangle = 4u_1 v_1 + u_2 v_2 + 3u_3 v_3 + 6u_4 v_4.$$
4. $V = \mathbb{R}^3$, and for $\vec{u}$, $\vec{v} \in \mathbb{R}^3$: $\langle \vec{u} | \vec{v} \rangle = T(\vec{u}) \circ T(\vec{v})$, where $T : \mathbb{R}^3 \to \mathbb{R}^3$ is given by:
$$[T] = \begin{bmatrix} 2 & 1 & 4 \\ 0 & 1 & 1 \\ 0 & 0 & -3 \end{bmatrix}.$$

5. $V = \mathbb{R}^3$, and for $\vec{u}, \vec{v} \in \mathbb{R}^3$: $\langle \vec{u} | \vec{v} \rangle = T(\vec{u}) \circ T(\vec{v})$, where $T : \mathbb{R}^3 \to \mathbb{R}^3$ is given by:

$$[T] = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}.$$

6. $V = \mathbb{R}^4$, and for $\vec{u}, \vec{v} \in \mathbb{R}^4$: $\langle \vec{u} | \vec{v} \rangle = T(\vec{u}) \circ T(\vec{v})$, where $T : \mathbb{R}^4 \to \mathbb{R}^4$ is given by:

$$[T] = \begin{bmatrix} -2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & -1 & 4 & 0 \\ 4 & 1 & -3 & -5 \end{bmatrix}.$$

7. $V = \mathbb{P}^2$, and for $p(x), q(x) \in \mathbb{P}^2$:

$$\langle p(x) | q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$

8. $V = \mathbb{P}^3$, and for $p(x), q(x) \in \mathbb{P}^3$:

$$\langle p(x) | q(x) \rangle = p(-1)q(-1) + p(1)q(1) + p(2)q(2) + p(4)q(4).$$

For Exercises (9) to (18): Find the inner product of each pair of vectors in the indicated space:

9. $\vec{u} = \langle 6,2,4 \rangle$, $\vec{v} = \langle -1,3,-2 \rangle$ under the inner product of Exercise 1.
10. $\vec{u} = \langle 6,2,4 \rangle$, $\vec{v} = \langle -1,3,-2 \rangle$ under the inner product of Exercise 2.
11. $\vec{u} = \langle 3,-7,2,1 \rangle$, $\vec{v} = \langle 1,2,-2,5 \rangle$ under the inner product of Exercise 3.
12. $\vec{u} = \langle 6,2,4 \rangle$, $\vec{v} = \langle -1,3,-2 \rangle$ under the inner product of Exercise 4.
13. $\vec{u} = \langle 6,2,4 \rangle$, $\vec{v} = \langle -1,3,-2 \rangle$ under the inner product of Exercise 5.
14. $\vec{u} = \langle 3,-7,2,1 \rangle$, $\vec{v} = \langle 1,2,-2,5 \rangle$ under the inner product of Exercise 6.
15. $p(x) = 5 + 2x - x^2$, $q(x) = 7 + 3x^2$ under the inner product of Exercise 7.
16. $p(x) = 5 + 2x + x^3$, $q(x) = 7 - 5x^2 - x^3$ under the inner product of Exercise 8.
17. $f(x) = \cos(x)$, $g(x) = \sin(x)$, under the inner product given by:

$$\langle f(x) | g(x) \rangle = \int_0^{\pi/2} f(x) \cdot g(x)\, dx.$$

18. $f(x) = \cos(x)$, $g(x) = \sin(x)$, under the inner product given by:

$$\langle f(x) | g(x) \rangle = \int_{-\pi/4}^{\pi/4} f(x) \cdot g(x)\, dx.$$

19. Consider the bilinear form on $\mathbb{P}^4$ given by:

$$\langle p(x) | q(x) \rangle = p(-1)q(-1) + p(1)q(1) + p(2)q(2) + p(4)q(4),$$

which is the bilinear form on $\mathbb{P}^3$ from Exercise 8. Show that this is ***not*** an inner product on $\mathbb{P}^4$ by finding a ***non-zero*** polynomial $r(x) \in \mathbb{P}^4$ such that $\langle r(x) | r(x) \rangle = 0$. Hint: think of a factored form for $r(x)$.

20. Decide whether or not the bilinear form on $\mathbb{R}^3$ given by:

$$\langle \vec{u} | \vec{v} \rangle = 4u_1v_3 + 5u_2v_1 + 7u_3v_2$$

is an inner product, where as usual, $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$.

21. Decide whether or not the bilinear form on $\mathbb{P}^2$ given by:

$$\langle p(x) | q(x) \rangle = p(-1)q(1) + p(0)q(0) + p(1)q(-1)$$

is an inner product. Look very carefully!

22. Decide whether or not the bilinear form on $\mathbb{P}^2$ given by:

$$\langle p(x) | q(x) \rangle = 3p(-2)q(-2) + p(0)q(0) + 2p(4)q(4).$$

is an inner product.

23. ***A "Weighted" Integral:***

 a. Show that if a ***fixed*** function $k(x) \in C([a, b])$ is always ***positive*** on $[a, b]$, then the bilinear form on $C([a, b])$ given by:

$$\langle f(x) | g(x) \rangle = \int_a^b f(x) \cdot g(x) \cdot k(x)\, dx$$

 is an inner product.

 b. Find $\langle \sin(x) | \cos(x) \rangle$ under the inner product given by:

$$\langle f(x) | g(x) \rangle = \int_0^\pi f(x) \cdot g(x)\, (x+1)\, dx$$

24. Show that the ***weighted*** inner product on $\mathbb{R}^n$ with positive weights γ_1 through γ_n is exactly the same as the inner product on $\mathbb{R}^n$ generated by the ***isomorphism:***

$$T(\langle x_1, x_2, \ldots, x_n \rangle) = \left(\sqrt{\gamma_1}\, x_1, \sqrt{\gamma_2}\, x_2, \ldots, \sqrt{\gamma_n}\, x_n\right).$$

25. Prove that if W is an inner product space under $\langle | \rangle_W$ and $T : V \to W$ is a ***one-to-one*** linear transformation of vector spaces, then we can also construct an inner product $\langle | \rangle_V$ on V by:

$$\langle \vec{u} | \vec{v} \rangle_V = \langle T(\vec{u}) | T(\vec{v}) \rangle_W.$$

(we use the subscripts V and W on the symbol $\langle | \rangle$ so as not to confuse the two inner products). Note that it is the ***codomain*** that needs to have an inner product, not the domain. Hint: you will need the additivity and homogeneity properties of T. Furthermore, if T is one-to-one, then its kernel . . .

Use Exercise (25) to prove Exercises (26) to (28):

26. Prove that if $T : V \to \mathbb{R}^n$ is a ***one-to-one*** linear transformation, then we can construct an inner product on V by:

$$\langle \vec{u} | \vec{v} \rangle_V = T(\vec{u}) \circ T(\vec{v}).$$

27. Prove that if B is a fixed ***basis*** of a finite dimensional vector space V, then we can construct an inner product on V by:

$$\langle \vec{u} | \vec{v} \rangle_V = \langle \vec{u} \rangle_B \circ \langle \vec{v} \rangle_B.$$

28. ***An Inner Product for Oz:*** Show that:

$$\langle x | y \rangle = \ln(x) \cdot \ln(y)$$

is an inner product on the vector space $\mathbb{R}^+$ of positive real numbers under ordinary multiplication and exponentiation, as seen in Section 3.1.

29. Suppose that $f(x)$ is continuous and $f(a) \neq 0$. Modify the proof in this Section to show that $\int_a^b [f(x)]^2\, dx$ is still ***positive***. You will need to write down the definitions of the terms ***right continuity*** and ***right limit***.

30. Suppose $f(x)$ is the piecewise function given as:

$$f(x) = \begin{cases} 1 & \text{if } x = 1/2 \\ 0 & \text{if } 0 \le x < 1/2 \text{ or } 1/2 < x \le 1 \end{cases}$$

a. Explain precisely (using the definitions stated in this Section) why $f(x)$ is ***not*** continuous on $[0,1]$.

b. What type of discontinuity does $f(x)$ have? Recall that the three common types are: (i) removable, (ii) jump and (iii) infinite discontinuities.

c. Show that $\int_0^1 [f(x)]^2\,dx = 0$, although $f(x)$ itself is ***not*** the zero function.

This shows that continuity is essential for our integral example to be an inner product. Note: since this is strictly speaking an ***improper integral***, you must compute it by breaking it up into two integrals and taking limits:

$$\int_0^1 [f(x)]^2\,dx = \lim_{a\to(1/2)^-} \int_0^a [f(x)]^2\,dx + \lim_{b\to(1/2)^+} \int_b^1 [f(x)]^2\,dx.$$

d. Show that for any positive integer n, we can create a function $f(x)$ that is not zero at n distinct points, but $\int_0^1 [f(x)]^2\,dx$ is still zero.

31. ***Inner Products on Spaces of Infinite Series:***

Recall that a series $\sum_{n=1}^{\infty} a_n$ is ***absolutely convergent*** if $\sum_{n=1}^{\infty} |a_n|$ converges.

A series $\sum_{n=1}^{\infty} a_n$ is only ***conditionally convergent*** if $\sum_{n=1}^{\infty} a_n$ converges but $\sum_{n=1}^{\infty} |a_n|$ diverges.

a. Suppose $\sum_{n=1}^{\infty} a_n$ and $\sum_{n=1}^{\infty} b_n$ are both absolutely convergent series. Prove that $\sum_{n=1}^{\infty} a_n b_n$ is also absolutely convergent.

Hint: Use the Ordinary Comparison Test, and explain ***why*** $|a_n| < 1$ if n is "big enough."

b. Part (a) shows that the bilinear form:

$$\left\langle \sum_{n=1}^{\infty} a_n \,\middle|\, \sum_{n=1}^{\infty} b_n \right\rangle = \sum_{n=1}^{\infty} a_n b_n$$

is a ***well defined*** quantity for two absolutely convergent infinite series, that is, the series on the right will converge. Show that it is in fact an ***inner product*** on the vector space of absolutely convergent infinite series. Notice that this looks like an ***infinite*** version of the ordinary dot product.

c. Find $\left\langle \sum_{n=1}^{\infty} \frac{1}{2^n} \,\middle|\, \sum_{n=1}^{\infty} \frac{1}{3^n} \right\rangle$. What kind of a series do you get?

d. Find $\left\langle \sum_{n=1}^{\infty} \frac{(-1)^n}{3^n} \,\middle|\, \sum_{n=1}^{\infty} \frac{1}{5^n} \right\rangle$.

e. Find an example of two ***conditionally convergent*** series whose inner product, under the definition above, is infinite, in other words, it is undefined. This shows that the inner product above is ***not*** well-defined on the vector space of convergent series.

Hint: use two alternating p-series of the form $\sum_{n=1}^{\infty} \frac{(-1)^n}{n^p}$, where $0 < p < 1$.

Why are these series only conditionally convergent?
How should you choose the p's so that the inner product of the two series is infinite?

7.2 Geometric Constructions in Inner Product Spaces

The four axioms of an inner product space are sufficient for us to prove properties similar to those of the dot product that we saw in Chapter 1. We can also construct the concept of the ***length*** of a vector, the ***distance*** between two vectors and the ***angle*** between two vectors in any inner product space, and not just Euclidean space.

Further Properties of Inner Products

Aside from the four basic properties that an inner product must possess, we summarize other properties that every inner product space must possess below, some of which we have already proven. The rest are left as Exercises.

Theorem: Let V be an inner product space under the bilinear form $\langle \, | \, \rangle$. Then the following properties also hold, for all vectors $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$, and for all $k \in \mathbb{R}$:

1. $\langle \vec{u} | k \cdot \vec{v} \rangle = k \cdot \langle \vec{u} | \vec{v} \rangle$
2. $\langle \vec{u} | \vec{v} + \vec{w} \rangle = \langle \vec{u} | \vec{v} \rangle + \langle \vec{u} | \vec{w} \rangle$
3. $\langle \vec{u} - \vec{v} |, \vec{w} \rangle = \langle \vec{u} | \vec{w} \rangle - \langle \vec{v} | \vec{w} \rangle$
4. $\langle \vec{u} | \vec{v} - \vec{w} \rangle = \langle \vec{u} | \vec{v} \rangle - \langle \vec{u} | \vec{w} \rangle$
5. $\langle \vec{u} + \vec{v} | \vec{u} + \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle + 2\langle \vec{u} | \vec{v} \rangle + \langle \vec{v} | \vec{v} \rangle$
6. $\langle \vec{u} - \vec{v} | \vec{u} - \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle - 2\langle \vec{u} | \vec{v} \rangle + \langle \vec{v} | \vec{v} \rangle$
7. $\langle \vec{u} + \vec{v} | \vec{u} - \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle - \langle \vec{v} | \vec{v} \rangle$
8. $\left\langle \vec{u} | \vec{\mathbf{0}}_V \right\rangle = 0 = \left\langle \vec{\mathbf{0}}_V | \vec{u} \right\rangle$

Notice that Properties 5, 6 and 7 look like formulas that we see in basic algebra, such as:

$$(a+b)(a+b) = a^2 + 2ab + b^2 \quad \text{and} \quad (a+b)(a-b) = a^2 - b^2.$$

Again, we are only allowed to use the ***four axioms*** of inner product spaces in order to prove the additional eight properties above.

Norms and Distances

The positivity property allows us to generalize the concept of the norm or length of a vector in an inner product space as well as distances between two vectors:

Definition: Let $\vec{v}, \vec{u} \in V$, an inner product space. Define the ***norm*** or the ***length*** of $\vec{v}$ by:

$$\|\vec{v}\| = \sqrt{\langle \vec{v} | \vec{v} \rangle}, \text{ in other words:}$$

$$\|\vec{v}\|^2 = \langle \vec{v} | \vec{v} \rangle.$$

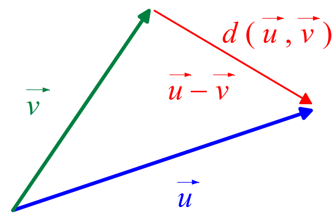
In particular, we say that $\vec{v}$ is a ***unit vector*** if $\|\vec{v}\| = 1$.

The set of all unit vectors in V is called the ***unit sphere*** or ***unit circle*** of V.

We can also define the ***distance*** between two vectors by:

$$d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|.$$

We can thus interpret the distance between two vectors geometrically as the separation between the ***heads*** of $\vec{u}$ and $\vec{v}$, when their tails are together, as we did in Chapter 1:



Geometric Interpretation of $d(\vec{u}, \vec{v})$

Example: Let $p(x) = 3x^2 - 5x + 2 \in \mathbb{P}^2$, under the inner product:

$$\langle p(x) | q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$

Since $p(-2) = 24$, $p(0) = 2$ and $p(1) = 0$, we get:

$$\langle p(x) | p(x) \rangle = 24^2 + 2^2 + 0^2 = 580, \text{ so}$$

$$\|p(x)\| = \sqrt{580}.$$

If we want a ***unit vector*** parallel to $p(x)$, we have the choice of either:

$$\frac{1}{\sqrt{580}}(3x^2 - 5x + 2) \quad \text{or} \quad -\frac{1}{\sqrt{580}}(3x^2 - 5x + 2). \square$$

Example: Consider the weighted inner product on $\mathbb{R}^2$ given by:

$$\langle \vec{u} | \vec{v} \rangle = 4u_1v_1 + 9u_2v_2.$$

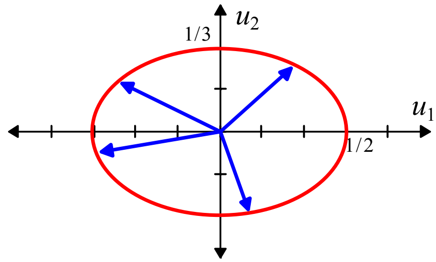
Let us describe the "unit sphere" in $\mathbb{R}^2$ under this new inner product. We must find all the unit vectors:

$$1 = \langle \vec{u} | \vec{u} \rangle = 4u_1^2 + 9u_2^2.$$

But notice we can rewrite this as the equation:

$$\frac{u_1^2}{(1/2)^2} + \frac{u_2^2}{(1/3)^2} = 1.$$

This is simply an ***ellipse*** in u_1 and u_2, with major axis along u_1 of length $1/2$, and minor axis along u_2 of length $1/3$. Thus, the "unit sphere" is the set of all vectors in $\mathbb{R}^2$ that, when drawn in standard position, have their endpoints on the ellipse:



Some Vectors on The "Unit Circle" of $\mathbb{R}^2$ under $\langle \vec{u} | \vec{v} \rangle = 4u_1v_1 + 9u_2v_2$

In general, therefore, we can say that the weighted inner product in $\mathbb{R}^n$ generates a unit sphere that is actually an ***ellipsoid***, as we call them in Multivariable Calculus.□

The norm and distance functions induced by an abstract inner product enjoy the following familiar properties. We leave their proofs as Exercises.

Theorem: For all vectors $\vec{u}$, $\vec{v}$ in an inner product space V, and all $k \in \mathbb{R}$:
1. $\| k \cdot \vec{u} \| = |k| \cdot \| \vec{u} \|$.
2. $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$.
3. $d(k \cdot \vec{u}, k \cdot \vec{v}) = |k| \cdot d(\vec{u}, \vec{v})$.

The Cauchy-Schwarz Inequality

The Theorem with this name that we saw in Chapter 1 is true in general for any inner product space, again, thanks to the four axioms:

Theorem — The Cauchy-Schwarz Inequality:
For all vectors $\vec{u}$ and $\vec{v}$ from an inner product space V:

$$|\langle \vec{u} | \vec{v} \rangle| \leq \| \vec{u} \| \cdot \| \vec{v} \|,$$

or equivalently: $\langle \vec{u} | \vec{v} \rangle^2 \leq \langle \vec{u} | \vec{u} \rangle \cdot \langle \vec{v} | \vec{v} \rangle$.

Proof: Although the proof is virtually identical in spirit to the version in Chapter 1, the Theorem's importance makes the proof worth revisiting and rewriting in terms of a general inner product and not just the dot product:

Case 1. Suppose ***either*** $\vec{u}$ or $\vec{v}$ is the zero vector. We know that $\left\| \vec{\mathbf{0}}_V \right\| = 0$, and from the additional properties of the inner product, that $\left\langle \vec{u} \,|\, \vec{\mathbf{0}}_V \right\rangle = 0$ and $\left\langle \vec{\mathbf{0}}_V \,|\, \vec{v} \right\rangle = 0$. Thus, both sides of the inequality are 0, and thus it is true.

Case 2. Suppose ***neither*** $\vec{u}$ nor $\vec{v}$ is the zero vector. Thus $\| \vec{u} \| > 0$ (i.e. its length is strictly positive). Let us construct the linear combination $\vec{w} = r\vec{u} + s\vec{v}$, where r and s are any two scalars. We know that for any r and s:

$$\begin{aligned} 0 \le \|\vec{w}\|^2 &= \langle r\vec{u} + s\vec{v} | r\vec{u} + s\vec{v} \rangle \\ &= \langle r\vec{u} | r\vec{u} \rangle + 2\langle r\vec{u} | s\vec{v} \rangle + \langle s\vec{v} | s\vec{v} \rangle \\ &= r^2 \cdot \langle \vec{u} | \vec{u} \rangle + 2rs \cdot \langle \vec{u} | \vec{v} \rangle + s^2 \cdot \langle \vec{v} | \vec{v} \rangle, \end{aligned}$$

again, by our eight additional properties. Now, suppose we let $s = \langle \vec{u} | \vec{u} \rangle$. We get:

$$0 \le r^2 \cdot \langle \vec{u} | \vec{u} \rangle + 2r \cdot \langle \vec{u} | \vec{u} \rangle \cdot \langle \vec{u} | \vec{v} \rangle + \langle \vec{u} | \vec{u} \rangle^2 \cdot \langle \vec{v} | \vec{v} \rangle,$$

and since $\langle \vec{u} | \vec{u} \rangle = \|\vec{u}\|^2$ is ***positive***, we can divide it out, and get:

$$0 \le r^2 + 2r \cdot \langle \vec{u} | \vec{v} \rangle + \langle \vec{u} | \vec{u} \rangle \cdot \langle \vec{v} | \vec{v} \rangle.$$

Finally, we let $r = -\langle \vec{u} | \vec{v} \rangle$, and we get:

$$\begin{aligned} 0 &\le \langle \vec{u} | \vec{v} \rangle^2 - 2 \cdot \langle \vec{u} | \vec{v} \rangle \cdot \langle \vec{u} | \vec{v} \rangle + \langle \vec{u} | \vec{u} \rangle \cdot \langle \vec{v} | \vec{v} \rangle \\ &= -\langle \vec{u} | \vec{v} \rangle^2 + \langle \vec{u} | \vec{u} \rangle \cdot \langle \vec{v} | \vec{v} \rangle, \end{aligned}$$

and this simplifies to:

$$\langle \vec{u} | \vec{v} \rangle^2 \le \langle \vec{u} | \vec{u} \rangle \cdot \langle \vec{v} | \vec{v} \rangle,$$

which is the second (equivalent) version of the Cauchy-Schwarz Inequality.■

Example: In Section 7.1, we saw that $\mathbb{P}^2$ is an inner product space under:

$$\langle p(x) | q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$

For $p(x) = 3x^2 - 5x + 2$, and $q(x) = 7x - 6$, we constructed the table of values:

c_i	$p(c_i)$	$q(c_i)$
-2	24	-20
0	2	-6
1	0	1

and found that $\langle p(x) | q(x) \rangle = -492$. The same table of values will give us:

$$\langle 3x^2 - 5x + 2 | 3x^2 - 5x + 2 \rangle = 24^2 + 2^2 + 0^2 = 580, \quad \text{and}$$

$$\langle 7x - 6 | 7x - 6 \rangle = (-20)^2 + (-6)^2 + 1^2 = 437.$$

Thus, we verify that: $|-492| \le \sqrt{580} \cdot \sqrt{437} \approx 503.45$ as guaranteed.□

The Angle Between Two Vectors

Recall that in Chapter 1, we used the ***Law of Cosines*** to obtain the alternative formula:

$$\vec{u} \circ \vec{v} = \|\vec{u}\| \|\vec{v}\| \cos(\theta)$$

in $\mathbb{R}^2$ or $\mathbb{R}^3$, where θ is the angle between the two non-zero vectors $\vec{u}$ and $\vec{v}$, when they are in standard position. This was because we visualize vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$ as directed line segments, and so $\vec{u}$ and $\vec{v}$ naturally form a triangle that we can actually see. Since we do not always have that ability in an arbitrary vector space, The Cauchy-Schwarz Inequality is the key that allows us to define an angle between any two vectors in an inner product space V.

First, if ***neither*** $\vec{u}$ nor $\vec{v}$ is $\vec{\mathbf{0}}_V$, then their lengths are both ***positive***, and the Cauchy-Schwarz Inequality says that:

$$\frac{|\langle \vec{u}|\vec{v}\rangle|}{\|\vec{u}\|\,\|\vec{v}\|} \leq 1, \quad \text{or equivalently:} \quad -1 \leq \frac{\langle \vec{u}|\vec{v}\rangle}{\|\vec{u}\|\,\|\vec{v}\|} \leq 1.$$

Since the cosine of any angle θ satisfies $-1 \leq \cos(\theta) \leq 1$, we can define the angle between $\vec{u}$ and $\vec{v}$ to be the angle θ between 0 and π, such that:

$$\cos(\theta) = \frac{\langle \vec{u}|\vec{v}\rangle}{\|\vec{u}\|\,\|\vec{v}\|}.$$

In particular, if $\cos(\theta) = 0 = \langle \vec{u}|\vec{v}\rangle$, then $\theta = \pi/2$, and we will say that $\vec{u}$ and $\vec{v}$ are ***orthogonal*** to each other.

Now, if either vector is $\vec{\mathbf{0}}_V$, then we saw that ***both sides*** of the Cauchy-Schwarz Inequality are zero. Thus, both the numerator and the denominator in our formula for $\cos(\theta)$ are zero. But still, we have:

$$\left\langle \vec{u}|\vec{\mathbf{0}}_V \right\rangle = 0 \text{ or } \left\langle \vec{\mathbf{0}}_V|\vec{v} \right\rangle = 0.$$

Since we defined two non-zero vectors $\vec{u}$ and $\vec{v}$ to be orthogonal if $\langle \vec{u}|\vec{v}\rangle = 0$, then for the sake of ***convention***, we will agree that $\vec{\mathbf{0}}_V$ is orthogonal to any other vector. Summarizing all this, we make the following:

Definitions: If $\vec{u}$ and $\vec{v}$ are non-zero vectors in V, we define the ***angle*** between them as the angle θ, where $0 \leq \theta \leq \pi$, such that:

$$\cos(\theta) = \frac{\langle \vec{u}|\vec{v}\rangle}{\|\vec{u}\|\,\|\vec{v}\|}.$$

Furthermore, we will say that $\vec{u}$ and $\vec{v}$ are ***orthogonal*** or ***perpendicular*** to each other ***if and only if*** $\langle \vec{u}|\vec{v}\rangle = 0$. We write this symbolically as:

$$\vec{u} \perp \vec{v} \textbf{\textit{ if and only if }} \langle \vec{u}|\vec{v}\rangle = 0.$$

In particular, $\vec{\mathbf{0}}_V$ is orthogonal to ***all*** vectors in V.

Example: Continuing with our previous example regarding $\mathbb{P}^2$ and the two indicated polynomials $p(x)$ and $q(x)$, we have:

$$\cos(\theta) = \frac{-492}{\sqrt{580} \cdot \sqrt{437}} \approx -0.977.$$

Note that this is quite close to -1, so in a sense, $p(x) = 3x^2 - 5x + 2$, and $q(x) = 7x - 6$ are almost "opposite" to each other. The angle between them is not π, but rather:

$$\theta \approx \cos^{-1}(-0.977) \approx 2.9267 \text{ radians or } 167.7 \text{ degrees.} \square$$

Other Consequences of the Cauchy-Schwarz Inequality

As in Euclidean spaces and the dot product, the Cauchy-Schwarz Inequality produces an inequality with a famous name, that we show below in two guises. We leave their proofs as Exercises:

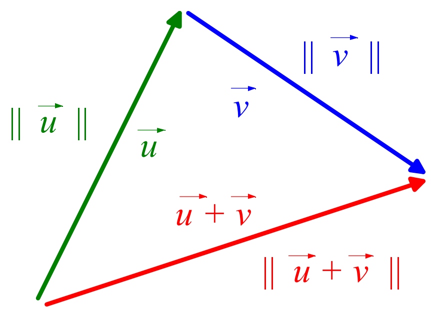
Theorem — The Triangle Inequality (Norm Version):
For any two vectors $\vec{u}$ and $\vec{v}$ in an inner product space V:

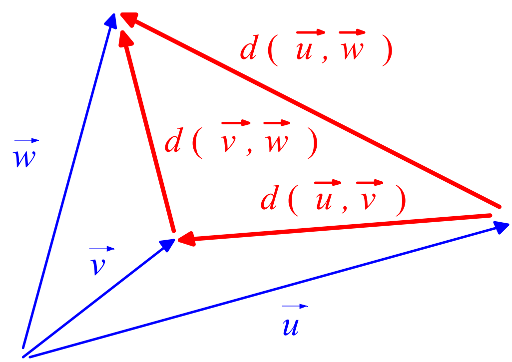
$$\|\vec{u}+\vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|.$$

Theorem — The Triangle Inequality (Distance Version):
For any three vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ in an inner product space V:

$$d(\vec{u},\vec{v}) \leq d(\vec{u},\vec{w}) + d(\vec{w},\vec{v}).$$

Again, the Theorem gets its name from the ordinary triangle, where any side has smaller measure than the sum of the measures of the two other sides. We can interpret these two versions as follows:





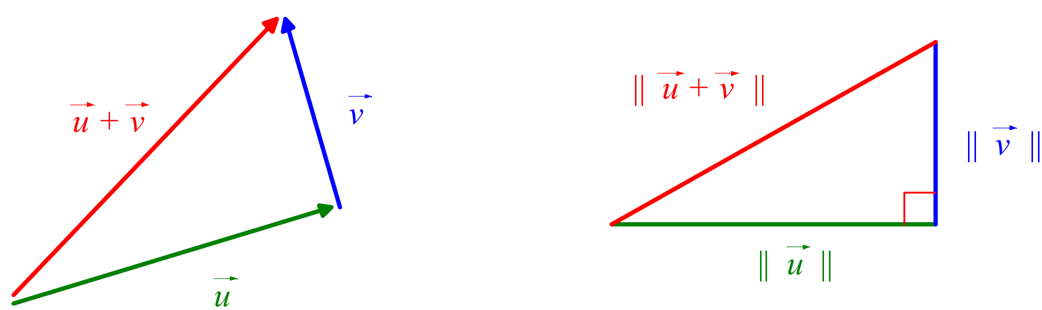
$\|\vec{u}+\vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$ $\qquad$ $d(\vec{u},\vec{v}) \leq d(\vec{u},\vec{w}) + d(\vec{w},\vec{v})$

The Two Triangle Inequalities

Similarly, as inspired by right triangles, we have a modern version of an ancient Greek's discovery, whose proof is also left as an Exercise:

Theorem — The Generalized Pythagorean Theorem:
Suppose that $\vec{u}$ and $\vec{v}$ are vectors from some inner product space V. Then: $\vec{u}$ and $\vec{v}$ are ***orthogonal*** to each other ***if and only if:***

$$\|\vec{u}\|^2 + \|\vec{v}\|^2 = \|\vec{u}+\vec{v}\|^2.$$



The Generalized Pythagorean Theorem

7.2 Section Summary

Let V be an inner product space under $\langle\ |\ \rangle$. Then the following properties also hold, for all vectors $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$:

1. $\langle \vec{u} | k \cdot \vec{v} \rangle = k \cdot \langle \vec{u} | \vec{v} \rangle$
2. $\langle \vec{u} | \vec{v} + \vec{w} \rangle = \langle \vec{u} | \vec{v} \rangle + \langle \vec{u} | \vec{w} \rangle$
3. $\langle \vec{u} - \vec{v} |, \vec{w} \rangle = \langle \vec{u} | \vec{w} \rangle - \langle \vec{v} | \vec{w} \rangle$
4. $\langle \vec{u} | \vec{v} - \vec{w} \rangle = \langle \vec{u} | \vec{v} \rangle - \langle \vec{u} | \vec{w} \rangle$
5. $\langle \vec{u} + \vec{v} | \vec{u} + \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle + 2\langle \vec{u} | \vec{v} \rangle + \langle \vec{v} | \vec{v} \rangle$
6. $\langle \vec{u} - \vec{v} | \vec{u} - \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle - 2\langle \vec{u} | \vec{v} \rangle + \langle \vec{v} | \vec{v} \rangle$
7. $\langle \vec{u} + \vec{v} | \vec{u} - \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle - \langle \vec{v} | \vec{v} \rangle$
8. $\left\langle \vec{u} | \vec{\mathbf{0}}_V \right\rangle = 0 = \left\langle \vec{\mathbf{0}}_V | \vec{u} \right\rangle$

Let $\vec{v}$, $\vec{u} \in V$. Define the ***norm*** or the ***length*** of $\vec{v}$ by:

$$\|\vec{v}\| = \sqrt{\langle \vec{v} | \vec{v} \rangle}, \text{ in other words, } \|\vec{v}\|^2 = \langle \vec{v} | \vec{v} \rangle.$$

In particular, we say that $\vec{v}$ is a ***unit vector*** if $\|\vec{v}\| = 1$. The set of all unit vectors in V is called the ***unit sphere*** or ***unit circle*** of V. We can also define the ***distance*** between two vectors by $d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|$.

For all vectors $\vec{u}$, $\vec{v}$ in an inner product space V, and all $k \in \mathbb{R}$:

1. $\|k \cdot \vec{u}\| = |k| \cdot \|\vec{u}\|$
2. $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$
3. $d(k \cdot \vec{u}, k \cdot \vec{v}) = |k| \cdot d(\vec{u}, \vec{v})$

The Cauchy-Schwarz Inequality: Let V be an inner product space with respect to $\langle\ |\ \rangle$. Then, for all vectors $\vec{u}$, $\vec{v} \in V$:

$$|\langle \vec{u} | \vec{v} \rangle| \leq \|\vec{u}\| \cdot \|\vec{v}\|,$$

or equivalently: $\langle \vec{u} | \vec{v} \rangle^2 \leq \langle \vec{u} | \vec{u} \rangle \cdot \langle \vec{v} | \vec{v} \rangle$.

If $\vec{u}$ and $\vec{v}$ are non-zero vectors in V, we define the angle between them as the angle θ such that $\cos(\theta) = \langle \vec{u} | \vec{v} \rangle / (\|\vec{u}\| \|\vec{v}\|)$, where $0 \leq \theta \leq \pi$.

Furthermore, we will say that $\vec{u}$ is ***orthogonal*** to $\vec{v}$ ***if and only if*** $\langle \vec{u} | \vec{v} \rangle = 0$. We write this symbolically as: $\vec{u} \perp \vec{v} \Leftrightarrow \langle \vec{u} | \vec{v} \rangle = 0$. In particular, $\vec{\mathbf{0}}_V$ is orthogonal to ***all*** vectors in V.

The Triangle Inequality (Norm Version): For any two vectors $\vec{u}$ and $\vec{v}$ in an inner product space V: $\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|$.

The Triangle Inequality (Distance Version): For any three vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ in an inner product space V: $d(\vec{u}, \vec{v}) \leq d(\vec{u}, \vec{w}) + d(\vec{w}, \vec{v})$.

The Generalized Pythagorean Theorem: Suppose that $\vec{u}$ and $\vec{v}$ are vectors from some inner product space V. Then: $\vec{u}$ and $\vec{v}$ are ***orthogonal*** to each other ***if and only if*** $\|\vec{u}\|^2 + \|\vec{v}\|^2 = \|\vec{u} + \vec{v}\|^2$.

7.2 Exercises

For Exercises (1) to (5): Find the length of the following vectors under the indicated inner product (these were seen in the Exercises in Section 7.1), and the two unit vectors that are parallel to them (Note: all vectors in $\mathbb{R}^n$ are represented as $\vec{u} = \langle u_1, u_2, \ldots, u_n \rangle$, as usual):

1. $\vec{u} = \langle 3, -4, 7 \rangle \in \mathbb{R}^3$, under the weighted dot product:
$$\langle \vec{u} | \vec{v} \rangle = 2u_1v_1 + u_2v_2 + 5u_3v_3.$$
2. $\vec{u} = \langle 4, -7, -2, 6 \rangle \in \mathbb{R}^4$, under the weighted dot product:
$$\langle \vec{u} | \vec{v} \rangle = 4u_1v_1 + u_2v_2 + 3u_3v_3 + 6u_4v_4.$$
3. $\vec{u} = \langle 6, 3, -4 \rangle \in \mathbb{R}^3$, under the inner product $\langle \vec{u} | \vec{v} \rangle = T(\vec{u}) \circ T(\vec{v})$, where $T : \mathbb{R}^3 \to \mathbb{R}^3$ is given by:
$$[T] = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}$$
4. $p(x) = 7 + 2x - 5x^2 \in \mathbb{P}^2$, under the inner product:
$$\langle p(x) | q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$
5. $p(x) = 5 - 7x + 3x^2 - x^3 \in \mathbb{P}^3$ under the inner product:
$$\langle p(x) | q(x) \rangle = p(-1)q(-1) + p(1)q(1) + p(2)q(2) + p(4)q(4).$$

For Exercises (6) to (10): Find the (approximate) angle between the two vectors and the distance between them, under the indicated inner product:

6. $\vec{u} = \langle 3, -4, 7 \rangle$, $\vec{v} = \langle 2, -5, 8 \rangle$, under the inner product of Exercise 1.
7. $\vec{u} = \langle 3, -7, 2, 1 \rangle$, $\vec{v} = \langle 1, 2, -2, 5 \rangle$, under the inner product of Exercise 2.
8. $\vec{u} = \langle 6, 2, 4 \rangle$, $\vec{v} = \langle -1, 3, -2 \rangle$, under the inner product of Exercise 3.
9. $p(x) = 7 + 2x - 5x^2$, $q(x) = 3 - 4x + 2x^2$, under the inner product of Exercise 4.
10. $p(x) = 5 + 2x + x^3$, $q(x) = 7 - 5x^2 - x^3$ under the inner product of Exercise 5.
11. We found the length of $p(x) = 3x^2 - 5x + 2 \in \mathbb{P}^2$ to be $\sqrt{580}$, under the inner product of the third Example. Find the length of $p(x)$ with respect to the inner product induced by the definite integral:
$$\langle p(x) | q(x) \rangle = \int_0^2 p(x)q(x)dx$$
(which is also valid, because polynomials are in $C([0, 2])$.)
12. Find the exact cosine and the approximate angle (in radians) between $f(x) = \sin(x)$ and $g(x) = \cos(x)$ under the inner product:
$$\langle f(x) | g(x) \rangle = \int_0^{\pi/4} f(x) \cdot g(x)\, dx.$$
13. Draw the graph of the unit circle in $\mathbb{R}^2$ under the weighted inner product:
$$\langle \vec{u} | \vec{v} \rangle = 49u_1v_1 + 25u_2v_2.$$

14. Draw the graph of the unit circle in $\mathbb{R}^2$ under the weighted inner product:

$$\langle \vec{u} | \vec{v} \rangle = \frac{1}{49} u_1 v_1 + \frac{1}{25} u_2 v_2.$$

15. Draw a graph of the unit sphere in $\mathbb{R}^3$ under the weighted inner product:

$$\langle \vec{u} | \vec{v} \rangle = 4u_1 v_1 + u_2 v_2 + 25u_3 v_3.$$

16. Suppose in an inner product space V, we have two vectors $\vec{u}$ and $\vec{v}$ such that $\langle \vec{u} | \vec{v} \rangle = -18$, $\|\vec{u}\| = 5$ and $\|\vec{v}\| = 7$. What is the length of $3\vec{u} - 11\vec{v}$?

17. Suppose someone told you that in an inner product space V, there are two vectors $\vec{u}$ and $\vec{v}$ such that $\langle \vec{u} | \vec{v} \rangle = -19$, $\|\vec{u}\| = 5$ and $\|\vec{v}\| = 3$. Should you believe them?

18. Suppose in an inner product space V, we have two vectors $\vec{u}$ and $\vec{v}$ such that $\|2\vec{u} - 5\vec{v}\| = \sqrt{981}$, $\|4\vec{u} + 3\vec{v}\| = \sqrt{3313}$, and $\langle \vec{u} | \vec{v} \rangle = 16$. Find $\|\vec{u}\|$ and $\|\vec{v}\|$.

For Exercises (19) to (25): Prove that the following properties are true in any inner product space V, for all $\vec{u}$, $\vec{v} \in V$ and all $k \in \mathbb{R}$:

19. $\langle \vec{u} + \vec{v} | \vec{u} + \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle + 2\langle \vec{u} | \vec{v} \rangle + \langle \vec{v} | \vec{v} \rangle$
20. $\langle \vec{u} - \vec{v} | \vec{u} - \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle - 2\langle \vec{u} | \vec{v} \rangle + \langle \vec{v} | \vec{v} \rangle$
21. $\langle \vec{u} + \vec{v} | \vec{u} - \vec{v} \rangle = \langle \vec{u} | \vec{u} \rangle - \langle \vec{v} | \vec{v} \rangle$
22. $\langle k \cdot \vec{u} | k \cdot \vec{v} \rangle = k^2 \cdot \langle \vec{u} | \vec{v} \rangle$
23. $\|k \cdot \vec{u}\| = |k| \cdot \|\vec{u}\|$. Hint: how do you simplify $\sqrt{k^2}$?
24. $d(\vec{u}, \vec{v}) = d(\vec{v}, \vec{u})$
25. $d(k \cdot \vec{u}, k \cdot \vec{v}) = |k| \cdot d(\vec{u}, \vec{v})$

For Exercises (26) to (32): The following are all ***manifestations*** of the ***Cauchy-Schwarz Inequality***. In other words, we can show they are true because we can find a vector space V and an inner product on V, such that the ***conclusion*** of the Cauchy-Schwarz Inequality is the indicated inequality. Your job is to prove these inequalities by finding the ***vector space*** V and the ***inner product*** on V, and then unwinding the conclusion of the Cauchy-Schwarz Inequality to obtain the indicated inequality. ***Do not*** try to prove these inequalities directly!

26. For any invertible $n \times n$ matrix A and $\vec{u}$, $\vec{v} \in \mathbb{R}^n$:

$$[A\vec{u} \circ A\vec{v}]^2 \leq [A\vec{u} \circ A\vec{u}] \cdot [A\vec{v} \circ A\vec{v}].$$

27. For any real numbers r, t and θ:

$$[r\cos(\theta) + t\sin(\theta)]^2 \leq r^2 + t^2.$$

28. For any two continuous functions $f(x)$ and $g(x)$ on an interval $[a, b]$:

$$\left[\int_a^b f(x) \cdot g(x)\, dx\right]^2 \leq \left[\int_a^b [f(x)]^2\, dx\right]\left[\int_a^b [g(x)]^2\, dx\right].$$

29. For any continuous function $f(x)$ on $[0, 2\pi]$:

$$\left[\int_0^{2\pi} f(x) \cdot \sin(x)\, dx\right]^2 \leq \pi \int_0^{2\pi} [f(x)]^2\, dx.$$

30. For any continuous function $f(x)$ on $[0, 1]$:

$$\left[\int_0^1 x \cdot f(x)\,dx\right]^2 \leq \frac{1}{3}\int_0^1 [f(x)]^2\,dx.$$

31. For any continuous function $f(x)$ on $[0, 1]$:

$$\left[\int_0^1 \sqrt{x} \cdot f(x)\,dx\right]^2 \leq \frac{1}{2}\int_0^1 [f(x)]^2\,dx$$

32. For any continuous function $f(x)$ on $[0, \pi/4]$:

$$\left[\int_0^{\pi/4} \frac{f(x)}{\cos(x)}\,dx\right]^2 \leq \int_0^{\pi/4} [f(x)]^2\,dx$$

For Exercises (33) and (34): Use the Cauchy-Schwarz Inequality to prove the two versions of ***The Triangle Inequalities:***

33. ***Norm Version:*** For any two vectors $\vec{u}$ and $\vec{v}$ in an inner product space V:

$$\|\vec{u} + \vec{v}\| \leq \|\vec{u}\| + \|\vec{v}\|.$$

Hint: Since both sides of the inequality involve non-negative quantities, it is equivalent to the inequality where both sides have been ***squared***.

34. ***Distance Version:*** For all vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ in an inner product space V:

$$d(\vec{u}, \vec{v}) \leq d(\vec{u}, \vec{w}) + d(\vec{w}, \vec{v}).$$

Hint: $\vec{u} - \vec{v} = \vec{u} - \vec{w} + \vec{w} - \vec{v}$.

For Exercises (35) and (36): Use the property that $\|\vec{x}\|^2 = \langle \vec{x} | \vec{x} \rangle$, for any vector $\vec{x} \in V$, to prove the following:

35. ***The Generalized Pythagorean Theorem:*** Suppose that $\vec{u}$ and $\vec{v}$ are vectors from some inner product space V. Then: $\vec{u}$ and $\vec{v}$ are ***orthogonal*** to each other ***if and only if:***

$$\|\vec{u}\|^2 + \|\vec{v}\|^2 = \|\vec{u} + \vec{v}\|^2.$$

36. ***The Parallelogram Principles:*** Prove that for any two vectors $\vec{u}$ and $\vec{v}$ in any inner product space V, we have:

$$\|\vec{u} + \vec{v}\|^2 + \|\vec{u} - \vec{v}\|^2 = 2\|\vec{u}\|^2 + 2\|\vec{v}\|^2,$$

and also:

$$\|\vec{u} + \vec{v}\|^2 - \|\vec{u} - \vec{v}\|^2 = 4\langle \vec{u} | \vec{v} \rangle.$$

See the diagram of the same name on page 29.

37. Use the previous Exercise to show that if $\vec{u}$ and $\vec{v}$ are two vectors in an inner product space V, then $\vec{u}$ and $\vec{v}$ are ***orthogonal*** to each other ***if and only if:***

$$\|\vec{u} + \vec{v}\| = \|\vec{u} - \vec{v}\|.$$

Draw a vector diagram that represents what this Exercise is saying if $V = \mathbb{R}^2$ under the ordinary dot product. What kind of a triangle do you get?

7.3 Orthonormal Sets and The Gram-Schmidt Algorithm

The standard basis $S = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ of $\mathbb{R}^n$ has the convenient property that any two distinct members of the set are ***orthogonal*** to each other (under the ordinary dot product), and every vector in S is a ***unit vector***. Such sets of vectors are very important, and so we call them by a special name:

Definition: Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be a set of vectors in an inner product space V. We say that S is an ***orthonormal set*** if:

$$\langle \vec{v}_i | \vec{v}_j \rangle = 0 \ \text{ if } i \neq j, \text{ and}$$

$$\langle \vec{v}_i | \vec{v}_i \rangle = 1 \ \text{ for } i = 1..k.$$

If we remove the condition that each member of S must be a unit vector but insist that all of the vectors be non-zero, we call S an ***orthogonal set***.

Notice that we can convert an orthogonal set to an orthonormal set by dividing each member by its length, so all that separates these two kinds of sets is an easy ***normalizing process***. Thus, for convenience, we will focus most of our attention on orthonormal sets. We will construct these sets in abstract inner product spaces in just a little bit, but first, let us describe the easiest case:

Example: Let us find all the orthonormal sets in $\mathbb{R}^2$ under the ordinary dot product. The unit vectors are the vectors:

$$\vec{u} = \langle x, y \rangle \ \text{ such that } \ x^2 + y^2 = 1.$$

In other words, their endpoints lie on the ***unit circle*** when they are in standard position. From Trigonometry, we can parametrize such vectors with an angle θ, as:

$$\vec{u} = \langle x, y \rangle = \langle \cos(\theta), \sin(\theta) \rangle.$$

Thus, there are two choices of a vector that will be orthogonal to $\vec{u}$, namely:

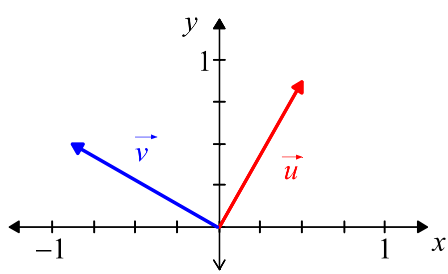
$$\vec{v} = \langle \cos(\theta + \pi/2), \sin(\theta + \pi/2) \rangle = \langle -\sin(\theta), \cos(\theta) \rangle \quad \text{or}$$

$$\vec{v} = \langle \cos(\theta - \pi/2), \sin(\theta - \pi/2) \rangle = \langle \sin(\theta), -\cos(\theta) \rangle.$$

For example, suppose $\vec{u} = \langle \cos(\pi/3), \sin(\pi/3) \rangle = \left\langle 1/2, \sqrt{3}/2 \right\rangle$. Then we can choose:

$$\vec{v} = \langle \cos(\pi/3 + \pi/2), \sin(\pi/3 + \pi/2) \rangle = \langle \cos(5\pi/6), \sin(5\pi/6) \rangle = \left\langle -\sqrt{3}/2, 1/2 \right\rangle.$$

Let us see these orthogonal vectors below:



$$\vec{v} = \left\langle -\sqrt{3}/2, 1/2 \right\rangle \ \text{ and } \ \vec{u} = \left\langle 1/2, \sqrt{3}/2 \right\rangle$$

They certainly look perpendicular to each other, and a quick check of the dot product:

$$\left\langle 1/2, \sqrt{3}/2 \right\rangle \circ \left\langle -\sqrt{3}/2, 1/2 \right\rangle = -\sqrt{3}/4 + \sqrt{3}/4 = 0$$

shows that these two vectors are indeed orthogonal to each other.□

Could we include a ***third vector*** in S such that the new set is still orthonormal? By our analysis above, the only other choice for a third vector would be $\vec{w} = \langle \cos(-\pi/6), \sin(-\pi/6) \rangle = \left\langle \sqrt{3}/2, -1/2 \right\rangle$, in order for $\vec{w}$ to be orthogonal to $\vec{u}$. But this is the ***negative*** of $\vec{v}$, and thus $\vec{v} \circ \vec{w} = -1$, so we cannot include this vector. Thus, S is ***as big as possible***, and the orthonormal sets in $\mathbb{R}^2$ are of the form:

$$\{\langle \cos(\theta), \sin(\theta) \rangle, \langle -\sin(\theta), \cos(\theta) \rangle\} \text{ or } \{\langle \cos(\theta), \sin(\theta) \rangle, \langle \sin(\theta), -\cos(\theta) \rangle\}.$$

for some number θ. It is of course no coincidence that $\mathbb{R}^2$ is 2-***dimensional*** and our orthonormal sets S could contain at most 2 vectors. The following Theorem says that this will be true in general:

Theorem: An orthonormal set S in an inner product space V is ***linearly independent***. Consequently, if $\dim(V) = n$, and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is an orthonormal set, then $k \leq n$, and any set with more than n vectors cannot be orthonormal.

A similar Theorem with the word "orthogonal" replacing "orthonormal" is still true.

Proof: Let S be the orthonormal set as written above, and suppose that:

$$c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k = \vec{\mathbf{0}}_V.$$

We must show that each coefficient c_i is zero. To exploit the orthonormal property, all we have to do is take the inner product of both sides of the equation with each $\vec{v}_i$, one at a time, starting with $\vec{v}_1$:

$$\langle \vec{v}_1 \,|\, c_1\vec{v}_1 + c_2\vec{v}_2 + \cdots + c_k\vec{v}_k \rangle = \left\langle \vec{v}_1 \,|\, \vec{\mathbf{0}}_V \right\rangle, \text{ and expanding as usual:}$$

$$c_1\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle + c_2\langle \vec{v}_1 \,|\, \vec{v}_2 \rangle + \cdots + c_k\langle \vec{v}_1 \,|\, \vec{v}_k \rangle = 0.$$

Now, since $\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle = 1$ and $\langle \vec{v}_1 \,|\, \vec{v}_2 \rangle = \cdots = \langle \vec{v}_1 \,|\, \vec{v}_k \rangle = 0$, we get $c_1 = 0$. We repeat this process with $\vec{v}_2$, and so on, thus concluding that each $c_i = 0$. The rest of the Theorem easily follows because the number of vectors in a linearly independent set cannot exceed the dimension of a vector space.■

Example: Let $\mathbb{P}^2$ be an inner product space under:

$$\langle p(x) \,|\, q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1),$$

as in our last two Sections. Let us construct an orthonormal set of 3 vectors from $\mathbb{P}^2$. The idea is to pick 3 quadratics which have value 0 at two out of three of the input values $\{-2, 0, 1\}$. In other words, each has a ***root*** at two of these points. Thus, we can pick:

$$p_1(x) = x(x-1), \;\; p_2(x) = (x+2)(x-1), \;\text{ and }\; p_3(x) = (x+2)x.$$

It is obvious that:

$$\langle p_1(x) \,|\, p_2(x) \rangle = \langle p_1(x) \,|\, p_3(x) \rangle = \langle p_2(x) \,|\, p_3(x) \rangle = 0$$

However, since $p_1(-2) = 6$, $p_2(0) = -2$, and $p_3(1) = 3$, the length of these vectors are 6, 2 and 3, respectively. Although we can choose to divide each vector by either its length or the ***negative*** of its length, we will choose the orthonormal basis $\{u_1(x), u_2(x), u_3(x)\}$, where:

$$u_1(x) = \frac{1}{6}x(x-1),\ \ u_2(x) = -\frac{1}{2}(x+2)(x-1),\ \text{ and }\ u_3(x) = \frac{1}{3}(x+2)x.$$

These quadratics have the property that they are zero at two out of the three input values $\{-2, 0, 1\}$, and they are ***normalized*** so that they have value 1 at the remaining input value, making them unit vectors. Thus, we can think of them as analogs to the standard basis vectors $\langle\, 1, 0, 0\,\rangle$, $\langle\, 0, 1, 0\,\rangle$ and $\langle\, 0, 0, 1\,\rangle$ of $\mathbb{R}^3$. □

Orthonormal Bases

Since an orthonormal set is automatically linearly independent, ***The Two-for-One-Theorem*** now implies that an orthonormal set B whose size is the ***dimension*** of V must be a ***basis*** for V. Our standard basis set $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ in $\mathbb{R}^n$ is of course such a basis, under the ordinary dot product. Notice that we can easily find the ***coordinates*** of any vector in $\mathbb{R}^n$ with respect to the standard basis. For any ordinary basis B, we would normally have to solve a system of equations to find the coordinates of an arbitrary vector with respect to B. However, if B is ***orthonormal***, all we need to do is compute some inner products (which is usually a lot more pleasant than solving systems of equations):

Definition/Theorem: Let V be a finite dimensional inner product space with $\dim(V) = n$. An orthonormal set $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ with n vectors is called an ***orthonormal basis*** for V. If $\vec{v}$ is an arbitrary member of V, and B is an orthonormal basis for V, and:

$$\langle\, \vec{v}\,\rangle_B = \langle c_1, c_2, \ldots, c_n\rangle, \quad \text{then:}$$

$$c_i = \langle\, \vec{v} \,|\, \vec{u}_i\,\rangle \ \text{ for } i = 1..n.$$

In other words:

$$\vec{v} = \langle\, \vec{v} \,|\, \vec{u}_1\,\rangle \cdot \vec{u}_1 + \langle\, \vec{v} \,|\, \vec{u}_2\,\rangle \cdot \vec{u}_2 + \cdots + \langle\, \vec{v} \,|\, \vec{u}_n\,\rangle \cdot \vec{u}_n.$$

Proof: If $\vec{v} \in V$, we know that we can express $\vec{v}$ as a linear combination of the member of any basis B in ***exactly one*** way, say:

$$\vec{v} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_n\vec{u}_n,$$

and these coefficients form $\langle\, \vec{v}\,\rangle_B = \langle c_1, c_2, \ldots, c_n\rangle$.

To compute each coefficient, we will use the same trick that we saw in the previous Theorem, that is, to take the inner product of both sides of the equation with each $\vec{u}_i$, one at a time. Thus:

$$\begin{aligned}\langle\, \vec{v} \,|\, \vec{u}_1\,\rangle &= \langle\, c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_n\vec{u}_n \,|\, \vec{u}_1\,\rangle \\ &= c_1\langle\, \vec{u}_1 \,|\, \vec{u}_1\,\rangle + c_2\langle\, \vec{u}_2 \,|\, \vec{u}_1\,\rangle + \cdots + c_n\langle\, \vec{u}_n \,|\, \vec{u}_1\,\rangle \\ &= c_1 \cdot 1 + c_2 \cdot 0 + \cdots + c_n \cdot 0 = c_1.\end{aligned}$$

Similarly, $\langle\, \vec{v} \,|\, \vec{u}_2\,\rangle = c_2$, and so on. ■

Note: In Mathematics, when a "trick" is used more than once, we promote it to a ***technique***.

Example: Consider the set:

$$B = \left\{ \left\langle \frac{5}{13}, \frac{12}{13} \right\rangle, \left\langle -\frac{12}{13}, \frac{5}{13} \right\rangle \right\}.$$

We can easily check that both are unit vectors, and their dot product is zero, thus B is an orthonormal set in $\mathbb{R}^2$. To find $\langle \vec{v} \rangle_B$ for $\vec{v} = \langle 8, -11 \rangle$, we compute:

$$\begin{aligned} \vec{v} \circ \vec{u}_1 &= 8\left(\frac{5}{13}\right) - 11\left(\frac{12}{13}\right) = -\frac{92}{13}, \text{ and} \\ \vec{v} \circ \vec{u}_2 &= 8\left(-\frac{12}{13}\right) - 11\left(\frac{5}{13}\right) = -\frac{151}{13}. \end{aligned}$$

Thus, $\langle \vec{v} \rangle_B = \left\langle -\frac{92}{13}, -\frac{151}{13} \right\rangle$. We can check that:

$$-\frac{92}{13}\left\langle \frac{5}{13}, \frac{12}{13} \right\rangle - \frac{151}{13}\left\langle -\frac{12}{13}, \frac{5}{13} \right\rangle = \left\langle \frac{-92(5) + 151(12)}{169}, \frac{-92(12) - 151(5)}{169} \right\rangle = \langle 8, -11 \rangle. \square$$

The Gram-Schmidt Algorithm

Now we are ready to construct orthonormal sets in any inner product space on an industrial scale. The idea behind the algorithm, called the Gram-Schmidt Algorithm, is basically the same idea that we used to compute the projection of a vector in $\mathbb{R}^2$ onto a line L through the origin.

Suppose that $\dim(V) = n$. Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be ***any*** basis for V. The input to the Gram-Schmidt Algorithm will be this basis B. The output will be an ***orthogonal*** set:

$$S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\},$$

with the special property that:

$$Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}) = Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}) \text{ for all } k = 1 \ldots n.$$

In other words:

$$\begin{aligned} Span(\{\vec{w}_1\}) &= Span(\{\vec{v}_1\}), \\ Span(\{\vec{w}_1, \vec{w}_2\}) &= Span(\{\vec{v}_1, \vec{v}_2\}), \\ Span(\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}) &= Span(\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}), \ldots \end{aligned}$$

and so on. By dividing each vector by its length, we can thus obtain an ***orthonormal*** basis S' for V.

The version of the Gram-Schmidt Algorithm that we show below varies slightly from the standard, but has the advantage of avoiding radicals, fractions, and large numbers, to some extent:

Step 1. Let $\vec{v}_1 = \vec{w}_1$. If $\vec{v}_1$ contains fractions or common factors, divide $\vec{v}_1$ by a suitable scalar to eliminate both. If $dim(V) = 1$, we normalize the set $\{\vec{v}_1\}$ to obtain an orthonormal basis for V, otherwise, proceed to Step 2.

Step 2. We want to produce a vector $\vec{v}_2$ that is orthogonal to $\vec{v}_1$. We will find $\vec{v}_2$ by subtracting a suitable multiple of $\vec{v}_1$ from $\vec{w}_2$. In other words, we ask: is there a scalar k such that:

$$\vec{v}_2 = \vec{w}_2 - k\vec{v}_1$$

is orthogonal to $\vec{v}_1$? But to find k, we can apply the inner product to both sides:

$$\langle \vec{v}_2 | \vec{v}_1 \rangle = \langle \vec{w}_2 - k\vec{v}_1 | \vec{v}_1 \rangle = \langle \vec{w}_2 | \vec{v}_1 \rangle - k\langle \vec{v}_1 | \vec{v}_1 \rangle.$$

Since we want $\langle \vec{v}_2 | \vec{v}_1 \rangle = 0$, this means that we can solve for k as:

$$k = \frac{\langle \vec{w}_2 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle}.$$

Note that this is well-defined because $\vec{v}_1 \neq \vec{\mathbf{0}}_V$, thus $\langle \vec{v}_1 | \vec{v}_1 \rangle$ is ***positive***. We get $\vec{v}_2$ as:

$$\vec{v}_2 = \vec{w}_2 - \frac{\langle \vec{w}_2 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle} \vec{v}_1.$$

For good measure, let us check that $\vec{v}_2$ is orthogonal to $\vec{v}_1$:

$$\langle \vec{v}_2 | \vec{v}_1 \rangle = \left\langle \vec{w}_2 - \frac{\langle \vec{w}_2 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle} \vec{v}_1 \middle| \vec{v}_1 \right\rangle = \langle \vec{w}_2 | \vec{v}_1 \rangle - \frac{\langle \vec{w}_2 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle} \langle \vec{v}_1 | \vec{v}_1 \rangle = 0.$$

It is important to point out two observations:

- $\vec{v}_2$ is a ***non-zero*** vector, because otherwise, we would get:
$$\vec{w}_2 = \frac{\langle \vec{w}_2 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle} \vec{v}_1.$$
  This is impossible, because $\vec{w}_2$ is not parallel to $\vec{w}_1$, which happens to be $\vec{v}_1$.
- $\vec{v}_2$ is a ***linear combination*** of $\vec{w}_2$ and $\vec{v}_1$ (and again, this happens to be $\vec{w}_1$). Thus:
$$Span(\{\vec{v}_1, \vec{v}_2\}) = Span(\{\vec{v}_1, \vec{w}_2\}) = Span(\{\vec{w}_1, \vec{w}_2\}).$$

Again, if $\vec{v}_2$ contains fractions or common factors, divide $\vec{v}_2$ by a suitable scalar to eliminate both. If $dim(V) = 2$, we normalize the orthogonal set $\{\vec{v}_1, \vec{v}_2\}$ to obtain an orthonormal basis for V, otherwise, proceed to Step 3.

Step 3. We want to produce a vector $\vec{v}_3$ that is orthogonal to ***both*** $\vec{v}_2$ and $\vec{v}_1$. Given the success of our formula for $\vec{v}_2$ above, let us ***guess*** that a reasonable formula for $\vec{v}_3$ would be:

$$\vec{v}_3 = \vec{w}_3 - \frac{\langle \vec{w}_3 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle} \vec{v}_1 - \frac{\langle \vec{w}_3 | \vec{v}_2 \rangle}{\langle \vec{v}_2 | \vec{v}_2 \rangle} \vec{v}_2.$$

This is not exactly an illegal act: like other sciences, Mathematics is also ***experimental***. But to validate our formula, we must ***check*** that indeed $\vec{v}_3$ is orthogonal to both $\vec{v}_2$ and $\vec{v}_1$:

$$\begin{aligned}
\langle \vec{v}_3 | \vec{v}_1 \rangle &= \left\langle \vec{w}_3 - \frac{\langle \vec{w}_3 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle} \vec{v}_1 - \frac{\langle \vec{w}_3 | \vec{v}_2 \rangle}{\langle \vec{v}_2 | \vec{v}_2 \rangle} \vec{v}_2 \middle| \vec{v}_1 \right\rangle \\
&= \langle \vec{w}_3 | \vec{v}_1 \rangle - \frac{\langle \vec{w}_3 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle} \langle \vec{v}_1 | \vec{v}_1 \rangle - \frac{\langle \vec{w}_3 | \vec{v}_2 \rangle}{\langle \vec{v}_2 | \vec{v}_2 \rangle} \langle \vec{v}_2 | \vec{v}_1 \rangle \\
&= \langle \vec{w}_3 | \vec{v}_1 \rangle - \langle \vec{w}_3 | \vec{v}_1 \rangle - 0 = 0.
\end{aligned}$$

Note that we used the fact that $\langle \vec{v}_2 | \vec{v}_1 \rangle = 0$. Similarly, we can check that:

$$\begin{aligned}\langle \vec{v}_3 \,|\, \vec{v}_2 \rangle &= \left\langle \vec{w}_3 - \frac{\langle \vec{w}_3 \,|\, \vec{v}_1 \rangle}{\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle}\vec{v}_1 - \frac{\langle \vec{w}_3 \,|\, \vec{v}_2 \rangle}{\langle \vec{v}_2 \,|\, \vec{v}_2 \rangle}\vec{v}_2 \,\middle|\, \vec{v}_2 \right\rangle \\ &= \langle \vec{w}_3 \,|\, \vec{v}_2 \rangle - \frac{\langle \vec{w}_3 \,|\, \vec{v}_1 \rangle}{\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle}\langle \vec{v}_1 \,|\, \vec{v}_2 \rangle - \frac{\langle \vec{w}_3 \,|\, \vec{v}_2 \rangle}{\langle \vec{v}_2 \,|\, \vec{v}_2 \rangle}\langle \vec{v}_2 \,|\, \vec{v}_2 \rangle \\ &= \langle \vec{w}_3 \,|\, \vec{v}_2 \rangle - 0 - \langle \vec{w}_3 \,|\, \vec{v}_2 \rangle = 0.\end{aligned}$$

As we did at the end of Step 2, we point out two observations:

- $\vec{v}_3$ is a ***non-zero*** vector, because otherwise, we would get:

$$\vec{w}_3 = \frac{\langle \vec{w}_3 \,|\, \vec{v}_1 \rangle}{\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle}\vec{v}_1 + \frac{\langle \vec{w}_3 \,|\, \vec{v}_2 \rangle}{\langle \vec{v}_2 \,|\, \vec{v}_2 \rangle}\vec{v}_2.$$

  This is impossible, because $\vec{w}_3$ is not a member of $Span(\{\vec{w}_1, \vec{w}_2\})$, and we noted that this was exactly the same as $Span(\{\vec{v}_1, \vec{v}_2\})$.
- $\vec{v}_3$ is a ***linear combination*** of $\vec{w}_3$, $\vec{v}_1$ and $\vec{v}_2$, therefore, by ***The Equality of Spans Theorem:***

$$Span(\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}) = Span(\{\vec{v}_1, \vec{v}_2, \vec{w}_3\}) = Span(\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}).$$

Again, if $\vec{v}_3$ contains fractions or common factors, divide $\vec{v}_3$ by a suitable scalar to eliminate both. If $dim(V) = 3$, we normalize the three vectors in $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ to obtain an orthonormal basis for V, otherwise, proceed to the next step.

Step k + 1. We are now ready to generalize the process of finding the next vector, $\vec{v}_{k+1}$, by Induction. Suppose we have already constructed $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$, an ***orthogonal*** set of vectors with the property that:

$$Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}) = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}).$$

Let us again guess that a reasonable formula for the next vector $\vec{v}_{k+1}$ would be:

$$\vec{v}_{k+1} = \vec{w}_{k+1} - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_1 \rangle}{\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle}\vec{v}_1 - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_2 \rangle}{\langle \vec{v}_2 \,|\, \vec{v}_2 \rangle}\vec{v}_2 - \cdots - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_k \rangle}{\langle \vec{v}_k \,|\, \vec{v}_k \rangle}\vec{v}_k.$$

Again, by computing $\langle \vec{v}_{k+1} \,|\, \vec{v}_1 \rangle$, $\langle \vec{v}_{k+1} \,|\, \vec{v}_2 \rangle$, $\ldots$, $\langle \vec{v}_{k+1} \,|\, \vec{v}_k \rangle$, we find that all of these inner products are ***zero***. For instance:

$$\begin{aligned}&\langle \vec{v}_{k+1} \,|\, \vec{v}_1 \rangle \\ &= \left\langle \vec{w}_{k+1} - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_1 \rangle}{\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle}\vec{v}_1 - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_2 \rangle}{\langle \vec{v}_2 \,|\, \vec{v}_2 \rangle}\vec{v}_2 - \cdots - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_k \rangle}{\langle \vec{v}_k \,|\, \vec{v}_k \rangle}\vec{v}_k \,\middle|\, \vec{v}_1 \right\rangle \\ &= \langle \vec{w}_{k+1} \,|\, \vec{v}_1 \rangle - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_1 \rangle}{\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle}\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_2 \rangle}{\langle \vec{v}_2 \,|\, \vec{v}_2 \rangle}\langle \vec{v}_2 \,|\, \vec{v}_1 \rangle - \cdots - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_k \rangle}{\langle \vec{v}_k \,|\, \vec{v}_k \rangle}\langle \vec{v}_k \,|\, \vec{v}_1 \rangle \\ &= \langle \vec{w}_{k+1} \,|\, \vec{v}_1 \rangle - \frac{\langle \vec{w}_{k+1} \,|\, \vec{v}_1 \rangle}{\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle}\langle \vec{v}_1 \,|\, \vec{v}_1 \rangle = 0.\end{aligned}$$

As before, we can see that $\vec{v}_{k+1}$ is ***not*** the zero vector, for otherwise we would get:

$$\vec{w}_{k+1} \in Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}) = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}),$$

and this would be impossible since B is a basis for V, and is thus linearly ***independent***.

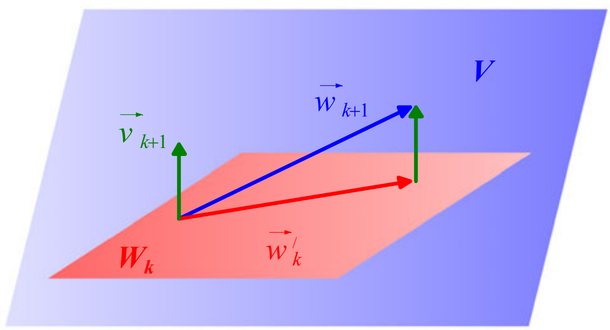
Finally, we also see that, by solving for $\vec{w}_{k+1}$ above, that this vector is a linear combination of $\vec{v}_1$ through $\vec{v}_{k+1}$. Similarly, $\vec{v}_{k+1}$ is also a linear combination of $\vec{v}_1$ through $\vec{v}_k$ and $\vec{w}_{k+1}$. But since we assumed that $\vec{v}_1$ through $\vec{v}_k$ are linear combinations of $\vec{w}_1$ through $\vec{w}_k$, then $\vec{v}_{k+1}$ is also a linear combination of $\vec{w}_1$ through $\vec{w}_{k+1}$. Thus:

$$Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k, \vec{v}_{k+1}\}) = Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k, \vec{w}_{k+1}\}).$$

If we denote by W_k the subspace $Span(\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\})$, we note that we can write $\vec{v}_{k+1}$ as:

$$\vec{v}_{k+1} = \vec{w}_{k+1} - \vec{w}^{\,/}_k, \quad \text{in other words:}$$
$$\vec{w}_{k+1} = \vec{v}_{k+1} + \vec{w}^{\,/}_k, \quad \text{where:}$$
$$\vec{w}^{\,/}_k = \frac{\langle \vec{w}_{k+1} | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle}\vec{v}_1 + \frac{\langle \vec{w}_{k+1} | \vec{v}_2 \rangle}{\langle \vec{v}_2 | \vec{v}_2 \rangle}\vec{v}_2 + \cdots + \frac{\langle \vec{w}_{k+1} | \vec{v}_k \rangle}{\langle \vec{v}_k | \vec{v}_k \rangle}\vec{v}_k \in W_k.$$

We can thus visualize the abstract meaning of $\vec{w}^{\,/}_k$ and $\vec{v}_{k+1}$ in the following diagram:



Constructing the Next Vector $\vec{v}_{k+1}$ in the Gram-Schmidt Algorithm

Notice that except for the labeling, this is essentially the same diagram that we used to visualize decomposing a vector $\vec{v} \in \mathbb{R}^3$ into a sum $\vec{v} = \vec{w}_1 + \vec{w}_2$ where $\vec{w}_2 \in \Pi$, a plane through the origin, and $\vec{w}_1 \in L$, its normal line.

We also see that once we choose $\vec{v}_1 = \vec{w}_1$, we need to compute the inner products:

$$\langle \vec{v}_1 | \vec{v}_1 \rangle, \langle \vec{w}_2 | \vec{v}_1 \rangle, \langle \vec{w}_3 | \vec{v}_1 \rangle, \ldots, \langle \vec{w}_n | \vec{v}_1 \rangle,$$

in Steps 2 through n, so we may as well do these computations after we perform Step 1. Similarly, after we obtain $\vec{v}_2$, we should ***check*** that $\langle \vec{v}_1 | \vec{v}_2 \rangle = 0$, but we also need to compute the inner products:

$$\langle \vec{v}_2 | \vec{v}_2 \rangle, \langle \vec{w}_3 | \vec{v}_2 \rangle, \langle \vec{w}_4 | \vec{v}_2 \rangle, \ldots, \langle \vec{w}_n | \vec{v}_2 \rangle,$$

for Steps 3 through n, so we should perform these after Step 2. We repeat these ideas after Step 3, and so on — checking that $\vec{v}_{k+1}$ is orthogonal to $\vec{v}_1$ through $\vec{v}_k$, and computing the inner products involving $\vec{v}_{k+1}$ that we will need in future Steps. Once we have $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$, we divide each vector by its ***length*** to produce an ***orthonormal basis*** for V.

Example: Let us construct an orthonormal basis for $\mathbb{R}^4$ under the ordinary dot product by starting with the basis:

$$B = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\} = \{\langle 1, -1, 1, 0 \rangle, \langle 1, 0, 1, -1 \rangle, \langle -1, 1, 1, 1 \rangle, \langle 1, 1, 1, -1 \rangle\}.$$

We can check with technology that $[B]$ has determinant 2, so B is a basis for $\mathbb{R}^4$. Since there are four vectors, there will be four major steps in the Gram-Schmidt Algorithm:

Step 1. $\vec{v}_1 = \vec{w}_1 = \langle 1,-1,1,0\rangle$.

To prepare for Steps 2, 3 and 4, we want to compute:

$$\begin{aligned}
\vec{v}_1 \circ \vec{v}_1 &= \langle 1,-1,1,0\rangle \circ \langle 1,-1,1,0\rangle = 1+1+1+0 = 3,\\
\vec{w}_2 \circ \vec{v}_1 &= \langle 1,0,1,-1\rangle \circ \langle 1,-1,1,0\rangle = 1+0+1+0 = 2,\\
\vec{w}_3 \circ \vec{v}_1 &= \langle -1,1,1,1\rangle \circ \langle 1,-1,1,0\rangle = -1-1+1+0 = -1, \quad \text{and}\\
\vec{w}_4 \circ \vec{v}_1 &= \langle 1,1,1,-1\rangle \circ \langle 1,-1,1,0\rangle = 1-1+1+0 = 1.
\end{aligned}$$

Step 2.

$$\begin{aligned}
\vec{v}_2 &= \vec{w}_2 - \frac{\vec{w}_2 \circ \vec{v}_1}{\vec{v}_1 \circ \vec{v}_1} \cdot \vec{v}_1\\
&= \langle 1,0,1,-1\rangle - \tfrac{2}{3}\langle 1,-1,1,0\rangle = \left\langle \tfrac{1}{3}, \tfrac{2}{3}, \tfrac{1}{3}, -1\right\rangle
\end{aligned}$$

To avoid fractions, we multiply this by 3 and get a better vector: $\vec{v}_2 = \langle 1,2,1,-3\rangle$. We verify that:

$$\vec{v}_1 \circ \vec{v}_2 = \langle 1,-1,1,0\rangle \circ \langle 1,2,1,-3\rangle = 1-2+1+0 = 0.$$

To prepare for Steps 3 and 4, we compute:

$$\begin{aligned}
\vec{v}_2 \circ \vec{v}_2 &= \langle 1,2,1,-3\rangle \circ \langle 1,2,1,-3\rangle = 1+4+1+9 = 15,\\
\vec{w}_3 \circ \vec{v}_2 &= \langle -1,1,1,1\rangle \circ \langle 1,2,1,-3\rangle = -1+2+1-3 = -1, \text{ and}\\
\vec{w}_4 \circ \vec{v}_2 &= \langle 1,1,1,-1\rangle \circ \langle 1,2,1,-3\rangle = 1+2+1+3 = 7.
\end{aligned}$$

Step 3.

$$\begin{aligned}
\vec{v}_3 &= \vec{w}_3 - \frac{\vec{w}_3 \circ \vec{v}_1}{\vec{v}_1 \circ \vec{v}_1} \cdot \vec{v}_1 - \frac{\vec{w}_3 \circ \vec{v}_2}{\vec{v}_2 \circ \vec{v}_2} \cdot \vec{v}_2\\
&= \langle -1,1,1,1\rangle - \frac{-1}{3}\langle 1,-1,1,0\rangle - \frac{-1}{15}\langle 1,2,1,-3\rangle\\
&= \left\langle -\frac{9}{15}, \frac{12}{15}, \frac{21}{15}, \frac{12}{15}\right\rangle = \left\langle -\frac{3}{5}, \frac{4}{5}, \frac{7}{5}, \frac{4}{5}\right\rangle.
\end{aligned}$$

We multiply this vector by 5, and get a better vector: $\vec{v}_3 = \langle -3,4,7,4\rangle$. We verify that:

$$\begin{aligned}
\vec{v}_1 \circ \vec{v}_3 &= \langle 1,-1,1,0\rangle \circ \langle -3,4,7,4\rangle = -3-4+7+0 = 0, \quad \text{and}\\
\vec{v}_2 \circ \vec{v}_3 &= \langle 1,2,1,-3\rangle \circ \langle -3,4,7,4\rangle = -3+8+7-12 = 0.
\end{aligned}$$

To prepare for Step 4, we compute:

$$\begin{aligned}
\vec{v}_3 \circ \vec{v}_3 &= \langle -3,4,7,4\rangle \circ \langle -3,4,7,4\rangle = 9+16+49+16 = 90, \text{ and}\\
\vec{w}_4 \circ \vec{v}_3 &= \langle 1,1,1,-1\rangle \circ \langle -3,4,7,4\rangle = -3+4+7-4 = 4.
\end{aligned}$$

Step 4.

$$\begin{aligned}
\vec{v}_4 &= \vec{w}_4 - \frac{\vec{w}_4 \circ \vec{v}_1}{\vec{v}_1 \circ \vec{v}_1} \cdot \vec{v}_1 - \frac{\vec{w}_4 \circ \vec{v}_2}{\vec{v}_2 \circ \vec{v}_2} \cdot \vec{v}_2 - \frac{\vec{w}_4 \circ \vec{v}_3}{\vec{v}_3 \circ \vec{v}_3} \cdot \vec{v}_3\\
&= \langle 1,1,1,-1\rangle - \frac{1}{3}\langle 1,-1,1,0\rangle - \frac{7}{15}\langle 1,2,1,-3\rangle - \frac{4}{90}\langle -3,4,7,4\rangle\\
&= \left\langle \frac{15}{45}, \frac{10}{45}, -\frac{5}{45}, \frac{10}{45}\right\rangle = \left\langle \frac{3}{9}, \frac{2}{9}, -\frac{1}{9}, \frac{2}{9}\right\rangle.
\end{aligned}$$

We multiply this vector by 9 and get the better vector: $\vec{v}_4 = \langle 3, 2, -1, 2\rangle$. We check that $\vec{v}_4$ is orthogonal to $\vec{v}_1$, $\vec{v}_2$ and $\vec{v}_3$:

$$\begin{aligned}
\vec{v}_1 \circ \vec{v}_4 &= \langle 1, -1, 1, 0\rangle \circ \langle 3, 2, -1, 2\rangle = 3 - 2 - 1 + 0 = 0, \quad \text{and} \\
\vec{v}_2 \circ \vec{v}_4 &= \langle 1, 2, 1, -3\rangle \circ \langle 3, 2, -1, 2\rangle = 3 + 4 - 1 - 6 = 0 \\
\vec{v}_3 \circ \vec{v}_4 &= \langle -3, 4, 7, 4\rangle \circ \langle 3, 2, -1, 2\rangle = -9 + 8 - 7 + 8 = 0.
\end{aligned}$$

The only missing piece of information is the length of $\vec{v}_4$, which we get from:

$$\vec{v}_4 \circ \vec{v}_4 = \langle 3, 2, -1, 2\rangle \circ \langle 3, 2, -1, 2\rangle = 9 + 4 + 1 + 4 = 18.$$

The length of our four output vectors are:

$$\|\vec{v}_1\| = \sqrt{3}, \quad \|\vec{v}_2\| = \sqrt{15}, \quad \|\vec{v}_3\| = \sqrt{90}, \quad \text{and } \|\vec{v}_4\| = \sqrt{18}.$$

To get an ***orthonormal basis*** for $\mathbb{R}^4$, we divide each vector by its length, and obtain:

$$S = \left\{ \frac{1}{\sqrt{3}}\langle 1, -1, 1, 0\rangle, \frac{1}{\sqrt{15}}\langle 1, 2, 1, -3\rangle, \frac{1}{\sqrt{90}}\langle -3, 4, 7, 4\rangle, \frac{1}{\sqrt{18}}\langle 3, 2, -1, 2\rangle \right\}. \square$$

Example: Let $\mathbb{P}^2$ be an inner product space under:

$$\left\langle p(x) \,|\, q(x) \right\rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1),$$

which we have already seen in several previous Examples. Let us construct an orthonormal basis for $\mathbb{P}^2$ starting with the standard basis:

$$B = \{1, x, x^2\}.$$

Since $\mathbb{P}^2$ is a 3-dimensional space, there will be three major steps, but the inner product is more complicated because we need to ***evaluate*** our polynomials at the three given points.

Step 1.

$$\vec{w}_1(x) = \vec{v}_1(x) = 1.$$

To prepare for Steps 2 and 3, we compute:

$$\begin{aligned}
&\text{For } \vec{v}_1(x) = 1\text{:} \quad \vec{v}_1(-2) = 1, \quad \vec{v}_1(0) = 1, \ \vec{v}_1(1) = 1; \\
&\text{For } \vec{w}_2(x) = x\text{:} \quad \vec{w}_2(-2) = -2, \ \vec{w}_2(0) = 0, \ \vec{w}_1(1) = 1; \\
&\text{For } \vec{w}_3(x) = x^2\text{:} \ \vec{w}_3(-2) = 4, \quad \vec{w}_3(0) = 0, \ \vec{w}_3(1) = 1; \\
&\langle \vec{v}_1(x) | \vec{v}_1(x)\rangle = \langle 1 | 1\rangle = 1 + 1 + 1 = 3; \\
&\langle \vec{w}_2(x) | \vec{v}_1(x)\rangle = \langle x | 1\rangle = -2 + 0 + 1 = -1; \quad \text{and} \\
&\langle \vec{w}_3(x) | \vec{v}_1(x)\rangle = \langle x^2 | 1\rangle = 4 + 0 + 1 = 5.
\end{aligned}$$

Step 2.

$$\begin{aligned}
\vec{v}_2(x) &= \vec{w}_2(x) - \frac{\langle \vec{w}_2(x) | \vec{v}_1(x)\rangle}{\langle \vec{v}_1(x) | \vec{v}_1(x)\rangle} \cdot \vec{v}_1(x) \\
&= x - \frac{-1}{3} \cdot 1 = x + \frac{1}{3}.
\end{aligned}$$

To avoid fractions, we multiply this by 3 and get a better: $\vec{v}_2(x) = 3x + 1$. We compute:

$$\vec{v}_2(-2) = -5,\ \vec{v}_2(0) = 1,\ \vec{v}_2(1) = 4;$$

$$\langle \vec{v}_1(x) | \vec{v}_2(x) \rangle = \langle 1 | 3x + 1 \rangle = -5 + 1 + 4 = 0;$$

$$\langle \vec{v}_2(x) | \vec{v}_2(x) \rangle = \langle 3x + 1 | 3x + 1 \rangle = 15 + 1 + 16 = 32;$$

$$\langle \vec{w}_3(x) | \vec{v}_2(x) \rangle = \langle x^2 | 3x + 1 \rangle = -20 + 0 + 4 = -16.$$

The 2nd line checks that $\{\vec{v}_1, \vec{v}_2\} = \{1, 3x + 1\}$ is an orthogonal set .

Step 3.

$$\vec{v}_3(x) = \vec{w}_3(x) - \frac{\langle \vec{w}_3(x) | \vec{v}_1(x) \rangle}{\langle \vec{v}_1(x) | \vec{v}_1(x) \rangle} \bullet \vec{v}_1(x) - \frac{\langle \vec{w}_3(x) | \vec{v}_2(x) \rangle}{\langle \vec{v}_2(x) | \vec{v}_2(x) \rangle} \bullet \vec{v}_2(x)$$
$$= x^2 - \frac{5}{3} \bullet 1 - \frac{-16}{42} \bullet (3x + 1) = x^2 + \frac{8}{7}x - \frac{9}{7}.$$

As with $\vec{v}_2$, we multiply $\vec{v}_3$ by 7, and get a better $\vec{v}_3(x) = 7x^2 + 8x - 9$. Although this is the final vector, we still need to evaluate it at the three points in order to compute its length, and to check that it is orthogonal to the first two output vectors:

$$\vec{v}_3(-2) = 3,\ \vec{v}_3(0) = -9,\ \vec{v}_3(1) = 6$$

$$\langle \vec{v}_1(x) | \vec{v}_3(x) \rangle = \langle 1 | 7x^2 + 8x - 9 \rangle = 1 \bullet 3 + 1 \bullet (-9) + 1 \bullet 6 = 0, \text{ and}$$

$$\langle \vec{v}_2(x) | \vec{v}_3(x) \rangle = \langle 3x + 1 | 7x^2 + 8x - 9 \rangle = (-5) \bullet 3 + 1 \bullet (-9) + 4 \bullet 6 = 0.$$

Thus, we get an orthogonal set $\{1, 3x + 1, 7x^2 + 8x - 9\}$. The length of the third vector is $\sqrt{9 + 81 + 36} = \sqrt{126}$. To get an orthonormal basis for $\mathbb{P}^2$, we just divide each vector by its length, and obtain:

$$S = \left\{ \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{42}}(3x + 1), \frac{1}{\sqrt{126}}(7x^2 + 8x - 9) \right\}. \square$$

We note that on a practical basis, we can assemble our data in a table, which will grow as we add more polynomials that result from our computations. For this Example, we will get the table:

	-2	0	1
$\vec{w}_1(x) = \vec{v}_1(x) = 1$	1	1	1
$\vec{w}_2(x) = x$	-2	0	1
$\vec{w}_3(x) = x^2$	4	0	1
$\vec{v}_2(x) = 3x + 1$	-5	1	4
$\vec{v}_3(x) = 7x^2 + 8x - 9$	3	-9	6

Generalization to Infinite Sets

The concept of an orthonormal or orthogonal set of vectors can be generalized to an infinite set of vectors. However, the Gram-Schmidt Algorithm can only be extended to any ***countable*** set of vectors.

Definition/Theorem: Let $S = \{\vec{v}_i | i \in I\}$ be an infinite set of vectors in an inner product space V, for some indexing set $I \subset \mathbb{R}$. We say that S is an ***orthonormal set*** if:

$$\langle \vec{v}_i | \vec{v}_j \rangle = 0 \ \text{ if } i \neq j, \text{ where } i, j \in I, \text{ and}$$
$$\langle \vec{v}_i | \vec{v}_i \rangle = 1 \ \text{ for any } i \in I.$$

If we remove the condition that each member of S must be a unit vector but insist that all of the vectors be ***non-zero***, we call S an ***orthogonal set***.

An infinite orthogonal set is also ***linearly independent***.

If $S = \{\vec{v}_i | i \in \mathbb{N}\}$ is any ***countable*** set of vectors in V, then we can apply the Gram-Schmidt Algorithm inductively in order to produce an infinite orthonormal set, $S' = \{\vec{w}_i | i \in \mathbb{N}\}$, with the special property that for every $k \in \mathbb{N}$:

$$Span(\{\vec{v}_0, \vec{v}_1, \ldots, \vec{v}_k\}) = Span(\{\vec{w}_0, \vec{w}_1, \ldots, \vec{w}_k\}).$$

Technically, since S is a ***countably infinite set***, the Gram-Schmidt algorithm will not terminate, although it would be possible to produce an orthonormal set of k vectors for any positive integer k.

In the Exercises, we will see how these concepts apply to $S = \{\sin(nx), \cos(nx) | n \in \mathbb{N}\}$ under an inner product that involves a definite integral. This set is central to the study of Fourier Series.

7.3 Section Summary

Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be a set of vectors in an inner product space V. We say that S is an ***orthonormal set*** if $\left\langle \vec{v}_i \,|\, \vec{v}_j \right\rangle = 0$ when $i \neq j$, and $\left\langle \vec{v}_i \,|\, \vec{v}_i \right\rangle = 1$ for $i = 1..k$.

If we remove the condition that each member of S is a unit vector but insist that all of the vectors be non-zero, we call S an ***orthogonal set***.

In $\mathbb{R}^2$, the orthonormal sets are of the form:

$$S = \{\langle \cos(\theta), \sin(\theta)\rangle, \langle -\sin(\theta), \cos(\theta)\rangle\} \ \text{ or } \ S = \{\langle \cos(\theta), \sin(\theta)\rangle, \langle \sin(\theta), -\cos(\theta)\rangle\},$$

for some number θ.

An orthonormal set S in an inner product space V is ***linearly independent***. Consequently, if $\dim(V) = n$, and $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is an orthonormal set, then $k \leq n$, and any set with more than n vectors cannot be orthonormal. A similar Theorem with the word "orthogonal" replacing "orthonormal" is still true.

Let V be a finite dimensional inner product space with $\dim(V) = n$. An orthonormal set $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ with n vectors is called an ***orthonormal basis*** for V. If $\vec{v}$ is an arbitrary member of V, and S is an orthonormal basis for V, and $\langle \vec{v} \rangle_S = \langle c_1, c_2, \ldots, c_n \rangle$, then $c_i = \left\langle \vec{v} \,|\, \vec{u}_i \right\rangle$ for $i = 1..n$.

In other words: $\vec{v} = \langle \vec{v} | \vec{u}_1 \rangle \cdot \vec{u}_1 + \langle \vec{v} | \vec{u}_2 \rangle \cdot \vec{u}_2 + \cdots + \langle \vec{v} | \vec{u}_n \rangle \cdot \vec{u}_n$.

The Gram-Schmidt Algorithm:

Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ be any ***basis*** for V.

1. Let $\vec{v}_1 = \vec{w}_1$. If $\dim(V) = 1$, proceed to Step 4, otherwise:
2. Let $\vec{v}_2 = \vec{w}_2 - \langle \vec{w}_2 | \vec{v}_1 \rangle \cdot \dfrac{\vec{v}_1}{\|\vec{v}_1\|^2}$.

Scale $\vec{v}_2$, if necessary, so that fractions or common factors do not appear.

If $\dim(V) = 2$, proceed to Step 4, otherwise:

3. For $k = 2$ to $n - 1$, construct:

$$\vec{v}_{k+1} = \vec{w}_{k+1} - \langle \vec{w}_{k+1} | \vec{v}_1 \rangle \cdot \frac{\vec{v}_1}{\|\vec{v}_1\|^2} - \langle \vec{w}_{k+1} | \vec{v}_2 \rangle \cdot \frac{\vec{v}_2}{\|\vec{v}_2\|^2} - \cdots - \langle \vec{w}_{k+1} | \vec{v}_k \rangle \cdot \frac{\vec{v}_k}{\|\vec{v}_k\|^2}.$$

Again, scale each $\vec{v}_{k+1}$ so that fractions do not appear.

4. Normalize the set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ by dividing each vector by its ***length*** to get an ***orthonormal basis*** S for V.

7.3 Exercises

For Exercises (1) to (15): Use the Gram-Schmidt Algorithm to construct an orthonormal basis beginning with the specified basis B for the indicated vector spaces V, under the specified inner product (you may safely assume that the sets B have been verified to be bases):

1. $B = \{\langle 1, 1, -1\rangle, \langle 0, -1, 1\rangle, \langle 2, 0, 1\rangle\}$ for $\mathbb{R}^3$ under the ordinary dot product.
2. $B = \{\langle 1, 0, 1\rangle, \langle 2, -1, 0\rangle, \langle 1, 1, -1\rangle\}$ for $\mathbb{R}^3$ under the ordinary dot product.
3. $B = \{\langle 1, 1, -1\rangle, \langle 0, -1, 1\rangle, \langle 2, 0, 1\rangle\}$ for $\mathbb{R}^3$ under the weighted dot product:

$$\langle \vec{u} | \vec{v} \rangle = 4u_1v_1 + 5u_2v_2 + 3u_3v_3.$$

Do you get the same answer as in Exercise 1?

4. $B = \{\langle 1, 0, 1\rangle, \langle 2, -1, 0\rangle, \langle 1, 1, -1\rangle\}$ for $\mathbb{R}^3$ under the weighted dot product of Exercise 3. Do you get the same answer as Exercise 2?
5. $B = \{\langle 1, 1, -1\rangle, \langle 0, -1, 1\rangle, \langle 2, 0, 1\rangle\}$ for $\mathbb{R}^3$ under the inner product generated by the isomorphism:

$$[T] = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}.$$

Do you get the same answer as in Exercise 1 or 3?

6. $B = \{\langle 1, 0, 1\rangle, \langle 2, -1, 0\rangle, \langle 1, 1, -1\rangle\}$ for $\mathbb{R}^3$ under the inner product of Exercise 5. Do you get the same answer as in Exercise 2 or 4?
7. $B = \{\langle 1, -1, 1, -1\rangle, \langle 1, 0, -1, 1\rangle, \langle 1, 1, 0, -1\rangle, \langle -1, 1, 1, -1\rangle\}$ for $\mathbb{R}^4$ under the ordinary dot product.
8. $B = \{\langle 1, -1, 0, 1\rangle, \langle 1, 0, -1, 1\rangle, \langle 1, 2, 0, -1\rangle, \langle -1, 1, -1, 1\rangle\}$ for $\mathbb{R}^4$ under the ordinary dot product.
9. $B = \{\langle 1, -1, 1, -1\rangle, \langle 1, 0, -1, 1\rangle, \langle 1, 1, 0, -1\rangle, \langle -1, 1, 1, -1\rangle\}$ for $\mathbb{R}^4$ under the weighted dot product:

$$\langle \vec{u} | \vec{v} \rangle = 4u_1v_1 + u_2v_2 + 3u_3v_3 + 6u_4v_4.$$

Do you get the same answer as in Exercise 7?

10. $B = \{\langle 1, -1, 0, 1\rangle, \langle 1, 0, -1, 1\rangle, \langle 1, 2, 0, -1\rangle, \langle -1, 1, -1, 1\rangle\}$ for $\mathbb{R}^4$ under the weighted dot product of Exercise 9. Do you get the same answer as in Exercise 8?

11. $B = \{x^2, x, 1\}$ for $\mathbb{P}^2$ under the inner product:

$$\langle\, p(x)\,|\,q(x)\,\rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$

12. $B = \{x^2 + 1, x - 1, 1\}$ for $\mathbb{P}^2$ under the inner product of Exercise 11.

13. $B = \{x^2, x, 1\}$ for $\mathbb{P}^2$ under the inner product: $\langle\, p(x)\,|\,q(x)\,\rangle = \int_0^1 p(x) \cdot q(x)dx$.

14. $B = \{1, x, x^2\}$ for $\mathbb{P}^2$ under the same inner product as Exercise 11. Do you get the same answer as in Exercise 11?

15. $B = \{x^3, x^2, x, 1\}$ for $\mathbb{P}^3$ under the inner product: $\langle\, p(x)\,|\,q(x)\,\rangle = \int_0^1 p(x) \cdot q(x)dx$.

For Exercises (16) to (30): Find the coordinates $\langle \vec{u}\,\rangle_S$ and $\langle \vec{v}\,\rangle_S$ for the two indicated vectors, with respect to the orthonormal basis S that was output by the Gram-Schmidt Algorithm in the indicated Exercise above. Use the formula in the 2nd Theorem of this Section. ***Do not*** solve a system of linear equations.

16. $\vec{u} = \langle 2, -4, 1\rangle$, $\vec{v} = \langle -3, 5, 8\rangle$; Exercise 1.

17. $\vec{u} = \langle 2, -4, 1\rangle, \vec{v} = \langle -3, 5, 8\rangle$; Exercise 2.

18. $\vec{u} = \langle 2, -4, 1\rangle, \vec{v} = \langle -3, 5, 8\rangle$; Exercise 3.

19. $\vec{u} = \langle 2, -4, 1\rangle, \vec{v} = \langle -3, 5, 8\rangle$; Exercise 4.

20. $\vec{u} = \langle 2, -4, 1\rangle, \vec{v} = \langle -3, 5, 8\rangle$; Exercise 5.

21. $\vec{u} = \langle 2, -4, 1\rangle, \vec{v} = \langle -3, 5, 8\rangle$; Exercise 6.

22. $\vec{u} = \langle 3, 6, -2, -4\rangle$, $\vec{v} = \langle 5, -2, 7, -3\rangle$; Exercise 7.

23. $\vec{u} = \langle 3, 6, -2, -4\rangle, \vec{v} = \langle 5, -2, 7, -3\rangle$; Exercise 8.

24. $\vec{u} = \langle 3, 6, -2, -4\rangle, \vec{v} = \langle 5, -2, 7, -3\rangle$; Exercise 9.

25. $\vec{u} = \langle 3, 6, -2, -4\rangle, \vec{v} = \langle 5, -2, 7, -3\rangle$; Exercise 10.

26. $\vec{u} = -x^2 + 2x - 5$, $\vec{v} = 3x^2 - 6x - 4$; Exercise 11.

27. $\vec{u} = -x^2 + 2x - 5$, $\vec{v} = 3x^2 - 6x - 4$; Exercise 12.

28. $\vec{u} = -x^2 + 2x - 5$, $\vec{v} = 3x^2 - 6x - 4$; Exercise 13.

29. $\vec{u} = -x^2 + 2x - 5$, $\vec{v} = 3x^2 - 6x - 4$; Exercise 14.

30. $\vec{u} = 2x^3 + 5x^2 - x + 3$, $\vec{v} = x^3 - x^2 + 7x - 5$; Exercise 15.

31. Consider $V = Span(B)$, where $B = \{\sin(x), \cos(x), \sin(2x)\}$, and V is an inner product space under: $\langle f(x)\,|\,g(x)\rangle = \int_0^{2\pi} f(x) \cdot g(x)\, dx$.
    a. Compute the inner product of all the pairs of functions $f(x)$ and $g(x)$ from B. There should be 6 such pairs. Note: You will need the techniques of Power Reduction Formulas, or Integration by Parts, or Product-to-Sum Formulas from Trigonometry.
    b. Use the Gram-Schmidt Algorithm to find an orthonormal basis S using the natural basis B as the input to the algorithm.
    c. Find the coordinates of:

    $$f(x) = 2\sin(x) + 7\cos(x) - 3\sin(2x) \quad \text{and} \quad g(x) = 5\sin(x) - 2\cos(x) + \sin(2x),$$

    with respect to the orthonormal basis S that you found in (b).

32. Suppose that somebody gave you a set of vectors $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_5\}$ from some $\mathbb{R}^n$, and you were told to apply the Gram-Schmidt Algorithm on B. After a few minutes, you found that $\vec{v}_4 = \vec{\mathbf{0}}_n$ (although $\vec{v}_1$, $\vec{v}_2$, and $\vec{v}_3$ were non-zero). What does this tell you about B? Be as specific as possible.

33. In Exercise 15 of Section 1.3 and Exercise 37 of Section 5.1, we saw the cross product of two vectors from $\mathbb{R}^3$. Use the properties and ideas in those Exercises to show that the orthonormal sets of vectors in $\mathbb{R}^3$ all have the form:

$$\{\vec{u}, \vec{v}, \vec{u} \times \vec{v}\} \text{ or } \{\vec{u}, \vec{v}, -\vec{u} \times \vec{v}\},$$

where $\vec{u}$ and $\vec{v}$ are orthonormal vectors from $\mathbb{R}^3$. In other words, $\vec{u}$ and $\vec{v}$ are already orthogonal to each other, and they are both unit vectors.

34. Consider $\mathbb{P}^3$ under the inner product:

$$\langle\, p(x) \,|\, q(x)\,\rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1) + p(2)q(2).$$

Construct an orthonormal set $\{u_1(x), u_2(x), u_3(x), u_4(x)\}$, where $u_1(-2) = 0$, $u_1(0) = u_1(1) = u_1(2)$, and $u_2(0) = 1$, $u_2(-2) = u_2(1) = u_2(2) = 0$, and so on. See the 2nd Example in this Section for a similar construction in $\mathbb{P}^2$.

35. ***Introduction to Fourier Series:*** In a course in Differential Equations, we encounter ***trigonometric polynomials***, which are linear combinations from the ***countable infinite set:***

$$S = \{\, \sin(nx),\ \cos(nx) \,|\, n \in \mathbb{N}\},$$

which includes the constant function 1. Although we allow only a finite number of terms in a linear combination, a Fourier Series is an ***infinite series*** involving these function, so it has the form:

$$c_0 + \sum_{n=1}^{\infty} (c_n \sin(nx) + d_n \cos(nx)).$$

The objective of this Exercise is to show that S is already an ***orthogonal*** set under the inner product:

$$\langle f(x) \,|\, g(x)\rangle = \int_0^{2\pi} f(x) \cdot g(x)\, dx,$$

which we saw in Exercise 31. In order to compute these integrals, though we need the following ***Product-to-Sum Identities*** from Trigonometry:

$$\begin{aligned}
\sin(\alpha)\cos(\beta) &= \tfrac{1}{2}[\sin(\alpha+\beta) + \sin(\alpha-\beta)] \\
\sin(\alpha)\sin(\beta) &= \tfrac{1}{2}[\cos(\alpha-\beta) - \cos(\alpha+\beta)] \\
\cos(\alpha)\cos(\beta) &= \tfrac{1}{2}[\cos(\alpha+\beta) + \cos(\alpha-\beta)]
\end{aligned}$$

a. Show that if $n \neq m$ are natural numbers:

$$\begin{aligned}
\langle\, \sin(nx) \,|\, \cos(mx)\rangle &= 0, \\
\langle\, \sin(nx) \,|\, \sin(mx)\rangle &= 0, \text{ and} \\
\langle\, \cos(nx) \,|\, \cos(mx)\rangle &= 0.
\end{aligned}$$

b. Show that for all natural number $n > 0$:

$$\begin{aligned}
\langle\, \sin(nx) \,|\, \sin(nx)\rangle &= \pi, \text{ and} \\
\langle\, \cos(nx) \,|\, \cos(nx)\rangle &= \pi.
\end{aligned}$$

c. Construct an orthonormal set of vectors S' from S.

7.4 Orthogonal Complements and Decompositions

We saw in Chapter 1 that a plane Π passing through the origin in $\mathbb{R}^3$ has a unique normal line L that also passes through the origin. If Π has equation $ax+by+cz=0$, then $L = Span(\{\langle a,b,c\rangle\})$. Notice that Π and L are ***subspaces*** of $\mathbb{R}^3$. More importantly, though, we can say that Π and L are ***exhaustive***: Any vector ***orthogonal*** to Π must be a member of L, and any vector ***orthogonal*** to L must be a member of Π. We also saw in Chapter 2 that we can ***decompose*** any vector uniquely as a sum of a vector on Π and a vector on L. In this Section, we will generalize these concepts.

Orthogonal Complements

We are now ready to generalize an important construction in inner product spaces:

> ***Definition/Theorem:*** Let W be any subspace of an inner product space V. We define the ***orthogonal complement*** of W, another ***subspace*** of V, by:
> $$W^{\perp} = \left\{\vec{v} \in V \,|\, \langle \vec{v}|\vec{w}\rangle = 0 \text{ for all } \vec{w} \in W\right\}.$$

Proof of the Subspace Property: We need to show that $W^{\perp}$ is non-empty and closed under vector addition and scalar multiplication. Since $\vec{\mathbf{0}}_V$ is orthogonal to any vector, $\vec{\mathbf{0}}_V \in W^{\perp}$, so W is non-empty. Now, if $\vec{v}_1$ and $\vec{v}_2$ are members of $W^{\perp}$, then:

$$\langle \vec{v}_1 + \vec{v}_2|\vec{w}\rangle = \langle \vec{v}_1|\vec{w}\rangle + \langle \vec{v}_2|\vec{w}\rangle = 0+0 = 0,$$

for any member $\vec{w}$ of W. Thus $\vec{v}_1+\vec{v}_2 \in W^{\perp}$. Similarly, $\langle k\cdot\vec{v}_1|\vec{w}\rangle = k\cdot 0 = 0$, so $k\cdot\vec{v}_1 \in W^{\perp}$. ■

We will also need the following computational device, whose proof we leave as an Exercise:

> ***Theorem:*** Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ be a ***basis*** for a finite dimensional subspace W of an inner product space V. Then $\vec{v} \in V$ is a member of $W^{\perp}$ ***if and only if***:
> $$\langle \vec{v}|\vec{w}_i\rangle = 0 \text{ for all } i = 1, 2, \ldots, k.$$
> Thus, it is both necessary and sufficient that we check that an arbitrary vector $\vec{v} \in V$ is orthogonal to every member of a basis B for W.

In Section 1.7, we saw that if W is a subspace of $\mathbb{R}^n$ under the usual dot product, then $W^{\perp}$ is the nullspace of the matrix A whose ***rows*** Span W. This is because for any matrix A, the rowspace and the nullspace are orthogonal complements of each other. By the Theorem above, we can again find a basis for the orthogonal complement of a subspace W of V by solving a ***homogeneous system of equations***. However, the ***nullspace*** of this system needs to be ***decoded*** to find a basis for $W^{\perp}$.

Example: Let $W = Span(\{\,2x-1,\ x^2-3x+4\})$ be a subspace of $\mathbb{P}^2$ under the inner product:

$$\langle\, p(x)|q(x)\rangle = p(-2)q(-2)+p(0)q(0)+p(1)q(1).$$

As before, we will need a table of values for $w_1(x) = 2x-1$ and $w_2(x) = x^2-3x+4$.
However, we will need the inner product of these polynomials with an ***arbitrary*** polynomial of $\mathbb{P}^2$:

$$p(x) = d_0 + d_1 x + d_2 x^2.$$

Thus, we will include the values of $p(c_i)$ in our table:

c_i	$w_1(c_i)$	$w_2(c_i)$	$p(c_i)$
-2	-5	14	$d_0 - 2d_1 + 4d_2$
0	-1	4	d_0
1	1	2	$d_0 + d_1 + d_2$

Thus, $p(x)$ is a member of $W^{\perp}$ ***if and only if:***

$$\langle p(x) | w_1(x) \rangle = 0, \ \ \textbf{\textit{and}} \ \ \langle p(x) | w_2(x) \rangle = 0.$$

Using our table of values above, these two conditions translate to:

$$0 = (d_0 - 2d_1 + 4d_2) \cdot (-5) + (d_0) \cdot (-1) + (d_0 + d_1 + d_2) \cdot 1 = -5d_0 + 11d_1 - 19d_2, \ \ \textbf{\textit{and}}$$
$$0 = (d_0 - 2d_1 + 4d_2) \cdot (14) + (d_0) \cdot (4) + (d_0 + d_1 + d_2) \cdot 2 = 20d_0 - 26d_1 + 58d_2.$$

Thus, $\langle d_0, d_1, d_2 \rangle$ is a member of the ***nullspace*** of the matrix:

$$\begin{bmatrix} -5 & 11 & -19 \\ 20 & -26 & 58 \end{bmatrix}, \ \text{ whose rref is: } \begin{bmatrix} 1 & 0 & 8/5 \\ 0 & 1 & -1 \end{bmatrix}.$$

Thus, d_2 is a free variable, and a basis for the nullspace is $\{\langle -8, 5, 5 \rangle\}$. But let us not forget to ***decode*** this basis, and conclude that:

$$W^{\perp} = Span(\{-8 + 5x + 5x^2\}).$$

We can check that $p(x) = -8 + 5x + 5x^2$ is orthogonal to our basis vectors for W. We will need the three values: $p(-2) = 2$, $p(0) = -8$, and $p(1) = 2$. Now we find:

$$\langle p(x) | w_1(x) \rangle = 2(-5) + (-8)(-1) + 2(1) = 0, \ \text{ and}$$
$$\langle p(x) | w_2(x) \rangle = 2(14) + (-8)(4) + 2(2) = 0.$$

Thus $p(x)$ is indeed orthogonal to W, and $dim(W^{\perp}) = 1$ □

Example: Let us stay with $\mathbb{P}^2$, but since polynomials are ***continuous***, we can use a definite integral to impose an inner product on $\mathbb{P}^2$, such as:

$$\langle p(x) | q(x) \rangle = \int_0^1 p(x) \cdot q(x)\, dx.$$

Let us consider the 1-dimensional subspace $W = Span(\{x^2 - 3\})$. As in the previous example, to find the orthogonal complement of W, we must consider an arbitrary member of $\mathbb{P}^2$:

$$p(x) = d_0 + d_1 x + d_2 x^2.$$

Since there is only one vector in the basis, we have to satisfy only one condition, namely:

$$0 = \int_0^1 p(x) \cdot (x^2 - 3)\, dx = \int_0^1 (d_0 + d_1 x + d_2 x^2) \cdot (x^2 - 3)\, dx$$
$$= \int_0^1 (d_0 x^2 - 3d_0 + d_1 x^3 - 3d_1 x + d_2 x^4 - 3d_2 x^2) dx$$

$$= d_0\frac{x^3}{3} - 3d_0x + d_1\frac{x^4}{4} - 3d_1\frac{x^2}{2} + d_2\frac{x^5}{5} - 3d_2\frac{x^3}{3}\Big|_0^1$$

$$= \frac{d_0}{3} - 3d_0 + \frac{d_1}{4} - \frac{3d_1}{2} + \frac{d_2}{5} - d_2 = -\frac{8}{3}d_0 - \frac{5}{4}d_1 - \frac{4}{5}d_2 = 0.$$

Thus we have two free variables, d_1 and d_2, and we have:

$$d_0 = -\frac{15}{32}d_1 - \frac{3}{10}d_2.$$

We substitute this condition ***back*** into $p(x)$ to get:

$$p(x) = \left(-\frac{15}{32}d_1 - \frac{3}{10}d_2\right) + d_1x + d_2x^2 = -\frac{1}{32}d_1(15 - 32x) - \frac{1}{10}d_2(3 - 10x^2).$$

Since the two polynomials in the set above have different degrees, they are linearly ***independent***. Thus, $W^\perp$ is a ***2-dimensional*** subspace, and:

$$W^\perp = Span(\{15 - 32x, 3 - 10x^2\}).$$

We can again check that these two polynomials are orthogonal to our single basis polynomial for W:

$$\int_0^1 (15 - 32x) \cdot (x^2 - 3)\,dx = \int_0^1 (15x^2 - 45 - 32x^3 + 96x)\,dx$$

$$= 15\frac{x^3}{3} - 45x - 32\frac{x^4}{4} + 96\frac{x^2}{2}\Big|_0^1 = 5 - 45 - 8 + 48 = 0,$$

and similarly, $\int_0^1 (3 - 10x^2) \cdot (x^2 - 3)\,dx = 0.$ □

The only vector in common between a plane Π through the origin and its normal line L is the zero vector. It should therefore come as no surprise that when it comes to subspaces and their orthogonal complements, this is true in general:

Theorem: Let W be any subspace of an inner product space V, with orthogonal complement $W^\perp$. Then: $W \cap W^\perp = \left\{\vec{\mathbf{0}}_V\right\}$.

Proof: Suppose that $\vec{w} \in W \cap W^\perp$. In other words, $\vec{w} \in W$ ***and*** $\vec{w} \in W^\perp$. But by definition, every member of W is ***orthogonal*** to every member of $W^\perp$. Since $\vec{w}$ is a member of ***both*** subspaces, it must therefore be orthogonal to ***itself***, that is: $\langle \vec{w} | \vec{w} \rangle = 0$.

But by the ***Positivity Property***, this means that $\vec{w} = \vec{\mathbf{0}}_V$. Thus, $W \cap W^\perp = \left\{\vec{\mathbf{0}}_V\right\}$. ■

The Dimension Theorem for Orthogonal Complements

When we found a basis for $W^\perp$ above, we solved a system of homogeneous equations in order to find all the vectors of V which were orthogonal to every member of the basis for W. The resulting basis that we found for $W^\perp$ is not necessarily an orthogonal set.

Fortunately, the Gram-Schmidt process has the added advantage of helping us find an ***orthonormal basis*** for W as well as for $W^\perp$. We will start with any basis $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ for W, where $\dim(W) = k$. Next, we enlarge B to a basis:

$$B^{/} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k, \vec{w}_{k+1}, \ldots, \vec{w}_n\}$$

for V, where $\dim(V) = n$. Now, applying the Gram-Schmidt Algorithm on $B^{/}$, we will get the set:

$$S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k, \vec{u}_{k+1}, \ldots, \vec{u}_n\}$$

which is an ***orthonormal basis*** for ***all*** of V. By our proof of the effectiveness of the Gram-Schmidt Algorithm, we know that the resulting subset $S_1 = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ***orthonormal basis*** for $W = W_k = Span(B)$.

We claim that $S_2 = \{\vec{u}_{k+1}, \ldots, \vec{u}_n\}$ is an ***orthonormal basis*** for $W^{\perp}$.

The proof relies again on our standard technique. Since S is an orthonormal basis for V, any vector $\vec{v} \in V$ can be written ***uniquely*** as:

$$\vec{v} = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k + c_{k+1}\vec{u}_{k+1} + \cdots + c_n\vec{u}_n.$$

For $\vec{v}$ to be a member of $W^{\perp}$, it is necessary and sufficient that $\langle \vec{u}_i | \vec{v} \rangle = 0$ for $i = 1, 2, \ldots, k$. But again, by taking the inner product of both sides of this equation with respect to $\vec{u}_1$, we get:

$$\begin{aligned}
\langle \vec{u}_1 | \vec{v} \rangle &= \langle \vec{u}_1 | c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k + c_{k+1}\vec{u}_{k+1} + \cdots + c_n\vec{u}_n \rangle \\
&= c_1\langle \vec{u}_1 | \vec{u}_1 \rangle + c_2\langle \vec{u}_1 | \vec{u}_2 \rangle + \cdots + c_k\langle \vec{u}_1 | \vec{u}_k \rangle + c_{k+1}\langle \vec{u}_1 | \vec{u}_{k+1} \rangle + \cdots + c_n\langle \vec{u}_1 | \vec{u}_n \rangle \\
&= c_1.
\end{aligned}$$

Thus, $\langle \vec{u}_1 | \vec{v} \rangle = 0$ ***if and only if*** $c_1 = 0$. Continuing in this manner with $\vec{u}_2$ through $\vec{u}_k$, we must have:

$$\vec{v} \in W^{\perp} \quad \textbf{\textit{if and only if}} \quad c_1 = c_2 = c_3 = \cdots = c_k = 0.$$

In other words, $\vec{v}$ has the form: $\vec{v} = c_{k+1}\vec{u}_{k+1} + \cdots + c_n\vec{u}_n$.

This shows that $S_2 = \{\vec{u}_{k+1}, \ldots, \vec{u}_n\}$ ***Spans*** $W^{\perp}$. Since S_2 is a subset of an orthonormal set, it is also an orthonormal set. Thus S_2 is also ***linearly independent***, and we can conclude that S_2 is a ***basis*** for $W^{\perp}$. We summarize the discussion above in the following:

Theorem — The Dimension Theorem for Orthogonal Complements:

Let W be a k-dimensional subspace of an n-dimensional inner product space V. Suppose $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a ***basis*** for W, and we enlarge B to $B^{/}$, a basis for V, where:

$$B^{/} = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k, \vec{w}_{k+1}, \ldots, \vec{w}_n\}.$$

Now, suppose we apply the Gram-Schmidt Algorithm to $B^{/}$, which outputs the set:

$$S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k, \vec{u}_{k+1}, \ldots, \vec{u}_n\},$$

an ***orthonormal basis*** for ***all*** of V. Let us denote by:

$$S_1 = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\} \quad \text{and} \quad S_2 = \{\vec{u}_{k+1}, \ldots, \vec{u}_n\}.$$

Then: S_1 is an ***orthonormal basis*** for $W = Span(B)$, and S_2 is an ***orthonormal basis*** for $W^{\perp}$. Consequently:

$$dim(W) + dim(W^{\perp}) = dim(V) = n.$$

More generally, if $S = S_1 \cup S_2$ is ***any*** orthonormal basis for V, where S_1 and S_2 are non-empty subsets of S with no member in common, then $W_1 = Span(S_1)$ and $W_2 = Span(S_2)$ are ***orthogonal complements*** of ***each other***. Thus:

$$(W^{\perp})^{\perp} = W.$$

Note: It is not necessarily true that $(W^{\perp})^{\perp} = W$ if W is a subspace of an ***infinite dimensional*** inner product space. However, it is ***always true*** that $W \subseteq (W^{\perp})^{\perp}$. This also generalizes our version for subspaces $W \trianglelefteq \mathbb{R}^n$ in Chapter 1 under the ordinary dot product: $dim(W) + dim(W^{\perp}) = n = dim(\mathbb{R}^n)$. The final paragraph in the Theorem will be proven in the Exercises.

Example: In the previous Section, we found an orthonormal basis for $\mathbb{R}^4$ (under the ordinary dot product) using the basis:

$$B = \{\vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4\} = \{\langle 1,-1, 1, 0\rangle, \langle 1, 0, 1,-1\rangle, \langle -1, 1, 1, 1\rangle, \langle 1, 1, 1,-1\rangle\},$$

as input to the Gram-Schmidt Algorithm. The output was:

$$S = \left\{\frac{1}{\sqrt{3}}\langle 1,-1, 1, 0\rangle, \frac{1}{\sqrt{15}}\langle 1, 2, 1,-3\rangle, \frac{1}{\sqrt{90}}\langle -3, 4, 7, 4\rangle, \frac{1}{\sqrt{18}}\langle 3, 2,-1,2\rangle\right\}.$$

Now, suppose $U = Span(\{\langle 1,-1, 1, 0\rangle\})$, where $\langle 1,-1, 1, 0\rangle$ is the first member of B. According to the Theorem above:

$$U^{\perp} = Span(\{\langle 1, 2, 1,-3\rangle, \langle -3, 4, 7, 4\rangle, \langle 3, 2,-1,2\rangle\}).$$

Notice that we can ignore the normalization. Similarly:

If $V = Span(\{\langle 1,-1, 1, 0\rangle, \langle 1, 0, 1,-1\rangle\})$, then $V^{\perp} = Span(\{\langle -3, 4, 7, 4\rangle, \langle 3, 2,-1,2\rangle\})$, and

if $W = Span(\{\langle 1,-1, 1, 0\rangle, \langle 1, 0, 1,-1\rangle, \langle -1, 1, 1, 1\rangle\})$, then $W^{\perp} = Span(\{\langle 3, 2,-1,2\rangle\})$ □

Example: In the previous Section, we also found an orthonormal basis for $\mathbb{P}^2$ under the inner product:

$$\langle p(x)|q(x)\rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1),$$

by starting with the standard basis $B = \{1, x, x^2\}$. Our final orthonormal basis was:

$$S = \left\{\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{42}}(3x + 1), \frac{1}{\sqrt{126}}(7x^2 + 8x - 9)\right\}.$$

Thus, if we let $W = Span(\{1, x\}) = Span(\{1, 3x + 1\})$, then we immediately get:

$$W^{\perp} = Span(\{7x^2 + 8x - 9\}).$$

Similarly, if we let $U = Span(\{1\})$, then:

$$U^{\perp} = Span(\{3x + 1, 7x^2 + 8x - 9\}). \square$$

We emphasize that this only works if W is Spanned by the first k ***consecutive*** vectors of the basis S.

Orthogonal Decompositions

In Section 2.2, we studied the projection and reflection of a vector in $\mathbb{R}^2$ onto and across a line L that passes through the origin. We showed that any vector $\vec{v} \in \mathbb{R}^2$ can be expressed (uniquely) as a sum:

$$\vec{v} = proj_L(\vec{v}) + proj_{L^{\perp}}(\vec{v}),$$

where $proj_L(\vec{v}) \in L$ and $proj_{L^{\perp}}(\vec{v}) \in L^{\perp}$, the unique line in $\mathbb{R}^2$ perpendicular to L and passing through the origin. We referred to this as an orthogonal decomposition because the two vectors in the sum are orthogonal to each other. We will now generalize this concept:

Theorem: Let W be a ***finite-dimensional subspace*** of an inner product space V. Then: any vector $\vec{v} \in V$ can be expressed ***uniquely*** as a sum:

$$\vec{v} = \vec{w}_1 + \vec{w}_2,$$

where $\vec{w}_1 \in W$ and $\vec{w}_2 \in W^{\perp}$.

We refer to this as the ***orthogonal decomposition*** of $\vec{v}$ with respect to W and $W^{\perp}$.

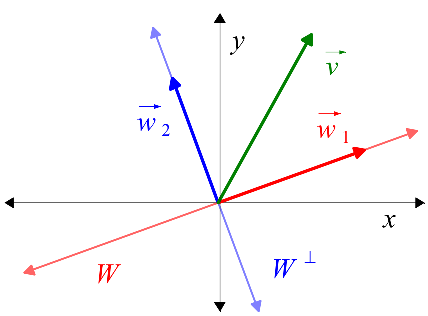
Moreover, we can explicitly construct $\vec{w}_1$ and $\vec{w}_2$ as follows:

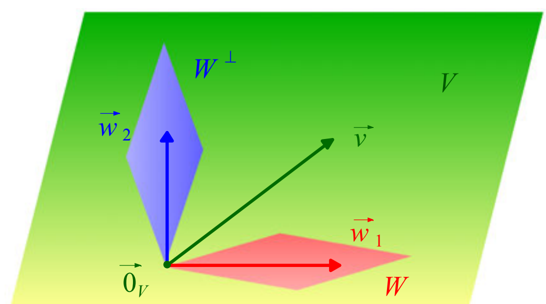
If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is any ***orthonormal basis*** for W, then:

$$\vec{w}_1 = \langle \vec{v} | \vec{u}_1 \rangle \cdot \vec{u}_1 + \langle \vec{v} | \vec{u}_2 \rangle \cdot \vec{u}_2 + \cdots + \langle \vec{v} | \vec{u}_k \rangle \cdot \vec{u}_k, \text{ and}$$
$$\vec{w}_2 = \vec{v} - \vec{w}_1.$$

We call $\vec{w}_1$ the ***orthogonal projection*** of $\vec{v}$ onto W, and $\vec{w}_2$ the ***orthogonal projection*** of $\vec{v}$ onto $W^{\perp}$. We write them as:

$$\vec{w}_1 = proj_W(\vec{v}) \quad \text{and} \quad \vec{w}_2 = proj_{W^{\perp}}(\vec{v}).$$





Orthogonal Decompositions in $\mathbb{R}^2$ and in a General Inner Product Space V

Proof: We can easily prove the ***uniqueness*** portion, so let us do that first. Suppose:

$$\vec{v} = \vec{w}_1 + \vec{w}_2 = \vec{z}_1 + \vec{z}_2,$$

where $\vec{w}_1$ and $\vec{z}_1$ are members of W, and $\vec{w}_2$ and $\vec{z}_2$ are members of $W^{\perp}$. Then:

$$\vec{w}_1 - \vec{z}_1 = \vec{z}_2 - \vec{w}_2.$$

But by the closure property, the vector on the left is a member of W, and the vector on the right is a member of $W^{\perp}$. The equals sign means that we are referring to the same vector, so this vector must be a member of $W \cap W^{\perp}$, which consists only of the zero vector. Thus:

$$\vec{w}_1 - \vec{z}_1 = \vec{0}_V, \quad \textbf{\textit{and}} \quad \vec{z}_2 - \vec{w}_2 = \vec{0}_V.$$

In other words, $\vec{w}_1 = \vec{z}_1$ and $\vec{z}_2 = \vec{w}_2$, so the decomposition is ***unique***.

Now, let us prove the ***existence*** portion. Let $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ be any orthonormal basis for W. Let us form the vector:

$$\vec{w}_1 = \left\langle \vec{v} \,|\, \vec{u}_1 \right\rangle \vec{u}_1 + \left\langle \vec{v} \,|\, \vec{u}_2 \right\rangle \vec{u}_2 + \cdots + \left\langle \vec{v} \,|\, \vec{u}_k \right\rangle \vec{u}_k$$

in our formula. Certainly, $\vec{w}_1 \in W$ by the basis property. Now, we must show that:

$$\vec{w}_2 = \vec{v} - \vec{w}_1$$

is a member of $W^{\perp}$, that is, it is orthogonal to every vector in $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$. We compute:

$$\begin{aligned}\langle \vec{u}_1 \,|\, \vec{w}_2 \rangle &= \langle \vec{u}_1 \,|\, \vec{v} - \vec{w}_1 \rangle = \langle \vec{u}_1 \,|\, \vec{v} - [\langle \vec{v} \,|\, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v} \,|\, \vec{u}_2 \rangle \vec{u}_2 + \cdots + \langle \vec{v} \,|\, \vec{u}_k \rangle \vec{u}_k] \rangle \\ &= \langle \vec{u}_1 \,|\, \vec{v} \rangle - \langle \vec{v} \,|\, \vec{u}_1 \rangle \cdot \langle \vec{u}_1 \,|\, \vec{u}_1 \rangle - \langle \vec{v} \,|\, \vec{u}_2 \rangle \cdot \langle \vec{u}_1 \,|\, \vec{u}_2 \rangle - \cdots - \langle \vec{v} \,|\, \vec{u}_k \rangle \cdot \langle \vec{u}_1 \,|\, \vec{u}_k \rangle \\ &= \langle \vec{u}_1 \,|\, \vec{v} \rangle - \langle \vec{v} \,|\, \vec{u}_1 \rangle = 0,\end{aligned}$$

and similarly, $\langle \vec{u}_2 \,|\, \vec{w}_2 \rangle = \langle \vec{u}_3 \,|\, \vec{w}_2 \rangle = \cdots = \langle \vec{u}_k \,|\, \vec{w}_2 \rangle = 0.$ ■

We wish to point out that this recipe for the decomposition works out for ***any*** orthonormal basis for W, not just the basis that we produce using the Gram-Schmidt Algorithm. Thus, the Theorem is actually more flexible, although in practice, we use the Gram-Schmidt Algorithm to construct this basis. Also, the Theorem only requires that W is finite dimensional, but the ambient space V could be infinite dimensional.

Example: Let us continue with our previous Example. We have the orthonormal basis:

$$B = \left\{ \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{42}}(3x+1), \frac{1}{\sqrt{126}}(7x^2+8x-9) \right\}$$

for $\mathbb{P}^2$ under the inner product:

$$\langle\, p(x) | q(x) \,\rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$

If $W = Span(\{1, x\})$, we found ***orthonormal*** bases for W and $W^{\perp}$, as:

$$W = Span\left\{ \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{42}}(3x+1) \right\} \quad \text{and} \quad W^{\perp} = Span\left(\left\{ \frac{1}{\sqrt{126}}(7x^2+8x-9) \right\} \right).$$

Now, let us find the orthogonal decomposition of the polynomial $r(x) = 5x^2 - 8x + 3$, with respect to W and $W^{\perp}$. To compute the inner product, we need:

$$r(-2) = 39,\ r(0) = 3, \text{ and } r(1) = 0.$$

First Solution:

To apply the formula directly, we compute the coefficients:

$$\left\langle r(x) \,\middle|\, \frac{1}{\sqrt{3}} \right\rangle = (39)\frac{1}{\sqrt{3}} + (3)\frac{1}{\sqrt{3}} + (0)\frac{1}{\sqrt{3}} = \frac{42}{\sqrt{3}}, \text{ and}$$

$$\left\langle r(x) \,\middle|\, \frac{1}{\sqrt{42}}(3x+1) \right\rangle = (39)\frac{-5}{\sqrt{42}} + (3)\frac{1}{\sqrt{42}} + (0)\frac{4}{\sqrt{42}} = \frac{-192}{\sqrt{42}}.$$

Thus:

$$\begin{aligned}\vec{w}_1(x) &= \frac{42}{\sqrt{3}} \cdot \frac{1}{\sqrt{3}} - \frac{192}{\sqrt{42}} \cdot \frac{1}{\sqrt{42}}(3x+1) \\ &= 14(1) - \frac{32}{7}(3x+1) = \frac{66}{7} - \frac{96}{7}x, \text{ and} \\ \vec{w}_2(x) &= r(x) - \vec{w}_1(x) = 5x^2 - 8x + 3 - \left[\frac{66}{7} - \frac{96}{7}x \right] \\ &= 5x^2 + \frac{40}{7}x - \frac{45}{7} = \frac{5}{7}(7x^2+8x-9).\end{aligned}$$

It is clear that $\vec{w}_2(x)$ is a member of $W^{\perp}$, so our decomposition is correct.

Second Solution:

Notice that $W^{\perp}$ is only 1-dimensional, so we can save time by finding $\vec{w}_2(x)$ ***first***:

$$\left\langle r(x) \mid \frac{1}{\sqrt{126}}(7x^2+8x-9) \right\rangle = (39)\frac{3}{\sqrt{126}} + (3)\frac{-9}{\sqrt{126}} + (0)\frac{6}{\sqrt{126}} = \frac{90}{\sqrt{126}}$$

and thus:

$$\vec{w}_2(x) = \frac{90}{\sqrt{126}} \cdot \frac{1}{\sqrt{126}}(7x^2+8x-9) = \frac{5}{7}(7x^2+8x-9).$$

Of course we will get the same $\vec{w}_1$, this time using $\vec{w}_1(x) = r(x) - \vec{w}_2(x)$.

Third Solution:

Now, here's the kicker. Let us completely ignore all the normalization, and consider the ***orthogonal basis*** $B' = \{1,\ 3x+1,\ 7x^2+8x-9\}$ for $\mathbb{P}^2$.

The first two vectors still form a basis for W, and the third vector forms a basis for $W^{\perp}$. But notice that the degrees of the members are in increasing order, so we can easily find the coefficients of any vector with respect to this basis, without too much difficulty, starting with the highest degree:

$$5x^2 - 8x + 3 = 14(1) - \frac{32}{7}(3x+1) + \frac{5}{7}(7x^2+8x-9).$$

So again $\vec{w}_1(x) = 14(1) - \frac{32}{7}(3x+1) = \frac{66}{7} - \frac{96}{7}x$, and $\vec{w}_2(x) = \frac{5}{7}(7x^2+8x-9)$. □

This Example was still relatively simple because our space is only 3-dimensional. If the space has a higher dimension, a good strategy would be to find the projection of $\vec{w}$ onto the subspace (either W or $W^{\perp}$) with the smaller dimension.

7.4 Section Summary

Let W be a subspace of an inner product space V. We define the ***orthogonal complement*** of W, which is also a ***subspace*** of V, by: $W^{\perp} = \left\{ \vec{v} \in V \mid \langle \vec{v} \mid \vec{w} \rangle = 0 \text{ for all } \vec{w} \in W \right\}$.

Let $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ be a ***basis*** for a finite dimensional subspace W of an inner product space V. Then $\vec{v}$ is a member of $W^{\perp}$ ***if and only if*** $\langle \vec{v} \mid \vec{w}_i \rangle = 0$ for all $i = 1, 2, ..k$.

Thus, it is both necessary and sufficient that we check that an arbitrary member $\vec{v}$ of V is orthogonal to every member of a basis for W.

The Dimension Theorem for Orthogonal Complements: Let $W = Span(B) \trianglelefteq V$, where $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a basis for W. If we enlarge B to:

$$B' = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k, \vec{w}_{k+1}, \ldots, \vec{w}_n\},$$

a basis for V, and $S = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k, \vec{u}_{k+1}, \ldots, \vec{u}_n\}$ is the orthonormal basis for V obtained after applying the Gram-Schmidt Algorithm to B', then $S_1 = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an ***orthonormal basis*** for W, and $S_2 = \{\vec{u}_{k+1}, \ldots, \vec{u}_n\}$ is an ***orthonormal basis*** for $W^{\perp}$.

Consequently: $\dim(W) + \dim(W^{\perp}) = n$.

More generally, if $S = S_1 \cup S_2$ is an orthonormal basis for V, where S_1 and S_2 are non-empty subsets of S with no member in common, then $W_1 = Span(S_1)$ and $W_2 = Span(S_2)$ are orthogonal complements of ***each other***. Thus $(W^{\perp})^{\perp} = W$.

Let W be a ***finite-dimensional subspace*** of an inner product space V. Then, any vector $\vec{v} \in V$ can be expressed ***uniquely*** as a sum: $\vec{v} = \vec{w}_1 + \vec{w}_2$, where $\vec{w}_1 \in W$ and $\vec{w}_2 \in W^{\perp}$.

We refer to this as the ***orthogonal decomposition*** of $\vec{v}$ with respect to W and $W^{\perp}$. Moreover, we can explicitly construct $\vec{w}_1$ and $\vec{w}_2$ as follows: If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is any orthonormal basis for W, then:

$$\vec{w}_1 = \langle \vec{v} \,|\, \vec{u}_1 \rangle \vec{u}_1 + \langle \vec{v} \,|\, \vec{u}_2 \rangle \vec{u}_2 + \cdots + \langle \vec{v} \,|\, \vec{u}_k \rangle \vec{u}_k, \quad \text{and} \quad \vec{w}_2 = \vec{v} - \vec{w}_1.$$

We call $\vec{w}_1$ the ***orthogonal projection*** of $\vec{v}$ onto W, and $\vec{w}_2$ the ***orthogonal projection*** of $\vec{v}$ onto $W^{\perp}$. We write these symbolically as $\vec{w}_1 = proj_W(\vec{v})$ and $\vec{w}_2 = proj_{W^{\perp}}(\vec{v})$.

7.4 Exercises

For Exercises (1) to (14): ***Without*** using the Gram-Schmidt Algorithm, find a basis for the orthogonal complement $W^{\perp}$ of the following subspaces W of the given ambient spaces, with respect to the given inner product:

1. $W = Span(\{\langle 1, 0, -1 \rangle, \langle -1, 1, 2 \rangle\}) \trianglelefteq \mathbb{R}^3$, under the ordinary dot product.
2. $W = Span(\{\langle 1, -1, 3 \rangle\}) \trianglelefteq \mathbb{R}^3$, under the ordinary dot product.
3. $W = Span(\{\langle 1, 0, -1 \rangle, \langle -1, 1, 2 \rangle\}) \trianglelefteq \mathbb{R}^3$, under the weighted dot product:
$$\langle \vec{u} | \vec{v} \rangle = 4u_1v_1 + 5u_2v_2 + 3u_3v_3.$$
4. $W = Span(\{\langle 1, -1, 3 \rangle\}) \trianglelefteq \mathbb{R}^3$, under the inner product of Exercise 3.
5. $W = Span(\{\langle 1, 0, -1, 1 \rangle, \langle -1, 1, 1, 1 \rangle\}) \trianglelefteq \mathbb{R}^4$, under the ordinary dot product.
6. $W = Span(\{\langle 1, 0, -1, 1 \rangle, \langle -1, 1, 1, 1 \rangle\}) \trianglelefteq \mathbb{R}^4$, under the weighted dot product:
$$\langle \vec{u} | \vec{v} \rangle = 4u_1v_1 + u_2v_2 + 3u_3v_3 + 6u_4v_4.$$
7. $W = Span(\{\langle 1, 0, -1, 1 \rangle, \langle -1, 1, 1, 1 \rangle, \langle -1, 1, 0, -1 \rangle\}) \trianglelefteq \mathbb{R}^4$, under the inner product of Exercise 6.
8. $W = Span(\{\langle 1, -1, 2 \rangle, \langle -1, 3, 1 \rangle\}) \trianglelefteq \mathbb{R}^3$, under the inner product generated by the isomorphism:
$$[T] = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}.$$
9. $W = Span(\{\langle 1, 1, -3 \rangle\}) \trianglelefteq \mathbb{R}^3$, under the inner product of Exercise 8.
10. $W = Span(\{x^2\}) \trianglelefteq \mathbb{P}^2$, under the inner product:
$$\langle p(x) | q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1).$$
11. $W = Span(\{x + 1, x^2 - 1\}) \trianglelefteq \mathbb{P}^2$, under the inner product of the Exercise 10.
12. $W = Span(\{4 - x, 2x + x^3\}) \trianglelefteq \mathbb{P}^3$, under the inner product:
$$\langle p(x) | q(x) \rangle = p(-1)q(-1) + p(1)q(1) + p(2)q(2) + p(4)q(4).$$
13. $W = Span(\{4 - x\}) \trianglelefteq \mathbb{P}^3$, under the inner product of Exercise 12.

14. $W = Span(\{x^2 - x + 1\}) \trianglelefteq \mathbb{P}^2$ under the inner product: $\langle\, p(x) | q(x) \,\rangle = \int_0^1 p(x)q(x)dx$.

For Exercises (15) to (25): Use your answers in the Section 7.3 Exercises (as indicated, or numbered in parentheses) to find an ***orthonormal basis*** for W and the orthogonal complement $W^\perp$ of the following subspaces of the indicated ambient spaces (there should be no need for further computations). Verify the Dimension Theorem for W and $W^\perp$.

15. $W = Span(\{\langle 1, 1, -1\rangle, \langle 0, -1, 1\rangle\}) \trianglelefteq \mathbb{R}^3$ under the ordinary dot product (Exercise 1).
16. $W = Span(\{\langle 1, 0, 1\rangle\}) \trianglelefteq \mathbb{R}^3$ under the ordinary dot product (Exercise 2).
17. $W = Span(\{\langle 1, 1, -1\rangle\}) \trianglelefteq \mathbb{R}^3$ under the weighted inner product of Exercise 3.
18. $W = Span(\{\langle 1, 0, 1\rangle, \langle 2, -1, 0\rangle\}) \trianglelefteq \mathbb{R}^3$ under the inner product of Exercise 4.
19. $W = Span(\{\langle 1, 0, 1\rangle, \langle 2, -1, 0\rangle\}) \trianglelefteq \mathbb{R}^3$ under the inner product of Exercise 6.
20. $W = Span(\{\langle 1, -1, 1, -1\rangle, \langle 1, 0, -1, 1\rangle\}) \trianglelefteq \mathbb{R}^4$ under the ordinary dot product (Exercise 7).
21. $W = Span(\{\langle 1, -1, 0, 1\rangle\}) \trianglelefteq \mathbb{R}^4$ under the ordinary dot product (Exercise 8).
22. $W = Span(\{\langle 1, -1, 1, -1\rangle, \langle 1, 0, -1, 1\rangle, \langle 1, 1, 0, -1\rangle\}) \trianglelefteq \mathbb{R}^4$ under the inner product of Exercise 9.
23. $W = Span(\{x^2\}) \trianglelefteq \mathbb{P}^2$ under the inner product of Exercise 11.
24. $W = Span(\{x^2 + 1, x - 1\}) \trianglelefteq \mathbb{P}^2$ under the inner product of Exercise 12.
25. $W = Span(\{x^2, x\}) \trianglelefteq \mathbb{P}^2$ under the inner product of Exercise 13.

For Exercises (26) to (31): Use ***The Extension Theorem***, Exercise 27 of Section 3.2, to enlarge the indicated basis for W, one vector at a time, using vectors from the standard basis S ***in the given order***, until you have a basis S' for the ambient vector space. Then use S' as input to the Gram-Schmidt Algorithm to find an orthonormal basis for V, W and $W^\perp$. Note: The Exercise numbers mentioned for certain items are from this section, 7.4.

26. $W = Span\{\langle 5, -2, 0\rangle\} \trianglelefteq \mathbb{R}^3 = Span(S)$ where $S = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, under the dot product.
27. $W = Span\{\langle 5, -2, 0\rangle\} \trianglelefteq \mathbb{R}^3 = Span(S)$ where $S = \{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$, under the weighted inner product of Exercise 3.
28. $W = Span\{\langle 1, -1, 0, 1\rangle, \langle 1, 0, -3, 1\rangle\} \trianglelefteq \mathbb{R}^4 = Span(S)$ where $S = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$, under the dot product.
29. $W = Span\{\langle 1, -1, 0, 1\rangle, \langle 1, 0, -3, 1\rangle\} \trianglelefteq \mathbb{R}^4 = Span(S)$ where $S = \{\vec{e}_1, \vec{e}_2, \vec{e}_3, \vec{e}_4\}$, under the weighted inner product of Exercise 6.
30. $W = Span\{x^2 + 5x\} \trianglelefteq \mathbb{P}^2 = Span(S)$, where $S = \{x^2, x, 1\}$ under the inner product of Exercise 10.
31. $W = Span\{x^2 - 3x, x\} \trianglelefteq \mathbb{P}^2 = Span(S)$, where $S = \{x^2, x, 1\}$ under the inner product of Exercise 10.
32. Explain why your answers in the Section 7.3 Exercise 11 will ***not*** help you find a basis for the orthogonal complement of $W = Span\{x^2, 1\}$ under the same inner product on $\mathbb{P}^2$.

For Exercises (33) to (41). Use your computations in the Section 7.3 Exercises and Exercises (15) to (25) above to find the orthogonal decomposition $\vec{w}_1 + \vec{w}_2$ of the indicated vector with respect to the indicated subspace W. Review the final Example of this Section and decide which strategy works best for each Exercise.

33. $\vec{u} = \langle 2, -4, 1\rangle$, $W = Span(\{\langle 1, 1, -1\rangle, \langle 0, -1, 1\rangle\}) \trianglelefteq \mathbb{R}^3$ under the ordinary dot product.
34. $\vec{v} = \langle -3, 5, 8\rangle$, $W = Span(\{\langle 1, 0, 1\rangle\}) \trianglelefteq \mathbb{R}^3$ under the ordinary dot product.
35. $\vec{u} = \langle 2, -4, 1\rangle$, $W = Span(\{\langle 1, 1, -1\rangle\}) \trianglelefteq \mathbb{R}^3$ under the weighted dot product of Exercise 3.

36. $\vec{v} = \langle -3, 5, 8\rangle$, $W = Span(\{\langle 1, 0, 1\rangle, \langle 2, -1, 0\rangle\}) \trianglelefteq \mathbb{R}^3$ under the inner product of Exercise 8.
37. $\vec{u} = \langle 3, 6, -2, -4\rangle$, $W = Span(\{\langle 1, -1, 1, -1\rangle, \langle 1, 0, -1, 1\rangle\}) \trianglelefteq \mathbb{R}^4$ under the dot product.
38. $\vec{v} = \langle 5, -2, 7, -3\rangle$, $W = Span(\{\langle 1, -1, 0, 1\rangle\}) \trianglelefteq \mathbb{R}^4$ under the ordinary dot product.
39. $\vec{u} = \langle 3, 6, -2, -4\rangle$, $W = Span(\{\langle 1, -1, 1, -1\rangle, \langle 1, 0, -1, 1\rangle, \langle 1, 1, 0, -1\rangle\}) \trianglelefteq \mathbb{R}^4$ under the weighted dot product of Exercise 6.
40. $\vec{u} = -x^2 + 2x - 5$, $W = Span(\{x^2\}) \trianglelefteq \mathbb{P}^2$ under the inner product of Exercise 10.
41. $\vec{v} = 3x^2 - 6x - 4$, $W = Span(\{x^2, x\}) \trianglelefteq \mathbb{P}^2$ under the inner product:

$$\langle p(x)|q(x)\rangle = \int_0^1 p(x)q(x)dx.$$

42. Prove that for any inner product space V: $V^\perp = \left\{\vec{\mathbf{0}}_V\right\}$, and $\left\{\vec{\mathbf{0}}_V\right\}^\perp = V$.
43. Prove that if $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_k\}$ is a ***basis*** for a subspace W of a finite dimensional inner product space V, then $\vec{v} \in V$ is a member of $W^\perp$ ***if and only if:***

$$\langle \vec{v}|\vec{w}_i\rangle = 0 \text{ for all } i = 1, 2, ..k.$$

44. Prove that for any subspace W of an inner product space V (even if V is infinite dimensional), then we must have: $W \subseteq (W^\perp)^\perp$.

Hint: to avoid confusion, let us write U for $W^\perp$. Thus, we want to show that $W \subseteq U^\perp$. Write down the definitions of $W^\perp$ ***and*** $U^\perp$. ***Stare*** at these two definitions for a while, and convincingly ***explain*** why these two definitions tell you that every member of W is also a member of $U^\perp$.

45. The goal of this Exercise is to show that if two non-empty sets $S_1 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ and $S_2 = \{\vec{v}_{k+1}, \vec{v}_{k+2}, \ldots, \vec{v}_n\}$ are ***disjoint*** (that is, they do not have any element in common) and furthermore $S_1 \cup S_2$ is an ***orthogonal basis*** for an inner product space V, then the two subspaces:

$$W = Span(S_1) \text{ and } U = Span(S_2)$$

are orthogonal complements of ***each other***, that is:

$$U = W^\perp \quad \textbf{\textit{and}} \quad W = U^\perp.$$

   a. Begin by writing any arbitrary vector $\vec{v} \in V$ as a linear combination of the vectors in $S_1 \cup S_2$.
   b. Use the standard technique to show that if $\vec{v} \in W^\perp$, then the coefficients of $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k$ must all be 0. Conclude that $\vec{v} \in U$, and thus $W^\perp \subseteq U$.
   c. Next, show that if $\vec{v} \in U$, (thus, it is only a linear combination of the vectors in S_2) then $\vec{v}$ is orthogonal to every vector in S_1 and thus $\vec{v} \in W^\perp$. Thus $U \subseteq W^\perp$ and this completes the proof that $U = W^\perp$.
   d. Use the same ideas in (b) and (c) to show that $W = U^\perp$.

46. Follow the construction of the basis for $W^\perp$ in the final Theorem of this section and explain why the two bases obtained in this Theorem (one for W and one for $W^\perp$) have the required properties for S_1 and S_2 that are found in the previous Exercise. This completes all the details of the proof that $(W^\perp)^\perp = W$.
47. ***De Morgan's Laws for Subspaces:*** In Chapter Zero, we encountered De Morgan's Laws for Logic, which states that if p and q are logical statements, then:

$$not(p\,\textbf{and}\,q) \Leftrightarrow not(p)\,\textbf{or}\,not(q), \quad \text{and}$$
$$not(p\,\textbf{or}\,q) \Leftrightarrow not(p)\,\textbf{and}\,not(q).$$

The goal of this Exercise is to establish analogous laws in the language of the join, intersection, and orthogonal complement of two subspaces V and W of a ***finite-dimensional*** inner product space U. Recall that:

$$V \vee W = \left\{ \vec{u} \in U \,|\, \vec{u} = \vec{v} + \vec{w} \text{ for some } \vec{v} \in V \text{ and some } \vec{w} \in W \right\}, \text{ and}$$
$$V \cap W = \left\{ \vec{u} \in U \,|\, \vec{u} \in V \text{ and } \vec{u} \in W \right\}.$$

These were discussed in detail in Section 4.1. Recall also that in the same Section, we proved the Dimension Theorem:

$$\dim(V \vee W) = \dim(V) + \dim(W) - \dim(V \cap W),$$

and the formula for the intersection in terms of the orthogonal complements:

$$V \cap W = (V^{\perp} \vee W^{\perp})^{\perp}.$$

We can rewrite this equation, and introduce a second equation, which will be the desired analogs to De Morgan's Laws:

$$(V \cap W)^{\perp} = V^{\perp} \vee W^{\perp}, \text{ and}$$
$$(V \vee W)^{\perp} = V^{\perp} \cap W^{\perp}.$$

Our goal is to prove these two equations. To begin the proof, let $\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\}$ be an orthonormal basis for $V \cap W$, where without loss of generality, we assume that $k \geq 1$.

a. Explain how to use the Extension Theorem and the Gram-Schmidt Algorithm to extend this set to the following orthonormal basis for all of V:

$$\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m\}.$$

b. Explain how to use the Extension Theorem and the Gram-Schmidt Algorithm to extend this set even further to the following orthonormal basis for $V \vee W$:

$$\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}.$$

c. In the notation above, explain why the following is a basis for W:

$$\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}.$$

d. Explain how to use the Extension Theorem and the Gram-Schmidt Algorithm to extend this set even further to the following orthonormal basis for all of U:

$$\{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k, \vec{v}_1, \vec{v}_2, \ldots, \vec{v}_m, \vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n, \vec{u}_1, \vec{u}_2, \ldots, \vec{u}_p\}.$$

e. In the notation above, what would be a basis for $(V \cap W)^{\perp}$? Explain your answer.

f. In the notation above, what would be a basis for $V^{\perp}$? Explain your answer.

g. In the notation above, what would be a basis for $W^{\perp}$? Explain your answer.

h. Combine your last two answers to find a basis for $V^{\perp} \vee W^{\perp}$, and show that you have the same answer as your basis for $(V \cap W)^{\perp}$. Thus, these two subspaces are exactly the same.

i. In the notation above, what would be a basis for $(V \vee W)^{\perp}$? Explain your answer.

j. Use the Dimension Theorem and parts (f) and (g) above to show that $\dim(V^{\perp} \cap W^{\perp}) = p$. See part (d) to see the role of p in the Proof.

k. Find a basis for $V^{\perp} \cap W^{\perp}$, and show that you have the same answer as your basis for $(V \vee W)^{\perp}$. Thus, these two subspaces are exactly the same.

7.5 Orthonormal Bases and Projection Operators

Orthonormal bases for a finite-dimensional inner product space V, and in particular from $\mathbb{R}^n$, possess many useful properties that can help us perform some computations in an easy way . We begin by seeing that inner products essentially reduce to an ordinary dot product in the finite-dimensional case through the use of coordinates with respect to an orthonormal basis, thus also enabling us to compute norms, angles and distances using simpler formulas:

Theorem: Let $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ be an ***orthonormal basis*** for an inner product space V. Let $\vec{v}$ and $\vec{w}$ be arbitrary members of V. If:

$$\langle \vec{v} \rangle_B = \langle v_1, v_2, \ldots, v_n \rangle, \text{ and } \langle \vec{w} \rangle_B = \langle w_1, w_2, \ldots, w_n \rangle, \quad \text{then:}$$

1. $\langle \vec{v} | \vec{w} \rangle = \langle \vec{v} \rangle_B \circ \langle \vec{w} \rangle_B = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$.
2. $\|\vec{v}\| = \|\langle \vec{v} \rangle_B\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.
3. $d(\vec{v}, \vec{w}) = \|\langle \vec{v} \rangle_B - \langle \vec{w} \rangle_B\| = \sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2 + \cdots + (v_n - w_n)^2}$.
4. $\cos(\theta) = \dfrac{\langle \vec{v} \rangle_B \circ \langle \vec{w} \rangle_B}{\|\langle \vec{v} \rangle_B\| \|\langle \vec{w} \rangle_B\|}$, where θ is the angle between $\vec{v}$ and $\vec{w}$, assuming that $\vec{v}$ and $\vec{w}$ are ***non-zero*** vectors.

Proof: We only need to prove part (1), because parts (2), (3) and (4) follow directly from the definitions of the norm, distance and angle between vectors. For example, if we already know that:

$$\langle \vec{v} | \vec{w} \rangle = \langle \vec{v} \rangle_B \circ \langle \vec{w} \rangle_B = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n, \text{ then:}$$

$$\langle \vec{v} | \vec{v} \rangle = \langle \vec{v} \rangle_B \circ \langle \vec{v} \rangle_B = v_1 v_1 + v_2 v_2 + \cdots + v_n v_n, \text{ and thus:}$$

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = \|\langle \vec{v} \rangle_B\|.$$

Proof of Property 1: Suppose B is an orthonormal basis as written above, and we have the coordinates $\langle \vec{v} \rangle_B$ and $\langle \vec{w} \rangle_B$ also as written above. By definition of coordinates, we have:

$$\vec{v} = v_1 \vec{u}_1 + v_2 \vec{u}_2 + \cdots + v_n \vec{u}_n, \quad \text{and} \quad \vec{w} = w_1 \vec{u}_1 + w_2 \vec{u}_2 + \cdots + w_n \vec{u}_n.$$

Then, using the additivity, homogeneity and orthonormality properties:

$$\begin{aligned}
\langle \vec{v} | \vec{w} \rangle &= \langle v_1 \vec{u}_1 + v_2 \vec{u}_2 + \cdots + v_n \vec{u}_n | w_1 \vec{u}_1 + w_2 \vec{u}_2 + \cdots + w_n \vec{u}_n \rangle \\
&= \langle v_1 \vec{u}_1 | w_1 \vec{u}_1 + w_2 \vec{u}_2 + \cdots + w_n \vec{u}_n \rangle + \\
&\quad\ \langle v_2 \vec{u}_2 | w_1 \vec{u}_1 + w_2 \vec{u}_2 + \cdots + w_n \vec{u}_n \rangle + \cdots + \\
&\quad\ \langle v_n \vec{u}_n | w_1 \vec{u}_1 + w_2 \vec{u}_2 + \cdots + w_n \vec{u}_n \rangle \\
&= v_1 w_1 \langle \vec{u}_1 | \vec{u}_1 \rangle + v_1 w_2 \langle \vec{u}_1 | \vec{u}_2 \rangle + \cdots + v_1 w_n \langle \vec{u}_1 | \vec{u}_n \rangle + \\
&\quad\ v_2 w_1 \langle \vec{u}_2 | \vec{u}_1 \rangle + v_2 w_2 \langle \vec{u}_2 | \vec{u}_2 \rangle + \cdots + v_2 w_n \langle \vec{u}_2 | \vec{u}_n \rangle + \cdots + \\
&\quad\ v_n w_1 \langle \vec{u}_n | \vec{u}_1 \rangle + v_n w_2 \langle \vec{u}_n | \vec{u}_2 \rangle + \cdots + v_n w_n \langle \vec{u}_n | \vec{u}_n \rangle \\
&= v_1 w_1 + v_2 w_2 + \cdots + v_n w_n.
\end{aligned}$$

We emphasize that these are the only surviving terms because $\langle \vec{u}_i | \vec{u}_j \rangle = 0 = \langle \vec{u}_j | \vec{u}_i \rangle$ if $i \neq j$, and so only the terms involving $\langle \vec{u}_i | \vec{u}_i \rangle = 1$ survive.■

Example: Let $\mathbb{P}^2$ be an inner product space under:

$$\langle\, p(x)\,|\,q(x)\,\rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1),$$

as in our last three Sections. In the previous Section, we constructed the orthonormal basis $B = \{u_1(x), u_2(x), u_3(x)\}$, where:

$$u_1(x) = \frac{1}{6}x(x-1), \;\; u_2(x) = -\frac{1}{2}(x+2)(x-1), \;\; \text{and} \;\; u_3(x) = \frac{1}{3}(x+2)x.$$

Let us consider two arbitrary polynomials of $\mathbb{P}^2$, say:

$$p(x) = 7x^2 - 3x + 4 \;\; \text{and} \;\; q(x) = -5x^2 + 2x + 6.$$

Let us construct the value table for our polynomials:

	-2	0	1
$u_1(x) = \frac{1}{6}x(x-1)$	1	0	0
$u_2(x) = -\frac{1}{2}(x+2)(x-1)$	0	1	0
$u_3(x) = \frac{1}{3}(x+2)x$	0	0	1
$p(x) = 7x^2 - 3x + 4$	38	4	8
$q(x) = -5x^2 + 2x + 6$	-18	6	3

Recall that to find the coordinates of any vector $\vec{v}$ with respect to an orthonormal basis B, we can use the general formula:

$$c_i = \langle \vec{v}\,|\,\vec{u}_i \rangle, \quad \text{where} \quad \langle \vec{v} \rangle_B = \langle c_1, c_2, \ldots, c_n \rangle.$$

Because we have chosen an orthonormal basis that behaves very similarly to $\{\vec{e}_1, \vec{e}_2, \vec{e}_3\}$ as seen in the table above, we immediately get:

$$\langle p(x) \rangle_B = \langle 38, 4, 8 \rangle, \text{ and}$$
$$\langle q(x) \rangle_B = \langle -18, 6, 3 \rangle.$$

Notice that if we were to directly find the inner product:

$$\langle\, p(x)\,|\,q(x)\,\rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1), \;\; \text{we get:}$$
$$\langle\, p(x)\,|\,q(x)\,\rangle = 38(-18) + 4(6) + 8(3) = -636.$$

But this is exactly the same as:

$$\langle p(x) \rangle_B \circ \langle q(x) \rangle_B = 38(-18) + 4(6) + 8(3) = -636.$$

On the other hand, consider the other orthonormal basis:

$$S = \{v_1(x), v_2(x), v_3(x)\} = \left\{ \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{42}}(3x+1), \frac{1}{\sqrt{126}}(7x^2 + 8x - 9) \right\}$$

that we obtained using the Gram-Schmidt Algorithm in the previous Section. Let us construct the value table for these polynomials, as well as the same $p(x)$ and $q(x)$:

	-2	0	1
$v_1(x) = \dfrac{1}{\sqrt{3}}$	$\dfrac{1}{\sqrt{3}}$	$\dfrac{1}{\sqrt{3}}$	$\dfrac{1}{\sqrt{3}}$
$v_2(x) = \dfrac{1}{\sqrt{42}}(3x+1)$	$\dfrac{-5}{\sqrt{42}}$	$\dfrac{1}{\sqrt{42}}$	$\dfrac{4}{\sqrt{42}}$
$v_3(x) = \dfrac{1}{\sqrt{126}}7x^2+8x-9$	$\dfrac{3}{\sqrt{126}}$	$\dfrac{-9}{\sqrt{126}}$	$\dfrac{6}{\sqrt{126}}$
$p(x) = 7x^2-3x+4$	38	4	8
$q(x) = -5x^2+2x+6$	-18	6	3

This time, finding the coordinates of $p(x)$ and $q(x)$ with respect to S is not as obvious. Applying the same formula for c_i that we have above, though, we get:

$$\begin{aligned}
\langle p(x)\rangle_S &= \left\langle \frac{38+4+8}{\sqrt{3}}, \frac{-5(38)+4+4(8)}{\sqrt{42}}, \frac{3(38)-9(4)+6(8)}{\sqrt{126}} \right\rangle \\
&= \left\langle \frac{50}{\sqrt{3}}, \frac{-154}{\sqrt{42}}, \frac{126}{\sqrt{126}} \right\rangle, \quad \text{and} \\
\langle q(x)\rangle_S &= \left\langle \frac{-18+6+3}{\sqrt{3}}, \frac{-5(-18)+6+4(3)}{\sqrt{42}}, \frac{3(-18)-9(6)+6(3)}{\sqrt{126}} \right\rangle \\
&= \left\langle \frac{-9}{\sqrt{3}}, \frac{108}{\sqrt{42}}, \frac{-90}{\sqrt{126}} \right\rangle
\end{aligned}$$

Despite the difference in their appearance, though, we still get:

$$\begin{aligned}
\langle p(x)\rangle_S \circ \langle q(x)\rangle_S &= \left\langle \frac{50}{\sqrt{3}}, \frac{-154}{\sqrt{42}}, \frac{126}{\sqrt{126}} \right\rangle \circ \left\langle \frac{-9}{\sqrt{3}}, \frac{108}{\sqrt{42}}, \frac{-90}{\sqrt{126}} \right\rangle \\
&= \frac{50}{\sqrt{3}}\left(\frac{-9}{\sqrt{3}}\right) - \frac{154}{\sqrt{42}}\left(\frac{108}{\sqrt{42}}\right) + \frac{126}{\sqrt{126}}\left(\frac{-90}{\sqrt{126}}\right) \\
&= -150 - 396 - 90 = -636. \;\square
\end{aligned}$$

We know in general from Exercise 22 in Section 3.8 that if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is ***any*** basis for a vector space V, then the linear transformation obtained by the coordinatization process defined by:

$$T : V \to \mathbb{R}^n, \quad \text{where} \quad T(\vec{v}) = \langle \vec{v} \rangle_B,$$

is an ***isomorphism of vector spaces***. However, the previous Theorem tells us something stronger when B is an orthonormal basis, and so we rephrase the ***same theorem*** in more elegant language:

Definition/Theorem: Let $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ be an ***orthonormal*** basis for an inner product space V. Then, the process of finding ***coordinates with respect to B:***

$$T : V \to \mathbb{R}^n, \quad \text{where} \quad T(\vec{v}) = \langle \vec{v} \rangle_B,$$

is an ***isomorphism of inner product spaces*** (also known as an ***isometry***), that is, for all vectors $\vec{u}$, $\vec{v} \in V$:

$$\langle \vec{v}|\vec{w}\rangle = T(\vec{v}) \circ T(\vec{w}) = \langle \vec{v}\rangle_B \circ \langle \vec{w}\rangle_B.$$

Consequently, T also ***preserves norms, distances and angles***, that is:

1. $\|\vec{v}\| = \|T(\vec{v})\| = \|\langle \vec{v}\rangle_B\|$,
   where the norm on the left is computed in V, and the norm on the right is computed in $\mathbb{R}^n$.
2. $d(\vec{v}, \vec{w}) = d(T(\vec{v}), T(\vec{w})) = d(\langle \vec{v}\rangle_B, \langle \vec{w}\rangle_B)$,
   where the distance on the left is computed in V, and the distance on the right is in $\mathbb{R}^n$.
3. The angle between $\vec{v}$ and $\vec{w}$ under the inner product in V is the same as the angle between $\langle \vec{v}\rangle_B$ and $\langle \vec{w}\rangle_B$ under the dot product of $\mathbb{R}^n$.

Projection Operators and their Matrices

In Section 2.2, we found the standard matrix of the projection operators onto a line L through the origin in $\mathbb{R}^2$ or $\mathbb{R}^3$, or a plane Π through the origin in $\mathbb{R}^3$. Let us now formally generalize the projection operator onto ***any*** subspace W of a Euclidean space, and find its standard matrix:

Definition/Theorem: Let W be a subspace of $\mathbb{R}^n$, under the ordinary dot product. Then the function $proj_W : \mathbb{R}^n \to \mathbb{R}^n$, given by:

$$proj_W(\vec{v}) = \vec{w}_1,$$

where $\vec{v} = \vec{w}_1 + \vec{w}_2$ is the orthogonal decomposition of $\vec{v}$ with respect to W and $W^\perp$, is a linear operator on $\mathbb{R}^n$, which we call ***the projection operator of $\mathbb{R}^n$ onto W***.
The ***range*** of $proj_W$ is W, hence the use of ***onto*** in the terminology is justified.
Furthermore, if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is any ***orthonormal basis*** for W, and $U = [\vec{u}_1|\vec{u}_2|\ldots|\vec{u}_k]$ is the $n \times k$ matrix with the vectors of B arranged in ***columns***, then:

$$[proj_W] = U \cdot U^\top.$$

We refer to $[proj_W]$ as the ***projection matrix*** representing $proj_W$.

Proof: First, let us show that $proj_W$ is indeed linear. Suppose $\vec{v}_1$ and $\vec{v}_2$ are vectors of V, and we have the decompositions:

$$\vec{v}_1 = \vec{w}_1 + \vec{w}_2, \quad \text{and} \quad \vec{v}_2 = \vec{w}_1' + \vec{w}_2',$$

where $\vec{w}_1$ and $\vec{w}_1'$ are vectors of W, and $\vec{w}_2$ and $\vec{w}_2'$ are members of $W^\perp$. Then:

$$\vec{v}_1 + \vec{v}_2 = \vec{w}_1 + \vec{w}_2 + \vec{w}_1' + \vec{w}_2' = \left(\vec{w}_1 + \vec{w}_1'\right) + \left(\vec{w}_2 + \vec{w}_2'\right).$$

Since the subspaces W and $W^\perp$ are closed under addition, and we know that the orthogonal decomposition of any vector is ***unique***, we can see that:

$$proj_W(\vec{v}_1 + \vec{v}_2) = \vec{w}_1 + \vec{w}_1' = proj_W(\vec{v}_1) + proj_W(\vec{v}_2),$$

so $proj_W$ is ***additive***. Similarly, we can easily show that $proj_W$ is ***homogeneous***.
If $\vec{w}$ is already a vector in W, then $proj_W(\vec{w}) = \vec{w}$, and thus the range of $proj_W$ must be all of W.
Now, let us find the matrix for $proj_W$. Suppose that $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is ***any*** orthonormal basis for W. Let $\vec{v} \in \mathbb{R}^n$. According to the projection formula from the previous Section:

$$\begin{aligned}\vec{w}_1 &= proj_W(\vec{v}) = \langle \vec{v}|\vec{u}_1\rangle\vec{u}_1 + \langle \vec{v}|\vec{u}_2\rangle\vec{u}_2 + \cdots + \langle \vec{v}|\vec{u}_k\rangle\vec{u}_k \\ &= (\vec{v} \circ \vec{u}_1)\vec{u}_1 + (\vec{v} \circ \vec{u}_2)\vec{u}_2 + \cdots + (\vec{v} \circ \vec{u}_k)\vec{u}_k.\end{aligned}$$

Since this is a ***linear combination*** of the vectors of B, we can write it as a ***matrix product:***

$$\vec{w}_1 = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_k] \begin{bmatrix} \vec{u}_1 \circ \vec{v} \\ \vec{u}_2 \circ \vec{v} \\ \vdots \\ \vec{u}_k \circ \vec{v} \end{bmatrix},$$

where we exchanged the order of the dot products, which we can do because the dot product is ***commutative***. But now, the column matrix on the right can again be expressed as another matrix product, this time, viewing $\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k$ as the ***rows*** of a matrix:

$$\vec{w}_1 = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_k] \begin{bmatrix} \vec{u}_1 \circ \vec{v} \\ \vec{u}_2 \circ \vec{v} \\ \vdots \\ \vec{u}_k \circ \vec{v} \end{bmatrix} = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_k] \left(\begin{bmatrix} \vec{u}_1 \\ \vec{u}_2 \\ \vdots \\ \vec{u}_k \end{bmatrix} \bullet \vec{v} \right).$$

If we denote by $U = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_k]$, the second matrix is now $U^\top$, and so by the ***associative property*** of matrix multiplication, we get:

$$\vec{w}_1 = (U \bullet U^\top) \bullet \vec{v}.$$

Thus, $\vec{w}_1 = proj_W(\vec{v})$ can be obtained from $\vec{v}$ by multiplying $\vec{v}$ by $U \bullet U^\top$. By ***The Uniqueness of the Standard Matrix*** of any linear transformation, this tells us that $[proj_W] = U \bullet U^\top$. ■

Example: Let $W = Span(\{\langle 1, -1, 0, 2\rangle, \langle -1, 3, 1, -1\rangle\})$. Let us use the Gram-Schmidt Algorithm to find an orthonormal basis for W:

$$\vec{v}_1 = \langle 1, -1, 0, 2\rangle$$

$$\vec{v}_2 = \langle -1, 3, 1, -1\rangle - \frac{\langle -1, 3, 1, -1\rangle \circ \langle 1, -1, 0, 2\rangle}{\langle 1, -1, 0, 2\rangle \circ \langle 1, -1, 0, 2\rangle} \langle 1, -1, 0, 2\rangle$$

$$= \langle -1, 3, 1, -1\rangle - \frac{-6}{6} \langle 1, -1, 0, 2\rangle = \langle 0, 2, 1, 1\rangle.$$

A quick check verifies that $\vec{v}_1 \circ \vec{v}_2 = 0$. Normalizing, we get our basis:

$$\left\{ \frac{1}{\sqrt{6}} \langle 1, -1, 0, 2\rangle, \frac{1}{\sqrt{6}} \langle 0, 2, 1, 1\rangle \right\}.$$

Thus:

$$[proj_W] = \begin{bmatrix} \frac{1}{\sqrt{6}} & 0 \\ -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \\ 0 & \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & 0 & \frac{2}{\sqrt{6}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 1 & -1 & 0 & 2 \\ -1 & 5 & 2 & 0 \\ 0 & 2 & 1 & 1 \\ 2 & 0 & 1 & 5 \end{bmatrix}.$$

To test that this matrix is correct, let us use $[proj_W]$ to orthogonally decompose $\vec{v} = \langle 7, -5, 3, 4\rangle$:

$$proj_W(\vec{v}) = \frac{1}{6}\begin{bmatrix} 1 & -1 & 0 & 2 \\ -1 & 5 & 2 & 0 \\ 0 & 2 & 1 & 1 \\ 2 & 0 & 1 & 5 \end{bmatrix}\begin{bmatrix} 7 \\ -5 \\ 3 \\ 4 \end{bmatrix} = \frac{1}{6}\begin{bmatrix} 20 \\ -26 \\ -3 \\ 37 \end{bmatrix} = \vec{w}_1, \text{ and}$$

$$\begin{aligned} \vec{w}_2 &= \vec{v} - \vec{w}_1 = \langle 7, -5, 3, 4\rangle - \frac{1}{6}\langle 20, -26, -3, 37\rangle \\ &= \frac{1}{6}\langle 22, -4, 21, -13\rangle. \end{aligned}$$

We verify that $\vec{w}_1 \circ \vec{w}_2 = (20 \bullet 22 + 26 \bullet 4 - 3 \bullet 21 - 37 \bullet 13)/36 = 0$, so at least for this vector, we obtain the correct decomposition.□

Notice that in this Example, $[proj_W]$ is a ***symmetric*** matrix. This is not an accident, and you will prove that this is true in general in the Exercises.

Example — Revisiting Projections and Reflections: Let us find yet another method to obtain the matrix of a projection operator onto a plane in $\mathbb{R}^3$. Consider Π, the plane with equation $3x - 5y + 2z = 0$, as in Section 2.2. We need an orthonormal basis for the vectors on this plane. Let us begin with an obvious basis for Π:

$$B = \{\langle 5, 3, 0\rangle, \langle 0, 2, 5\rangle\},$$

as we have done before. Now, we apply the Gram-Schmidt Algorithm to B:

$$\vec{v}_1 = \langle 5, 3, 0\rangle, \text{ and}$$

$$\vec{v}_2 = \langle 0, 2, 5\rangle - \frac{\langle 0, 2, 5\rangle \circ \langle 5, 3, 0\rangle}{\langle 5, 3, 0\rangle \circ \langle 5, 3, 0\rangle}\langle 5,3,0\rangle = \langle 0, 2, 5\rangle - \frac{6}{34}\langle 5, 3, 0\rangle = \frac{5}{17}\langle -3, 5, 17\rangle,$$

or we can use $\vec{v}_2 = \langle -3, 5, 17\rangle$. Thus we get the orthonormal basis:

$$\left\{\frac{1}{\sqrt{34}}\langle 5,3,0\rangle, \frac{1}{\sqrt{323}}\langle -3,5,17\rangle\right\}.$$

From this, we obtain the matrix:

$$\begin{aligned} [proj_\Pi] &= \begin{bmatrix} \frac{5}{\sqrt{34}} & -\frac{3}{\sqrt{323}} \\ \frac{3}{\sqrt{34}} & \frac{5}{\sqrt{323}} \\ 0 & \frac{17}{\sqrt{323}} \end{bmatrix}\begin{bmatrix} \frac{5}{\sqrt{34}} & \frac{3}{\sqrt{34}} & 0 \\ -\frac{3}{\sqrt{323}} & \frac{5}{\sqrt{323}} & \frac{17}{\sqrt{323}} \end{bmatrix} \\ &= \frac{1}{38}\begin{bmatrix} 29 & 15 & -6 \\ 15 & 13 & 10 \\ -6 & 10 & 34 \end{bmatrix}. \end{aligned}$$

This is exactly the same matrix we obtained in Section 2.2.□

7.5 Section Summary

Let $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ be an orthonormal basis for an inner product space V. Let $\vec{v}$ and $\vec{w}$ be arbitrary members of V. If:

$$\langle \vec{v} \rangle_B = \langle v_1, v_2, \ldots, v_n \rangle, \text{ and } \langle \vec{w} \rangle_B = \langle w_1, w_2, \ldots, w_n \rangle, \text{ then:}$$

1. $\langle \vec{v} | \vec{w} \rangle = \langle \vec{v} \rangle_B \circ \langle \vec{w} \rangle_B = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$.
2. $\|\vec{v}\| = \|\langle \vec{v} \rangle_B\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.
3. $d(\vec{v}, \vec{w}) = \|\langle \vec{v} \rangle_B - \langle \vec{w} \rangle_B\| = \sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2 + \cdots + (v_n - w_n)^2}$.
4. $\cos(\theta) = \dfrac{\langle \vec{v} \rangle_B \circ \langle \vec{w} \rangle_B}{\|\langle \vec{v} \rangle_B\| \|\langle \vec{w} \rangle_B\|}$, where θ is the angle between $\vec{v}$ and $\vec{w}$, ***non-zero*** vectors.

Let W be a subspace of $\mathbb{R}^n$, under the ordinary dot product. Then the function $proj_W : \mathbb{R}^n \to \mathbb{R}^n$, given by: $proj_W(\vec{v}) = \vec{w}_1$, where $\vec{v} = \vec{w}_1 + \vec{w}_2$ is the orthogonal decomposition of $\vec{v}$, is a linear operator of $\mathbb{R}^n$, which we call ***the projection operator of*** $\mathbb{R}^n$ ***onto*** W.

Furthermore, if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an orthonormal basis for W, and $U = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_k]$ is the $n \times k$ matrix with the vectors of B arranged in ***columns***, then: $[proj_W] = U \cdot U^\top$.

7.5 Exercises

For Exercises (1) to (5): Use the coordinate vectors $\langle \vec{u} \rangle_S$ and $\langle \vec{v} \rangle_S$ for the two indicated vectors which you found in Exercise (18) to (28) in Section 7.3, with respect to the indicated orthonormal basis S, to find: (a) $\langle \vec{u} | \vec{v} \rangle$, (b) $\|\vec{u}\|$, (c) $\|\vec{v}\|$, (d) $d(\vec{u}, \vec{v})$, and (e) the cosine of the angle θ between $\vec{u}$ and $\vec{v}$; (f) check (a) by directly computing $\langle \vec{u} | \vec{v} \rangle$ using the definition of the indicated inner product.

1. $\vec{u} = \langle 2, -4, 1 \rangle$ and $\vec{v} = \langle -3, 5, 8 \rangle$ from Exercise 18 of Section 7.3; S is the output of the Gram-Schmidt Algorithm using the basis $B = \{\langle 1, 1, -1 \rangle, \langle 0, -1, 1 \rangle, \langle 2, 0, 1 \rangle\}$ for $\mathbb{R}^3$, under the weighted inner product: $\langle \vec{u} | \vec{v} \rangle = 4u_1 v_1 + 5u_2 v_2 + 3u_3 v_3$.
2. $\vec{u} = \langle 2, -4, 1 \rangle$ and $\vec{v} = \langle -3, 5, 8 \rangle$ from Exercise 21 of Section 7.3; S is the output of the Gram-Schmidt Algorithm using the basis $B = \{\langle 1, 0, 1 \rangle, \langle 2, -1, 0 \rangle, \langle 1, 1, -1 \rangle\}$ for $\mathbb{R}^3$ under the inner product generated by the isomorphism:

$$[T] = \begin{bmatrix} 1 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -1 & -1 \end{bmatrix}.$$

3. $\vec{u} = \langle 3, 6, -2, -4 \rangle$ and $\vec{v} = \langle 5, -2, 7, -3 \rangle$ from Exercise 24 of Section 7.3; S is the output of the Gram-Schmidt Algorithm using the basis:

$$B = \{\langle 1, -1, 1, -1 \rangle, \langle 1, 0, -1, 1 \rangle, \langle 1, 1, 0, -1 \rangle, \langle -1, 1, 1, -1 \rangle\}$$

for $\mathbb{R}^4$ under the weighted inner product: $\langle \vec{u} | \vec{v} \rangle = 4u_1 v_1 + u_2 v_2 + 3u_3 v_3 + 6u_4 v_4$.

4. $\vec{u} = -x^2 + 2x - 5$ and $\vec{v} = 3x^2 - 6x - 4$ from Exercise 26 of Section 7.3; S is the output of the Gram-Schmidt Algorithm using the basis $B = \{x^2, x, 1\}$ for $\mathbb{P}^2$ under the inner product: $\langle p(x) | q(x) \rangle = p(-2)q(-2) + p(0)q(0) + p(1)q(1)$.

5. $\vec{u} = -x^2 + 2x - 5$ and $\vec{v} = 3x^2 - 6x - 4$ from Exercise 28 of Section 7.3; S is the output of the Gram-Schmidt Algorithm using the basis $B = \{x^2, x, 1\}$ for $\mathbb{P}^2$ under the inner product:

$$\langle\, p(x) \,|\, q(x) \,\rangle = \int_0^1 p(x)q(x)dx.$$

For Exercises (6) to (9): Find the matrix of the projection operator with respect to the subspace W. You may use your answers from the given Exercises. Check your answer by computing $\vec{w}_1 = proj_W(\vec{v})$ for the indicated $\vec{v}$, and verifying that $\vec{w}_2 = \vec{v} - \vec{w}_1$ is orthogonal to $\vec{w}_1$ under the ordinary dot product.

6. $W = Span(\{\langle 1, 1, -1\rangle, \langle 0, -1, 1\rangle\}) \trianglelefteq \mathbb{R}^3$ from Exercise 15, Section 7.4; $\vec{v} = \langle 2, -5, 3\rangle$.

7. $W = Span(\{\langle 1, 0, 1\rangle\}) \trianglelefteq \mathbb{R}^3$ from Exercise 16, Section 7.4; $\vec{v} = \langle 2, -5, 3\rangle$.

8. $W = Span(\{\langle 1, -1, 1, -1\rangle, \langle 1, 0, -1, 1\rangle\}) \trianglelefteq \mathbb{R}^4$ from Exercise 20, Section 7.4; $\vec{v} = \langle 2, -3, -7, 4\rangle$.

9. $W = Span(\{\langle 1, -1, 0, 1\rangle\}) \trianglelefteq \mathbb{R}^4$ from Exercise 21, Section 7.4; $\vec{v} = \langle 2, -3, -7, 4\rangle$.

10. Find the matrix of the projection operator of $\mathbb{R}^4$ with respect to:

$$W = Span(\{\langle 1, -1, 1, 0\rangle, \langle 1, 0, -1, 1\rangle, \langle 0, 1, 1, 1\rangle\})$$

Check your answer by computing $\vec{w}_1 = proj_W(\vec{v})$, where $\vec{v} = \langle 2, -3, -7, 4\rangle$, and verifying that $\vec{w}_2 = \vec{v} - \vec{w}_1$ is orthogonal to $\vec{w}_1$ under the ordinary dot product.

11. Use the technique shown in this Section to find the standard matrix of $proj_\Pi$ where Π is the plane in $\mathbb{R}^3$ with Cartesian equation: $5x - 3y - 7z = 0$.

12. Repeat the previous Exercise for the plane Π with Cartesian equation $7y - 4z = 0$. Be careful how you choose B.

13. Use the formula for $[proj_W]$ to prove that $[proj_W]$ is ***symmetric*** for every projection operator on $\mathbb{R}^n$. Hint: what is the formula for $(A \bullet B)^\top$?

14. ***Idempotent Matrices:*** An $n \times n$ matrix A is called ***idempotent*** if $A^2 = A$. The word "idempotent" comes from the Latin words "idem," which means "the same," and "potent," which means "power."

   a. Let A by an $n \times n$ idempotent matrix. Show that $A^3 = A$ also.

   b. (continuation) Use induction to show that $A^k = A$ for all positive integers k. Notice how the word "idempotent" perfectly describes such a matrix.

   c. Now, let W be any subspace of $\mathbb{R}^n$. Show that $[proj_W]$ is always an ***idempotent*** matrix. Hint: for any $\vec{v} \in \mathbb{R}^n$, let $\vec{v} = \vec{w}_1 + \vec{w}_2$ be the orthogonal decomposition of $\vec{v}$ with respect to W and $W^\perp$. Use the linearity properties. What is $proj_W(\vec{w}_1)$? What is $proj_W(\vec{w}_2)$? Show that $proj_W\left(proj_W(\vec{v})\right) = proj_W(\vec{v})$, and thus $[proj_W]^2 = [proj_W]$.

   In the previous Exercise, we also showed that $[proj_W]$ is ***symmetric*** for every subspace W. Now, we have the converse:

   d. Suppose that A is a ***symmetric*** and ***idempotent*** $n \times n$ matrix. Prove that there exists a subspace W of $\mathbb{R}^n$ such that $A = [proj_W]$. Hints: Review Section 1.8, The Four Fundamental Matrix Spaces of A. In particular, review which pairs of these spaces are orthogonal complements of each other. Which one of these spaces should W be? Prove that with this choice of W, $[proj_W] = A$.

15. ***Norm-Preserving Transformation:*** Suppose that $T: V \to W$ is a linear transformation from one inner product space to another, such that $\|\vec{v}\|_V = \|T(\vec{v})\|_W$. In other words, the length of a vector in V (under the inner product of V) is the same as the length of its image in W (under the inner product of W). We also say that T is a ***norm-preserving*** transformation. Prove that T is ***one-to-one***. Hint: look at the kernel of T.

16. Use the previous Exercise to explain why rotations in $\mathbb{R}^2$, reflections across lines through the origin in $\mathbb{R}^2$, and reflections across lines or planes through the origin in $\mathbb{R}^3$ are ***all*** isomorphisms.

17. Prove that every projection operator $proj_W$ on $\mathbb{R}^n$ is ***diagonalizable***, and describe the diagonal matrix for $proj_W$ and the basis for $\mathbb{R}^n$ under which the matrix is diagonal. Assume that W has dimension k. Hint: think about what happens to the vectors of W and $W^\perp$. You must also prove that the union of a basis for W and a basis for $W^\perp$ yields a basis for $\mathbb{R}^n$.

18. ***Non-orthogonal Decompositions:*** Suppose that U is an n-dimensional vector space (not necessarily an inner product space), and V and W are both subspaces of U (not necessarily orthogonal to each other, even if U is an inner product space), satisfying the two conditions:

$$V \cap W = \left\{\vec{\mathbf{0}}_U\right\} \quad \text{and} \quad dim(V) + dim(W) = dim(U).$$

 a. Let $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be a basis for V, and let $B' = \{\vec{w}_{k+1}, \ldots, \vec{w}_n\}$ be a basis for W (note that the dimension equation above tells us that these subscripts are correct). Prove that $B \cup B'$ is a ***linearly independent*** set, and therefore a ***basis*** for U. Hint: use an idea from the ***uniqueness*** portion of the proof of the orthogonal decomposition theorem.

 b. Use (a) to show that any vector $\vec{u} \in U$ can be decomposed ***uniquely*** as: $\vec{u} = \vec{v} + \vec{w}$, where $\vec{v} \in V$ and $\vec{w} \in W$. Hint: use the ordinary coordinatization process from Section 3.5.

 c. Show that the plane $V = Span(\{\langle 1, -1, 2\rangle, \langle 0, 3, 1\rangle\})$ and the line $W = Span\{\langle 1, 2, -1\rangle\}$, both subspaces of $\mathbb{R}^3$, satisfy the two conditions for V and W above. Hint: to make this easy, find a Cartesian equation for V and show that W is not on V.

 d. Find the decomposition of $\vec{u} = \langle 3, -8, 5\rangle$ with respect to the subspaces V and W in (c).

 e. Make a sketch of the plane V, the line W and the three vectors in your decomposition. Do they form a right triangle?

19. ***Flashback to Section 7.1:*** Prove that the set of $m \times n$ matrices is an ***inner product space*** under the bilinear form $\langle A|B\rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{i,j} b_{i,j}$, that is, take all the corresponding pairs of products of the entries of the two matrices and add them together, just like the ordinary dot product.

20. ***Matrix Decompositions:*** Recall that a square matrix A is ***symmetric*** if $A = A^\top$. We showed in Exercise 16 and 17 in Section 3.4 that the set of all $n \times n$ symmetric matrices, which we shall denote $Symm(n)$, is a subspace of $Mat(n,n)$. Furthermore we found that:

$$dim(Symm(n)) = \frac{n(n+1)}{2} \quad \text{and} \quad dim(Mat(n,n)) = n^2.$$

Similarly, a square matrix B is called ***skew-symmetric*** if $B = -B^\top$.

 a. Warm-up: Verify that $B = \begin{bmatrix} 0 & 5 & -3 \\ -5 & 0 & 2 \\ 3 & -2 & 0 \end{bmatrix}$ is a ***skew-symmetric*** matrix.

 b. Prove in general that the ***diagonal*** of a skew symmetric matrix must consist entirely of zeroes, as demonstrated in (a).

c. Prove that the subset $Skew(n)$ of all $n \times n$ skew-symmetric matrices is also a ***subspace*** of $Mat(n, n)$, that is, $Skew(n)$ is closed under matrix addition and scalar multiplication.

d. Prove in general that every symmetric matrix is ***orthogonal*** to every skew-symmetric matrix, under the inner product of the previous Exercise. Hint: consider the terms which are on the diagonal, and consider ***in pairs*** the terms which are off the diagonal.

e. Rephrase the last two parts as: $Skew(n, n) \trianglelefteq Symm(n, n)^{\perp}$.

Our next goal is to show that these two subspaces are actually ***equal***.

f. Show that:

$$\left\{ \begin{bmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{bmatrix} \right\}$$

is a basis for $Skew(3)$, and thus this space is 3-dimensional.

g. Prove in general that $dim(Skew(n)) = \dfrac{n(n-1)}{2}$ by constructing a ***general basis*** for $Skew(n)$ using (f).

h. Use the formula $dim(W) + dim\left(W^{\perp}\right) = dim(V)$ to show that $Skew(n) = Symm(n)^{\perp}$.

i. ***Prove using a theorem*** that any $n \times n$ matrix C can be written as $C = A + B$, where A is ***symmetric*** and B is ***skew-symmetric***.

j. ***Show explicitly*** that any $n \times n$ matrix C can be written as $C = A + B$, where A is symmetric and B is skew-symmetric. This means, ***find a formula*** for A and B, in terms of the arbitrary matrix C. Hint: ***Assume*** that you can find A and B, as above, take the transpose of both sides, then ***solve*** for A and B, but ***prove*** that your formula works.

k. Demonstrate your formula (that is, find A and B) on:

$$C = \begin{bmatrix} 7 & 5 & -3 \\ -3 & 4 & -9 \\ -8 & -1 & 2 \end{bmatrix}.$$

21. ***Decompositions of Continuous Functions:*** Consider the space of continuous functions $C([-a, a])$ for some positive number a (or $+\infty$). These functions are also integrable on $[-a, a]$. Recall that a function $f(x)$ is ***even*** if $f(-x) = f(x)$, and a function $g(x)$ is ***odd*** if $g(x) = -g(-x)$.

a. Prove that every even function in $C([-a, a])$ is orthogonal to every odd function in $C([-a, a])$, under the inner product: $\langle f(x), g(x) \rangle = \int_{-a}^{a} f(x) \cdot g(x)\, dx$.

Hint: what kind of function is $p(x) = f(x) \cdot g(x)$? Prove it.

b. Prove that any continuous function $h(x)$ in $C([-a, a])$ can be decomposed as:

$$h(x) = f(x) + g(x),$$

where $f(x)$ is a continuous ***even*** function, and $g(x)$ is a continuous ***odd*** function. Hint: Use the idea of part (j) in the previous Exercise.

c. Demonstrate part (b) on the polynomial $h(x) = 8x^5 - 7x^4 - 2x^3 + 5x^2 + 6x - 1$.

7.6 Orthogonal Matrices

The orthonormal sets of $\mathbb{R}^n$ (under the ordinary dot product) obtained through the Gram-Schmidt Algorithm allow us to construct matrices that are easily inverted:

Definition: An $n \times n$ matrix Q is called ***orthogonal*** if:

$$QQ^\top = Q^\top Q = \boldsymbol{I}_n.$$

Equivalently, this means that Q is ***invertible***, and:

$$Q^{-1} = Q^\top.$$

The dot product formula for the matrix product tells us that the entry in row i, column j of $Q^\top Q$ is the dot product of row i of $Q^\top$ with column j of Q. But row i of $Q^\top$ is column i of Q, so by matching this entry of $Q^\top Q$ with the corresponding entry of $\boldsymbol{I}_n$, we get:

$$\begin{aligned} \vec{c}_i \circ \vec{c}_j &= 0 \text{ if } i \neq j, \quad \text{and} \\ \vec{c}_i \circ \vec{c}_i &= 1 \text{ for } i = 1..n, \end{aligned}$$

where $\vec{c}_i$ and $\vec{c}_j$ are columns i and j of Q. But this is exactly the definition of an ***orthonormal*** set in $\mathbb{R}^n$ with respect to the ordinary dot product. Since $QQ^\top = \boldsymbol{I}_n$ as well, this proves that:

Theorem: The following conditions are equivalent for an $n \times n$ matrix Q:

1. Q is an ***orthogonal*** matrix.
2. The ***columns*** of Q form an ***orthonormal set*** in $\mathbb{R}^n$ with respect to the dot product.
3. The ***rows*** of Q form an ***orthonormal set*** in $\mathbb{R}^n$ with respect to the dot product.

It is an unfortunate matter of nomenclature that we call these matrices orthogonal rather than the more appropriate "orthonormal matrices," but this is the accepted terminology.

Example: In Section 7.3, we saw that the orthonormal sets in $\mathbb{R}^2$ are of the form:

$$S = \{\langle \cos(\theta), \sin(\theta)\rangle, \langle -\sin(\theta), \cos(\theta)\rangle\} \quad \textbf{\textit{or}} \quad S = \{\langle \cos(\theta), \sin(\theta)\rangle, \langle \sin(\theta), -\cos(\theta)\rangle\},$$

for some number $\theta \in [0, 2\pi)$. Thus, by assembling the two vectors in S into the columns of Q, we get two kinds of 2×2 orthogonal matrices:

$$Q = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \quad \textbf{\textit{or}} \quad Q = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ \sin(\theta) & -\cos(\theta) \end{bmatrix}.$$

These matrices might appear to be indistinguishable, but notice their ***determinants*** are:

$$\cos^2(\theta) + \sin^2(\theta) = 1 \quad \text{or} \quad -\cos^2(\theta) - \sin^2(\theta) = -1.$$

We say that the orthogonal matrices Q with $det(Q) = 1$ are ***proper***, and those with $det(Q) = -1$ are ***improper***. We will see very soon that these are the ***only*** possible determinants for an orthogonal matrix.

For example:

$$\begin{bmatrix} \sqrt{3}/2 & -1/2 \\ 1/2 & \sqrt{3}/2 \end{bmatrix} \text{ is proper, but } \begin{bmatrix} -3/5 & 4/5 \\ 4/5 & 3/5 \end{bmatrix} \text{ is improper.}$$

Notice also that the first kind are the matrices of the ***rotation operators*** that we saw in Section 2.2. In fact, the first example above is the matrix of the counterclockwise rotation by $\pi/3$. □

Example: It is more difficult to produce orthogonal matrices in $\mathbb{R}^3$, but this is where the Gram-Schmidt Algorithm is useful. In Section 7.3, we constructed the orthonormal set:

$$S = \left\{ \frac{1}{\sqrt{2}}\langle 1, 0, 1\rangle, \frac{1}{\sqrt{6}}\langle -1, -2, 1\rangle, \frac{1}{\sqrt{3}}\langle -1, 1, 1\rangle \right\}.$$

To form an orthogonal matrix, we assemble the three vectors in S into the ***columns*** of Q :

$$Q = \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{6} & -1/\sqrt{3} \\ 0 & -2/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}.$$

We can easily check, of course, that $QQ^\top = Q^\top Q = \boldsymbol{I}_3$. More generally, starting with any basis for $\mathbb{R}^3$, we can construct an orthonormal basis, and thus an orthogonal matrix. □

Next, let us revisit another old friend from Section 2.2:

Example: Let Q be the matrix of the ***reflection*** across the plane $3x - 5y + 2z = 0$ that we saw in Section 2.2:

$$Q = [\mathit{refl}_\Pi] = \frac{1}{19}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix}.$$

Observe that Q is ***symmetric***, so:

$$Q^\top Q = Q^2 = \frac{1}{19^2}\begin{bmatrix} 10 & 15 & -6 \\ 15 & -6 & 10 \\ -6 & 10 & 15 \end{bmatrix}^2 = \frac{1}{361}\begin{bmatrix} 361 & 0 & 0 \\ 0 & 361 & 0 \\ 0 & 0 & 361 \end{bmatrix} = \boldsymbol{I}_3.$$

But this should come as no surprise, because when we reflect a vector ***twice*** across the same plane, we are back to our original vector. Thus $[\mathit{refl}_\Pi]^2 = \boldsymbol{I}_3$, and Q is an ***orthogonal*** matrix. □

In the Exercises, you will see that this Example can be generalized to an arbitrary subspace W of $\mathbb{R}^n$, and not just a plane through the origin in $\mathbb{R}^3$.

Further Properties of Orthogonal Matrices

We observed in the first Example that the determinants of our orthogonal matrices were either 1 or -1. Our second Example is not as pleasant, but since it is only a 3×3 matrix, it would not take us long to see that its determinant is in fact -1. It turns out that no other determinant is possible, and this is just one property of several that orthogonal matrices possess:

> ***Theorem:*** Let Q and P be $n \times n$ ***orthogonal*** matrices. Then:
> 1. $det(Q) = 1$ or -1.
> 2. Q^{-1} is also orthogonal.
> 3. PQ and QP are also orthogonal.

Notice that these are basically ***multiplicative*** and not additive properties. In particular, it is definitely ***not*** true that orthogonal matrices form a subspace of the vector space of $n \times n$ matrices.

Proof of (1): We have:

$$\begin{aligned} QQ^\top &= I_n, && \text{thus} \\ det(QQ^\top) &= det(I_n), && \text{or:} \\ det(Q)det(Q^\top) &= 1. && \text{But since } det(Q) = det(Q^\top)\text{:} \\ (det(Q))^2 &= 1, && \text{and so:} \\ det(Q) &= \pm 1. \end{aligned}$$

In general, we call orthogonal matrices Q with $det(Q) = 1$ ***proper***, and those with $det(Q) = -1$ ***improper***. We will leave the other two parts as Exercises. They follow easily from the formulas for $(A^{-1})^\top$ and $(AB)^\top$ that we saw in Chapter 2.■

In the previous Section, we saw that the linear transformation:

$$T : V \to \mathbb{R}^n, \text{ given by:}$$
$$T(\vec{v}) = \langle \vec{v} \rangle_B,$$

that finds the ***coordinates*** of a vector in an inner product space with respect to an orthonormal basis $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is an ***isometry***, and thus it also ***preserves the length*** of vectors. Notice that if $V = \mathbb{R}^n$, then T becomes an ***operator*** of $\mathbb{R}^n$, and we can talk about $[T]$, that is:

$$[T] = [\, T(\vec{e}_1) \,|\, T(\vec{e}_2) \,|\, \ldots \,|\, T(\vec{e}_n) \,].$$

But these columns are the ***coordinates*** of each $\vec{e}_i$ with respect to the basis B, so we have to solve a system of equations:

$$[\, \vec{u}_1 \; \vec{u}_2 \; \ldots \; \vec{u}_n \,|\, \vec{e}_1 \; \vec{e}_2 \; \ldots \; \vec{e}_n \,],$$

and the resulting matrix on the ***right*** hand side after we finish the Gauss-Jordan reduction will be $[T]$. But this is exactly the same set-up to find the ***inverse*** of B. Since B is an orthonormal set, $[B]$ and $[B]^{-1}$ are both ***orthogonal*** matrices. Thus, $[T] = [B]^{-1}$ is an orthogonal matrix.

It should therefore come as no surprise that the properties of orthogonality, preservation of the dot product, and preservation of length are highly interconnected:

Theorem: Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator. Then, the following conditions on T are ***equivalent:***

1. $[T]$ is an ***orthogonal*** matrix.
2. T preserves the ***dot product:*** for all $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$: $\vec{u} \circ \vec{v} = T(\vec{u}) \circ T(\vec{v})$.
3. T preserves ***length:*** for all $\vec{v} \in \mathbb{R}^n$: $\|\vec{v}\| = \|T(\vec{v})\|$.

Proof: Before we begin the proof, let us establish some useful formulas. We saw that the entries of the matrix product $Q^\top Q$ can be interpreted as the dot product of two columns of Q. Let us use this idea to think of the dot product as a ***matrix product***. To avoid confusion, let us denote by $[\vec{v}]$ the $n \times 1$ (column) matrix whose entries are those of $\vec{v}$, in the natural order. Similarly, the column matrix $[T(\vec{v})]$ contains the entries of $T(\vec{v})$. Thus, we can compute a linear transformation using a matrix product:

$$[T(\vec{v})] = [T] \cdot [\vec{v}].$$

Now, let us write the dot product as:

$$\vec{u} \circ \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} = [\vec{v}]^\top \cdot [\vec{u}] = [\vec{u}]^\top \cdot [\vec{v}],$$

where the last equality is justified because we can exchange $\vec{u}$ and $\vec{v}$. Thus:

$$\begin{aligned} T(\vec{u}) \circ T(\vec{v}) &= [T(\vec{v})]^\top \cdot [T(\vec{u})] \\ &= ([T][\vec{v}])^\top \cdot ([T][\vec{u}]) \\ &= ([\vec{v}]^\top [T]^\top) \cdot ([T][\vec{u}]) \\ &= [\vec{v}]^\top \cdot ([T]^\top \cdot [T][\vec{u}]) && \text{by the Associative Property,} \\ &= (([T]^\top [T])[\vec{u}]) \circ \vec{v} && \text{again using } \vec{w} \circ \vec{v} = [\vec{v}]^\top \cdot [\vec{w}]. \end{aligned}$$

We remark that the resulting column matrix inside the parentheses in the final step should be regarded as a ***vector*** for this dot product to make sense.

Now, let us show that $(1) \Rightarrow (2)$. If $[T]$ is orthogonal, then $[T]^\top [T] = \boldsymbol{I_n}$, so the formula above becomes:

$$T(\vec{u}) \circ T(\vec{v}) = \boldsymbol{I_n} \vec{u} \circ \vec{v} = \vec{u} \circ \vec{v},$$

so T preserves the dot product.

To show that $(2) \Rightarrow (3)$, let us assume that T preserves the dot product. Then:

$$\|T(\vec{u})\|^2 = T(\vec{u}) \circ T(\vec{u}) = \vec{u} \circ \vec{u} = \|\vec{u}\|^2,$$

so T also preserves lengths. Now, let us work our way backwards:

Let us show that $(3) \Rightarrow (2)$. Suppose T preserves length. Let us show that T also preserves the dot product. But Exercise 35 in Section 7.2 says that for all $\vec{u}$, $\vec{v}$ in an inner product space V:

$$\|\vec{u} + \vec{v}\|^2 - \|\vec{u} - \vec{v}\|^2 = 4\langle \vec{u} | \vec{v} \rangle, \quad \text{so under the dot product:}$$

$$\vec{u} \circ \vec{v} = \frac{1}{4}\|\vec{u} + \vec{v}\|^2 - \frac{1}{4}\|\vec{u} - \vec{v}\|^2.$$

Thus, if a vector always has the same length as its image under T, then we have (applying the formula above to $T(\vec{u})$ and $T(\vec{v})$ instead):

$$\begin{aligned} T(\vec{u}) \circ T(\vec{v}) &= \frac{1}{4}\|T(\vec{u}) + T(\vec{v})\|^2 - \frac{1}{4}\|T(\vec{u}) - T(\vec{v})\|^2 \\ &= \frac{1}{4}\|T(\vec{u} + \vec{v})\|^2 - \frac{1}{4}\|T(\vec{u} - \vec{v})\|^2 \\ &= \frac{1}{4}\|\vec{u} + \vec{v}\|^2 - \frac{1}{4}\|\vec{u} - \vec{v}\|^2 = \vec{u} \circ \vec{v}. \end{aligned}$$

Thus, T preserves the dot product.

Finally, let us show that (2) $\Rightarrow$ (1). If T preserves the dot product, then, using our results above, we have:

$$\begin{aligned} \vec{u} \circ \vec{v} &= T(\vec{u}) \circ T(\vec{v}) = (([T]^{\top}[T])[\vec{u}]) \circ \vec{v}, && \text{in other words} \\ 0 &= (([T]^{\top}[T])[\vec{u}]) \circ \vec{v} - \vec{u} \circ \vec{v}, && \text{or upon factoring} \\ 0 &= ([T]^{\top}[T] - \boldsymbol{I}_n)[\vec{u}] \circ \vec{v}. \end{aligned}$$

However, this equation is true for ***all*** vectors $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$, and thus it is true in particular if we let:

$$\vec{v} = ([T]^{\top}[T] - \boldsymbol{I}_n)[\vec{u}].$$

Thus we get:

$$([T]^{\top}[T] - \boldsymbol{I}_n)[\vec{u}] \circ ([T]^{\top}[T] - \boldsymbol{I}_n)[\vec{u}] = 0.$$

But by the ***positive*** property of the dot product, we must have $([T]^{\top}[T] - \boldsymbol{I}_n)[\vec{u}] = \vec{\mathbf{0}}_n$.

This equation is true, again, for ***all*** vectors $\vec{u} \in \mathbb{R}^n$, and thus $[T]^{\top}[T] - \boldsymbol{I}_n$ must be the ***zero transformation***. In other words, $[T]^{\top}[T] = \boldsymbol{I}_n$, and thus $[T]$ is ***orthogonal***.■

We mention that it is also possible to define an orthogonal operator T on an infinite dimensional inner product space V by requiring that T preserves norms:

> ***Definition:*** An operator $T : V \to V$ on an ***infinite-dimensional*** inner product space V is ***orthogonal*** if for all $\vec{v} \in V$: $\|\vec{v}\| = \|T(\vec{v})\|$.

Change of Basis Matrices for Orthonormal Bases

We saw in Section 6.4 that if B and $B^{/}$ are any two bases for a vector space V, then we can find an invertible basis $C_{B,B^{/}}$ for which:

$$\langle \vec{v} \rangle_{B^{/}} = C_{B,B^{/}} \bullet \langle \vec{v} \rangle_B.$$

for any vector $\vec{v} \in V$. But if ***both*** bases are orthonormal, $C_{B,B^{/}}$ turns out to be quite special:

> ***Theorem:*** Let B and $B^{/}$ both be ***orthonormal*** bases for an inner product space V. Then the change of basis matrix $C_{B,B^{/}}$ is an ***orthogonal*** matrix.

Proof: We saw in Section 6.4 that:

$$C_{B,B^{/}} = [B^{/}]^{-1}[B],$$

where $[B]$ is the matrix whose columns are the vectors of B, and analogously for $B^{/}$. However, we know that both of these matrices are ***orthogonal*** matrices, and since the product of two orthogonal matrices is again orthogonal, so is $C_{B,B^{/}}$. ■

Example: Let $B = \{\langle 4/5, -3/5\rangle, \langle 3/5, 4/5\rangle\}$ and $B^{/} = \{\langle 5/13, 12/13\rangle, \langle 12/13, -5/13\rangle\}$. Notice that both are orthonormal sets. We assemble the orthogonal matrices:

$$[B] = \begin{bmatrix} 4/5 & 3/5 \\ -3/5 & 4/5 \end{bmatrix}, \quad \text{and} \quad [B^{/}] = \begin{bmatrix} 5/13 & 12/13 \\ 12/13 & -5/13 \end{bmatrix}.$$

Notice also that $[B^{/}]$ is ***symmetric***, and so it is its own ***inverse***. Thus:

$$C_{B,B^{/}} = [B^{/}]^{-1}[B] = [B^{/}][B]$$

$$= \begin{bmatrix} 5/13 & 12/13 \\ 12/13 & -5/13 \end{bmatrix}\begin{bmatrix} 4/5 & 3/5 \\ -3/5 & 4/5 \end{bmatrix} = \begin{bmatrix} -16/65 & 63/65 \\ 63/65 & 16/65 \end{bmatrix},$$

and this is indeed an (improper) orthogonal matrix which also happens to be symmetric.□

7.6 Key Concepts

An $n \times n$ matrix Q is called ***orthogonal*** if $QQ^{\top} = Q^{\top}Q = \boldsymbol{I}_n$.
Equivalently, this means that Q is ***invertible***, and $Q^{-1} = Q^{\top}$.

The following are equivalent for an $n \times n$ matrix Q:

1. Q is an ***orthogonal*** matrix.
2. The ***columns*** of Q form an ***orthonormal set*** in $\mathbb{R}^n$ with respect to the dot product.
3. The ***rows*** of Q form an ***orthonormal set*** in $\mathbb{R}^n$ with respect to the dot product.

Let Q and P be ***orthogonal*** matrices. Then:

1. $det(Q) = 1$ or -1.
2. Q^{-1} is also orthogonal.
3. PQ and QP are also orthogonal.

We call Q ***proper*** if $|Q| = 1$ and ***improper*** if $|Q| = -1$.

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator. Then, the following are equivalent:

1. $[T]$ is an ***orthogonal*** matrix.
2. T preserves the ***dot product:*** for all $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$: $\vec{u} \circ \vec{v} = T(\vec{u}) \circ T(\vec{v})$.
3. T preserves ***length:*** for all $\vec{v} \in \mathbb{R}^n$: $\|\vec{v}\| = \|T(\vec{v})\|$.

Orthogonal matrices appear in many situations:

- as matrices whose ***rows*** and ***columns*** form ***orthonormal*** sets in $\mathbb{R}^n$.
- as a ***rotation*** matrix in $\mathbb{R}^2$.
- as the ***change of basis matrix*** from one orthonormal basis to another.
- as the matrix of a ***reflection*** across a line through the origin in $\mathbb{R}^2$, a line or a plane through the origin in $\mathbb{R}^3$ (and more generally, across a subspace W of $\mathbb{R}^n$, as will be seen in the Exercises).
- as a ***diagonalizing matrix*** for a ***symmetric*** matrix (as will be seen in the next Section).

7.6 Exercises

1. Create two 2×2 orthogonal matrices whose first columns are parallel to the vector $\langle -8, 15 \rangle$. Classify them as proper or improper.
2. Create two 2×2 orthogonal matrices whose first rows are parallel to the vector $\langle 20, -21 \rangle$. Classify them as proper or improper.
3. Use the results of Exercise 1, Section 7.3, to create a 3×3 orthogonal matrix. Is it proper or improper?
4. Use the results of Exercise 7, Section 7.3, to create a 4×4 orthogonal matrix. Is it proper or improper?
5. Let $B = \{\langle -20/29, 21/29 \rangle, \langle 21/29, 20/29 \rangle\}$ and $B' = \{\langle 15/17, 8/17 \rangle, \langle -8/17, 15/17 \rangle\}$.
   a. Verify that B and B' are orthonormal sets.
   b. Form the orthogonal matrices Q and Q' whose columns are the vectors, respectively, of B and B'.
   c. Classify Q and Q' as proper or improper.
   d. Compute QQ' and verify that it is also orthogonal.
   e. Is QQ' proper or improper?
   f. Find the change of basis matrix $C_{B,B'}$.
   g. Is $C_{B,B'}$ proper or improper?
6. Prove that if $Q = [\vec{c}_1 | \vec{c}_2 | \ldots | \vec{c}_n]$ is an orthogonal matrix, so is $[\pm\vec{c}_1 | \pm\vec{c}_2 | \ldots | \pm\vec{c}_n]$ for every possible choice of the sign of each column. How many possible combinations are there, including the original matrix?
7. Prove that if Q is an $n \times n$ orthogonal matrix, then the matrices obtained from Q by rearranging its columns in any order, is again an orthogonal matrix. How many possible rearrangements are there?
8. Let $B = \{\vec{u}, \vec{v}\}$ be any orthonormal basis for $\mathbb{R}^2$ such that the matrix $Q = [\vec{u} | \vec{v}]$ is a ***proper*** orthogonal matrix. Suppose rot_θ is the rotation in $\mathbb{R}^2$ by the counterclockwise angle θ. Show that $[rot_\theta]_B$ is exactly the same matrix as $[rot_\theta]$. Hint: express $\vec{u}$ and $\vec{v}$ in terms of the same angle α, and use the description of proper orthogonal matrices from the first Example.
9. Show that every Type II elementary matrix (i.e. obtained by exchanging two rows of $\boldsymbol{I}_n$) is an orthogonal matrix. Are they proper or improper?
10. Show that every ***improper*** 2×2 orthogonal matrix Q can be factored in the form:

$$Q = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} Q',$$

   where Q' is a ***proper*** 2×2 orthogonal matrix.

For Exercises (11) to (13): Use the multiplicative property of the determinant function to prove the following statements:

11. The product of two proper $n \times n$ orthogonal matrices is again proper.
12. The product of two improper $n \times n$ orthogonal matrices is proper.
13. The product of a proper and an improper $n \times n$ orthogonal matrix is improper.

14. ***Reflections as Orthogonal Matrices:*** We saw in Section 2.2 that we can compute the reflection across the plane Π using what we refer to now as the orthogonal decomposition of any vector $\vec{v} \in \mathbb{R}^3$: $\vec{v} = \vec{w}_1 + \vec{w}_2$, where $\vec{w}_1 \in \Pi$ and $\vec{w}_2$ is orthogonal to Π. We defined:

$$refl_{\Pi}(\vec{v}) = \vec{w}_1 - \vec{w}_2 = 2\vec{w}_1 - \vec{v} = 2proj_W(\vec{v}) - \boldsymbol{I}_3\vec{v}.$$

More generally, if W is a subspace of a finite-dimensional inner product space V, we can likewise define the ***reflection operator across W*** via:

$$\begin{aligned} refl_W(\vec{v}) &= (2proj_W - \boldsymbol{I}_V)(\vec{v}) \\ &= 2proj_W(\vec{v}) - \vec{v}, \end{aligned}$$

where $\boldsymbol{I}_V$ is the identity operator on V. However, for the rest of this Exercise, let us assume that $V = \mathbb{R}^n$ under the ordinary dot product, and $W \trianglelefteq \mathbb{R}^n$:

a. Use this definition to show that $[refl_W]$ is ***symmetric***. Hint: Use Exercise 13 in Section 7.5.

b. Prove that for any projection operator: $proj_W \circ proj_W = proj_W$.
Hint: where does $\vec{w}_1$ live?

c. Use (b) to show that $[refl_W]^2 = \boldsymbol{I}_n$

d. Put (a) and (c) together to prove that $[refl_W]$ is ***orthogonal***. This allows us to create orthogonal matrices that do not contain ***radicals***.

e. Construct the matrix of $refl_W$ for the subspace W of $\mathbb{R}^4$ in Exercise 8 Section 7.5, using your answer for $[proj_W]$ from that same Exercise. Check mentally that the columns are orthonormal.

15. ***Right Handed vs. Left Handed Orthonormal Bases:*** Recall from Section 1.3 that if $\vec{u} = \langle u_1, u_2, u_3 \rangle$ and $\vec{v} = \langle v_1, v_2, v_3 \rangle$ are vectors from $\mathbb{R}^3$, we defined the ***cross product:***

$$\begin{aligned} &\vec{u} \times \vec{v} \\ &= (u_2v_3 - u_3v_2)\vec{i} - (u_1v_3 - u_3v_1)\vec{j} + (u_1v_2 - u_2v_1)\vec{k}. \\ &= \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{vmatrix} \end{aligned}$$

according to the formula from Exercise 25 of Section 5.1 (we will need the other results from that Exercise).

We will ***define*** an ordered orthonormal basis $\{\vec{u}, \vec{v}, \vec{w}\}$ for $\mathbb{R}^3$ to be ***right handed*** if:

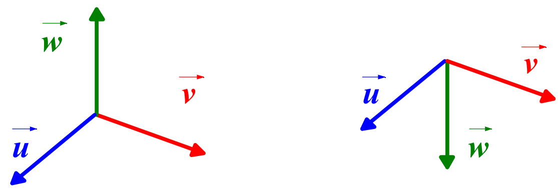
$$\vec{u} \times \vec{v} = \vec{w}.$$

a. Check that $\langle 3, 5, -2 \rangle$ is orthogonal to $\langle 4, -2, -1 \rangle$, and use the cross-product to create a third vector. Normalize the three vectors and produce a right handed orthonormal basis for $\mathbb{R}^3$.

b. Use the vectors from (a) to construct an orthogonal matrix.

c. Verify that $\left\{\vec{i}, \vec{j}, \vec{k}\right\}$ is right handed.

d. Prove that if $\{\vec{u}, \vec{v}, \vec{w}\}$ is any ***orthonormal*** basis for $\mathbb{R}^3$, then either $\vec{u} \times \vec{v} = \vec{w}$ or $-\vec{w}$. If the second possibility occurs, we call the basis ***left handed***. We see the two possibilities below:



A Right Handed Basis versus A Left Handed Basis

e. Is $\left\{\vec{k}, \vec{j}, \vec{i}\right\}$ left-handed or right-handed?

f. Prove that an orthonormal basis $\{\vec{u}, \vec{v}, \vec{w}\}$ is ***right handed*** ***if and only if*** the 3×3 orthogonal matrix:

$$Q = \left[\,\vec{u}\ \vec{v}\ \vec{w}\,\right]$$

is ***proper***. Notice that $\vec{i}$ and $\vec{j}$ are on the xy-plane and $\vec{k}$ is on the z-axis, which is the normal line to the xy-plane.

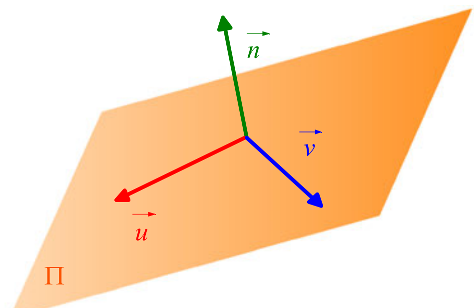
The goal for the rest of this Exercise is to create a right handed orthonormal basis using two vectors from an arbitrary plane:

$$\Pi : ax + by + cz = 0,$$

and its normal vector $\vec{n} = \langle a, b, c\rangle$, and we assume this to be a ***unit vector***. Let us also first assume, for simplicity, that ***none*** of the components of $\vec{n}$ are zero.

g. Verify that $S = \{\langle -b, a, 0\rangle, \langle -c, 0, a\rangle\}$ contains vectors on Π, and apply the Gram-Schmidt Algorithm on this set. Check your answers at the back of the book before proceeding to the next step.

h. Suppose that $\{\vec{u}, \vec{v}\}$ are the two unit vectors that you obtained from the previous step (in the same order). Prove that $\vec{u} \times \vec{v} = \vec{n}$, and thus $\{\vec{u}, \vec{v}, \vec{n}\}$ is a right handed coordinate system.



A Right Handed Orthonormal Basis Associated to Π

i. Apply steps (g) and (h) to the plane $3x - 2y + 6z = 0$ to obtain an orthonormal basis for $\mathbb{R}^3$ consisting of two vectors on this plane and a vector orthogonal to this plane.

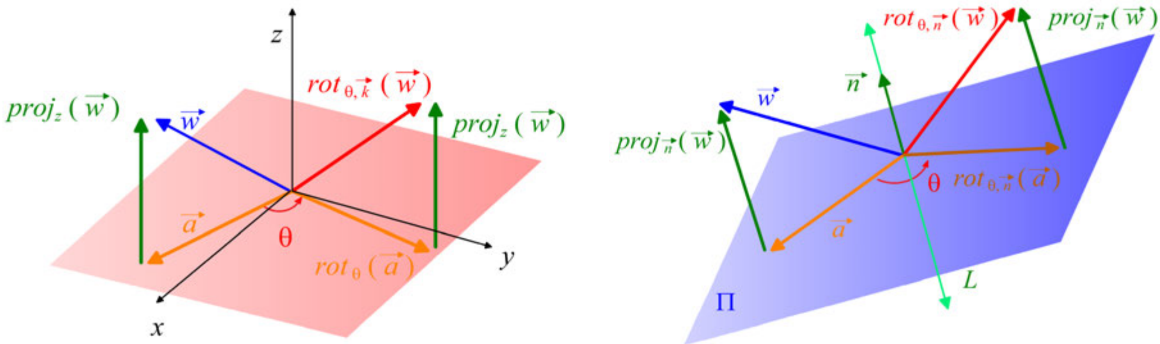
j. Modify the construction above in steps (g) and (h) if ***exactly one*** of the components of $\vec{n}$, say a, is zero.

k. Modify the construction above in steps (g) and (h) if ***exactly two*** of the components of $\vec{n}$, say b and c, are zero.

16. ***Rotations in Space:*** We know from Section 2.2 that the standard matrix of the linear transformation rot_θ given by the counterclockwise rotation in $\mathbb{R}^2$ by the positive angle θ is:

$$[rot_\theta] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}.$$

Notice that if $\vec{a} \in \mathbb{R}^2$, we can visualize this rotation taking place from the vantage point of the positive z-axis, by placing a spindle at the origin, pointing up (in other words, in the direction of $\vec{k}$), and spinning $\vec{a}$ counterclockwise by θ, as seen on the left below:



The Counterclockwise Rotation of $\vec{w}$ by θ About $\vec{k}$, and About an Arbitrary Vector $\vec{n}$

Suppose $\vec{w} = \langle x, y, z \rangle$ is now a vector in $\mathbb{R}^3$ in standard position. We will ***define*** the counterclockwise rotation of $\vec{w}$ by θ about $\vec{k}$, denoted $rot_{\theta,\vec{k}}$, to be:

$$rot_{\theta,\vec{k}}(\vec{w}) = \langle rot_\theta(\langle x, y \rangle), z \rangle.$$

Let us explain this notation: the z-coordinate of the rotated vector is the same as that of $\vec{w}$, and if $\vec{a} = \langle x, y \rangle$ is the projection of $\vec{w}$ onto the xy-plane, we rotate $\vec{a}$ by θ, and the resulting vector gives the x and y coordinates of the rotated vector.

a. Warm-up: ***Explain*** why the standard matrix of $rot_{\theta,\vec{k}}$ is:

$$\left[rot_{\theta,\vec{k}}\right] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & 0 \\ \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The goal for the rest of this Exercise is to generalize this rotation about an arbitrary ***unit vector*** $\vec{n} = \langle a, b, c \rangle$, denoted $rot_{\theta,\vec{n}}$, and to find $[rot_{\theta,\vec{n}}]$. As in the previous Exercise, we will assume for simplicity that none of the components of $\vec{n}$ is zero. Note also that $\vec{n}$ uniquely determines a plane Π passing through the origin with unit normal $\vec{n}$.

Our next step is to find a more convenient ***basis*** for $\mathbb{R}^3$:

b. Use your work from Exercise 15 (g), to show that:

$$B = \{\vec{u}, \vec{v}, \vec{n}\}$$

$$= \left\{ \frac{1}{\sqrt{a^2+b^2}} \langle -b, a, 0 \rangle, \frac{1}{\sqrt{a^2+b^2}} \langle -ac, -bc, a^2+b^2 \rangle, \langle a, b, c \rangle \right\}$$

is a ***right handed*** orthonormal basis for $\mathbb{R}^3$.

The natural way to geometrically describe $rot_{\theta,\vec{n}}$ is to imagine the unit normal $\vec{n}$ to be the spindle instead of $\vec{k}$, and we rotate $\vec{w}$ about $\vec{n}$ using its projection $\vec{a}$ onto Π, as seen in the diagram on the right above.

c. Let B be the orthonormal basis in (b). ***Explain*** why:

$$[rot_{\theta,\vec{n}}]_B = \left[rot_{\theta,\vec{k}}\right],$$

and this equation ***does not depend*** on the choice of the two orthonormal vectors $\vec{u}$ and $\vec{v}$ of Π. Note: there are no computations involved. Find an explanation by staring at the diagram.

To find the standard matrix $[rot_{\theta,\vec{n}}]$, we will need the change of basis formula from Section 6.4:

$$[rot_{\theta,\vec{n}}] = [B][rot_{\theta,\vec{n}}]_B[B]^{-1},$$

where B is our basis from part (b).

d. ***Explain*** why $[B]$ is an ***orthogonal*** matrix, and thus $[B]^{-1} = [B]^{\top}$.

e. Perform the multiplications in the formula $[rot_{\theta,\vec{n}}] = [B][rot_{\theta,\vec{n}}]_B[B]^{\top}$ and show that by simplifying:

$$[rot_{\theta,\vec{n}}] = \begin{bmatrix} a^2\tau + \cos\theta & ab\tau - c\sin\theta & ac\tau + b\sin\theta \\ ab\tau + c\sin\theta & b^2\tau + \cos\theta & bc\tau - a\sin\theta \\ ac\tau - b\sin\theta & bc\tau + a\sin\theta & c^2\tau + \cos\theta \end{bmatrix},$$

where $\tau = 1 - \cos\theta$. Notice the perfectly balanced occurrences of a, b and c in this matrix.

f. Explain why $[rot_{\theta,\vec{n}}]$ is an orthogonal matrix.

g. The formula $[rot_{\theta,\vec{n}}] = [B][rot_{\theta,\vec{n}}]_B[B]^{-1}$ involves a composition of three linear operators. Write a short paragraph ***explaining geometrically*** what each of these operators does, in the correct order of operations.

h. Find the standard matrix of the rotation in $\mathbb{R}^3$ about the unit normal vector $\vec{n} = \frac{1}{7}\langle 3, -2, 6\rangle$ counterclockwise by the angle $\theta = \sin^{-1}(3/5)$. Use your computations from Exercise 15 (i).

i. Use this matrix to compute $rot_{\theta,\vec{n}}(\vec{w})$, where $\vec{w} = \langle 5, 8, -9\rangle$ and $\vec{n} = \frac{1}{7}\langle 3, -2, 6\rangle$ is the unit vector from the previous part.

j. Draw a diagram showing $\vec{n}$, the plane W with normal $\vec{n}$, the vector $\vec{w}$ from part (i), and $rot_{\theta,\vec{n}}(\vec{w})$. Convince yourself that you have indeed rotated $\vec{w}$ by $\sin^{-1}(3/5)$.

k. Modify the construction above in steps (b) through (e) if ***exactly one*** of the components of $\vec{n}$, say a, is zero. Show that you still get exactly the same matrix $[rot_{\theta,\vec{n}}]$.

l. Modify the construction above in steps (b) through (e) if ***exactly two*** of the components of $\vec{n}$, say b and c, are zero. Show that you still get exactly the same matrix $[rot_{\theta,\vec{n}}]$.

17. ***Matrices in Block Diagonal Form:*** Suppose that $Q_1, Q_2, \ldots, Q_k$ are all square matrices, not necessarily of the same size, with $k \geq 2$. Show that the direct sum:

$$Q = Q_1 \oplus Q_2 \oplus \cdots \oplus Q_k,$$

as defined in the Exercises of Section 2.8, is an orthogonal matrix ***if and only if*** every Q_i is also an orthogonal matrix.

7.7 Orthogonal Diagonalization of Symmetric Matrices

There are two magical properties that symmetric matrices possess, and one of them is related to orthogonal matrices. The other property, though, involves only their eigenvalues and eigenvectors, so we begin with the following:

> ***Theorem — Orthogonality of Distinct Eigenspaces:***
> Let A be a ***symmetric*** matrix. Then all of the ***eigenvalues*** of A are ***real numbers***. Furthermore, if λ_1 and λ_2 are two ***distinct*** eigenvalues and $\vec{v}_1$ and $\vec{v}_2$ are corresponding eigenvectors, then:
> $$\vec{v}_1 \circ \vec{v}_2 = 0.$$

Proof: The proof that the eigenvalues of A are real will be shown in Chapter 8, and not-surprisingly will involve the study of matrices with ***complex*** entries. However, we can prove the second statement concerning eigenvectors.

Suppose λ_1 and λ_2 are two distinct eigenvalues with corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$. Instead of studying $\vec{v}_1 \circ \vec{v}_2$ directly, let us look instead at the expression $\vec{v}_1 \circ A\vec{v}_2$. We have:

$$\begin{aligned}
\vec{v}_1 \circ A\vec{v}_2 &= [A\vec{v}_2]^{\top} \cdot [\vec{v}_1\,] && \text{(using our matrix product formula for the dot product)}\\
&= [\vec{v}_2\,]^{\top} A^T \cdot [\vec{v}_1\,] && \text{(using } (CD)^T = D^T C^T)[\vec{v}\,]\\
&= [\vec{v}_2\,]^{\top} A \cdot [\vec{v}_1\,] && \text{(since } A \text{ is symmetric)}\\
&= [\vec{v}_2\,]^{\top} \cdot A[\vec{v}_1\,] && \text{(by the associative property of matrix multiplication)}\\
&= A\vec{v}_1 \circ \vec{v}_2. && \text{(again by the matrix product formula for the dot product)}
\end{aligned}$$

Thus we get:

$$\vec{v}_1 \circ A\vec{v}_2 = A\vec{v}_1 \circ \vec{v}_2.$$

Notice that, like an escape artist, the matrix A magically moved from right to left. This was only possible precisely because A is ***symmetric***. Now, since $\vec{v}_1$ and $\vec{v}_2$ are ***eigenvectors***, we can rewrite this equation as:

$$\begin{aligned}
\vec{v}_1 \circ \lambda_2\vec{v}_2 &= \lambda_1\vec{v}_1 \circ \vec{v}_2, && \text{or:}\\
\lambda_2(\vec{v}_1 \circ \vec{v}_2) &= \lambda_1(\vec{v}_1 \circ \vec{v}_2) && \text{by homogeneity. Thus:}\\
(\lambda_2 - \lambda_1)(\vec{v}_1 \circ \vec{v}_2) &= 0.
\end{aligned}$$

But since λ_1 and λ_2 are ***distinct*** eigenvalues, $\lambda_2 - \lambda_1 \neq 0$, so we must have $\vec{v}_1 \circ \vec{v}_2 = 0$. ■

Example: Consider the symmetric matrix:

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

The characteristic polynomial is:

$$\left| \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda \end{bmatrix} - \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \right| = \begin{vmatrix} \lambda & -1 & -1 \\ -1 & \lambda & -1 \\ -1 & -1 & \lambda \end{vmatrix}$$

$$= \lambda^3 - 3\lambda - 2 = (\lambda - 2)(\lambda + 1)^2,$$

and thus the eigenvalues are $\lambda = 2$ and -1, which are indeed real numbers.$\square$

Orthogonal Diagonalization

We are now ready for the second magical property. The previous Theorem says that two distinct ***eigenspaces*** for A must be ***orthogonal*** to each other (this does not mean that they are orthogonal complements of each other). However, if an eigenspace is more than 1-dimensional, we can use the ***Gram-Schmidt Algorithm*** to construct an ***orthonormal basis*** for it. It is not clear, however, that we will get n linearly independent vectors, where A is an $n \times n$ matrix. However, Chapter 8 will show us that this will indeed be possible, and thus we will state the following, whose complete proof will be in Chapter 8:

> ***Theorem — The Spectral Theorem for Symmetric Matrices:***
> Let A be a ***symmetric*** matrix. Then we can find an ***orthogonal*** matrix Q such that Q ***diagonalizes*** A, that is:
>
> $$D = Q^{-1}AQ = Q^{\top}AQ,$$
>
> where $D = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix containing the eigenvalues of A, and these eigenvalues are all ***real*** numbers.

The term "Spectral Theorem" comes from the word ***spectrum***, which means the set of ***eigenvalues*** of a matrix. This Theorem says that the spectrum of a symmetric matrix A consists of real numbers, and A can be diagonalized not just by an ordinary invertible matrix, but by an ***orthogonal*** matrix Q. To reiterate, the columns of Q are the ***unit eigenvectors*** produced by the Gram-Schmidt Algorithm on ***each*** eigenspace, if necessary.

Example: Let us diagonalize the symmetric matrix from our previous Example:

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

We found the eigenvalues $\lambda = 2$ and -1. Next, we must find the nullspaces of the matrices:

$$2I_3 - A = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}, \text{ and}$$

$$-1I_3 - A = \begin{bmatrix} -1 & -1 & -1 \\ -1 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus we find the eigenspaces:

$$Eig(A, 2) = Span(\{\langle 1, 1, 1\rangle\}), \quad \text{and}$$

$$Eig(A, -1) = Span(\{\langle -1, 1, 0\rangle, \langle -1, 0, 1\rangle\}).$$

Notice that the second eigenspace is 2-dimensional, but the two basis vectors are unfortunately not orthogonal. Thus, we must perform the Gram Schmidt Algorithm on the indicated basis:

$$\begin{aligned} \vec{v}_1 &= \langle -1, 1, 0\rangle, \quad \text{and} \\ \vec{v}_2 &= \langle -1, 0, 1\rangle - \frac{\langle -1, 0, 1\rangle \circ \langle -1, 1, 0\rangle}{\langle -1, 1, 0\rangle \circ \langle -1, 1, 0\rangle}\langle -1, 1, 0\rangle \\ &= \langle -1, 0, 1\rangle - \frac{1}{2}\langle -1, 1, 0\rangle = \left\langle -\frac{1}{2}, -\frac{1}{2}, 1\right\rangle. \end{aligned}$$

Clearing fractions, we use $\vec{v}_2 = \langle -1, -1, 2\rangle$. A quick check of three dot products verifies that:

$$S = \{\langle -1, 1, 0\rangle, \langle -1, -1, 2\rangle, \langle 1, 1, 1\rangle\}$$

is indeed an orthogonal set. Next, we ***normalize*** each vector, obtaining our ***orthonormal*** basis:

$$B = \left\{ \frac{1}{\sqrt{2}}\langle -1, 1, 0\rangle, \frac{1}{\sqrt{6}}\langle -1, -1, 2\rangle, \frac{1}{\sqrt{3}}\langle 1, 1, 1\rangle \right\},$$

corresponding to $\lambda = -1, -1, 2$. As usual, we want our eigenvalues to be in ***increasing*** order. Thus, our diagonalizing orthogonal matrix is:

$$Q = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}.$$

We can check that:

$$\begin{aligned} &Q^\top A Q \\ &= \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix} \\ &= \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{6} & -1/\sqrt{6} & 2/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & 2/\sqrt{3} \\ -1/\sqrt{2} & 1/\sqrt{6} & 2/\sqrt{3} \\ 0 & -2/\sqrt{6} & 2/\sqrt{3} \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{bmatrix}. \end{aligned}$$

As promised, the resulting matrix is diagonal, with diagonals the increasing eigenvalues. $\square$

7.7 Section Summary

Let A be a symmetric matrix. Then all of the eigenvalues of A are real numbers. Furthermore, if λ_1 and λ_2 are two ***distinct*** eigenvalues with corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$, then $\vec{v}_1$ is orthogonal to $\vec{v}_2$.

The Spectral Theorem for Symmetric Matrices: Let A be a ***symmetric*** matrix. Then we can find an ***orthogonal*** matrix Q such that Q ***diagonalizes*** A, that is:

$$D = Q^{\top}AQ,$$

where $D = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ is a diagonal matrix containing the ***eigenvalues*** of A, and these eigenvalues are all ***real*** numbers.

To find this orthogonal matrix Q, we find a basis for each eigenspace of A, apply the Gram-Schmidt algorithm to these bases, if there is more than one vector in a basis, and assemble the resulting ***unit vectors*** into the ***columns*** of Q.

7.7 Exercises

1. Let $Q = \begin{bmatrix} 1/\sqrt{3} & -1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 1/\sqrt{2} & -1/\sqrt{6} \\ 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{bmatrix}$,

   the diagonalizing orthogonal matrix that we found in the last Example in this Section. Determine if Q is proper or improper.

   For Exercises (2) to (22): Find an orthogonal matrix Q and a diagonal matrix D such that $D = Q^{\top}AQ$ for the following symmetric matrices A. For the sake of convention, place the eigenvalues in increasing order in D.

2. $\begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

4. $\begin{bmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{bmatrix}$

5. $\begin{bmatrix} 0 & 0 & 2 \\ 0 & 3 & 0 \\ 2 & 0 & 0 \end{bmatrix}$

6. $\begin{bmatrix} 0 & 5 & 5 \\ 5 & 0 & 5 \\ 5 & 5 & 0 \end{bmatrix}$

7. $\begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$

8. $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

9. $\begin{bmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}$

10. $\begin{bmatrix} 5 & -3 & -3 \\ -3 & 5 & -3 \\ -3 & -3 & 5 \end{bmatrix}$

11. $\begin{bmatrix} 5 & 0 & 1 \\ 0 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix}$

12. $\begin{bmatrix} -6 & 1 & 1 \\ 1 & -6 & 1 \\ 1 & 1 & -6 \end{bmatrix}$

13. $\begin{bmatrix} 0 & 0 & 1 \\ 0 & 5 & 0 \\ 1 & 0 & 0 \end{bmatrix}$

14. $\begin{bmatrix} -4 & 1 & 1 & 1 \\ 1 & -4 & 1 & 1 \\ 1 & 1 & -4 & 1 \\ 1 & 1 & 1 & -4 \end{bmatrix}$ 15. $\begin{bmatrix} -4 & 3 & 3 & 3 \\ 3 & -4 & 3 & 3 \\ 3 & 3 & -4 & 3 \\ 3 & 3 & 3 & -4 \end{bmatrix}$ 16. $\begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & -2 & 1 & 2 \\ 0 & 1 & 0 & 1 \\ 1 & 2 & 1 & -2 \end{bmatrix}$

17. $\begin{bmatrix} 5 & 0 & 0 & 1 \\ 0 & 5 & 1 & 0 \\ 0 & 1 & 5 & 0 \\ 1 & 0 & 0 & 5 \end{bmatrix}$ 18. $\begin{bmatrix} -5 & 0 & 0 & 2 \\ 0 & -5 & 2 & 0 \\ 0 & 2 & -5 & 0 \\ 2 & 0 & 0 & -5 \end{bmatrix}$ 19. $\begin{bmatrix} 0 & 0 & 0 & 5 \\ 0 & 0 & -2 & 0 \\ 0 & -2 & 0 & 0 \\ 5 & 0 & 0 & 0 \end{bmatrix}$

20. $\begin{bmatrix} 0 & 0 & 0 & 0 & 5 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 7 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{bmatrix}$ 21. $\begin{bmatrix} 7 & 0 & 0 & 0 & 1 \\ 0 & 7 & 0 & 1 & 0 \\ 0 & 0 & 7 & 0 & 0 \\ 0 & 1 & 0 & 7 & 0 \\ 1 & 0 & 0 & 0 & 7 \end{bmatrix}$ 22. $\begin{bmatrix} 7 & 0 & 0 & 0 & 1 \\ 0 & 7 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 7 & 0 \\ 1 & 0 & 0 & 0 & 7 \end{bmatrix}$

Did you notice a difference ***and*** similarity in the answers to 21 and 22?

23. Let $a, b \in \mathbb{R}$. Find the eigenvalues, with their multiplicities, of the matrices:

a. $\begin{bmatrix} a & b & b \\ b & a & b \\ b & b & a \end{bmatrix}$ b. $\begin{bmatrix} a & b & b & b \\ b & a & b & b \\ b & b & a & b \\ b & b & b & a \end{bmatrix}$

24. Suppose $c_1, c_2 \in \mathbb{R}$.

a. Find the eigenvalues, with their multiplicities, of the matrix:

$$\begin{bmatrix} 0 & 0 & 0 & c_1 \\ 0 & 0 & c_2 & 0 \\ 0 & c_2 & 0 & 0 \\ c_1 & 0 & 0 & 0 \end{bmatrix}.$$

b. Let us generalize part (a). Suppose $c_1, c_2, \ldots, c_k \in \mathbb{R}$. Let $n = 2k$, and ***even*** integer. Find the eigenvalues, with multiplicities, of the $n \times n$ matrix that is zero everywhere, except for the entries:

$$a_{1,n} = c_1,\ a_{2,n-1} = c_2,\ \ldots,\ a_{k,n-k+1} = c_k,\ a_{k+1,n-k} = c_k,\ \ldots,\ a_{n-1,2} = c_2,\ a_{n,1} = c_1.$$

In other words, the entries in the "reverse diagonal" are:

$$c_1,\ c_2, \ldots,\ c_k,\ c_k \ldots,\ c_2,\ c_1.$$

25. Suppose $c_1,\ c_2,\ c_3 \in \mathbb{R}$.

a. Find the eigenvalues, with their multiplicities, of the matrix:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & c_1 \\ 0 & 0 & 0 & c_2 & 0 \\ 0 & 0 & c_3 & 0 & 0 \\ 0 & c_2 & 0 & 0 & 0 \\ c_1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

b. Let us generalize part (a). Suppose $c_1,\ c_2, \ldots, c_k \in \mathbb{R}$. Let $n = 2k - 1$, an ***odd*** integer. Find the eigenvalues, with multiplicities, of the $n \times n$ matrix that is zero everywhere, except for the entries: $a_{1,n} = c_1,\ \ a_{2,n-1} = c_2,\ \ldots,\ a_{k-1,n-k} = c_{k-1},\ a_{k,n-k+1} = c_k,\ a_{k+1,n-k} = c_{k-1},\ \ldots,$ $a_{n-1,2} = c_2,\ a_{n,1} = c_1$. In other words, the entries in the "reverse diagonal" are:

$$c_1,\ c_2, \ldots,\ c_{k-1},\ c_k,\ c_{k-1},\ \ldots,\ c_2,\ c_1.$$

26. Suppose $a,\ b \in \mathbb{R}$.

a. Find the eigenvalues, with their multiplicities, of the matrix:

$$\begin{bmatrix} a & 0 & 0 & b \\ 0 & a & b & 0 \\ 0 & b & a & 0 \\ b & 0 & 0 & a \end{bmatrix}.$$

b. Let us generalize part (a). Suppose n is an ***even*** integer. Find the eigenvalues, with multiplicities, of the $n \times n$ matrix that is zero everywhere, except each diagonal entry is a, and each "reverse diagonal" entry is b. Note that since n is even, the diagonal and reverse diagonal entries do not intersect (as seen above in the 4×4 case).

27. Suppose $a,\ b \in \mathbb{R}$.

a. Find the eigenvalues, with their multiplicities, of the matrix:

$$\begin{bmatrix} a & 0 & 0 & 0 & b \\ 0 & a & 0 & b & 0 \\ 0 & 0 & a & 0 & 0 \\ 0 & b & 0 & a & 0 \\ b & 0 & 0 & 0 & a \end{bmatrix}.$$

b. Change the matrix in part (a) so that the entry in row 3, column 3 is $\boldsymbol{b}$ instead of a. Find the eigenvalues, with their multiplicities, of the new matrix.

c. Let us generalize part (a). Suppose $n = 2k - 1$ is an ***odd*** integer. Find the eigenvalues, with multiplicities, of the $n \times n$ matrix that is zero everywhere, except each diagonal entry is a, and each "reverse diagonal" entry is b, except for the a in row k, column k.

d. Similarly, let us generalize part (b). Again suppose $n = 2k - 1$ is an ***odd*** number. Find the eigenvalues, with multiplicities, of the $n \times n$ matrix that is zero everywhere, except each diagonal entry is a, and each "reverse diagonal" entry is b, including the entry in row k, column k.

28. The Converse of the Spectral Theorem for Symmetric Matrices: Show that if A is an $n \times n$ matrix, and we can find an orthogonal matrix Q and a diagonal matrix D such that $D = Q^{\top}AQ$, then A is a symmetric matrix.

29. Suppose that A is a symmetric matrix, and A has ***exactly two*** eigenspaces, W_1 and W_2, corresponding to two distinct eigenvalues, λ_1 and λ_2. Prove that $W_1^{\perp} = W_2$, and vice versa.

For Exercises (30) to (44): Find an orthogonal matrix Q and a diagonal matrix D such that $D = Q^{\top}AQ$ for the following symmetric matrices A. For the sake of convention, place the eigenvalues in increasing order in D. Use technology if permitted by your instructor. Each matrix has at least one eigenspace which is at least 2-dimensional, so the Gram-Schmidt Algorithm will be necessary.

30. $\begin{bmatrix} 8 & 10 & -10 \\ 10 & -7 & -5 \\ -10 & -5 & -7 \end{bmatrix}$

31. $\begin{bmatrix} 1 & 8 & -8 \\ 8 & -11 & -4 \\ -8 & -4 & -11 \end{bmatrix}$

32. $\begin{bmatrix} -1 & -1 & 2 \\ -1 & -1 & -2 \\ 2 & -2 & 2 \end{bmatrix}$

33. $\begin{bmatrix} -22 & 3 & 2 \\ 3 & -30 & -6 \\ 2 & -6 & -25 \end{bmatrix}$

34. $\begin{bmatrix} -11 & -4 & 8 \\ -4 & -11 & -8 \\ 8 & -8 & 1 \end{bmatrix}$

35. $\begin{bmatrix} 45 & 33 & 22 \\ 33 & -43 & -66 \\ 22 & -66 & 12 \end{bmatrix}$

36. $\begin{bmatrix} 7 & 3 & 2 \\ 3 & -1 & -6 \\ 2 & -6 & 4 \end{bmatrix}$

37. $\begin{bmatrix} 9 & 3 & 2 \\ 3 & 1 & -6 \\ 2 & -6 & 6 \end{bmatrix}$

38. $\begin{bmatrix} -5 & -7 & 14 \\ -7 & -5 & -14 \\ 14 & -14 & 16 \end{bmatrix}$

39. $\begin{bmatrix} 36 & 9 & 27 & 36 \\ 9 & 25 & -44 & 23 \\ 27 & -44 & 1 & -1 \\ 36 & 23 & -1 & -20 \end{bmatrix}$

40. $\begin{bmatrix} -51 & 24 & -12 & 12 \\ 24 & 13 & 4 & -4 \\ -12 & 4 & -23 & -40 \\ 12 & -4 & -40 & -23 \end{bmatrix}$

41. $\begin{bmatrix} 3 & 6 & -3 & 3 \\ 6 & 19 & 1 & -1 \\ -3 & 1 & 10 & -10 \\ 3 & -1 & -10 & 10 \end{bmatrix}$

42. $\begin{bmatrix} 5 & 10 & -5 & 5 \\ 10 & 29 & 5 & 7 \\ -5 & 5 & 30 & -10 \\ 5 & 7 & -10 & 6 \end{bmatrix}$

43. $\begin{bmatrix} -14 & -21 & 14 & 7 \\ -21 & -54 & -6 & -3 \\ 14 & -6 & -59 & 2 \\ 7 & -3 & 2 & -62 \end{bmatrix}$

44. $\begin{bmatrix} 15 & -7 & 9 & -22 & 1 \\ -7 & 7 & 7 & 14 & 7 \\ 9 & 7 & 39 & -2 & 23 \\ -22 & 14 & -2 & 36 & 6 \\ 1 & 7 & 23 & 6 & 15 \end{bmatrix}$

7.8 The Method of Least Squares

In this Section, we return to one of the core problems of Linear Algebra: solving a system of linear equations. We know that any linear system is either ***consistent*** (it has at least one solution) or ***inconsistent*** (it has no solutions). Furthermore, we learned from Chapter 1 that the linear system:

$$A\vec{x} = \vec{b}$$

is consistent if and only if $\vec{b}$ is in the ***columnspace*** of A. Thus, if A is an $m \times n$ matrix, unless the n columns of A Span ***all*** of $\mathbb{R}^m$, then there will definitely be vectors $\vec{b} \in \mathbb{R}^m$ such that $A\vec{x} = \vec{b}$ is ***inconsistent***.

Suppose, therefore, that $\vec{b}$ is not in $W = colspace(A)$. The system $A\vec{x} = \vec{b}$ is thus inconsistent. However, $proj_W\left(\vec{b}\right)$ is ***always*** in W, and therefore if we replace $\vec{b}$ with $proj_W\left(\vec{b}\right)$, this new system will now be ***consistent***. We give this ***new*** system a special name:

Definition/Theorem: Let A be an $m \times n$ matrix, and $\vec{b}$ an $m \times 1$ column matrix. If $\vec{b}$ is ***not*** a member of $W = colspace(A)$, then we call the system:

$$A\vec{x} = b_1 = proj_W\left(\vec{b}\right),$$

the ***least squares system associated to*** $A\vec{x} = \vec{b}$.

This system is always ***consistent***, and we call any solution $\vec{x}_1 \in \mathbb{R}^n$ to this system a ***least squares solution*** or ***best approximation*** to the original (inconsistent) system $A\vec{x} = \vec{b}$. Any such solution $\vec{x}_1$ has the property that if $\vec{x}$ is any other vector in $\mathbb{R}^n$ that is ***not*** a least squares solution, then:

$$\left\| A\vec{x}_1 - \vec{b} \right\| < \left\| A\vec{x} - \vec{b} \right\|.$$

In other words, the ***distance*** from $A\vec{x}_1$ to $\vec{b}$ is ***as small as possible***.

Before we prove the last part of this Definition/Theorem, let us see why this consolation prize makes sense. If $\vec{b} \notin W$, then $\vec{b}$ can be orthogonally decomposed as:

$$\vec{b} = \vec{b}_1 + \vec{b}_2, \quad \text{where } \vec{b}_1 = proj_W\left(\vec{b}\right) \in W, \quad \text{and } \vec{\mathbf{0}}_n \neq \vec{b}_2 \in W^{\perp}.$$

Therefore the linear system $A\vec{x} = b_1 \in W$ must be consistent.

Proof of Theorem: We will prove below that the solutions $\vec{x}_1$ to the least square system $A\vec{x} = b_1$ result in a vector $A\vec{x}_1$ that are as close to $\vec{b}$ as possible. But this system is also the ideal one to solve, because $\vec{b}_1$ itself is the closest vector from W to $\vec{b}$ To see this, let $\vec{w}$ be any other vector in W except $\vec{b}_1$. Then:

$$\begin{aligned}
\left\| \vec{b} - \vec{w} \right\|^2 &= \left\| \vec{b} + \left(-\vec{b}_1 + \vec{b}_1\right) - \vec{w} \right\|^2 = \left\| \left(\vec{b} - \vec{b}_1\right) + \left(\vec{b}_1 - \vec{w}\right) \right\|^2 \\
&= \left\| \vec{b}_2 + \left(\vec{b}_1 - \vec{w}\right) \right\|^2 = \left\| \vec{b}_2 \right\|^2 + \left\| \vec{b}_1 - \vec{w} \right\|^2 \\
&> \left\| \vec{b}_1 - \vec{w} \right\|^2,
\end{aligned}$$

where we were able to use the General Pythagorean Theorem, since $\vec{b}_2 \in W^{\perp}$, $\vec{b}_2 \neq \vec{\mathbf{0}}_n$, and $\vec{b}_1 - \vec{w} \in W$. Thus $\left\| \vec{b} - \vec{w} \right\| > \left\| \vec{b}_1 - \vec{w} \right\|$.

This also explains the terminology: the formula for the distance between two vectors $\vec{u}$ and $\vec{v}$ is:

$$d(\vec{u},\vec{v}) = \|\vec{u}-\vec{v}\| = \sqrt{(u_1-v_1)^2+(u_2-v_2)^2+\cdots+(u_n-v_n)^2}\,,$$

and therefore the least squares solution is the solution for which the ***squares*** of the ***differences*** between the coordinates of $A\vec{x}_1$ and the coordinates of $\vec{b}$ have the ***least*** sum that is possible.

Let us now show that if $\vec{x}$ is any vector in $\mathbb{R}^n$ that is not a least squares solution, then:

$$\left\|A\vec{x}_1-\vec{b}\right\| < \left\|A\vec{x}-\vec{b}\right\|.$$

But the idea of the proof is exactly the same idea that we saw above:

$$\begin{aligned}\left\|A\vec{x}-\vec{b}\right\|^2 &= \left\|A\vec{x}-A\vec{x}_1+A\vec{x}_1-\vec{b}\right\|^2\\ &= \left\|A(\vec{x}-\vec{x}_1)+\left(A\vec{x}_1-\vec{b}\right)\right\|^2\\ &= \|A(\vec{x}-\vec{x}_1)\|^2+\left\|A\vec{x}_1-\vec{b}\right\|^2\\ &> \left\|A\vec{x}_1-\vec{b}\right\|^2,\end{aligned}$$

where again we were able to use the General Pythagorean Theorem because $A(\vec{x}-\vec{x}_1)\in W = colspace(A)$, and $A\vec{x}_1-\vec{b} = \vec{b}_1-\vec{b} = -\vec{b}_2 \in W^{\perp}$. We were also able to use a strict inequality because $\vec{x}$ is not a least square solution, and thus $A\vec{x} \neq \vec{b}_1$, so $A\vec{x}-A\vec{x}_1$ is not the zero vector, and thus has a ***positive*** length.■

Computational Issues

Let us now focus our attention on solving the system:

$$A\vec{x} = \vec{b}_1 = proj_W\left(\vec{b}\right).$$

First, let us recall that if $\vec{x}_1$ is any solution to this system, then any other solution is of the form:

$$\vec{x}_1+\vec{x}_0, \quad \text{where } \vec{x}_0 \in nullspace(A).$$

Thus, to find all the solutions, we only need to find at least one, and also find a basis for the nullspace so that we can use it to find all the other solutions.

Our first step will therefore be to find the reduced row echelon form R of the matrix A. This of course allows us to find a basis for $nullspace(A)$. But there is another good reason for finding R. We saw in Chapter 1 that the columns of A corresponding to the columns of R that contain a leading one form a ***basis*** for $colspace(A)$.

Next, let us assemble the matrix C, whose columns consist of those columns of A that correspond to the columns of R that contain a leading 1. Thus, $colspace(C) = colspace(A)$, and if A has k linearly independent columns, with $k \leq n$, then C is $m \times k$. Thus, the system:

$$C\vec{x} = \vec{b}_1.$$

is still consistent. It is possible, of course, that the columns of A are already linearly independent, in which case $C = A$. By way of example, suppose that A has 5 columns, and columns 1, 3 and 4 correspond to the leading 1's of R. Thus:

$$A = [\,\vec{a}_1\,|\,\vec{a}_2\,|\,\vec{a}_3\,|\,\vec{a}_4\,|\,\vec{a}_5\,], \quad \text{and} \quad C = [\,\vec{a}_1\,|\,\vec{a}_3\,|\,\vec{a}_4\,].$$

Now, if $\vec{y} = \langle y_1, y_2, y_3 \rangle$ is a solution to the system:

$$C\vec{x} = \vec{b}_1,$$

then $\vec{x}_1 = \langle y_1, 0, y_2, y_3, 0 \rangle$ is a solution to our system:

$$A\vec{x} = \vec{b}_1.$$

We will call this operation ***padding with zeroes***. Thus, the system $C\vec{x} = \vec{b}_1$ yields a solution to our system of interest $A\vec{x} = \vec{b}_1$.

Thus, we only need to solve the (possibly smaller) system $C\vec{x} = \vec{b}_1$. All we really need is $\vec{b}_1$ and the Gauss-Jordan Algorithm. We can of course find $\vec{b}_1 = proj_W\left(\vec{b}\right)$ by applying the Gram-Schmidt Algorithm on the columns of C (which form a basis for W), and using the resulting orthonormal basis $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ to find:

$$proj_W\left(\vec{b}\right) = \left\langle \vec{b}|\vec{u}_1 \right\rangle \vec{u}_1 + \left\langle \vec{b}|\vec{u}_2 \right\rangle \vec{u}_2 + \cdots + \left\langle \vec{b}|\vec{u}_k \right\rangle \vec{u}_k.$$

As we saw in the previous Sections, this is a rather tedious algorithm, so let us find an alternative process, and in so doing, we will also find another way to construct the matrix of $proj_W$. Suppose:

$$C\vec{x} = \vec{b}_1. \quad \text{Then :}$$

$$\vec{b} - C\vec{x} = \vec{b} - \vec{b}_1 = \vec{b}_2 \in W^{\perp}.$$

This means that if $\vec{x}$ is a solution to our new system, then $\vec{b} - C\vec{x}$ must be in the orthogonal complement of $colspace(C)$. But this last space is exactly the same as $rowspace(C^{\top})$. Thus:

$$C^{\top}\left(\vec{b} - C\vec{x}\right) = \vec{\mathbf{0}}_{\boldsymbol{k}}, \text{ or in other words,}$$

$$C^{\top}C\vec{x} = C^{\top}\vec{b}.$$

We call this the ***normal system*** associated to $A\vec{x} = \vec{b}$. This system is again consistent, because we know that $C\vec{x} = \vec{b}_1$ is consistent. But the big surprise is that it is more than just consistent:

Theorem: Let C be an $m \times k$ matrix with ***linearly independent*** columns. Then: $C^{\top}C$ is an ***invertible*** $k \times k$ matrix, and therefore the ***normal system:***

$$C^{\top}C\vec{x} = C^{\top}\vec{b}$$

has ***exactly one solution***.

This solution, of course, has to be a solution to the system $C\vec{x} = \vec{b}_1$.

Proof: Let C be an $m \times k$ matrix with linearly independent columns. Thus, $C^{\top}C$ must be a $k \times k$ matrix. To show that $C^{\top}C$ is invertible, let us show that the only solution $\vec{x}$ to the homogenous system:

$$(C^{\top}C)\vec{x} = \vec{\mathbf{0}}_{\boldsymbol{k}},$$

is the ***trivial*** solution. By the ***associativity*** of matrix multiplication, we can write this as:

$$C^{\top}(C\vec{x}) = \vec{\mathbf{0}}_{\boldsymbol{k}}.$$

This tells us that $C\vec{x}$ is a member of the ***nullspace*** of $C^{\top}$. But recall that the nullspace of a matrix is the orthogonal complement of its rowspace. Thus, the nullspace of $C^{\top}$ is the orthogonal complement of the rowspace of $C^{\top}$. But since the rowspace of $C^{\top}$ is the same as the columnspace of C, we can finally conclude that $C\vec{x}$ is a member of the orthogonal complement of the columnspace of C.

But by the definition of matrix multiplication, $C\vec{x}$ is also a member of the columnspace of C. Since $W \cap W^{\perp} = \left\{\vec{\mathbf{0}}_{m}\right\}$, we must have $C\vec{x} = \vec{\mathbf{0}}_{m}$.

Finally, since the columns of C are linearly independent, we must have $\vec{x} = \vec{\mathbf{0}}_{k}$. Thus the system $(C^{\top}C)\vec{x} = \vec{\mathbf{0}}_{k}$ only has the trivial solution, so $C^{\top}C$ is invertible.■

Thus, the normal system associated to our original system has unique solution:

$$\vec{x} = (C^{\top}C)^{-1} \cdot C^{\top}\vec{b}.$$

As a well-earned bonus, we also get a new formula for $[proj_W]$:

Theorem: Let C be an $m \times k$ matrix with linearly independent columns, and let W be the ***columnspace*** of C. Then: $C^{\top}C$ is an ***invertible*** $k \times k$ matrix, and:

$$[proj_W] = C \cdot (C^{\top}C)^{-1} \cdot C^{\top}.$$

Proof: Using the notation we have established so far, we have:

$$proj_W\left(\vec{b}\right) = \vec{b}_1 = C\vec{x} = C \cdot (C^{\top}C)^{-1} \cdot C^{\top}\vec{b},$$

and thus:

$$[proj_W] = C \cdot (C^{\top}C)^{-1} \cdot C^{\top}.■$$

One would argue that this is a major improvement, because it is usually easier to find the ***inverse*** of a matrix rather than apply the Gram-Schmidt Algorithm.

Revisiting Projections and Reflections

Let us present a fourth method to find the matrix for the projection operator onto a plane Π through the origin in $\mathbb{R}^3$ using the formula above for $[proj_W]$. The same formula can be used in the simpler case when $W = L$, a line through the origin.

Example: Suppose that Π is the plane with equation: $5x + 2y - 6z = 0$.

This is the same plane that we studied in Sections 3.6 and 6.4. Recall that we chose for our basis for Π the set $B = \{\langle 2, -5, 0\rangle, \langle 0, 3, 1\rangle\}$. We assemble these vectors into the columns of a matrix C :

$$C = \begin{bmatrix} 2 & 0 \\ -5 & 3 \\ 0 & 1 \end{bmatrix}.$$

Thus:

$$C^{\top} \cdot C = \begin{bmatrix} 2 & -5 & 0 \\ 0 & 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 \\ -5 & 3 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 29 & -15 \\ -15 & 10 \end{bmatrix}, \quad \text{with inverse:}$$

$$(C^{\top} \cdot C)^{-1} = \begin{bmatrix} \frac{10}{65} & \frac{15}{65} \\ \frac{15}{65} & \frac{29}{65} \end{bmatrix}.$$

Now, we get:

$$\begin{aligned} [proj_{\Pi}] &= C \cdot (C^{\top} C)^{-1} \cdot C^{\top} \\ &= \begin{bmatrix} 2 & 0 \\ -5 & 3 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \frac{10}{65} & \frac{15}{65} \\ \frac{15}{65} & \frac{29}{65} \end{bmatrix} \begin{bmatrix} 2 & -5 & 0 \\ 0 & 3 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \frac{20}{65} & \frac{30}{65} \\ -\frac{5}{65} & \frac{12}{65} \\ \frac{15}{65} & \frac{29}{65} \end{bmatrix} \begin{bmatrix} 2 & -5 & 0 \\ 0 & 3 & 1 \end{bmatrix} = \frac{1}{65} \begin{bmatrix} 40 & -10 & 30 \\ -10 & 61 & 12 \\ 30 & 12 & 29 \end{bmatrix}, \end{aligned}$$

which is the same answer we obtained the last two times.□

The Best Approximation Algorithm

We are now ready to summarize our discussion above in order to construct an algorithm to find the best approximation to an inconsistent system:

Theorem — The Best Approximation Algorithm:

Let $A\vec{x} = \vec{b}$ be an ***inconsistent*** system with A an $m \times n$ coefficient matrix, whose reduced row echelon form is R. Let C be the $m \times k$ matrix whose columns are the columns of A corresponding to the leading 1's of R, in the same order. Then: $C^{\top} C$ is ***invertible***, and:

$$\vec{x} = (C^{\top} C)^{-1} \cdot C^{\top} \vec{b}$$

is a solution to:

$$C\vec{x} = \vec{b}_1 = proj_W\left(\vec{b}\right).$$

By padding $\vec{x}$ with zeroes, we obtain a solution $\vec{x}_1$ to the least squares system $A\vec{x} = \vec{b}_1$, and all the best approximation solutions are of the form:

$$\vec{x}_1 + \vec{x}_0, \quad \text{where } \vec{x}_0 \in nullspace(A).$$

Consequently:

$$[proj_W] = C \cdot (C^{\top} C)^{-1} \cdot C^{\top}.$$

We will also call $\left\| \vec{b} - \vec{b}_1 \right\| = \left\| \vec{b}_2 \right\|$ the ***common error*** of all our best approximation solutions.

Example: Let us go back, almost full circle, to the very first system of equations that we encountered in Section 1.4, but let us modify the entries of the right-hand side $\vec{b}$ ever so slightly:

$$\begin{array}{rcrcrcrcr} 4x_1 & - & 8x_2 & + & 3x_3 & + & 9x_4 & = & 7 \\ 3x_1 & - & 6x_2 & - & 4x_3 & + & 13x_4 & = & 15 \\ -2x_1 & + & 4x_2 & + & 3x_3 & - & 9x_4 & = & -9 \end{array}$$

Let us see if this system is now inconsistent. The augmented matrix is:

$$\left[\begin{array}{rrrr|r} 4 & -8 & 3 & 9 & 7 \\ 3 & -6 & -4 & 13 & 15 \\ -2 & 4 & 3 & -9 & -9 \end{array}\right], \text{ with rref } \left[\begin{array}{rrrr|r} 1 & -2 & 0 & 3 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right],$$

and thus, our new system is indeed inconsistent. This was not a complete waste of time, though, because we also get the rref of the coefficient matrix A on the left-hand side. Thus, we see that the four columns of A are dependent (as they should be, since they are vectors from $\mathbb{R}^3$), and in fact columns 1 and 3 form a basis for $colspace(A)$. Thus, our matrix C is:

$$C = \begin{bmatrix} 4 & 3 \\ 3 & -4 \\ -2 & 3 \end{bmatrix}.$$

Applying our recipe, we first need:

$$C^\top C = \begin{bmatrix} 4 & 3 & -2 \\ 3 & -4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 3 & -4 \\ -2 & 3 \end{bmatrix} = \begin{bmatrix} 29 & -6 \\ -6 & 34 \end{bmatrix}, \text{ with inverse:}$$

$$(C^\top C)^{-1} = \frac{1}{29 \cdot 34 - 36} \begin{bmatrix} 34 & 6 \\ 6 & 29 \end{bmatrix} = \begin{bmatrix} \frac{17}{475} & \frac{3}{475} \\ \frac{3}{475} & \frac{29}{950} \end{bmatrix}.$$

We are now ready to find our solution vector:

$$\begin{aligned} \vec{x} &= (C^\top C)^{-1} \cdot C^\top \vec{b} \\ &= \begin{bmatrix} \frac{17}{475} & \frac{3}{475} \\ \frac{3}{475} & \frac{29}{950} \end{bmatrix} \begin{bmatrix} 4 & 3 & -2 \\ 3 & -4 & 3 \end{bmatrix} \begin{bmatrix} 7 \\ 15 \\ -9 \end{bmatrix} \\ &= \begin{bmatrix} \frac{77}{475} & \frac{39}{475} & -\frac{1}{19} \\ \frac{111}{950} & -\frac{49}{475} & \frac{3}{38} \end{bmatrix} \begin{bmatrix} 7 \\ 15 \\ -9 \end{bmatrix} = \begin{bmatrix} \frac{71}{25} \\ -\frac{36}{25} \end{bmatrix}. \end{aligned}$$

Notice that we performed the matrix multiplication starting on the ***left***, since we will need this product later when we find $[proj_W]$. We pad this with zeroes in the 2nd and 4th entries, to get:

$$\vec{x}_1 = \left\langle \frac{71}{25}, 0, -\frac{36}{25}, 0 \right\rangle,$$

which should be a best-approximation solution to the system:

$$A\vec{x} = \vec{b}_1 = proj_W\left(\vec{b}\right).$$

Thus:

$$A\vec{x}_1 = \begin{bmatrix} 4 & -8 & 3 & 9 \\ 3 & -6 & -4 & 13 \\ -2 & 4 & 3 & -9 \end{bmatrix}\begin{bmatrix} \frac{71}{25} \\ 0 \\ -\frac{36}{25} \\ 0 \end{bmatrix} = \begin{bmatrix} \frac{176}{25} \\ \frac{357}{25} \\ -10 \end{bmatrix}.$$

Looking at the rref R, we can easily find a basis for the nullspace of A, and see that all the best approximation solutions have the form:

$$\vec{x}_1 + \vec{x}_0 = \left\langle \frac{71}{25}, 0, -\frac{36}{25}, 0\right\rangle + r\langle 2, 1, 0, 0\rangle + s\langle -3, 0, 1, 1\rangle.$$

To find the common error of these solutions, we find:

$$\vec{b}_2 = \vec{b} - \vec{b}_1 = \langle 7, 15, -9\rangle - \left\langle \frac{176}{25}, \frac{357}{25}, -10\right\rangle = \left\langle -\frac{1}{25}, \frac{18}{25}, 1\right\rangle,$$

and so the common error of our best approximations is:

$$\left\|\vec{b}_2\right\| = \left\|\left\langle -\frac{1}{25}, \frac{18}{25}, 1\right\rangle\right\| = \sqrt{\left(-\frac{1}{25}\right)^2 + \left(\frac{18}{25}\right)^2 + 1^2} = \frac{1}{5}\sqrt{38} \approx 1.2329,$$

which is really not too bad, since we changed the entries on the right side $\vec{b}$ of our original system in Section 1.4 by 1, 2 and 3 respectively. We can also find the matrix of $proj_W$ using our intermediate computation above:

$$\begin{aligned}
[proj_W] &= C \cdot (C^\top C)^{-1} \cdot C^\top \\
&= \begin{bmatrix} 4 & 3 \\ 3 & -4 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} \frac{17}{475} & \frac{3}{475} \\ \frac{3}{475} & \frac{29}{950} \end{bmatrix}\begin{bmatrix} 4 & 3 & -2 \\ 3 & -4 & 3 \end{bmatrix} \\
&= \begin{bmatrix} 4 & 3 \\ 3 & -4 \\ -2 & 3 \end{bmatrix}\begin{bmatrix} \frac{77}{475} & \frac{39}{475} & -\frac{1}{19} \\ \frac{111}{950} & -\frac{49}{475} & \frac{3}{38} \end{bmatrix} = \begin{bmatrix} \frac{949}{950} & \frac{9}{475} & \frac{1}{38} \\ \frac{9}{475} & \frac{313}{475} & -\frac{9}{19} \\ \frac{1}{38} & -\frac{9}{19} & \frac{13}{38} \end{bmatrix}.
\end{aligned}$$

We can find $\vec{b}_1$ using this matrix, and verify our answer above:

$$\begin{aligned}
\vec{b}_1 &= proj_W\left(\vec{b}\right) \\
&= \begin{bmatrix} \frac{949}{950} & \frac{9}{475} & \frac{1}{38} \\ \frac{9}{475} & \frac{313}{475} & -\frac{9}{19} \\ \frac{1}{38} & -\frac{9}{19} & \frac{13}{38} \end{bmatrix}\begin{bmatrix} 7 \\ 15 \\ -9 \end{bmatrix} = \begin{bmatrix} \frac{176}{25} \\ \frac{357}{25} \\ -10 \end{bmatrix},
\end{aligned}$$

which is indeed our answer.□

7.8 Section Summary

Let A be an $m \times n$ matrix, and $\vec{b}$ an $m \times 1$ column matrix. If $\vec{b}$ is not a member of W, the ***columnspace*** of A, then we call the system $A\vec{x} = b_1 = proj_W(\vec{b})$ the ***least squares system associated to*** $A\vec{x} = \vec{b}$.

This system is always ***consistent***, and we call any solution $\vec{x}_1 \in \mathbb{R}^n$ to this system as a ***least squares solution*** or ***best approximation*** to the original (inconsistent) system $A\vec{x} = \vec{b}$.

Any such solution $\vec{x}_1$ has the property that if $\vec{x}$ is any other vector in $\mathbb{R}^n$ that is ***not*** a least squares solution, then $\left\| A\vec{x}_1 - \vec{b} \right\| < \left\| A\vec{x} - \vec{b} \right\|$.

In words, the ***distance*** from $A\vec{x}_1$ to $\vec{b}$ is ***as small as possible***.

Let C be an $m \times k$ matrix with linearly independent columns. Then $C^\top C$ is an ***invertible*** matrix, and therefore the system $C^\top C\vec{x} = C^\top \vec{b}$ has ***exactly one solution***.

Let $A\vec{x} = \vec{b}$ be an inconsistent system with A an $m \times n$ coefficient matrix, whose rref is R. Let C be the $m \times k$ matrix whose columns are the columns of A corresponding to the leading 1's of R, in the same order. Then: $C^\top C$ is ***invertible***, and:

$$\vec{x} = (C^\top C)^{-1} \cdot C^\top \vec{b}$$

is a solution to $C\vec{x} = \vec{b}_1 = proj_W(\vec{b})$.

By padding $\vec{x}$ with zeroes, we obtain a solution $\vec{x}_1$ to the least squares system $A\vec{x} = \vec{b}_1$, and all the best approximation solutions are of the form $\vec{x}_1 + \vec{x}_0$, where $\vec{x}_0 \in nullspace(A)$.

Consequently, $[proj_W] = C \cdot (C^\top C)^{-1} \cdot C^\top$.

We will also call $\left\| \vec{b} - \vec{b}_1 \right\| = \left\| \vec{b}_2 \right\|$ the ***common error*** of all our best approximation solutions.

7.8 Exercises

For Exercises (1) to (5): Perform the following: (a) Verify that the system $A\vec{x} = \vec{b}$ is inconsistent; (b) Find all the best approximation solutions to this inconsistent system; (c) Find the common error of all the best approximations solutions; (d) Find $[proj_W]$ using our new formula, where W is the columnspace of A :

1. $A = \begin{bmatrix} 3 & -1 & -6 \\ -2 & 1 & 3 \\ 5 & -2 & -9 \end{bmatrix}, \vec{b} = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$

2. $A = \begin{bmatrix} 1 & 1 & -1 \\ 1 & -2 & 5 \\ 2 & -1 & 4 \\ 2 & 1 & 0 \end{bmatrix}, \vec{b} = \begin{bmatrix} -2 \\ 9 \\ 5 \\ 4 \end{bmatrix}$

3. $A = \begin{bmatrix} 3 & -15 & -6 & 2 \\ -2 & 10 & 4 & -4 \\ 5 & -25 & -10 & -1 \end{bmatrix}, \vec{b} = \begin{bmatrix} 28 \\ -26 \\ 13 \end{bmatrix}$

4. $A = \begin{bmatrix} 3 & -2 & 19 & 4 \\ 4 & -1 & 22 & -3 \\ -1 & 5 & -15 & 2 \\ 1 & 2 & 1 & 4 \end{bmatrix}, \vec{b} = \begin{bmatrix} 38 \\ 5 \\ -28 \\ 2 \end{bmatrix}$

5. $A = \begin{bmatrix} 4 & -3 & 1 \\ -2 & 0 & -5 \\ 3 & 1 & 2 \\ -1 & 5 & 6 \\ 0 & 3 & 2 \end{bmatrix}, \vec{b} = \begin{bmatrix} 11 \\ -9 \\ 5 \\ -7 \\ -4 \end{bmatrix}$

6. Let C be an $m \times k$ matrix, and let $W = colspace(C)$. Suppose the columns of C already form an ***orthonormal set***. Show that:

$$C^{\top} C = \boldsymbol{I_k},$$

and consequently $[proj_W] = C \cdot C^{\top}$. This shows that our new formula for $[proj_W]$ is a ***generalization*** of our formula in Section 7.5.

7. ***Converse of the Second Theorem:*** Prove that if C is an $m \times k$ matrix, and if $C^{\top} C$ is ***invertible***, then the ***columns*** of C are linearly ***independent***.

8. In Exercise 14 (d) of Section 7.6, we mentioned that $[proj_W]$, and thus $[refl_W]$, will not contain any radicals, though we tacitly assumed that the entries of the basis for W consist only of rational numbers. Use our new formula for $[proj_W]$ to show that its entries are indeed rational numbers if the columns of A contain only rational numbers.

9. Use a simple basis for the plane $\Pi : 3x - 7y + 4z = 0$ and our new formula for $[proj_W]$ from this Section in order to find $[proj_{\Pi}]$.

10. Use the direction vector for the line $L : Span(\{\langle 5, -2, 3 \rangle\})$ and our new formula for $[proj_W]$ from this Section in order to find $[proj_L]$.

7.9 The QR-Decomposition

In this Section, we present a factorization method that arises from the application of the Gram-Schmidt Algorithm, and can be used to solve the Least-Squares Problem when the coefficient matrix A has linearly independent columns.

Recall that we have as our input to the Gram-Schmidt Algorithm a basis $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ for some inner product space V. So let us suppose that B is a set of vectors from some $\mathbb{R}^m$, and therefore the columns of the matrix:

$$A = [\vec{w}_1\ \vec{w}_2\ \ldots\ \vec{w}_n]$$

are linearly independent. According to the Algorithm, we produce a sequence of orthogonal vectors:

$$\begin{aligned}
\vec{v}_1 &= \vec{w}_1, \\
\vec{v}_2 &= \vec{w}_2 - (\vec{w}_2 \circ \vec{u}_1)\vec{u}_1, \\
\vec{v}_3 &= \vec{w}_3 - (\vec{w}_3 \circ \vec{u}_1)\vec{u}_1 - (\vec{w}_3 \circ \vec{u}_2)\vec{u}_2, \ \ldots \\
\vec{v}_n &= \vec{w}_n - (\vec{w}_n \circ \vec{u}_1)\vec{u}_1 - (\vec{w}_n \circ \vec{u}_2)\vec{u}_2 - \cdots - (\vec{w}_n \circ \vec{u}_{n-1})\vec{u}_{n-1},
\end{aligned}$$

where the unit vectors $\vec{u}_i$ are obtained from the $\vec{v}_i$ via: $\vec{u}_i = \vec{v}_i / \|\vec{v}_i\|$. However, this also says that $\vec{v}_i = \|\vec{v}_i\| \bullet \vec{u}_i$. Since $\vec{u}_i$ is a unit vector, though, we get:

$$\begin{aligned}
\vec{v}_i \circ \vec{u}_i &= (\|\vec{v}_i\| \bullet \vec{u}_i) \circ \vec{u}_i = \|\vec{v}_i\|(\vec{u}_i \circ \vec{u}_i) = \|\vec{v}_i\|, \text{ and so:} \\
\vec{v}_i &= (\vec{v}_i \circ \vec{u}_i)\vec{u}_i.
\end{aligned}$$

If we solve for the original vectors $\vec{w}_i$, we obtain:

$$\begin{aligned}
\vec{w}_1 &= \vec{v}_1, \\
\vec{w}_2 &= (\vec{w}_2 \circ \vec{u}_1)\vec{u}_1 + \vec{v}_2, \\
\vec{w}_3 &= (\vec{w}_3 \circ \vec{u}_1)\vec{u}_1 + (\vec{w}_3 \circ \vec{u}_2)\vec{u}_2 + \vec{v}_3, \ \ldots \\
\vec{w}_n &= (\vec{w}_n \circ \vec{u}_1)\vec{u}_1 + (\vec{w}_n \circ \vec{u}_2)\vec{u}_2 + \cdots + (\vec{w}_n \circ \vec{u}_{n-1})\vec{u}_{n-1} + \vec{v}_n.
\end{aligned}$$

Now, by the orthonormality of $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$, we obtain, for all $i = 1 \ldots n$:

$$\begin{aligned}
\vec{w}_i \circ \vec{u}_i &= (\vec{w}_i \circ \vec{u}_1) \bullet (\vec{u}_1 \circ \vec{u}_i) + (\vec{w}_i \circ \vec{u}_2) \bullet (\vec{u}_2 \circ \vec{u}_i) \\
&\qquad + \cdots + (\vec{w}_i \circ \vec{u}_{i-1}) \bullet (\vec{u}_{i-1} \circ \vec{u}_i) + (\vec{v}_i \circ \vec{u}_i) \\
&= \vec{v}_i \circ \vec{u}_i.
\end{aligned}$$

Since we know from above that $\vec{v}_i = (\vec{v}_i \circ \vec{u}_i)\vec{u}_i$, we now obtain:

$$\vec{v}_i = (\vec{v}_i \circ \vec{u}_i)\vec{u}_i = (\vec{w}_i \circ \vec{u}_i)\vec{u}_i.$$

We can now substitute this expression for each $\vec{v}_i$ in our sums for $\vec{w}_i$ above to get:

$$\begin{aligned}
\vec{w}_1 &= (\vec{w}_1 \circ \vec{u}_1)\vec{u}_1, \\
\vec{w}_2 &= (\vec{w}_2 \circ \vec{u}_1)\vec{u}_1 + (\vec{w}_2 \circ \vec{u}_2)\vec{u}_2, \ \ldots \\
\vec{w}_n &= (\vec{w}_n \circ \vec{u}_1)\vec{u}_1 + (\vec{w}_n \circ \vec{u}_2)\vec{u}_2 + \cdots + (\vec{w}_n \circ \vec{u}_{n-1})\vec{u}_{n-1} + (\vec{w}_n \circ \vec{u}_n)\vec{u}_n.
\end{aligned}$$

These n equations can now be written more compactly as a matrix equation:

$$[\vec{w}_1\ \vec{w}_2\ \ldots\ \vec{w}_n] = [\vec{u}_1\ \vec{u}_2\ \ldots\ \vec{u}_n]\begin{bmatrix} \vec{w}_1 \circ \vec{u}_1 & \vec{w}_2 \circ \vec{u}_1 & \cdots & \cdots & \vec{w}_n \circ \vec{u}_1 \\ 0 & \vec{w}_2 \circ \vec{u}_2 & \cdots & \cdots & \vec{w}_n \circ \vec{u}_2 \\ 0 & 0 & \ddots & & \vdots \\ 0 & 0 & 0 & \ddots & \vec{w}_n \circ \vec{u}_{n-1} \\ 0 & 0 & 0 & 0 & \vec{w}_n \circ \vec{u}_n \end{bmatrix}.$$

Since $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is an orthonormal set, the matrix $Q = [\vec{u}_1\ \vec{u}_2\ \ldots\ \vec{u}_n]$ has orthonormal columns. In particular, if $m = n$, Q is an ***orthogonal matrix***. Furthermore:

$$\vec{w}_i \circ \vec{u}_i = \vec{v}_i \circ \vec{u}_i = \|\vec{v}_i\| \neq 0,$$

because $\vec{w}_i$ is not a member of $Span(\{\vec{w}_1, \vec{w}_2, \ldots \vec{w}_{i-1}\})$. Therefore, the upper triangular matrix R on the right side above is ***invertible***, since none of the diagonal entries are 0. Let us summarize our discussion above in the following:

Theorem — The QR-Decomposition:
Suppose that $A = [\vec{w}_1\ \vec{w}_2\ \ldots\ \vec{w}_n]$ is an $m \times n$ matrix with ***linearly independent columns***. Then we can factor A into:

$$A = QR,$$

where Q is an $m \times n$ matrix with ***orthonormal columns***, and R is an $n \times n$ ***invertible upper triangular*** matrix. In the case when A is an ***invertible*** $n \times n$ matrix, Q is an ***orthogonal*** $n \times n$ matrix.

Example: Consider the matrix:

$$A = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix}.$$

We can easily verify that A has linearly independent columns. However, this will not really be necessary because if the Gram-Schmidt Algorithm ever gives us the zero vector, we know that the columns are dependent. Now, let us apply the Algorithm to the columns:

$$\vec{v}_1 = \langle 1, 2, -1, 0\rangle;$$

$$\begin{aligned}\vec{v}_2 &= \langle 0, 1, 1, -2\rangle - \frac{\langle 0, 1, 1, -2\rangle \circ \langle 1, 2, -1, 0\rangle}{\langle 1, 2, -1, 0\rangle \circ \langle 1, 2, -1, 0\rangle}\langle 1, 2, -1, 0\rangle \\ &= \langle 0, 1, 1, -2\rangle - \frac{1}{6}\langle 1, 2, -1, 0\rangle \\ &= \frac{1}{6}\langle -1, 4, 7, -12\rangle, \quad \text{and so we will use:}\end{aligned}$$

$$\vec{v}_2 = \langle -1, 4, 7, -12\rangle. \qquad \text{Lastly:}$$

$$\vec{v}_3 = \langle -1, 0, 1, 1\rangle - \frac{\langle -1, 0, 1, 1\rangle \circ \langle 1, 2, -1, 0\rangle}{\langle 1, 2, -1, 0\rangle \circ \langle 1, 2, -1, 0\rangle}\langle 1, 2, -1, 0\rangle$$
$$- \frac{\langle -1, 0, 1, 1\rangle \circ \langle -1, 4, 7, -12\rangle}{\langle -1, 4, 7, -12\rangle \circ \langle -1, 4, 7, -12\rangle}\langle -1, 4, 7, -12\rangle$$
$$= \langle -1, 0, 1, 1\rangle - \frac{-2}{6}\langle 1, 2, -1, 0\rangle - \frac{-4}{210}\langle -1, 4, 7, -12\rangle = \frac{1}{35}\langle -24, 26, 28, 27\rangle,$$

and so we will use $\vec{v}_3 = \langle -24, 26, 28, 27\rangle$. Thus, the columns are indeed linearly independent. Normalizing these three vectors and assembling them into Q, we get:

$$Q = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{210}} & -\frac{24}{\sqrt{2765}} \\ \frac{2}{\sqrt{6}} & \frac{4}{\sqrt{210}} & \frac{26}{\sqrt{2765}} \\ -\frac{1}{\sqrt{6}} & \frac{7}{\sqrt{210}} & \frac{28}{\sqrt{2765}} \\ 0 & -\frac{12}{\sqrt{210}} & \frac{27}{\sqrt{2765}} \end{bmatrix}.$$

Now, we compute the six dot products that we need for R :

$$\vec{w}_1 \circ \vec{u}_1 = \langle 1, 2, -1, 0\rangle \circ \frac{1}{\sqrt{6}}\langle 1, 2, -1, 0\rangle = \sqrt{6},$$
$$\vec{w}_2 \circ \vec{u}_1 = \langle 0, 1, 1, -2\rangle \circ \frac{1}{\sqrt{6}}\langle 1, 2, -1, 0\rangle = \frac{1}{\sqrt{6}},$$
$$\vec{w}_3 \circ \vec{u}_1 = \langle -1, 0, 1, 1\rangle \circ \frac{1}{\sqrt{6}}\langle 1, 2, -1, 0\rangle = \frac{-2}{\sqrt{6}},$$
$$\vec{w}_2 \circ \vec{u}_2 = \langle 0, 1, 1, -2\rangle \circ \frac{1}{\sqrt{210}}\langle -1, 4, 7, -12\rangle = \frac{35}{\sqrt{210}},$$
$$\vec{w}_3 \circ \vec{u}_2 = \langle -1, 0, 1, 1\rangle \circ \frac{1}{\sqrt{210}}\langle -1, 4, 7, -12\rangle = \frac{-4}{\sqrt{210}}, \text{ and}$$
$$\vec{w}_3 \circ \vec{u}_3 = \langle -1, 0, 1, 1\rangle \circ \frac{1}{\sqrt{2765}}\langle -24, 26, 28, 27\rangle = \frac{79}{\sqrt{2765}}.$$

Thus, we obtain the QR-decomposition:

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 1 \\ 0 & -2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{210}} & -\frac{24}{\sqrt{2765}} \\ \frac{2}{\sqrt{6}} & \frac{4}{\sqrt{210}} & \frac{26}{\sqrt{2765}} \\ -\frac{1}{\sqrt{6}} & \frac{7}{\sqrt{210}} & \frac{28}{\sqrt{2765}} \\ 0 & -\frac{12}{\sqrt{210}} & \frac{27}{\sqrt{2765}} \end{bmatrix} \begin{bmatrix} \sqrt{6} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \\ 0 & \frac{35}{\sqrt{210}} & -\frac{4}{\sqrt{210}} \\ 0 & 0 & \frac{79}{\sqrt{2765}} \end{bmatrix}.$$

Connection with the Least Squares Problem

Suppose that A is an $m \times n$ matrix. We saw in the previous Section that if the matrix equation $A\vec{x} = \vec{b}$ is inconsistent, we can still find an approximate solution to this system by solving the associated Least Squares System $A\vec{x} = proj_W\left(\vec{b}\right)$, where $W = colspace(A)$. The key to solving this new system is to create the (possibly smaller) $m \times k$ matrix C whose columns are original columns of A that form a basis for W. The ***normal system:***

$$C^\top C\vec{x} = C^\top\vec{b}$$

is not only consistent, but has ***exactly one*** solution $\vec{x}$. By "padding" $\vec{x}$ with zeroes, we can obtain a solution $\vec{x}_1$ to our Least Squares System.

Now, here is where the QR-decomposition comes in: Since C has linearly independent columns, we can find its QR-decomposition:

$$C = QR.$$

But then, the normal system becomes:

$$\begin{aligned}
(QR)^\top(QR)\vec{x} &= (QR)^\top\vec{b} && \Rightarrow \\
R^\top Q^\top QR\vec{x} &= R^\top Q^\top\vec{b} && \Rightarrow \\
R^\top R\vec{x} &= R^\top Q^\top\vec{b} && (\text{since } Q^\top Q = \boldsymbol{I}_k) \quad \Rightarrow \\
R\vec{x} &= Q^\top\vec{b} && (\text{since } R^\top \text{ is invertible}).
\end{aligned}$$

Thus, the solution to our normal system is:

$$\vec{x} = R^{-1}Q^\top\vec{b}$$

Example: Let us consider once again the inconsistent system:

$$\begin{array}{rcrcrcrcr}
4x_1 & - & 8x_2 & + & 3x_3 & + & 9x_4 & = & 7 \\
3x_1 & - & 6x_2 & - & 4x_3 & + & 13x_4 & = & 15 \\
-2x_1 & + & 4x_2 & + & 3x_3 & - & 9x_4 & = & -9
\end{array}$$

that we saw in the previous Section. We saw that the 1st and 3rd columns of the coefficient matrix A form a basis for $colspace(A)$, and so:

$$C = \begin{bmatrix} 4 & 3 \\ 3 & -4 \\ -2 & 3 \end{bmatrix}.$$

Applying Gram-Schmidt on the columns, we get:

$$
\begin{aligned}
\vec{v}_1 &= \langle 4, 3, -2\rangle;\\
\vec{v}_2 &= \langle 3, -4, 3\rangle - \frac{\langle 3, -4, 3\rangle \circ \langle 4, 3, -2\rangle}{\langle 4, 3, -2\rangle \circ \langle 4, 3, -2\rangle}\langle 4, 3, -2\rangle\\
&= \langle 3, -4, 3\rangle - \frac{-6}{29}\langle 4, 3, -2\rangle\\
&= \frac{1}{29}\langle 111, -98, 75\rangle, \quad \text{so we use:}\\
\vec{v}_2 &= \langle 111, -98, 75\rangle.
\end{aligned}
$$

Thus, our matrix Q is:

$$
Q = \begin{bmatrix} \frac{4}{\sqrt{29}} & \frac{111}{\sqrt{27550}} \\ \frac{3}{\sqrt{29}} & \frac{-98}{\sqrt{27550}} \\ \frac{-2}{\sqrt{29}} & \frac{75}{\sqrt{27550}} \end{bmatrix}.
$$

The three entries in our upper triangular matrix are:

$$
\begin{aligned}
\vec{w}_1 \circ \vec{u}_1 &= \langle 4, 3, -2\rangle \circ \frac{1}{\sqrt{29}}\langle 4, 3, -2\rangle = \sqrt{29},\\
\vec{w}_2 \circ \vec{u}_1 &= \langle 3, -4, 3\rangle \circ \frac{1}{\sqrt{29}}\langle 4, 3, -2\rangle = \frac{-6}{\sqrt{29}}, \quad \text{and}\\
\vec{w}_2 \circ \vec{u}_2 &= \langle 3, -4, 3\rangle \circ \frac{1}{\sqrt{27550}}\langle 111, -98, 75\rangle = \frac{950}{\sqrt{27550}}.
\end{aligned}
$$

Thus:

$$
R = \begin{bmatrix} \sqrt{29} & \frac{-6}{\sqrt{29}} \\ 0 & \frac{950}{\sqrt{27550}} \end{bmatrix}.
$$

We obtain our solution to the normal equation:

$$
\begin{aligned}
\vec{x} &= R^{-1}Q^{\mathsf{T}}\vec{b}\\
&= \begin{bmatrix} \sqrt{29} & \frac{-6}{\sqrt{29}} \\ 0 & \frac{950}{\sqrt{27550}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{4}{\sqrt{29}} & \frac{3}{\sqrt{29}} & \frac{-2}{\sqrt{29}} \\ \frac{111}{\sqrt{27550}} & \frac{-98}{\sqrt{27550}} & \frac{75}{\sqrt{27550}} \end{bmatrix} \begin{bmatrix} 7 \\ 15 \\ -9 \end{bmatrix}\\
&= \frac{1}{25}\begin{bmatrix} 7 \\ -36 \end{bmatrix},
\end{aligned}
$$

which is of course the same solution for $\vec{x}$ that we obtained from the previous Section. We can then proceed to the complete solution to the Least Squares problem from this point as we did in that problem.□

We remark that this is certainly more work than computing the solution:

$$\vec{x} = (C^{\top}C)^{-1} \cdot C^{\top}\vec{b},$$

but if the QR-decomposition of C is handy, then our new method is certainly more convenient because it is much easier to invert an invertible upper-triangular matrix than the matrix $C^{\top}C$.

The QR-decomposition can also be used to simplify the computation of the Singular Value Decomposition (which will be discussed in Chapter 8) in the case when m is much larger than n. This improvement, though will be beyond the scope of this book. For certain kinds of matrices, this decomposition is also helpful in finding its eigenvalues. These issues are usually found in advanced courses in Numerical Linear Algebra.

7.9 Section Summary

If the $m \times n$ matrix $A = [\vec{w}_1\ \vec{w}_2 \ldots \vec{w}_n]$ has linearly independent columns, then we can find its ***QR-decomposition***, given by:

$$A = QR, \quad \text{or}$$

$$[\vec{w}_1\ \vec{w}_2 \ldots \vec{w}_n] = [\vec{u}_1\ \vec{u}_2 \ldots \vec{u}_n]\begin{bmatrix} \vec{w}_1 \circ \vec{u}_1 & \vec{w}_2 \circ \vec{u}_1 & \cdots & \cdots & \vec{w}_n \circ \vec{u}_1 \\ 0 & \vec{w}_2 \circ \vec{u}_2 & \cdots & \cdots & \vec{w}_n \circ \vec{u}_2 \\ 0 & 0 & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \ddots & \vec{w}_n \circ \vec{u}_{n-1} \\ 0 & 0 & 0 & 0 & \vec{w}_n \circ \vec{u}_n \end{bmatrix},$$

where the orthonormal set $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is obtained from the set of columns $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ via the ***Gram-Schmidt Algorithm***.

Suppose that $A\vec{x} = \vec{b}$ is an inconsistent system. Let C be the $m \times k$ matrix whose columns form a basis for $W = colspace(A)$ (thus if A already has linearly independent columns, then $C = A$). If $C = QR$ is the QR-decomposition of C, then the normal system associated to A, $C^{\top}C\vec{x} = C^{\top}\vec{b}$, has unique solution given by:

$$\vec{x} = R^{-1}Q^{\top}\vec{b}.$$

The general solutions to the Least Squares System can then be computed as in the previous Section.

7.9 Exercises

For Exercises (1) to (9): Find the QR-decomposition of the following matrices. Note that the matrix in Exercise 5 is an extension of the matrix in Exercise 3, so you may build on your work in Exercise 3 to solve Exercise 5. Similarly, Exercise 7 is an extension of Exercise 6, which is an extension of Exercise 4.

1. $\begin{bmatrix} 1 & 3 \\ 2 & 5 \end{bmatrix}$ 2. $\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}$ 3. $\begin{bmatrix} 1 & -1 \\ 0 & 2 \\ 1 & 0 \end{bmatrix}$

4. $\begin{bmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 1 \\ 3 & 0 \end{bmatrix}$ 5. $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 3 \end{bmatrix}$ 6. $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 1 \\ -1 & 1 & 1 \\ 3 & 0 & -1 \end{bmatrix}$

7. $\begin{bmatrix} 1 & -1 & 0 & 1 \\ 0 & 2 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 3 & 0 & -1 & 2 \end{bmatrix}$ 8. $\begin{bmatrix} 1 & -1 & 0 \\ 0 & 2 & 1 \\ -1 & 1 & 1 \\ 1 & 0 & -1 \\ 2 & -1 & -1 \end{bmatrix}$ 9. $\begin{bmatrix} 1 & -1 & 0 & 1 \\ 0 & 2 & 1 & 1 \\ -1 & 1 & 1 & -1 \\ 1 & 0 & -1 & 2 \\ 2 & -1 & -1 & 0 \end{bmatrix}$

For Exercises (10) to (14): Solve the Least Squares System in the corresponding Exercise using the QR-decomposition.

10. Exercise 1, Section 7.8
11. Exercise 2, Section 7.8
12. Exercise 3, Section 7.8
13. Exercise 4, Section 7.8
14. Exercise 5, Section 7.8

A Summary of Chapter 7

Let V be a vector space. An ***inner product*** on V is a ***bilinear form*** $\langle\,|\,\rangle$ on V, that is, a function that takes two vectors $\vec{u}$, $\vec{v} \in V$, and produces a ***scalar***, denoted $\langle\vec{u}|\vec{v}\rangle$, such that the following properties are satisfied by all vectors $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$:

1. ***The Symmetric Property:*** $\langle\vec{u}|\vec{v}\rangle = \langle\vec{v}|\vec{u}\rangle$;

2. ***The Homogenous Property:*** $\langle k \cdot \vec{u}|\vec{v}\rangle = k\langle\vec{u}|\vec{v}\rangle$;

3. ***The Additive Property:*** $\langle\vec{u}+\vec{v}|\vec{w}\rangle = \langle\vec{u}|\vec{w}\rangle + \langle\vec{v}|\vec{w}\rangle$;

4. ***The Positive Property:*** If $\vec{v} \neq \vec{\mathbf{0}}_V$, then $\langle\vec{v}|\vec{v}\rangle > 0$.

For any $\vec{v} \in V$, $\left\langle\vec{v}|\vec{\mathbf{0}}_V\right\rangle = \left\langle\vec{\mathbf{0}}_V|\vec{v}\right\rangle = 0$. In particular, $\left\langle\vec{\mathbf{0}}_V|\vec{\mathbf{0}}_V\right\rangle = 0$.

Let $\vec{v}$, $\vec{u} \in V$. Define the ***norm*** or the ***length*** of $\vec{v}$ by $\|\vec{v}\| = \sqrt{\langle\vec{v}|\vec{v}\rangle}$. In other words, $\|\vec{v}\|^2 = \langle\vec{v}|\vec{v}\rangle$.

In particular, we say that $\vec{v}$ is a ***unit vector*** if $\|\vec{v}\| = 1$.
The set of all unit vectors in V is called the ***unit sphere*** or unit circle of V.
We can also define the ***distance*** between two vectors by $d(\vec{u}, \vec{v}) = \|\vec{u} - \vec{v}\|$.

The Cauchy-Schwarz Inequality: Let V be an inner product space with respect to $\langle\,|\,\rangle$. Then, for all vectors $\vec{u}, \vec{v} \in V$: $|\langle\vec{u}|\vec{v}\rangle| \leq \|\vec{u}\| \cdot \|\vec{v}\|$.

If $\vec{u}$ and $\vec{v}$ are non-zero vectors in V, we define the angle between them as the angle θ such that: $\cos(\theta) = \langle\vec{u}|\vec{v}\rangle/(\|\vec{u}\|\,\|\vec{v}\|)$, where $0 \leq \theta \leq \pi$.

Furthermore, we will say that $\vec{u}$ is orthogonal to $\vec{v}$ ***if and only if*** $\langle \vec{u} | \vec{v} \rangle = 0$.
In particular, $\vec{\mathbf{0}}_V$ is orthogonal to all vectors in V.

Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be a set of vectors in an inner product space V.
We say that S is an ***orthonormal set*** if: $\langle \vec{v}_i | \vec{v}_j \rangle = 0$ if $i \neq j$, and $\langle \vec{v}_i | \vec{v}_i \rangle = 1$ for $i = 1..k$.

If we remove the condition that each member of S is a unit vector but insist that all of the vectors be non-zero, we call S an ***orthogonal set***.

The Gram-Schmidt Algorithm: Start with any basis $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ for V.

1. Let $\vec{v}_1 = \vec{w}_1$. If $dim(V) = 1$, go to Step 3, otherwise:
2. For $k = 1$ to $n - 1$, let:

$$\vec{v}_{k+1} = \vec{w}_{k+1} - \langle \vec{w}_{k+1} | \vec{v}_1 \rangle \cdot \frac{\vec{v}_1}{\|\vec{v}_1\|^2} - \langle \vec{w}_{k+1} | \vec{v}_2 \rangle \cdot \frac{\vec{v}_2}{\|\vec{v}_2\|^2} - \cdots - \langle \vec{w}_{k+1} | \vec{v}_k \rangle \cdot \frac{\vec{v}_k}{\|\vec{v}_k\|^2}.$$

3. Normalize $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ by dividing each vector by its length.

Let $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ be an orthonormal basis for an inner product space V.
Let $\vec{v}, \vec{w} \in V$. If $\langle \vec{v} \rangle_B = \langle v_1, v_2, \ldots, v_n \rangle$, and $\langle \vec{w} \rangle_B = \langle w_1, w_2, \ldots, w_n \rangle$, then:

1. $\langle \vec{v} | \vec{w} \rangle = \langle \vec{v} \rangle_B \circ \langle \vec{w} \rangle_B = v_1 w_1 + v_2 w_2 + \cdots + v_n w_n$.
2. $\|\vec{v}\| = \|\langle \vec{v} \rangle_B\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$.
3. $d(\vec{v}, \vec{w}) = \|\langle \vec{v} \rangle_B - \langle \vec{w} \rangle_B\| = \sqrt{(v_1 - w_1)^2 + (v_2 - w_2)^2 + \cdots + (v_n - w_n)^2}$.
4. $\cos(\theta) = \langle \vec{v} \rangle_B \circ \langle \vec{w} \rangle_B / (\|\langle \vec{v} \rangle_B\| \|\langle \vec{w} \rangle_B\|)$, where θ is the angle between $\vec{v}$ and $\vec{w}$ (nonzero vectors).

Let W be a subspace of an inner product space V. We define the ***orthogonal complement*** of W, which is also a subspace of V, by: $W^\perp = \left\{\vec{v} \in V | \langle \vec{v} | \vec{w} \rangle = 0 \text{ for all } \vec{w} \in W\right\}$.
Consequently, $W \cap W^\perp = \left\{\vec{\mathbf{0}}_V\right\}$, and if V is finite dimensional: $dim(W) + dim(W^\perp) = dim(V)$, and $(W^\perp)^\perp = W$.

Let W be a ***finite-dimensional subspace*** of an inner product space V. Then, any vector $\vec{v} \in V$ can be expressed ***uniquely*** as a sum $\vec{v} = \vec{w}_1 + \vec{w}_2$, where $\vec{w}_1 \in W$ and $\vec{w}_2 \in W^\perp$. We refer to this as an ***orthogonal decomposition*** with respect to W and $W^\perp$.

Moreover we can explicitly construct $\vec{w}_1$ and $\vec{w}_2$ as follows: If $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is any orthonormal basis for W, then: $\vec{w}_1 = c_1\vec{u}_1 + c_2\vec{u}_2 + \cdots + c_k\vec{u}_k$, and $\vec{w}_2 = \vec{v} - \vec{w}_1$, where $c_i = \langle \vec{v} | \vec{u}_i \rangle$ for $i = 1 \ldots n$.

We call $\vec{w}_1$ the ***orthogonal projection*** of $\vec{v}$ onto W, and $\vec{w}_2$ the orthogonal projection of $\vec{v}$ onto $W^\perp$. We write this as $\vec{w}_1 = proj_W(\vec{v})$ and $\vec{w}_2 = proj_{W^\perp}(\vec{v})$.

Let W be a subspace of $\mathbb{R}^n$, under the ordinary dot product. Then the function: $proj_W : \mathbb{R}^n \to \mathbb{R}^n$, given by: $proj_W(\vec{v}) = \vec{w}_1$, where $\vec{v} = \vec{w}_1 + \vec{w}_2$ is the orthogonal decomposition of $\vec{v}$, is a linear operator of $\mathbb{R}^n$, which we call ***the projection operator onto*** W.

Furthermore, if $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ is an orthonormal basis for W, and $U = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_k]$ is the $n \times k$ matrix with the vectors of B arranged in columns, then $[proj_W] = UU^\top$.

An $n \times n$ matrix Q is called ***orthogonal*** if $QQ^\top = Q^\top Q = I_n$. Equivalently, this means that Q is invertible, and $Q^{-1} = Q^\top$. The following are equivalent for an $n \times n$ matrix Q:

1. Q is an orthogonal matrix.
2. The columns of Q form an orthonormal set in $\mathbb{R}^n$ with respect to the dot product.

3. The rows of Q form an orthonormal set in $\mathbb{R}^n$ with respect to the dot product.

Analogously, let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear operator. Then, the following are equivalent:
1. $[T]$ is an orthogonal matrix.
2. T preserves the dot product — For all $\vec{u}$ and $\vec{v} \in \mathbb{R}^n$: $\vec{u} \circ \vec{v} = T(\vec{u}) \circ T(\vec{v})$.
3. T preserves length — For all $\vec{v} \in \mathbb{R}^n$: $\|\vec{v}\| = \|T(\vec{v})\|$.

Let Q and P be orthogonal matrices. Then:
1. $det(Q) = 1$ or -1.
2. Q^{-1} is also orthogonal.
3. PQ and QP are also orthogonal.

Let A be a symmetric matrix. Then all of the eigenvalues of A are real numbers. Furthermore, if λ_1 and λ_2 are two ***distinct*** eigenvalues with corresponding eigenvectors $\vec{v}_1$ and $\vec{v}_2$, then $\vec{v}_1$ is orthogonal to $\vec{v}_2$.

We can find an orthogonal matrix Q such that Q diagonalizes a symmetric matrix A, that is, $D = Q^\top AQ$, where D is a diagonal matrix with real entries.

Let A be an $m \times n$ matrix, and $\vec{b}$ an $m \times 1$ column matrix. If $\vec{b}$ is not a member of W, the columnspace of A, then we call the system $A\vec{x} = b_1 = proj_W\left(\vec{b}\right)$ the ***least squares system associated to*** $A\vec{x} = \vec{b}$.

This system is always consistent, and we call any solution $\vec{x}_1 \in \mathbb{R}^n$ to this system as a ***least squares solution*** or ***best approximation*** to the original (inconsistent) system $A\vec{x} = \vec{b}$.

Any such solution $\vec{x}_1$ has the property that if $\vec{x}$ is any other vector in $\mathbb{R}^n$ that is ***not*** a least squares solution, then $\left\|A\vec{x}_1 - \vec{b}\right\| < \left\|A\vec{x} - \vec{b}\right\|$.

Let C be the $m \times k$ matrix whose columns form a basis for $colspace(A)$. Then $C^\top C$ is an ***invertible*** matrix, and therefore the ***normal system associated to*** A, defined by $C^\top C\vec{x} = C^\top\vec{b}$, has ***exactly one solution***, namely $\vec{x} = (C^\top C)^{-1} \cdot C^\top\vec{b}$.

This is a solution to $C\vec{x} = \vec{b}_1 = proj_W\left(\vec{b}\right)$. By padding $\vec{x}$ with zeroes, if necessary, we obtain a solution $\vec{x}_1$ to the least squares system $A\vec{x} = \vec{b}_1$, and all the best approximation solutions are of the form: $\vec{x}_1 + \vec{x}_0$, where $\vec{x}_0 \in nullspace(A)$. Consequently, $[proj_W] = C \cdot (C^\top C)^{-1} \cdot C^\top$.

The ***common error*** of all our best approximation solutions is $\left\|\vec{b} - \vec{b}_1\right\| = \left\|\vec{b}_2\right\|$.

If the $m \times n$ matrix $A = [\vec{w}_1\ \vec{w}_2\ \ldots\ \vec{w}_n]$ has linearly independent columns, then we can find its ***QR-decomposition***, given by: $A = QR$, or

$$[\vec{w}_1\ \vec{w}_2\ \ldots\ \vec{w}_n] = [\vec{u}_1\ \vec{u}_2\ \ldots\ \vec{u}_n]\begin{bmatrix} \vec{w}_1 \circ \vec{u}_1 & \vec{w}_2 \circ \vec{u}_1 & \cdots & \cdots & \vec{w}_n \circ \vec{u}_1 \\ 0 & \vec{w}_2 \circ \vec{u}_2 & \cdots & \cdots & \vec{w}_n \circ \vec{u}_2 \\ 0 & 0 & \ddots & \cdots & \vdots \\ 0 & 0 & 0 & \ddots & \vec{w}_n \circ \vec{u}_{n-1} \\ 0 & 0 & 0 & 0 & \vec{w}_n \circ \vec{u}_n \end{bmatrix},$$

where the orthonormal set $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is obtained from the set of columns $\{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ via the Gram-Schmidt Algorithm.

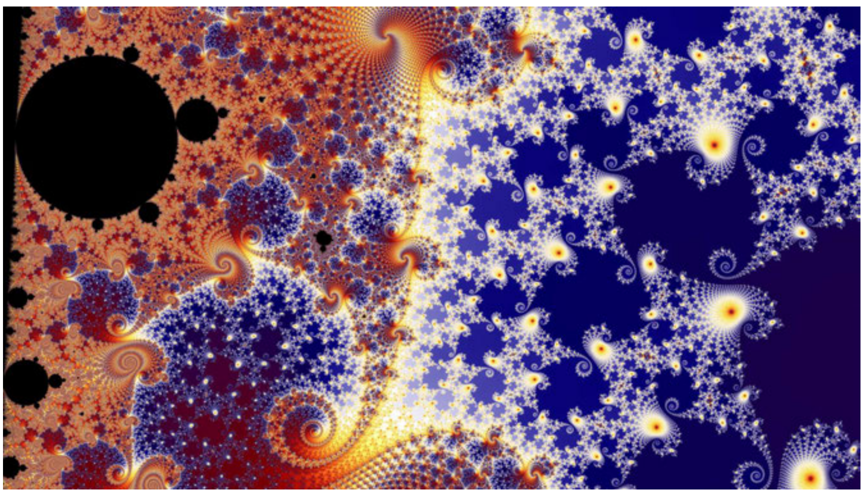
Suppose that $A\vec{x} = \vec{b}$ is an inconsistent system. Let C be the $m \times k$ matrix whose columns form a basis for $W = colspace(A)$ as above. If $C = QR$ is the QR-decomposition of C, then the normal system associated to A, $C^\top C\vec{x} = C^\top\vec{b}$, has unique solution given by $\vec{x} = R^{-1}Q^\top\vec{b}$.

Chapter 8

Imagine That:

Complex Spaces and The Spectral Theorems

In this Chapter, we will generalize the concept of real numbers to an abstract mathematical structure called a ***field***. The most important field we will work with is the field $\mathbb{C}$ of ***complex numbers***. These numbers have the form $a + bi$, where $i = \sqrt{-1}$ is the imaginary unit, and a and b are real numbers. We will review the arithmetic of complex numbers and their properties. Central in our analysis will be understanding the role of the ***complex conjugate:*** $\overline{a + bi} = a - bi$. Complex numbers can be used to construct fascinating mathematical objects called ***fractals***, such as the ***Mandelbrot set:***



From ***Mandelbrot Set*** by Jan Thor. Copyright (c) 2010 by Jan Thor. Reprinted by permission.

We will generalize the concept of Euclidean spaces over the set of real numbers to vector spaces over any field, and in particular, over the complex numbers. Thanks to the operation of complex conjugation, we will be able to construct a ***complex inner product*** on complex Euclidean spaces that have slightly different properties from the dot product. We will of course also generalize ***matrices*** and ***linear transformations*** over complex Euclidean spaces.

One of our main objectives is to prove the Spectral Theorem for Symmetric Matrices from Chapter 6, that states that all symmetric matrices have real eigenvalues, and furthermore, every symmetric matrix can be diagonalized using an orthogonal matrix. To do so, we need to introduce matrices that generalize the concept of symmetric, skew-symmetric and orthogonal matrices, namely, the ***Hermitian***, ***skew-Hermitian*** and ***unitary*** matrices. These three kinds of matrices fall into a general category called ***normal matrices***.

The ***Spectral Theorem for Normal Matrices*** tells us that these normal matrices are precisely the matrices that have very special diagonalizability properties. In particular, all $n \times n$ normal matrices possess a complete set of n linearly independent eigenvectors, and any two vectors from distinct eigenspaces are ***orthogonal*** to each other, under the complex inner product mentioned above. The Spectral Theorem for Symmetric Matrices is a particular consequence of this general Spectral Theorem.

8.1 The Field of Complex Numbers

In basic algebra, we learn that the square root of a positive number is another positive number, but the square root of a negative number is not a real number. To get around this, we construct the ***imaginary unit***, namely:

$$i = \sqrt{-1}.$$

This new object i has the magical property, therefore, that $i^2 = -1$. This means that it ***cannot*** be a real number, since the square of any real number cannot be negative. In other words, we are dealing with a ***new*** mathematical quantity, one that is definitely not a real number. From this, we can create the square root of any negative number, with the rule:

$$\sqrt{-a} = \sqrt{a} \bullet i, \text{ where } a > 0.$$

Thus:

$$\sqrt{-9} = \sqrt{9}\sqrt{-1} = 3i.$$

Notice that we instinctively used some kind of a product rule for radicals, but we need to be careful:

> ***Definition:*** The square root function can be extended to the negative numbers, with the property:
>
> $$\sqrt{a \bullet b} = \sqrt{a} \bullet \sqrt{b},$$
>
> if a and b are both non-negative, or exactly ***one*** of them is negative. In other words, the product rule is invalid if both a and b are negative. Likewise, if a and b are ***both positive***, then:
>
> $$\sqrt{\frac{-a}{b}} = \sqrt{\frac{a}{-b}} = \sqrt{-\frac{a}{b}} = \sqrt{\frac{a}{b}} \bullet i.$$

Let us see why this product rule makes sense: We know that $\sqrt{36} = 6$. If we allow the product rule to work when both factors are negative, we will have:

$$6 = \sqrt{36} = \sqrt{(-4)(-9)} = \sqrt{-4}\sqrt{-9} = (2i)(3i) = 6i^2 = 6(-1) = -6.$$

giving us the ***wrong*** sign. We also need to be extra careful with quotients, as you will see in the Exercises. Using the set of real numbers and the imaginary unit i, we can construct the set of all ***complex numbers:***

> ***Definition:*** The ***field of complex numbers*** is denoted by:
>
> $$\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}.$$
>
> We call a the ***real part***, and b the ***imaginary part***, respectively, of $a + bi$.
>
> A complex number with $b = 0$ is called a ***pure real number***, or simply a real number, whereas a complex number with $a = 0$ is called a ***pure imaginary number***.
>
> The real number 0 is extended to the ***complex zero:*** $0_{\mathbb{C}} = 0 + 0i$.
>
> We say that $a + bi$ and $c + di$ are ***equal if and only if*** $a = c$ and $b = d$.

Thus 7, $-8/3$, $\sqrt{2}$ and π are pure real numbers and $3i$, $\sqrt{5}\,i$ and $e^2 i$ are pure imaginary numbers.

Under this definition of equality, the only complex number that is both pure real and pure imaginary is $0_{\mathbb{C}}$,.because if $a + 0i = 0 + bi$, then $a = 0 = b$. We also write $1 \cdot i$ and $-1 \cdot i$ as i and $-i$, in the same way we write $-1 \cdot x$ as $-x$.

We can now extend the arithmetic operations to the set of complex numbers, as we see in intermediate algebra. The underlying idea in defining these operations as we will is to preserve the familiar basic properties of the arithmetic of real numbers, namely, the commutative, associative and distributive properties. Complex variables are often denoted by a letter such as z or w. We also include the ***complex conjugate*** $\bar{z}$ of z, pronounced "***z bar***":

Definition: Operations on Complex Numbers:

Let $z = a + bi$ and $w = c + di$ be two complex numbers. We define the operations of:

1. ***Addition:*** $$z + w = (a + bi) + (c + di) = (a + c) + (b + d)i.$$
2. ***Negation:*** $$-w = -(c + di) = -c - di.$$
3. ***Subtraction:*** $$z - w = z + (-w) = (a + bi) + (-c - di) = (a - c) + (b - d)i.$$
4. ***Multiplication:*** $$z \cdot w = (a + bi) \cdot (c + di) = ac + bci + adi + bdi^2 = (ac - bd) + (bc + ad)i.$$
5. ***Complex Conjugation:*** $$\bar{z} = \overline{a + bi} = a - bi.$$
6. ***Norm*** or ***Length:*** $$\|z\| = \sqrt{z \cdot \bar{z}} = \sqrt{(a + bi)(a - bi)} = \sqrt{a^2 + b^2}\,.$$
7. ***Division:*** If $w \neq 0_{\mathbb{C}}$, then: $$\frac{z}{w} = \frac{z \cdot \bar{w}}{w \cdot \bar{w}} = \frac{z \cdot \bar{w}}{\|w\|^2} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}i.$$

In particular, the property $i^2 = -1$ can be used to compute higher powers of i, thus:

$$i^3 = i^2 \cdot i = -1 \cdot i = -i,$$
$$i^4 = i^2 \cdot i^2 = -1 \cdot -1 = 1, \ldots,$$

and so on. Thus, the pattern repeats in groups of ***four***.

Examples: Let us simplify the following expressions:

$$(6 - 5i) + (-2 - 7i) = 4 - 12i,$$
$$(3 - i)(4 + 5i) = 12 - 4i + 15i - 5i^2 = 17 + 11i,$$
$$i^{523} = i^{520} \cdot i^3 = 1 \cdot (-i) = -i,$$
$$\|5 - 12i\| = \sqrt{5^2 + 12^2} = 13, \text{ and}$$

$$\frac{7+4i}{2-3i} = \frac{(7+4i)(2+3i)}{2^2+3^2}$$
$$= \frac{14+8i+21i+12i^2}{4+9} = \frac{2}{13} + \frac{29}{13}i. \square$$

Unfortunately, the concept of ***order*** does not extend to naturally to the set of complex numbers. However, it is possible to compare the ***norms*** of two complex numbers. Thus, $\|3-7i\| > \|2+5i\|$.

The Field Properties

The set of complex numbers and the operations of addition and multiplication in this set possess the same familiar properties of the arithmetic of real numbers:

Theorem — The Field Properties for the Set of Complex Numbers:
Let $z, w, u \in \mathbb{C}$. Then the following properties are true:

1. ***The Closure Property of Addition*** $z + w \in \mathbb{C}$.
2. ***The Closure Property of Multiplication*** $z \cdot w \in \mathbb{C}$.
3. ***The Commutative Property of Addition*** $z + w = w + z$.
4. ***The Commutative Property of Multiplication*** $z \cdot w = w \cdot z$.
5. ***The Associative Property of Addition*** $z + (w + u) = (z + w) + u$.
6. ***The Associative Property of Multiplication*** $z \cdot (w \cdot u) = (z \cdot w) \cdot u$.
7. ***The Distributive Property of Addition over Multiplication:*** $z \cdot (w + u) = (z \cdot w) + (z \cdot u)$.
8. ***The Existence of the Additive Identity:*** There exists $0_{\mathbb{C}} = 0 + 0i \in \mathbb{C}$ such that $z + 0_{\mathbb{C}} = z$.
9. ***The Existence of the Multiplicative Identity:*** There exists $1_{\mathbb{C}} = 1 + 0i \in \mathbb{C}$ such that $z \cdot 1_{\mathbb{C}} = z$.
10. ***The Existence of Additive Inverses:*** There exists $-z \in \mathbb{C}$, such that $z + (-z) = 0_{\mathbb{C}} = (-z) + z$.
11. ***The Existence of Multiplicative Inverses:*** If $z \neq 0_{\mathbb{C}}$, then there exists $z^{-1} \in \mathbb{C}$, such that $z \cdot z^{-1} = 1_{\mathbb{C}} = z^{-1} \cdot z$.

The proof of the 11 properties above all directly follow from the definitions of the basic operations made earlier. They should remind you of the 11 Field Axioms for the Real Numbers that we saw in Chapter Zero.

More generally, a ***field*** $(\mathbf{F},+,\cdot)$ is any non-empty set $\mathbf{F}$, together with an addition $+$ and multiplication $\cdot$ defined on pairs of members of $\mathbf{F}$, that satisfy the eleven properties above. These properties are naturally called ***The Field Axioms***.

Recall that when we discussed abstract vector spaces, the zero vector may not look at all like the number zero, so it is worthwhile to note that the additive and multiplicative identities mentioned above (which we denote $0_{\mathbf{F}}$ and $1_{\mathbf{F}}$) may not look at all like the number 0 or 1, respectively. Again, keep an open mind!

Aside from the field of complex numbers $\mathbb{C}$ and the field of real numbers $\mathbb{R}$, you should also be familiar with the field of ***rational numbers*** mentioned in Chapter Zero:

$$\mathbb{Q} = \{ a/b \,|\, a,b \in \mathbb{Z},\, b \neq 0 \},$$

all under the usual operations of addition and multiplication of fractions. This follows easily because the sum and product of two rational numbers is again a rational number, the negative of a rational number is also a rational number, and the reciprocal of a non-zero rational number is a rational number.

Let us next look at an example of a field that is just slightly irrational:

Example: Consider the set:

$$\mathbf{F} = \left\{ x \in \mathbb{R} \,|\, x = p + q\sqrt{2}\,, \text{ where } p,\, q \in \mathbb{Q} \right\}.$$

Some members of $\mathbf{F}$ are $3/4$, $\sqrt{2}/9$ and $5/4 - 7\sqrt{2}/3$. However, $\pi \notin \mathbf{F}$.

Notice the similarity between $\mathbf{F}$ and the set of complex numbers $a + bi = a + b\sqrt{-1}$, although this time p and q are restricted to be ***rational*** numbers. We define the addition and multiplication of members of $\mathbf{F}$ to be the usual ones for real numbers.

Addition is clearly closed, because:

$$\left(p + q\sqrt{2}\right) + \left(r + s\sqrt{2}\right) = (p + r) + (q + s)\sqrt{2}$$

by using the usual properties of arithmetic. Since $p + r$ and $q + s$ are again rational, addition is closed. Multiplication is likewise closed, because:

$$\begin{aligned}\left(p + q\sqrt{2}\right)\left(r + s\sqrt{2}\right) &= pr + ps\sqrt{2} + qr\sqrt{2} + 2qs \\ &= (pr + 2qs) + (ps + qr)\sqrt{2},\end{aligned}$$

again has the required form. Addition and multiplication enjoy the commutative, associative and distributive properties, because they are ***inherited*** from these operations as they apply to ***all*** real numbers. The identity elements are:

$$0_{\mathbf{F}} = 0 + 0 \cdot \sqrt{2} \quad \text{and} \quad 1_{\mathbf{F}} = 1 + 0 \cdot \sqrt{2}$$

which is very similar to how we defined them for $\mathbb{C}$. The negative of a member of $\mathbf{F}$ clearly has the same form, so it remains to show that so does its reciprocal. To do so, we ***rationalize the denominator*** as we do in algebra:

$$\begin{aligned}\frac{1}{p+q\sqrt{2}} &= \frac{1}{p+q\sqrt{2}} \cdot \frac{p-q\sqrt{2}}{p-q\sqrt{2}} \\ &= \frac{p-q\sqrt{2}}{p^2-2q^2} \\ &= \frac{p}{p^2-2q^2} - \frac{q}{p^2-2q^2}\sqrt{2}\end{aligned}$$

Since p and q are rational numbers, the two fractions above will also simplify into rational numbers, and thus the reciprocal is again a member of $\mathbf{F}$. We note that the denominator of the two fractions above is ***not zero*** because otherwise, we would get $p^2 = 2q^2$, in other words:

$$\sqrt{2} = \pm\frac{p}{q}.$$

However, we showed in the Exercises of Chapter Zero that $\sqrt{2}$ is ***irrational***. This field is known as $\mathbb{Q}\left(\sqrt{2}\right)$, pronounced "$\mathbb{Q}$ ***adjoined with*** $\sqrt{2}$" and it is the ***smallest field*** that contains the rational numbers as well as $\sqrt{2}$. We also note that you can view the members of $\mathbb{Q}\left(\sqrt{2}\right)$ as ***linear combinations*** of the numbers 1 and $\sqrt{2}$ with ***coefficients*** from $\mathbb{Q}$.□

Properties of the Conjugate

Notice that the complex conjugate is not mentioned at all in the 11 field axioms above. We list below the properties of the complex conjugate separately, since the field axioms do not require the existence of such an operation:

Theorem — Properties of the Complex Conjugate:
Let $z, w, u \in \mathbb{C}$. Then the following properties are true:

1. ***The Double Conjugate Property:*** $\quad \bar{\bar{z}} = z.$
2. ***Additivity:*** $\quad \overline{z+w} = \bar{z} + \bar{w}.$
3. ***Multiplicativity:*** $\quad \overline{z \bullet w} = \bar{z} \bullet \bar{w}.$
4. ***The Test for Pure Real Numbers:*** $\quad z = \bar{z}$ ***if and only if*** z is a pure real number.
5. ***The Test for Pure Imaginary Numbers:*** $\quad z = -\bar{z}$ ***if and only if*** z is a pure imaginary number.
6. ***The Positivity of Multiplication by the Conjugate:*** $\quad z \bullet \bar{z} \geq 0$, and $z \bullet \bar{z} > 0$ ***if and only if*** $z \neq 0_{\mathbb{C}}$.

Again, the proofs of these properties are fairly straightforward and follow directly from the definitions (as long as one follows the order of operations that are implied in each equation). Notice also that the positivity property above is very similar to the positivity property of inner products that we saw in Chapter 5.

$\mathbb{C}$ *as a 2-Dimensional Real Vector Space*

Notice that we can write any complex number in the form $a + bi$, where a and b are real numbers. But we can also think of this as a ***linear combination*** of the numbers 1 and i, with real coefficients a and b. In other words:

$$\mathbb{C} = Span(\{1, i\}).$$

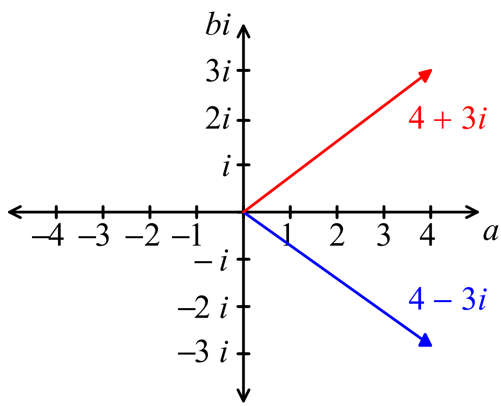
Furthermore, using our definition of equality:

$$a + bi = 0 + 0i \ \textit{\textbf{if and only if}} \ a = 0 \ \textit{\textbf{and}} \ b = 0.$$

In other words, the set $\{1, i\}$ is ***linearly independent***! Thus we have proven:

> ***Theorem:*** The set of complex numbers $\mathbb{C}$ is a 2-dimensional ***vector space*** over the set of real numbers, with ***basis*** $\{1, i\}$.

Continuing with this line of thinking, we can thus conclude that $\mathbb{C}$ is ***isomorphic*** to $\mathbb{R}^2$. Indeed, it is easy to check that $T : \mathbb{C} \to \mathbb{R}^2$, where$T(a + bi) = \langle a, b\rangle$, is an isomorphism. But this is very convenient because we can now visualize complex numbers on the Cartesian plane, as we do in the following diagram, where we plot a complex number and its conjugate:



$z = 4 + 3i$ and its conjugate $\bar{z} = 4 - 3i$

Traditionally, we use the x-axis to symbolize the real part, hence it is also called the ***real axis***, and similarly the y-axis is also called the ***imaginary axis***. We showed above the complex number $4 + 3i$ and its complex conjugate $4 - 3i$. Notice that they are ***reflections*** of each other across the real axis. Also, note that:

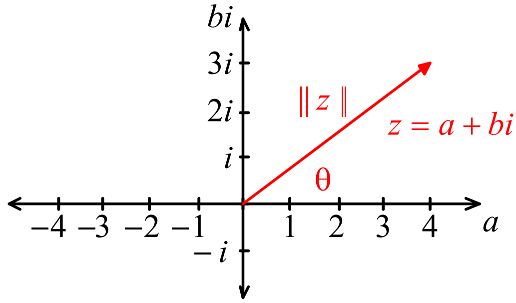
$$\|4 + 3i\| = 5 = \|\langle 4, 3\rangle\|,$$

where the norm on the left is that of complex numbers and the norm on the right is that of $\mathbb{R}^2$. Clearly, in general, the norm of a complex number corresponds to the length of the vector that represents it in $\mathbb{R}^2$, under the usual dot product.

With the concept of length, we can express ***non-zero*** complex numbers in ***polar form:***

$$z = \|z\|(\cos(\theta) + \sin(\theta) \cdot i)$$

where, for convention, $\theta \in [0, 2\pi)$ is the angle made by z with respect to the positive real axis, which is also called the ***argument*** of z, or $arg(z)$.



The Norm $\|z\|$ and Argument $\theta = arg(z)$ of a Complex Number z

Notice what happens when we square the expression in the parentheses:

$$\begin{aligned}(\cos(\theta) + \sin(\theta) \cdot i)^2 &= \cos^2(\theta) + 2\sin(\theta)\cos(\theta) \cdot i + \sin^2(\theta) \cdot i^2 \\ &= \cos^2(\theta) - \sin^2(\theta) + 2\sin(\theta)\cos(\theta) \cdot i \\ &= \cos(2\theta) + \sin(2\theta) \cdot i.\end{aligned}$$

Proceeding by induction and using the familiar addition formulas from trigonometry, we can prove:

Theorem — De Moivre's Theorem: For any real number θ:

$$(\cos(\theta) + \sin(\theta) \cdot i)^n = \cos(n\theta) + \sin(n\theta) \cdot i.$$

Consequently, we can find powers of a complex number in polar form:

$$z^n = \|z\|^n(\cos(n\theta) + \sin(n\theta) \cdot i).$$

The polar form is also very useful to compute square-roots or general *nth* roots. We state the following Theorem and leave its proof as an Exercise:

Theorem: Let $z = \|z\|(\cos(\theta) + \sin(\theta) \cdot i)$ be a complex number expressed in polar form. Then all the solutions w of the equation:

$$w^n = z$$

have the form:

$$w = \|z\|^{1/n}(\cos(\alpha) + \sin(\alpha) \cdot i), \text{ where}$$

$$\alpha = \frac{\theta + 2\pi k}{n}, \text{ and } k = 0, 1, 2, \ldots, n-1.$$

In particular, the solutions w of $w^2 = z$ are:

$$w = \pm\sqrt{\|z\|}\,(\cos(\alpha/2) + \sin(\alpha/2) \cdot i).$$

The solution with $k = 0$ is also called the ***principal nth root of*** z. We will see an example of how this Theorem can be used in the next sub-section.

The Fundamental Theorem of Algebra

The algorithm in the previous Theorem basically tells us that the complex polynomial:

$$p(w) = w^n - z$$

has exactly n roots. Notice that this is a polynomial in w of degree exactly $n,$ and the formula for the solutions indicate that these n roots are all ***distinct***.

We can generalize this phenomenon to any polynomial of degree n with complex coefficients, provided that we take into account the multiplicity of the roots:

Theorem — The Fundamental Theorem of Algebra:
Let $c_0, c_1, \ldots, c_n$ be fixed complex numbers, with $c_n \neq 0_C$. Then the polynomial:

$$p(w) = c_n w^n + \cdots + c_1 w + c_0$$

can be factored completely as:

$$p(w) = c_n(w - w_1)^{n_1} \cdot (w - w_2)^{n_2} \cdot \cdots \cdot (w - w_j)^{n_j},$$

where $w_1, w_2, \ldots, w_j$ are the ***distinct complex roots*** of $p(w),$ and $n_1 + n_2 + \cdots + n_j = n.$ We call n_i the ***multiplicity*** of the root w_i.

It took many attempts before a complete and correct proof of the Fundamental Theorem of Algebra was produced. It is often attributed to ***Carl Friedrich Gauss***, but certainly many mathematicians worked on this Theorem. It has several known proofs, and any two of them could use completely different strategies.

Unfortunately, the Fundamental Theorem of Algebra does not tell us ***how*** to find the roots of a complex polynomial, only that they ***exist***, and their multiplicities add up to the degree of the polynomial. In fact, a field of mathematics called ***Galois Theory*** tells us that we can only find a general formula to solve a ***linear***, ***quadratic*** or a ***cubic*** equation. In other words, there is ***no general solution*** to find the roots of a complex polynomial of degree 4 or higher, which uses only ***radicals***, addition, subtraction, multiplication and division.

However, we know how to obtain the quadratic formula by the method of ***completing the square***. Every step in this derivation is still valid in the field of complex numbers (but not all fields), so the quadratic formula can still be used to solve quadratic equations with complex coefficients.

Example: Let us find the solutions of the quadratic equation:

$$9z^2 + (12 - 30i)z - 21 - 16i = 0.$$

The discriminant of this quadratic is:

$$\begin{aligned} b^2 - 4ac &= (12 - 30i)^2 - 4(9)(-21 - 16i) \\ &= -144i \\ &= 144(-i) \\ &= 144\left(\cos\left(\frac{3\pi}{2}\right) + \sin\left(\frac{3\pi}{2}\right) \cdot i\right). \end{aligned}$$

Now we find the principal square root of the discriminant (with $n = 2$ and $k = 0$ in our formula from the last sub-section):

$$\begin{aligned}\sqrt{b^2-4ac} &= 12\left(\cos\left(\frac{3\pi}{4}\right)+\sin\left(\frac{3\pi}{4}\right)\cdot i\right)\\ &= 12\left(-\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{2}}i\right)\\ &= -6\sqrt{2}+6\sqrt{2}\,i.\end{aligned}$$

Finally, we can insert this into the quadratic formula to obtain two solutions:

$$\begin{aligned}z &= \frac{-b\pm\sqrt{b^2-4ac}}{2a}\\ &= \frac{-12+30i\pm\left(-6\sqrt{2}+6\sqrt{2}\,i\right)}{2\cdot 9}\\ &= \frac{-\sqrt{2}-2}{3}+\frac{5+\sqrt{2}}{3}i \quad\text{or}\quad \frac{\sqrt{2}-2}{3}+\frac{5-\sqrt{2}}{3}i.\ \square\end{aligned}$$

8.1 Section Summary

The set of all complex numbers:

$$\mathbb{C} = \{a+bi\,|\,a,b\in\mathbb{R}\}$$

is constructed using the ***imaginary unit*** $i=\sqrt{-1}$. We can add, negate, subtract, and multiply two complex numbers, and divide a complex numbers by a non-zero complex number. We can also find the ***complex conjugate*** and ***norm*** of a complex number. A complex number $a+bi$ is ***pure real*** if $b=0$, and ***pure imaginary*** if $a=0$.

The set of complex numbers is an example of a ***field***, a non-empty set upon which we define an addition and a multiplication, such that special properties are satisfied. A field is closed under these two operations, and obey the commutative and associative properties for both. Addition distributes over multiplication. A field also possesses two special elements, $0_{\mathbf{F}}$ and $1_{\mathbf{F}}$, which are ***additive*** and ***multiplicative identities*** respectively. Every member of a field has a ***negative***, and every non-zero member has a ***reciprocal***.

The complex conjugate also satisfies special properties. It can be used to test when a complex number is pure real or pure imaginary.

De Moivre's Theorem: For any real number θ:

$$(\cos(\theta)+\sin(\theta)\cdot i)^n = \cos(n\theta)+\sin(n\theta)\cdot i.$$

Consequently, we can find powers of a complex number in polar form:

$$z^n = \|z\|^n(\cos(n\theta)+\sin(n\theta)\cdot i).$$

Let $z=\|z\|(\cos(\theta)+\sin(\theta)\cdot i)$ be a complex number expressed in polar form. Then all the solutions of the equation $w^n=z$ have the form:

$$\begin{aligned}w &= \|z\|^{1/n}(\cos(\alpha)+\sin(\alpha)\cdot i),\ \text{ where}\\ \alpha &= \frac{\theta+2\pi k}{n},\ \text{ and } k=0,\ 1,\ 2,\ldots,\ n-1.\end{aligned}$$

Every polynomial $p(w)$ with degree n and complex coefficients has exactly n complex roots counting up to multiplicities.

The quadratic formula can still be used to solve quadratic equations with complex coefficients.

8.1 Exercises

For Exercises (1) to (5): Simplify the following expressions:

1. $(5 - 3i)(2 + 7i)$
2. $\dfrac{4 + 7i}{3 - 2i}$
3. $\|7 - 24i\|$
4. i^{755}
5. i^{-327}

For Exercises (6) to (8): Use De Moivre's Theorem to compute the following powers, and write the final answer in the standard form $a + bi$.

6. $\left(\sqrt{3} - i\right)^{12}$
7. $(-3 - 3i)^{17}$
8. $(3 - 4i)^8$. Hint: use the formulas for $\cos(2\theta)$ and $\sin(2\theta)$ ***three*** times.
9. Prove de Moivre's Theorem by using Mathematical Induction on n.
10. ***Negative Powers:*** Prove that:

$$\frac{1}{\cos(\theta) + \sin(\theta) \cdot i} = \cos(-\theta) + \sin(-\theta) \cdot i = \cos(\theta) - \sin(\theta) \cdot i,$$

and then use induction to prove that in general:

$$(\cos(\theta) + \sin(\theta) \cdot i)^{-n} = \cos(n\theta) - \sin(n\theta) \cdot i,$$

where n is a positive integer. Together with de Moivre's Theorem, this proves that for any non-zero complex number z and any integer m:

$$z^m = \|z\|^m(\cos(m\theta) + \sin(m\theta) \cdot i)$$

11. Use the previous exercise to simplify $\left(\sqrt{3} + i\right)^{-7}$.
12. What is wrong with the following argument?

$$\sqrt{\frac{-4}{9}} = \frac{\sqrt{-4}}{\sqrt{9}} = \frac{2i}{3}, \text{ but}$$

$$\sqrt{\frac{4}{-9}} = \frac{\sqrt{4}}{\sqrt{-9}} = \frac{2}{3i} = \frac{2}{3i} \cdot \frac{i}{i} = \frac{2i}{-3} = -\frac{2i}{3}.$$

$$\text{Thus } \frac{2i}{3} = -\frac{2i}{3}.$$

For Exercises (13) to (16): Solve the following equations:

13. $z^2 = -8 - 8\sqrt{3}\,i$
14. $z^3 = 4\sqrt{2}(-1 + i)$
15. $z^4 + 81 = 0$
16. $z^2 = -7 - 24i$. Hint: Use the half-angle formulas from trigonometry:

$$\cos(\alpha/2) = \pm\sqrt{\frac{1 + \cos(\alpha)}{2}} \quad \textit{and} \quad \sin(\alpha/2) = \pm\sqrt{\frac{1 - \cos(\alpha)}{2}}$$

where the choice of sign depends on the location of $\alpha/2$ on the unit circle.

For Exercises (17) to (19): Use the quadratic formula to solve the following equations.

17. $2z^2 + z - iz + 9 + 19i = 0$
18. $6z^2 - z - 16iz = 23 - 14i$
19. $z^2 + 7iz^2 + 7z = 11iz + 10i$
20. Let $z = \|z\|(\cos(\theta) + \sin(\theta) \cdot i)$ be a ***fixed*** complex number expressed in polar form. We will assume for simplicity that $\theta \in [0, 2\pi)$. Our goal in this Exercise is to show that all the solutions w of the equation $w^n = z$ have the form:
$$w = \|z\|^{1/n}(\cos(\alpha) + \sin(\alpha) \cdot i), \text{ where}$$
$$\alpha = \frac{\theta + 2\pi k}{n}, \text{ and } k = 0,\ 1,\ 2, \ldots,\ n-1.$$
    a. Consider the polynomial $p(w) = w^n - z$. Explain why $w^n = z$ ***if and only if*** $p(w) = 0$.
    b. What does the Fundamental Theorem of Algebra say about the number of roots of $p(w)$?
    c. Explain why for every $k = 0,\ 1,\ 2, \ldots,\ n-1$, the angle $\dfrac{\theta + 2\pi k}{n}$ is also in the interval $[0, 2\pi)$.
    d. Explain why each k yields a unique value for α, and thus, there are n unique values for w in the formula above.
    e. Use De Moivre's Theorem to show that $w^n = z$ for each of the w in the formula above.
    f. Use the Fundamental Theorem of Algebra to explain why these must be the only solutions to the equation $w^n = z$.
21. In particular, the solutions w of $w^2 = z$ are:
$$w = \pm\sqrt{\|z\|}\,(\cos(\alpha/2) + \sin(\alpha/2) \cdot i).$$
22. Let $p(w)$ be a polynomial in a complex variable w with complex coefficients, say:
$$p(w) = c_n w^n + \cdots + c_1 w + c_0.$$
Denote by $\overline{p}(w)$ the polynomial: $\overline{p}(w) = \overline{c_n} w^n + \cdots + \overline{c_1} w + \overline{c_0}$, that is, it is the polynomial in w whose coefficients are the conjugates of those of p. Prove that: $\overline{p}(\overline{w}) = \overline{p(w)}$.
23. Use the previous Exercise to prove that if $p(w)$ is a polynomial with ***real coefficients***, and z is a complex root of p, i.e. $p(z) = 0$, then $\bar{z}$ is also a complex root. This is the familiar Theorem from algebra that says that the imaginary roots of a polynomial with real coefficients come in complex conjugate ***pairs***.
24. Use De Moivre's Theorem to show that the complex numbers w_i, as defined in the Theorem to find the n^{th}-roots of z, all satisfy the equation: $w^n = z$.
25. Prove that the set of all rational functions:
$$\mathbf{F} = \left\{ \frac{p(x)}{q(x)} \,|\, p(x) \text{ and } q(x) \text{ are polynomials with real coefficients, } q(x) \neq 0 \right\}$$
form a field under the natural addition and multiplication:
$$\frac{p(x)}{q(x)} \oplus \frac{r(x)}{s(x)} = \frac{p(x)s(x) + r(x)q(x)}{q(x)s(x)}, \quad \text{and}$$
$$\frac{p(x)}{q(x)} \otimes \frac{r(x)}{s(x)} = \frac{p(x)r(x)}{q(x)s(x)}$$
Notice that this field is analogous to the field of rational numbers $\mathbb{Q}$.

26. Explain why the set of integers $\mathbb{Z}$ is ***not*** a field. What are missing?

27. Prove that $\sqrt{3}$ is not a member of $\mathbb{Q}\left(\sqrt{2}\right)$. Hint: $\sqrt{2}$ is irrational.

28. ***The Field of Oz:*** Recall that in Chapter 3, we saw a very unusual vector space $\mathbb{R}^+$, the set of positive real numbers. We defined ***addition*** there by:

$$x \oplus y = x \cdot y \text{ (ordinary multiplication of positive numbers).}$$

We will not need scalar multiplication. However, suppose we define ***multiplication*** of two positive numbers by:

$$x \odot y = x^{\ln(y)} \quad (x \text{ raised to the power of } \ln(y)).$$

For example: $5 \odot e^2 = 5^{\ln\left(e^2\right)} = 5^2 = 25$.

Prove that $\mathbb{R}^+$ under this addition and multiplication satisfy all the eleven field axioms. As part of your proof, you need to identify the additive and multiplicative identities, the additive inverse of a positive number, and the reciprocal of a positive number, under these two operations. As in Chapter 3, keep an open mind!

For Exercises (29) to (35): Using only the 11 Field Axioms, prove the following properties of ***any*** field $\mathbf{F}$:

29. ***The Uniqueness of the Additive Identity*** $0_{\mathbf{F}}$: If $z \in \mathbf{F}$ such that $w + z = w$ for ***all*** $w \in \mathbf{F}$, then $z = 0_{\mathbf{F}}$. Hint: simplify $0_{\mathbf{F}} + z$ in two different ways, and explain each simplification.

30. ***The Uniqueness of the Multiplicative Identity*** $1_{\mathbf{F}}$: If $u \in \mathbf{F}$ such that $w \cdot u = w$ for all $w \in \mathbf{F}$, then $u = 1_{\mathbf{F}}$.

31. ***The Multiplicative Property of*** $0_{\mathbf{F}}$: For all $u \in \mathbf{F}$: $0_{\mathbf{F}} \cdot u = 0_{\mathbf{F}}$.
Hint: first explain why $0_{\mathbf{F}} + 0_{\mathbf{F}} = 0_{\mathbf{F}}$.

32. ***The Cancellation Property for Addition:*** For all $u, w, z \in \mathbf{F}$:

$$\text{If } z + u = w + u, \text{ then } z = w.$$

33. ***The Cancellation Property for Multiplication:*** For all $u, w, z \in \mathbf{F}$:

$$\text{If } z \cdot u = w \cdot u, \text{ then } z = w, \text{ provided } u \neq 0_{\mathbf{F}}.$$

34. ***The Zero-Factors Theorem:*** For all $u, w, z \in \mathbf{F}$:

$$w \cdot z = 0_{\mathbf{F}} \textit{ if and only if } \text{either } w = 0_{\mathbf{F}} \text{ or } z = 0_{\mathbf{F}}.$$

35. ***The Absence of Zero Divisors:*** If ***neither*** w nor z is $0_{\mathbf{F}}$, then $w \cdot z \neq 0_{\mathbf{F}}$. Hint: how is this related to the previous Exercise?

Mini-Project: Finite Fields.

All of the examples of fields that we have seen so far — the set of rational numbers $\mathbb{Q}$, the set of real numbers $\mathbb{R}$, the set of complex numbers $\mathbb{C}$, and so on, all have an ***infinite*** number of elements. We introduce now a field with only ***five*** elements, and thus is an example of a ***finite field:***

Let $\mathbf{F} = \{0, 1, 2, 3, 4\}$. Define an addition and multiplication on $\mathbf{F}$ via:

$$a \oplus b = (a + b) \bmod 5, \quad \text{and}$$

$$a \otimes b = (a \cdot b) \bmod 5,$$

where the phrase "mod 5" means that we find the ***remainder*** of the result after it is divided by 5. For example:

$$2 \oplus 4 = 6 \bmod 5 = 1, \quad \text{and} \quad 3 \otimes 4 = 12 \bmod 5 = 2.$$

1. Complete an "addition table" for **F**. We have entered $2 \oplus 4 = 1$ below.

$\oplus$	0	1	2	3	4
0					
1					
2					1
3					
4					

2. Convince yourself that 0 and 1 are indeed the natural additive and multiplicative identities.
3. Give the "negative" of 0, 1, 2, 3 and 4.
4. Construct a "multiplication table" for **F**.
5. Guessing game: Stare at the two tables you obtained. What word would elegantly describe them? What property accounts for this?
6. Notice that the number 1 appears on each row of your multiplication table except on the row containing 0 (where all the entries are 0, of course). Give the "reciprocal" of 1, 2, 3 and 4.

   Remarks: The fact that all 11 field axioms are indeed satisfied follow from the inherited properties from "modular arithmetic," as the processes above are called. The set **F** as defined above is known as $\mathbb{Z}$ ***mod 5***, and is written as $\mathbb{Z}/(\mathbf{5})$, or $\mathbb{Z}/\mathbf{5}\mathbb{Z}$, or $\mathbf{F}_5$, although sometimes it is written improperly as $\mathbb{Z}_5$.

   The reason for the notation $\mathbb{Z}/(\mathbf{5})$ has to do with the idea of ***cosets*** that we saw in Chapter 4, but we will not go into it.

   In general, if p is a ***prime number***, then the set:

$$\mathbb{Z}/(\boldsymbol{p}) = \mathbb{Z}/\boldsymbol{p}\mathbb{Z} = \mathbf{F}_{\boldsymbol{p}} = \{0, 1, 2, 3, \ldots, p-1\}$$

   under addition and multiplication mod p, form a field consisting of exactly p elements. The only tricky property to prove is that of the existence of reciprocals.

7. Repeat Exercises 1 to 6 above for $\boldsymbol{Z}/(\mathbf{7}) = \{0, 1, 2, 3, 4, 5, 6\}$. This time, of course, the addition and multiplication are modulo 7, and thus, for example: $4 + 6 = 10 \bmod 7 = 3$. In Exercise 6, you should give the reciprocals of 1 through 6.
8. Show that the set $\boldsymbol{Z}/(\mathbf{6}) = \{0, 1, 2, 3, 4, 5\}$ under addition and multiplication modulo 6 does ***not*** satisfy all the field axioms. Hint: look at the multiplication table.

   As a final note, it is well known that all fields that contain only a ***finite*** number of elements must have exactly p^n elements, for some prime number p. Conversely, it is also known that there exists a field with exactly p^n elements for every prime number p and every integer power n, and any two such fields are ***isomorphic*** to each other (in some natural sense). Thus, there is a field with $27 = 3^3$ elements, but there cannot be a field with 26 or 28 elements, since these are not pure prime powers. These finite fields are denoted $\mathbf{F}_{p^n}$.

   More precisely, within each field $\mathbf{F}_{p^k}$, there is ***exactly one*** subfield isomorphic to $\mathbf{F}_{p^{k-1}}$, for every $k > 1$. Thus, there exists a unique descending chain of ***subfields*** within each $\mathbf{F}_{p^n}$:

$$\mathbf{F}_{p^n} \supset \mathbf{F}_{p^{n-1}} \supset \cdots \supset \mathbf{F}_{p^2} \supset \mathbf{F}_p.$$

8.2 Complex Vector Spaces

We can naturally generalize the definition that we saw in Chapter 3 of a vector space over $\mathbb{R}$ to a vector space over an arbitrary field $\mathbf{F}$.

Definition — The Axioms of a Vector Space over a Field:
Let $(\mathbf{F}, +, \cdot)$ be any field. A ***vector space*** $(V, \oplus, \odot)$ ***over*** $\mathbf{F}$ is a non-empty set V, along with two operations: $\oplus$ (vector addition), and $\odot$ (scalar multiplication), that satisfy: for all $\vec{u}$, $\vec{v}$ and $\vec{w} \in V$ and all scalars $r, s \in \mathbf{F}$, we have:

1. ***The Closure Property of Vector Addition:*** $\quad \vec{u} \oplus \vec{v} \in V$
2. ***The Closure Property of Scalar Multiplication:*** $\quad r \odot \vec{u} \in V$
3. ***The Commutative Property of Vector Addition:*** $\quad \vec{u} \oplus \vec{v} = \vec{v} \oplus \vec{u}$
4. ***The Associative Property of Vector Addition:*** $\quad (\vec{u} \oplus \vec{v}) \oplus \vec{w} = \vec{u} \oplus (\vec{v} \oplus \vec{w})$
5. ***The Existence of a Zero Vector:*** There exists $\vec{\mathbf{0}}_V \in V$, such that: $\vec{\mathbf{0}}_V \oplus \vec{v} = \vec{v} = \vec{v} \oplus \vec{\mathbf{0}}_V$
6. ***The Existence of Additive Inverses:*** There exists $-\vec{v} \in V$ such that: $\vec{v} \oplus (-\vec{v}) = \vec{\mathbf{0}}_V = (-\vec{v}) \oplus \vec{v}$
7. ***The Distributive Property of Ordinary Addition over Scalar Multiplication:*** $\quad (r+s) \odot \vec{v} = (r \odot \vec{v}) \oplus (s \odot \vec{v})$
8. ***The Distributive Property of Vector Addition over Scalar Multiplication:*** $\quad r \odot (\vec{u} \oplus \vec{v}) = (r \odot \vec{u}) \oplus (r \odot \vec{v})$
9. ***The Associative Property of Scalar Multiplication:*** $\quad r \odot (s \odot \vec{v}) = s \odot (r \odot \vec{v}) = (rs) \odot \vec{v}$
10. ***The Unitary Property of Scalar Multiplication:*** $\quad 1_{\mathbf{F}} \odot \vec{v} = \vec{v}$

We constructed the Euclidean n-space $\mathbb{R}^n$ by constructing n-tuples of real numbers. We can analogously construct the ***Complex Euclidean n-space***, denoted $\mathbb{C}^n$, by constructing n-tuples of complex numbers:

$$\mathbb{C}^n = \{\langle z_1, z_2, \ldots, z_n \rangle \mid z_i \in \mathbb{C}\}.$$

As before, we will indicate that a variable represents a vector by putting an arrow above the variable. Thus, for example, $\vec{z} = \langle -5i, 2+i, 7\rangle \in \mathbb{C}^3$. To avoid confusion, we will refer to $\mathbb{R}^n$ as ***Real Euclidean n-space***.

In the same way that we refer to real numbers as scalars, we will also refer to complex numbers as ***complex scalars***. It is easy to see that $\mathbb{C}^n$ is a vector space over the field of complex scalars under the natural operations of:

$$\textit{Vector Addition:} \quad \langle z_1, z_2, \ldots, z_n \rangle \oplus \langle w_1, w_2, \ldots, w_n \rangle = \langle z_1 + w_1, z_2 + w_2, \ldots, z_n + w_n \rangle, \quad \text{and}$$

$$\textit{Complex Scalar Multiplication:} \quad u \odot \langle z_1, z_2, \ldots, z_n \rangle = \langle u \cdot z_1, u \cdot z_2, \ldots, u \cdot z_n \rangle.$$

Under these two operations, we can easily verify that $\mathbb{C}^n$ satisfies the Ten Axioms of a Vector Space over the field of complex numbers. When verifying these axioms, our scalars will now be complex scalars. Obviously, the ***zero vector*** is:

$$\vec{0}_{\mathbb{C}^n} = \langle 0_{\mathbb{C}}, 0_{\mathbb{C}}, \ldots, 0_{\mathbb{C}} \rangle.$$

and the ***negative of a vector*** is:

$$-\vec{z} = -\langle z_1, z_2, \ldots, z_n \rangle = \langle -z_1, -z_2, \ldots, -z_n \rangle.$$

Aside from the two basic operations, though, it is also useful to define the ***complex conjugate of a complex vector***:

$$\overline{\vec{z}} = \overline{\langle z_1, z_2, \ldots, z_n \rangle} = \langle \overline{z_1}, \overline{z_2}, \ldots, \overline{z_n} \rangle.$$

We will now proceed with the gigantic but completely natural task of generalizing and extending all of the Examples, Definitions and Theorems that we have seen in the previous Chapters to vector spaces over arbitrary fields, but in particular, to vector spaces over the complex field.

Examples of Complex Vector Spaces

Some of the examples that we saw in Chapter 3 can be naturally extended to create complex vector spaces. The set of all ***polynomials:***

$$p(z) = c_n z^n + c_{n-1} z^{n-1} + \cdots + c_1 z + c_0$$

of degree n or less in the complex variable z and with complex coefficients c_i, under the natural operations of polynomial addition and scalar multiplication by a complex number, is an example of a complex vector space. We will denote this vector space as $\mathbb{P}^n(\mathbb{C})$, the ***polynomials of degree at most*** *n* ***over*** $\mathbb{C}$, to distinguish it from the polynomials with real coefficients, which we denote as $\mathbb{P}^n(\mathbb{R})$. We can construct $m \times n$ matrices with complex entries, and the set of all such matrices is denoted $Mat(\mathbb{C}, m, n)$, the $m \times n$ ***matrices*** over $\mathbb{C}$. For example:

$$A = \begin{bmatrix} 5 & 3-i & 4i \\ 2i & 7 & \sqrt{5} \end{bmatrix}$$

is a matrix in $Mat(\mathbb{C}, 2, 3)$. The operations of addition and scalar multiplication naturally follow from the arithmetic of complex numbers. The zero matrix and the negative of a matrix are what we would naturally expect. The set $Mat(\mathbb{C}, m, n)$ can thus be shown to be a vector space over $\mathbb{C}$. The rules for matrix addition, scalar multiplication, matrix multiplication, and the methods to find the inverse (when possible) and determinant of a square matrix, are exactly the same, except we perform the arithmetic on complex numbers. If we review the proofs of all the properties of the determinant, we notice that they still hold for matrices over an arbitrary field, since they depend only on field arithmetic and the properties of permutations.

Example: Consider the 2×2 complex matrices:

$$A = \begin{bmatrix} 2+3i & 7 \\ 5i & 4-i \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} 3-i & -2i \\ 1+i & 2+3i \end{bmatrix}.$$

Then:

$$A + B = \begin{bmatrix} 2+3i & 7 \\ 5i & 4-i \end{bmatrix} + \begin{bmatrix} 3-i & -2i \\ 1+i & 2+3i \end{bmatrix}$$

$$= \begin{bmatrix} 2+3i+3-i & 7+(-2i) \\ 5i+1+i & 4-i+2+3i \end{bmatrix} = \begin{bmatrix} 5+2i & 7-2i \\ 1+6i & 6+2i \end{bmatrix},$$

$$(3-2i)A = (3-2i)\begin{bmatrix} 2+3i & 7 \\ 5i & 4-i \end{bmatrix}$$

$$= \begin{bmatrix} (3-2i)(2+3i) & (3-2i)7 \\ (3-2i)5i & (3-2i)(4-i) \end{bmatrix} = \begin{bmatrix} 12+5i & 21-14i \\ 10+15i & 10-11i \end{bmatrix},$$

$$AB = \begin{bmatrix} 2+3i & 7 \\ 5i & 4-i \end{bmatrix}\begin{bmatrix} 3-i & -2i \\ 1+i & 2+3i \end{bmatrix}$$

$$= \begin{bmatrix} (2+3i)(3-i)+(7)(1+i) & (2+3i)(-2i)+(7)(2+3i) \\ (5i)(3-i)+(4-i)(1+i) & (5i)(-2i)+(4-i)(2+3i) \end{bmatrix}$$

$$= \begin{bmatrix} 16+14i & 20+17i \\ 10+18i & 21+10i \end{bmatrix},$$

$$det(B) = (3-i)(2+3i) - (-2i)(1+i) = 7+9i, \quad \text{and}$$

$$B^{-1} = \frac{1}{7+9i}\begin{bmatrix} 2+3i & 2i \\ -1-i & 3-i \end{bmatrix} = \frac{7-9i}{7^2+9^2}\begin{bmatrix} 2+3i & 2i \\ -1-i & 3-i \end{bmatrix}$$

$$= \frac{1}{130}\begin{bmatrix} 41+3i & 18+14i \\ -16+2i & 12-34i \end{bmatrix}. \square$$

We have left out some of the details in the computations above, since they all involve the basic arithmetic of complex numbers. Notice also that we found B^{-1} using the formula for the inverse of a 2×2 matrix and division of complex numbers.

Linear Combinations, Spans and Independence

The definitions in Chapter 3 for real vector spaces can easily be extended to analogous definitions for a vector space $(V, \oplus, \odot)$ over an arbitrary field $(\mathbf{F}, +, \cdot)$.

Suppose $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n \in V$, and $r_1, r_2, \ldots, r_n \in \mathbf{F}$. Then, a ***linear combination*** of the vectors $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ with ***coefficients*** $r_1, r_2, \ldots, r_n$ has the form:

$$(r_1 \odot \vec{v}_1) \oplus (r_2 \odot \vec{v}_2) \oplus \cdots \oplus (r_n \odot \vec{v}_n).$$

The ***Span*** of the set of vectors $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is the set of all possible linear combinations of these vectors:

$$\begin{aligned} Span(S) &= Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}) \\ &= \{(r_1 \odot \vec{v}_1) \oplus (r_2 \odot \vec{v}_2) \oplus \cdots \oplus (r_n \odot \vec{v}_n) \,|\, r_1, r_2, \ldots, r_n \in \mathbf{F}\}. \end{aligned}$$

We will generalize this concept to infinite sets of vectors which are indexed by a subset of $\mathbb{R}$. As before, we write: $S = \{\vec{v}_i \,|\, i \in I\} \subset (V, \oplus, \odot)$, where $I \subset \mathbb{R}$ is some non-empty indexing set. We will define the Span of S as the set of all possible linear combinations of ***every finite subset*** of S. In other words, we form a finite subset $\{\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_n}\}$ of S, form all the possible linear combinations of this subset, and repeat this process for ***all*** the finite subsets of S.

Let S be a (possibly infinite) subset of a vector space $(V, \oplus, \odot)$. We say that S is ***linearly independent*** if any linear combination of vectors from S results in the zero vector if and only if all the coefficients of these vectors are 0. In other words, the only solution to the ***dependence test equation:***

$$(c_1 \odot \vec{v}_{i_1}) \oplus (c_2 \odot \vec{v}_{i_2}) \oplus \cdots \oplus (c_n \odot \vec{v}_{i_n}) = \vec{0}_V.$$

is the ***trivial solution*** $c_1 = 0, c_2 = 0, \ldots, c_n = 0$, for every ***finite list*** of vectors $\vec{v}_{i_1}, \vec{v}_{i_2}, \ldots, \vec{v}_{i_n} \in V$. If we have a non-trivial solution, we say that S is ***linearly dependent***, and an equation above with at least one non-zero coefficient is called a ***dependence equation*** for S.

The ***Gauss-Jordan Algorithm*** can again be used to solve a system of linear equations with coefficients from the field **F.** Our goal will be the same: to find the ***rref*** of an augmented matrix with entries from **F**. In general, we replace "0" and "1" with $0_{\mathbf{F}}$ and $1_{\mathbf{F}}$ in the definition of the rref. This time, we will normalize the pivot row by dividing it by the pivot entry. As before, we will be able to decide if a set of vectors S is linearly dependent or independent by finding the rref of the matrix whose columns are the coordinates of the vectors in S.

Example: Consider the set of three vectors:

$$S = \{\langle i, 1+i, 2\rangle, \langle -5i, -8-4i, -5+5i\rangle, \langle -1, 2i, 1-i\rangle\},$$

a subset of $\mathbb{C}^3$. Let us decide if S is dependent or independent. As in Chapter 1, we write the vectors in S as the ***columns*** of a matrix, and in this case we obtain the 3×3 complex matrix:

$$\begin{bmatrix} i & -5i & -1 \\ 1+i & -8-4i & 2i \\ 2 & -5+5i & 1-i \end{bmatrix}.$$

We will need to apply the Gauss-Jordan Algorithm to this matrix, which thanks to the nature of complex arithmetic, would be a very messy process. To obtain a leading 1 in the first column, we can either exploit the 2 in the third row or the i in the first row. In this case, dividing the first row by i will not introduce fractions, unlike dividing the third row by 2, so we get:

$$\begin{bmatrix} 1 & -5 & i \\ 1+i & -8-4i & 2i \\ 2 & -5+5i & 1-i \end{bmatrix}.$$

Now we multiply the first row by $1+i$ and subtract this from row 2, and similarly multiply the first row by 2 and subtract this from row 3, obtaining:

$$\begin{bmatrix} 1 & -5 & i \\ 0 & -3+i & 1+i \\ 0 & 5+5i & 1-3i \end{bmatrix}.$$

To get a leading 1 in row 2, let us just divide row 2 by $-3+i$:

$$\begin{bmatrix} 1 & -5 & i \\ 0 & 1 & -\frac{1}{5}-\frac{2}{5}i \\ 0 & 5+5i & 1-3i \end{bmatrix}.$$

We multiply the 2nd row by $5+5i$ and subtract this from row 3:

$$\begin{bmatrix} 1 & -5 & i \\ 0 & 1 & -\frac{1}{5}-\frac{2}{5}i \\ 0 & 0 & 0 \end{bmatrix},$$

and since we get a row of zeroes, we can now conclude that the set is ***dependent***, and the third column (i.e., the third vector) is a linear combination of the first and second. To find a dependence equation, though, we complete the process of finding the rref, which in this case is just one step away, by adding 5 times row 2 to row 1:

$$\begin{bmatrix} 1 & 0 & -1-i \\ 0 & 1 & -\frac{1}{5}-\frac{2}{5}i \\ 0 & 0 & 0 \end{bmatrix}.$$

The third column gives us the coefficients for a dependence equation involving our original columns:

$$(-1-i)\langle i, 1+i, 2\rangle + \left(-\frac{1}{5}-\frac{2}{5}i\right)\langle -5i, -8-4i, -5+5i\rangle = \langle -1, 2i, 1-i\rangle. \square$$

If we are given a "standard" basis for an abstract complex vector space, we can test for the independence of a set of vectors by ***encoding*** the coordinates of these vectors with respect to this standard basis, and applying the Gauss-Jordan Algorithm to the matrix obtained by assembling these coordinates into columns, as we did in Chapter 3. We will see this in the Exercises.

Subspaces and Basis

A non-empty subset W of a vector space $(V, \oplus, \odot)$ over a field $(\mathbf{F}, +, \cdot)$ is called a ***subspace*** of V if $(W, \oplus, \odot)$ is also a vector space over $\mathbf{F}$, that is, under the ***same*** vector addition and scalar multiplication as V. We will write $(W, \oplus, \odot) \trianglelefteq (V, \oplus, \odot)$ or just $W \trianglelefteq V$. This equivalent to saying that W is a non-empty subset of V, and W is ***closed*** under vector addition and scalar multiplication. The basic example of a subspace of a vector space $(V, \oplus, \odot)$ is:

$$W = Span(S),$$

where S is a (possibly infinite) subset of V.

A set of vectors B from a vector space $(V, \oplus, \odot)$ is a ***basis*** for V if it is ***linearly independent*** and ***Spans*** V. Equivalently, a set S (which is possibly infinite) is a basis for $(V, \oplus, \odot)$ if and only if every vector $\vec{v} \in V$ can be represented ***uniquely*** as a linear combination of a finite set of members $\vec{w}_{i_1}$, $\vec{w}_{i_2}$, $\ldots$, $\vec{w}_{i_k}$ from S :

$$\vec{v} = c_1\vec{w}_{i_1} + c_2\vec{w}_{i_2} + \cdots + c_k\vec{w}_{i_k}.$$

Every vector space $(V, \oplus, \odot)$ has a (possibly infinite) basis B. Thus, every subspace W of V is in fact the Span of a set of vectors B. A vector space $(V, \oplus, \odot)$ is called ***finite dimensional*** if we can find a ***finite basis*** B for V, otherwise we say that V is ***infinite dimensional***. (Again, to prove the existence of a basis completely in the infinite dimensional case, we need The Axiom of Choice or Zorn's Lemma.)

The theorems needed to construct a basis, namely the Extension Theorem, the Elimination Theorem, and the Dependent vs. Spanning Sets Theorem are all still true, and can easily be proven using complex scalars.

Consequently, any two bases for a finite-dimensional vector space $(V, \oplus, \odot)$ have exactly the same number of elements, and we call this common number the ***dimension*** of V or $dim(V)$. For any subspace W of V, $dim(W) \leq dim(V)$. In the finite dimensional case, $dim(W) = dim(V)$ ***if and only if*** $W = V$.

Example: The standard basis $S = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$, where $\vec{e}_i$ has zeroes in all components except for $1_{\mathbb{C}} = 1 + 0 \cdot i$ in the i^{th} component, is a basis for $\mathbb{C}^n$ as a vector space over $\mathbb{C}$. Thus $\mathbb{C}^n$ is n –dimensional over $\mathbb{C}$, in the same way that $\mathbb{R}^n$ is n –dimensional over $\mathbb{R}$. □

Example: Every polynomial $p(z)$ of degree n with complex coefficients can be written as:

$$p(z) = c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^n.$$

Thus the set $S = \{1, z, z^2, \ldots, z^n\}$ ***Spans*** $\mathbb{P}^n(\mathbb{C})$. However, the Fundamental Theorem of Algebra again tells us that S is ***linearly independent***, since a non-constant polynomial of degree n with coefficients from $\mathbb{C}$ can have at most n complex roots. Thus $\mathbb{P}^n(\mathbb{C})$ is an $(n+1)$-dimensional vector space over $\mathbb{C}$, in the same way that $\mathbb{P}^n(\mathbb{R})$ has dimension $n+1$ over $\mathbb{R}$. □

Example: We saw in the previous subsection that:

$$S = \{\langle i, 1+i, 2\rangle, \langle -5i, -8-4i, -5+5i\rangle, \langle -1, 2i, 1-i\rangle\}$$

is a dependent subset of $\mathbb{C}^3$, with all three vectors appearing in a dependence equation. However, no two of these vectors are parallel to each other. Thus $W = Span(S)$ is ***2-dimensional***, and we can choose any two of the vectors in S to serve as a basis. One choice would be:

$$B = \{\langle i, 1+i, 2\rangle, \langle -1, 2i, 1-i\rangle\}. \square$$

8.2 Section Summary

Most of the definitions, Theorems, and constructions that we saw in the first Chapters 1 to 6 extend naturally to vector spaces over the field of complex numbers, and even more generally, to vector spaces over an arbitrary field $(\mathbf{F}, +, \cdot)$.

Complex Euclidean n-space, denoted $\mathbb{C}^n$, is the set of all n-tuples of complex numbers:

$$\mathbb{C}^n = \{\langle z_1, z_2, \ldots, z_n\rangle | z_i \in \mathbb{C}\}.$$

The set of polynomials of degree at most n with complex coefficients is denoted $\mathbb{P}^n(\mathbb{C})$, and is an $(n+1)$-dimensional vector space over $\mathbb{C}$.

We can construct $m \times n$ matrices with complex entries, and the set of all such matrices is denoted $Mat(\mathbb{C}, m, n)$. This is an $(m \times n)$-dimensional vector space over $\mathbb{C}$.

The following terms, concepts and constructions are defined for complex vector spaces as they are to real vector spaces, with the notable exception that scalars appearing in definitions are now complex numbers:

- the ***linear combinations*** of a set of vectors;
- the ***Span*** of a set of vectors;
- linear ***dependence*** or ***independence*** of a set of vectors;
- a ***basis*** for a vector space;
- the ***dimension*** of a vector space;
- a ***subspace*** of a vector space;
- ***complex matrices*** and their ***arithmetic:*** addition, subtraction, multiplication, finding determinants and inverses of square matrices, when they exist.

8.2 Exercises

For Exercises (1) to (5): Determine whether or not the indicated vector $\vec{b}$ is a member of $Span(S)$, and if so, express $\vec{b}$ as a linear combination of the vectors of S.

1. $\vec{b} = \langle i, 1, i\rangle$; $S = \{\langle 1-i, 2i, 3\rangle, \langle 2+2i, -4, 6i\rangle, \langle -i, 3i, 2\rangle\}$.
2. $\vec{b} = \langle 6-4i, -3+10i, 8+7i\rangle$; $S = \{\langle 1-i, 2i, 3\rangle, \langle 2+2i, -4, 6i\rangle, \langle -i, 3i, 2\rangle\}$.
3. $\vec{b} = \langle 9+9i, -12+8i, 13+27i, 10-i\rangle$; $S = \{\langle 1-i, i, 3, -2i\rangle, \langle 2i, -1, i, 3i\rangle, \langle 2, 3i, 6+i, -i\rangle\}$.
4. $\vec{b} = 2-4i+(-10+7i)z-(5+9i)z^2$;
   $S = \{1-i+2iz-3z^2,\ 2i+3z+(i-1)z^2,\ -1-i+5iz-(4+i)z^2\}$.
5. $\vec{b} = 11-5i+(-28+13i)z-(12+29i)z^2$;
   $S = \{1-i+2iz-3z^2,\ 2i+3z+(i-1)z^2,\ -1-i+5iz-(4+i)z^2\}$.

For Exercises (6) to (9): Determine whether or not the indicated subset of the respective vector space is linearly dependent or independent. If it is dependent, give an example of a dependence relation among the vectors. If we let $W = Span(S)$, find $dim(W)$.

6. $S = \{\langle 1-i, 2i, 3\rangle, \langle 2+2i, -4, 6i\rangle, \langle -i, 3i, 2\rangle\} \subseteq \mathbb{C}^3$.
7. $S = \{\langle 1-i, i, 3, -2i\rangle, \langle 2i, -1, i, 3i\rangle, \langle 2, 3i, 6+i, -i\rangle\} \subseteq \mathbb{C}^4$.
8. $S = \left\{ \begin{bmatrix} i & -2 \\ 1 & 1-i \end{bmatrix}, \begin{bmatrix} 3 & 2i \\ -i & 1+i \end{bmatrix}, \begin{bmatrix} -1+i & -2-2i \\ 1+i & 2 \end{bmatrix}, \begin{bmatrix} -1+3i & -2-2i \\ 1+i & 2i \end{bmatrix} \right\}$
   $\subseteq Mat(2,2,\mathbb{C})$.
9. $S = \{1-i+2iz-3z^2,\ 2i+3z+(i-1)z^2,\ -1-i+5iz-(4+i)z^2\} \subseteq \mathbb{P}^2(\mathbb{C})$.

For Exercises (10) to (15): Determine whether or not the indicated subset of the respective vector space is a subspace of that space. If so, then find a basis for the subspace, and find its dimension.

10. $W = \left\{ p(z) \in \mathbb{P}^2(\mathbb{C}) \,|\, p(2i) = 0 \text{ and } p(1+i) = 0 \right\}$
11. $W = \left\{ p(z) \in \mathbb{P}^2(\mathbb{C}) \,|\, p(2i) = 0 \text{ and } p(1-i) = 2+i \right\}$
12. $W = \left\{ p(z) \in \mathbb{P}^3(\mathbb{C}) \,|\, p(1+2i) = 0 \text{ and } p(1-i) = p(1+i) \right\}$
13. $W = \left\{ p(z) \in \mathbb{P}^3(\mathbb{C}) \,|\, p(-i) = 0 \text{ and } p'(2-i) = 0 \right\}$

    Note: Although the concept of a "limit" is different for the field of complex numbers, we can ***formally differentiate*** a polynomial with complex coefficients using the same formula as in ordinary Calculus:

    $$\text{If } p(z) = c_n z^n + c_{n-1}z^{n-1} + \cdots + c_1 z + c_0, \text{ then:}$$
    $$\frac{d}{dz}p(z) = p'(z) = n(c_n)z^n + (n-1)(c_{n-1})z^{n-1} + \cdots + 2(c_2)z + c_1.$$

14. $W = \left\{ p(z) \in \mathbb{P}^3(\mathbb{C}) \,|\, p'(1+2i) = 0 \text{ and } p''(1-i) = 0 \right\}$
15. $W = \{A \in Mat(n,n,\mathbb{C}) \,|\, tr(A) = 0\}$,
    where $tr(A) = a_{1,1} + a_{2,2} + \cdots + a_{n,n}$ is the trace function.

For Exercises (16) to (20): For each of the following matrices A, find: (a) A^2; (b) $det(A)$; (c) A^{-1}, if it exists.

16. $\begin{bmatrix} 1-i & 2i \\ 1+i & -i \end{bmatrix}$

17. $\begin{bmatrix} -i & 2-i \\ 1+3i & i \end{bmatrix}$

18. $\begin{bmatrix} -i & 2-i \\ 1-i & 3+i \end{bmatrix}$

19. $\begin{bmatrix} i & 1-i & -i \\ 1+i & -i & 1-i \\ 0 & 2 & i \end{bmatrix}$

20. $\begin{bmatrix} i & 1+i & 1+i \\ 2i & -i & 1+i \\ 1-i & 3i & -2i \end{bmatrix}$

21. In the same way we defined $\mathbb{R}^n$ and $\mathbb{C}^n$, define by $\mathbf{F}^n$ the set of all n –tuples of $\mathbf{F}$:

$$\mathbf{F}^n = \{\langle f_1, f_2, \ldots, f_n \rangle | f_i \in \mathbf{F}\}$$

and define the natural vector addition via addition of pairs of components and similarly for scalar multiplication.

   a. Prove that $\mathbf{F}^n$ is a vector space over $\mathbf{F}$.
   b. Show that the standard basis $S = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ where $\vec{e}_i$ has $0_\mathbf{F}$ in all components except for $1_\mathbf{F}$ in the i^{th} component, is a basis for $\mathbf{F}^n$ as a vector space over $\mathbf{F}$.

22. Define $Mat(m,n,\mathbf{F})$ to be the set of all $m \times n$ matrices with entries from $\mathbf{F}$, under the natural addition operation for matrices and the multiplication of such a matrix by a scalar from $\mathbf{F}$.

   a. Prove that $Mat(m,n,\mathbf{F})$ is a vector space over $\mathbf{F}$.
   b. Generalize the construction in Chapter 3 to create a basis for $Mat(m,n,\mathbf{F})$ over $\mathbf{F}$.
   c. Find the dimension of $Mat(m,n,\mathbf{F})$ over $\mathbf{F}$.

23. Define $\mathbb{P}^n(\mathbf{F})$ to be the set of all polynomials:

$$p(z) = c_n z^n + c_{n-1} z^{n-1} + \cdots + c_1 z + c_0$$

of degree n or less in the variable z and with coefficients c_i from $\mathbf{F}$, under the natural operations of polynomial addition and scalar multiplication by $\mathbf{F}$. Prove that $\mathbb{P}^n(\mathbf{F})$ is a vector space over $\mathbf{F}$. (Note: it is not true in general that $\mathbb{P}^n(\mathbf{F})$ has dimension $n+1$ over $\mathbf{F}$).

24. Show that $\mathbb{C}^n$ is a vector space over $\mathbb{R}$ with dimension $2n$. As part of your proof, find a basis for $\mathbb{C}^n$ over $\mathbb{R}$.

25. Show that $\mathbf{F} = \mathbb{Q}(\sqrt{2})$, from the previous Section, is a 2-dimensional vector space over $\mathbb{Q}$. What is a basis for $\mathbf{F}$ over $\mathbb{Q}$?

26. Use the definition of the determinant function and the properties of the conjugate to prove that for any $n \times n$ matrix A : $\overline{det(A)} = det(\overline{A})$.

In this formula, $\overline{A}$ is the $n \times n$ matrix whose entries are the complex conjugates of the corresponding entries of A, and $\overline{det(A)}$ denotes the complex conjugate of $det(A)$.

8.3 Complex Inner Products

We shall now generalize the concept of an inner product to complex Euclidean spaces and the associated concept of orthogonality.

The Complex Euclidean Inner Product

We defined the dot product on Real Euclidean n-space by taking the sum of the matching pairs of products of the components. Although there is nothing stopping us, computationally, from doing the same thing to two complex vectors, we need to modify this definition so that the inner product of a complex vector ***with itself*** still satisfies the ***positivity*** property. It turns out that we only need to make a small adjustment:

> ***Definition:*** Let $\vec{z} = \langle z_1, z_2, \ldots, z_n\rangle$, and $\vec{w} = \langle w_1, w_2, \ldots, w_n\rangle$ be vectors from $\mathbb{C}^n$. We define their ***Complex Euclidean inner product***, or simply their ***inner product***, by:
>
> $$\langle \vec{z} | \vec{w} \rangle = \vec{z} \circ \overline{\vec{w}} = z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}.$$

Example: Let $\vec{z} = \langle 1-i, 2+3i, 4\rangle$ and $\vec{w} = \langle 2i, 3-i, 1+2i\rangle \in \mathbb{C}^3$. Then:

$$\begin{aligned}\langle \vec{z} | \vec{w} \rangle &= \vec{z} \circ \overline{\vec{w}} \\ &= \langle 1-i, 2+3i, 4\rangle \circ \langle -2i, 3+i, 1-2i\rangle \\ &= (1-i)(-2i) + (2+3i)(3+i) + 4(1-2i) \\ &= 5+i. \square\end{aligned}$$

There is a very good reason why we require the ***conjugate*** of $\vec{w}$ in this definition. We mentioned in the Exercises of Section 6.1 that if V is any n-dimensional vector space over $\mathbb{R}$, then any isomorphism (in fact, just a one-to-one linear transformation) $T : V \to \mathbb{R}^n$ will induce an inner product:

$$\langle \vec{u} | \vec{v} \rangle_V = T(\vec{u}) \circ T(\vec{v}).$$

Consider now $V = \mathbb{C}$. We saw that the function:

$$\begin{aligned} T &: \mathbb{C} \to \mathbb{R}^2, \quad \text{where} \\ T(a+bi) &= \langle a, b\rangle \end{aligned}$$

is an ***isomorphism*** of vector spaces. Thus we can induce an inner product on $\mathbb{C}$ via:

$$\langle a+bi | c+di \rangle = \langle a, b\rangle \circ \langle c, d\rangle = ac + bd,$$

and in particular:

$$\langle a+bi | a+bi \rangle = \langle a, b\rangle \circ \langle a, b\rangle = a^2 + b^2.$$

But recall that

$$(a+bi)\overline{(a+bi)} = (a+bi)(a-bi) = a^2 + b^2.$$

Thus, it is natural to define in general:

$$\langle z | w \rangle = \langle a+bi | c+di \rangle = (a+bi)\overline{(c+di)} = z \cdot \overline{w}$$

This definition has the added advantage that if $\vec{z}$ and $\vec{w}$ were actually ***real*** vectors, then the new definition coincides with the old one. We can now verify the ***positivity*** property:

Theorem — The Positivity of the Complex Inner Product:
Under the inner product above:
$$\left\langle \vec{0}_{\mathbb{C}^n} | \vec{z} \right\rangle = 0_{\mathbb{C}},$$
for any complex vector $\vec{z}$, and thus in particular $\left\langle \vec{0}_{\mathbb{C}^n} | \vec{0}_{\mathbb{C}^n} \right\rangle = 0_{\mathbb{C}}$.
Furthermore, if $\vec{z} \in \mathbb{C}^n$, and $\vec{z} \neq \vec{0}_{\mathbb{C}^n}$, then $\langle \vec{z} | \vec{z} \rangle > 0$. Thus:
$$\langle \vec{z} | \vec{z} \rangle > 0 \textit{ if and only if } \vec{z} \neq \vec{0}_{\mathbb{C}^n}, \text{ and } \left\langle \vec{0}_{\mathbb{C}^n} | \vec{0}_{\mathbb{C}^n} \right\rangle = 0_{\mathbb{C}}.$$
Consequently, we can define the ***length*** of a vector $\vec{z}$ via:
$$\|\vec{z}\| = \sqrt{\langle \vec{z} | \vec{z} \rangle},$$
and the ***distance*** between two vectors $\vec{z}, \vec{w} \in \mathbb{C}^n$ via:
$$d(\vec{z}, \vec{w}) = \|\vec{z} - \vec{w}\|.$$

Proof: It is obvious from the definition that $\left\langle \vec{0}_{\mathbb{C}^n} | \vec{z} \right\rangle = 0_{\mathbb{C}}$ for any complex vector $\vec{z}$. Now, if $\vec{z} \in \mathbb{C}^n$ is not the zero vector, we have:
$$\begin{aligned} \langle \vec{z} | \vec{z} \rangle &= z_1\overline{z_1} + z_2\overline{z_2} + \cdots + z_n\overline{z_n} \\ &= \|z_1\|^2 + \|z_2\|^2 + \cdots + \|z_n\|^2 > 0, \end{aligned}$$
since at least one of the components z_i is nonzero.■

Example: Let $\vec{z} = \langle 2i, 3-i, 1+2i \rangle \in \mathbb{C}^3$ from the previous Example. Then:
$$\begin{aligned} \langle \vec{z} | \vec{z} \rangle &= \langle 2i, 3-i, 1+2i \rangle \circ \langle -2i, 3+i, 1-2i \rangle \\ &= (2i)(-2i) + (3-i)(3+i) + (1+2i)(1-2i) \\ &= 4 + 9 + 1 + 1 + 4 = 19. \end{aligned}$$
As expected, this is a positive real number, and thus $\|\vec{z}\| = \sqrt{19}$.□

Unfortunately, nothing comes for free. Our new definition has the consequence that our inner product is no longer ***symmetric***, that is, in general: $\langle \vec{z} | \vec{w} \rangle \neq \langle \vec{w} | \vec{z} \rangle$.

Example: Let $\vec{z} = \langle 3-2i, 1+5i \rangle$ and $\vec{w} = \langle 2+i, 4-3i \rangle$. Then:
$$\begin{aligned} \langle \vec{z} | \vec{w} \rangle &= (3-2i)(2-i) + (1+5i)(4+3i) = -7+16i, \quad \text{but} \\ \langle \vec{w} | \vec{z} \rangle &= (2+i)(3+2i) + (4-3i)(1-5i) = -7-16i. \end{aligned}$$□

Notice, however, that the two complex numbers we obtained above are ***complex conjugates*** of each other. This is obviously not a coincidence:

Theorem: Let $\vec{z}, \vec{w} \in \mathbb{C}^n$. Under the inner product defined above:
$$\langle \vec{z} | \vec{w} \rangle = \overline{\langle \vec{w} | \vec{z} \rangle}.$$

Proof: We have:

$$\begin{aligned}
\langle \vec{z}|\vec{w}\rangle &= \vec{z}\circ\overline{\vec{w}} = z_1\overline{w_1}+z_2\overline{w_2}+\cdots+z_n\overline{w_n}, \quad \text{but}\\
\overline{\langle \vec{w}|\vec{z}\rangle} &= \overline{w_1\overline{z_1}+w_2\overline{z_2}+\cdots+w_n\overline{z_n}}\\
&= \overline{w_1\overline{z_1}}+\overline{w_2\overline{z_2}}+\cdots+\overline{w_n\overline{z_n}} && \text{by the additivity of the conjugate,}\\
&= \overline{w_1}\overline{\overline{z_1}}+\overline{w_2}\overline{\overline{z_2}}+\cdots+\overline{w_n}\overline{\overline{z_n}} && \text{by the multiplicativity of the conjugate,}\\
&= \overline{w_1}z_1+\overline{w_2}z_2+\cdots+\overline{w_n}z_n && \text{by the double conjugate property,}\\
&= z_1\overline{w_1}+z_2\overline{w_2}+\cdots+z_n\overline{w_n} && \text{by the commutative property of multiplication,}
\end{aligned}$$

and hence the two sides are the same.■

This new property is called ***Hermitian-symmetry***, named after French mathematician ***Charles Hermite*** (1822-1901). He is known for being the first to prove that Euler's number e (which is approximately $2.7182818\ldots$) is a ***transcendental number***, that is, there is no polynomial $p(x)$ with ***integer*** coefficients such that $p(e)=0$. Another well-known transcendental number is π. We will encounter the adjective ***Hermitian*** again in future Sections.

The Axioms of a Complex Inner Product Space

Recall that the ordinary dot product of real Euclidean n−spaces has special properties that we later used as the axioms for an inner product space. In the same way, let us summarize the properties that we saw above for our new complex inner product:

Theorem — Axioms of a Complex Inner Product Space:
Let $\vec{z}$, $\vec{w}$, and $\vec{u}\in\mathbb{C}^n$, and $k\in\mathbb{C}$. Under the complex inner product, the following properties are true:

1. ***The Hermitian-Symmetry Property:*** $\langle \vec{z}|\vec{w}\rangle = \overline{\langle \vec{w}|\vec{z}\rangle}$.
2. ***The Left Homogeneity Property:*** $\langle k\cdot\vec{z}|\vec{w}\rangle = k\cdot\langle \vec{z}|\vec{w}\rangle$.
3. ***The Left Additivity Property:*** $\langle \vec{u}+\vec{z}|\vec{w}\rangle = \langle \vec{u}|\vec{w}\rangle+\langle \vec{z}|\vec{w}\rangle$.
4. ***The Positivity Property:*** If $\vec{z}\neq\vec{0}_{\mathbb{C}^n}$, then $\langle \vec{z}|\vec{z}\rangle>0$.

Proof: We already proved the first and fourth properties. The second and third properties follow directly from the definition and the properties of the addition and multiplication of complex numbers. We remark, though, that the Hermitian-Symmetry Property also implies, using $\vec{z}=\vec{w}$, that $\langle \vec{z}|\vec{z}\rangle = \overline{\langle \vec{z}|\vec{z}\rangle}$. Thus $\langle \vec{z}|\vec{z}\rangle$ must be a ***pure real number***. The Positivity Property says that not only is this a pure real number, it must be ***positive*** if $\vec{z}$ is not the zero vector.■

More generally, if $(V,\oplus,\odot)$ is a ***complex*** vector space on which we can define a ***bilinear form*** $\langle\,|\,\rangle$ that satisfies the four properties above, we call V a ***complex inner product space*** under the bilinear form $\langle\,|\,\rangle$. We naturally refer to these four properties as the ***Axioms of a Complex Inner Product Space***. We also get the following bonus properties:

Theorem: Let $(V, \oplus, \odot)$ be a complex inner product space. Let $\vec{z}, \vec{w}, \vec{u} \in V$, and $k \in \mathbb{C}$. Then, the following properties also hold:

1. ***The Right Additivity Property:*** $\langle \vec{z} | \vec{w} \oplus \vec{u} \rangle = \langle \vec{z} | \vec{w} \rangle + \langle \vec{z} | \vec{u} \rangle$.
2. ***The Right Conjugate-Homogeneity Property:*** $\langle \vec{z} | k \odot \vec{w} \rangle = \bar{k} \cdot \langle \vec{z} | \vec{w} \rangle$.
3. ***The Inner Product with the Zero Vector Property:*** $\left\langle \vec{z} | \vec{\mathbf{0}}_V \right\rangle = 0_{\mathbb{C}} = \left\langle \vec{\mathbf{0}}_V | \vec{z} \right\rangle$, and in particular: $\left\langle \vec{\mathbf{0}}_V | \vec{\mathbf{0}}_V \right\rangle = 0_{\mathbb{C}}$.

The proofs of these properties are left as Exercises. Again, we are only allowed to use the four axioms for a complex inner product space, as well as the axioms (and consequent properties) of a complex vector space in order to prove these properties, but their proofs are very similar to the analogous properties that we saw in Chapter 7.

Orthogonality and Orthogonal Complements

The complex inner product allows us to recast definitions and Theorems from real inner product spaces in Chapter 7 in the setting of complex vector spaces. Their proofs will be left as Exercises:

Definitions/Theorems: We say that two vectors $\vec{v}$ and $\vec{w}$ in a complex inner product space are ***orthogonal*** ***if and only if*** $\langle \vec{v} | \vec{w} \rangle = 0_{\mathbb{C}}$. In particular, $\vec{\mathbf{0}}_V$ is orthogonal to all $\vec{v} \in V$.

A set of vectors S which does not contain the zero vector is ***orthogonal*** if any two distinct vectors in S are orthogonal to each other, and S is ***orthonormal*** if it is an orthogonal set consisting of ***unit vectors***. Orthogonal sets are automatically ***linearly independent***.

If W is a subspace of a (possibly infinite dimensional) complex inner product space V, the ***orthogonal complement*** $W^{\perp}$ of W is:

$$W^{\perp} = \left\{ \vec{v} \in V | \langle \vec{v} | \vec{w} \rangle = 0_{\mathbb{C}} \text{ for all } \vec{w} \in W \right\}.$$

$W^{\perp}$ is again a subspace of V. If B is a basis for W, then:

$$W^{\perp} = \left\{ \vec{v} \in V | \langle \vec{v} | \vec{w} \rangle = 0_{\mathbb{C}} \text{ for all } \vec{w} \in B \right\},$$

that is, it is necessary and sufficient that we check if $\vec{v}$ is orthogonal to every member of a basis for W.

The ***Gram-Schmidt Algorithm*** extends naturally to complex inner product spaces, and we can again use it to construct an orthonormal basis for a subspace W as well as its orthogonal complement $W^{\perp}$. Needless to say, the computations are messier when dealing with complex vectors.

If V is a ***finite dimensional*** complex inner product space, and W is any subspace of V, then by the Gram-Schmidt Algorithm, $dim(W) + dim(W^{\perp}) = dim(V)$, and $(W^{\perp})^{\perp} = W$, as before.

Example: Let $B = \{ \langle 1 - i, i, 1 \rangle, \langle -i, 2, 1 + i \rangle, \langle -1, 2 + i, i \rangle \} \subset \mathbb{C}^3$. Let us apply the Gram-Schmidt Algorithm to B, in the given order:

$$
\begin{aligned}
\vec{v}_1 &= \vec{z}_1 = \langle 1-i, i, 1\rangle. \\
\vec{v}_2 &= \vec{z}_2 - \frac{\langle \vec{z}_2 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle}\vec{v}_1 \\
&= \langle -i, 2, 1+i\rangle - \frac{\langle -i, 2, 1+i\rangle \circ \langle 1+i, -i, 1\rangle}{\langle 1-i, i, 1\rangle \circ \langle 1+i, -i, 1\rangle}\langle 1-i, i, 1\rangle \\
&= \langle -i, 2, 1+i\rangle - \frac{-i(1+i)-2i+1+i}{(1-i)(1+i)-i^2+1}\langle 1-i, i, 1\rangle \\
&= \langle -i, 2, 1+i\rangle - \frac{2-2i}{4}\langle 1-i, i, 1\rangle = \langle -i, 2, 1+i\rangle + \frac{-1+i}{2}\langle 1-i, i, 1\rangle \\
&= \frac{1}{2}[\langle -2i, 4, 2+2i\rangle + \langle 2i, -1-i, -1+i\rangle] = \frac{1}{2}\langle 0, 3-i, 1+3i\rangle,
\end{aligned}
$$

so we will use instead $\vec{v}_2 = \langle 0, 3-i, 1+3i\rangle$. We easily check that:

$$
\begin{aligned}
\langle \vec{v}_1 | \vec{v}_2 \rangle &= \langle 1-i, i, 1\rangle \circ \langle 0, 3+i, 1-3i\rangle \\
&= 0 + i(3+i) + (1-3i) = 3i + i^2 + 1 - 3i = 0,
\end{aligned}
$$

so $\vec{v}_2$ is indeed orthogonal to $\vec{v}_1$. Lastly, we get:

$$
\begin{aligned}
\vec{v}_3 &= \vec{z}_3 - \frac{\langle \vec{z}_3 | \vec{v}_1 \rangle}{\langle \vec{v}_1 | \vec{v}_1 \rangle}\vec{v}_1 - \frac{\langle \vec{z}_3 | \vec{v}_2 \rangle}{\langle \vec{v}_2 | \vec{v}_2 \rangle}\vec{v}_2 \\
&= \langle -1, 2+i, i\rangle - \frac{\langle -1, 2+i, i\rangle \circ \langle 1+i, -i, 1\rangle}{\langle 1-i, i, 1\rangle \circ \langle 1+i, -i, 1\rangle}\langle 1-i, i, 1\rangle \\
&\quad - \frac{\langle -1, 2+i, i\rangle \circ \langle 0, 3+i, 1-3i\rangle}{\langle 0, 3-i, 1+3i\rangle \circ \langle 0, 3+i, 1-3i\rangle}\langle 0, 3-i, 1+3i\rangle \\
&= \langle -1, 2+i, i\rangle - \frac{-(1+i)-i(2+i)+i}{4}\langle 1-i, i, 1\rangle \\
&\quad - \frac{(2+i)(3+i)+i(1-3i)}{(3-i)(3+i)+(1+3i)(1-3i)}\langle 0, 3-i, 1+3i\rangle \\
&= \langle -1, 2+i, i\rangle - \frac{-1-i-2i-i^2+i}{4}\langle 1-i, i, 1\rangle \\
&\quad - \frac{6+5i+i^2+i-3i^2}{9-i^2+1-9i^2}\langle 0, 3-i, 1+3i\rangle \\
&= \langle -1, 2+i, i\rangle - \frac{-2i}{4}\langle 1-i, i, 1\rangle - \frac{8+6i}{20}\langle 0, 3-i, 1+3i\rangle \\
&= \langle -1, 2+i, i\rangle + \frac{i}{2}\langle 1-i, i, 1\rangle - \frac{4+3i}{10}\langle 0, 3-i, 1+3i\rangle \\
&= \frac{1}{10}[\langle -10, 20+10i, 10i\rangle + \langle 5+5i, -5, 5i\rangle - \langle 0, 15+5i, -5+15i\rangle] \\
&= \frac{1}{10}\langle -5+5i, 5i, 5\rangle = \frac{1}{2}\langle -1+i, i, 1\rangle,
\end{aligned}
$$

so we will use $\vec{v}_3 = \langle -1+i, i, 1\rangle$ instead. Let us verify that $\vec{v}_3$ is orthogonal to both $\vec{v}_1$ and $\vec{v}_2$:

$$
\begin{aligned}
\langle \vec{v}_1 | \vec{v}_3 \rangle &= \langle 1-i, i, 1\rangle \circ \langle -1-i, -i, 1\rangle \\
&= (1-i)(-1-i) + i(-i) + 1 \\
&= -1 + i - i + i^2 - i^2 + 1 = 0, \quad \text{and}
\end{aligned}
$$

$$\begin{aligned}\langle \vec{v}_2 | \vec{v}_3 \rangle &= \langle 0, 3-i, 1+3i\rangle \circ \langle -1-i, -i, 1\rangle \\ &= (3-i)(-i) + (1+3i) \\ &= -3i + i^2 + 1 + 3i = 0,\end{aligned}$$

so indeed we get an orthogonal set. We already know from above that $\langle \vec{v}_1 | \vec{v}_1 \rangle = 4$ and $\langle \vec{v}_2 | \vec{v}_2 \rangle = 20$, so all we need is:

$$\langle \vec{v}_3 | \vec{v}_3 \rangle = \langle -1+i, i, 1\rangle \circ \langle -1-i, -i, 1\rangle = 1+1+1+1 = 4.$$

Finally, we normalize the three vectors and get the orthonormal basis for $\mathbb{C}^3$:

$$\left\{ \frac{1}{2}\langle 1-i, i, 1\rangle, \frac{1}{2\sqrt{5}}\langle 0, 3-i, 1+3i\rangle, \frac{1}{2}\langle -1+i, i, 1\rangle \right\}. \square$$

8.3 Section Summary

Let $\vec{z} = \langle z_1, z_2, \ldots, z_n\rangle$, and $\vec{w} = \langle w_1, w_2, \ldots, w_n\rangle$ be vectors from $\mathbb{C}^n$. We define their ***Complex Euclidean inner product***, or simply their ***inner product***, by:

$$\langle \vec{z} | \vec{w} \rangle = \vec{z} \circ \overline{\vec{w}} = z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}.$$

Let $\vec{z}$, $\vec{w}$, and $\vec{u} \in \mathbb{C}^n$, and $k \in \mathbb{C}$. Under the complex inner product, the following properties are true:

1. ***The Hermitian-Symmetry Property:*** $\langle \vec{z} | \vec{w} \rangle = \overline{\langle \vec{w} | \vec{z} \rangle}$.
2. ***The Left Homogeneity Property:*** $\langle k \cdot \vec{z} | \vec{w} \rangle = k \cdot \langle \vec{z} | \vec{w} \rangle$.
3. ***The Left Additivity Property:*** $\langle \vec{u} + \vec{z} | \vec{w} \rangle = \langle \vec{u} | \vec{w} \rangle + \langle \vec{z} | \vec{w} \rangle$.
4. ***The Positivity Property:*** If $\vec{z} \neq \vec{0}_{\mathbb{C}^n}$, then $\langle \vec{z} | \vec{z} \rangle > 0$.

A complex vector space V is a ***complex inner product space*** under a bilinear form $\langle\ |\ \rangle$ if the above four axioms are satisfied by $\langle\ |\ \rangle$. The following properties also hold in such a space:

1. ***The Right Additivity Property:*** $\langle \vec{z} | \vec{w} \oplus \vec{u} \rangle = \langle \vec{z} | \vec{w} \rangle + \langle \vec{z} | \vec{u} \rangle$.
2. ***The Right Conjugate-Homogeneity Property:*** $\langle \vec{z} | k \odot \vec{w} \rangle = \bar{k} \cdot \langle \vec{z} | \vec{w} \rangle$.
3. ***The Inner Product with the Zero Vector Property:*** $\left\langle \vec{z} | \vec{\mathbf{0}}_V \right\rangle = 0_{\mathbb{C}} = \left\langle \vec{\mathbf{0}}_V | \vec{z} \right\rangle$, and in particular: $\left\langle \vec{\mathbf{0}}_V | \vec{\mathbf{0}}_V \right\rangle = 0_{\mathbb{C}}$.

We can define the following concepts and constructions in V:

- the ***length*** of a vector: $\|\vec{z}\| = \sqrt{\langle \vec{z} | \vec{z} \rangle}$.
- the ***distance*** between two vectors: $d(\vec{z}, \vec{w}) = \|\vec{z} - \vec{w}\|$.
- orthogonality: $\vec{z}$ is ***orthogonal*** to $\vec{w}$ ***if and only if*** $\langle \vec{z} | \vec{w} \rangle = 0$.
- ***orthogonal*** and ***orthonormal sets*** of vectors.
- the applicability of the ***Gram-Schmidt Algorithm***.
- the ***orthogonal complement*** $W^{\perp}$ of a subspace W of V.

8.3 Exercises

1. Let A be an ***invertible*** $n \times n$ complex matrix.

   a. Prove that the bilinear form on $\mathbb{C}^n$ induced by A, denoted $\langle | \rangle_A$, and given by:
   $$\langle \vec{z} | \vec{w} \rangle_A = \langle A[\vec{z}] | A[\vec{w}] \rangle,$$
   is also an inner product on $\mathbb{C}^n$, where the inner product on the right is the ordinary complex inner product on $\mathbb{C}^n$.

   b. Compute $\langle \langle i, 2-i \rangle | \langle 4, 1-2i \rangle \rangle_A$, where:
   $$A = \begin{bmatrix} 1-i & 3+i \\ 2+i & 5-3i \end{bmatrix}.$$

   c. Find the length of $\langle 4, 1-2i \rangle$ under the inner product induced by A from (b).

2. Let $z_1, z_2, \ldots, z_n$ be n fixed (distinct) complex numbers. Let us define the bilinear form $\langle p(z) | q(z) \rangle$ on $\mathbb{P}^n(\mathbb{C})$ via:
   $$\langle p(z) | q(z) \rangle = p(z_1)\overline{q(z_1)} + p(z_2)\overline{q(z_2)} + \cdots + p(z_n)\overline{q(z_n)}.$$

   a. Prove that this bilinear form is an inner product on $\mathbb{P}^n(\mathbb{C})$.

   b. Compute $\langle 3z^2 - 2iz + 1 - 5i | (1-i)z^2 + 3z - 2i \rangle$, where:
   $$\langle p(z) | q(z) \rangle = p(2i)\overline{q(2i)} + p(1-i)\overline{q(1-i)} + p(3)\overline{q(3)}.$$
   is an inner product on $\mathbb{P}^2(\mathbb{C})$.

   c. Find $\|3z^2 - 2iz + 1 - 5i\|$ under the inner product above. Note that you may use some of your computations from (b).

For Exercises (3) to (5): Apply the Gram-Schmidt algorithm on the following sets of vectors from the corresponding $\mathbb{C}^n$, under the ordinary complex inner product on $\mathbb{C}^n$:

3. $\{\langle i, 2-i \rangle, \langle 1-i, 3i \rangle\}$
4. $\{\langle 1+i, 2, 3-i \rangle, \langle 1-i, 3+i, 2-i \rangle, \langle i, 5-2i, -i \rangle\}$
5. $\{\langle i, 2+i, -i \rangle, \langle 1+i, 3, 2+i \rangle, \langle i, 2i, 1-i \rangle\}$
6. Apply the Gram-Schmidt algorithm on the set:
   $$\{\langle i, 2-i \rangle, \langle 1-i, 3i \rangle\}$$
   under the inner product of Exercise 1. Do you get the same answer as Exercise 3?
7. Apply the Gram-Schmidt algorithm on the set:
   $$\{1, z, z^2\}$$
   under the inner product of Exercise 2.

For Exercises (8) to (12): Use your answers to Exercises (3) to (7) in order to find the an orthonormal basis for the orthogonal complement of the following subspaces W of the corresponding vector space. There should be no further computations necessary.

8. $W = Span(\{\langle i, 2-i \rangle\}) \trianglelefteq \mathbb{C}^2$, under the ordinary inner product.

9. $W = Span(\{\langle 1+i, 2, 3-i\rangle, \langle 1-i, 3+i, 2-i\rangle\}) \trianglelefteq \mathbb{C}^3$, under the ordinary inner product.

10. $W = Span(\{\langle i, 2+i, -i\rangle\}) \trianglelefteq \mathbb{C}^3$, under the ordinary inner product.

11. $W = Span(\{\langle i, 2-i\rangle\}) \trianglelefteq \mathbb{C}^2$ under the inner product of Exercise 1.

12. $W = Span(\{1, z\}) \trianglelefteq \mathbb{P}^2(\mathbb{C})$ under the inner product of Exercise 2.

For Exercises (13) to (17): Let $(V, \oplus, \odot)$ be a complex inner product space. Let $\vec{z}, \vec{w}, \vec{u} \in V$, and $k \in \mathbb{C}$. Using only the Axioms of a Complex Inner Product Space, prove that the following properties also hold:

13. ***The Right Additivity Property:***

$$\langle \vec{z} | \vec{w} + \vec{u} \rangle = \langle \vec{z} | \vec{w} \rangle + \langle \vec{z} | \vec{u} \rangle.$$

14. ***The Double-Conjugate Property:***

$$\overline{\left\langle \overline{\vec{z}} | \overline{\vec{w}} \right\rangle} = \langle \vec{z} | \vec{w} \rangle.$$

15. ***The Inner Product with the Zero Vector Property:***

$$\left\langle \vec{z} | \vec{\mathbf{0}}_V \right\rangle = 0_{\mathbb{C}} = \left\langle \vec{\mathbf{0}}_V | \vec{z} \right\rangle,$$

and in particular:

$$\left\langle \vec{\mathbf{0}}_V | \vec{\mathbf{0}}_V \right\rangle = 0_{\mathbb{C}}.$$

16. ***The Right Conjugate-Homogeneity Property:***

$$\langle \vec{z} | k \bullet \vec{w} \rangle = \bar{k} \bullet \langle \vec{z} | \vec{w} \rangle.$$

17. ***The Left Conjugate-Homogeneity Property:***

$$\langle k \bullet \vec{z} | \vec{w} \rangle = \left\langle \vec{z} | \bar{k} \bullet \vec{w} \right\rangle.$$

For Exercises (18) to (23): The following statements have analogous counterparts in Chapter 7. Prove them by mimicking the proofs of the counterparts in Chapter 7.

18. Let S be an orthogonal set of vectors from some complex inner product space V. Prove that S is linearly independent.

19. Let W be a subspace of a (possibly infinite dimensional) complex inner product space V, and let $W^\perp$ be its orthogonal complement. Prove that $W^\perp$ is again a subspace of V.

20. Let V be a finite dimensional complex inner product space, and let W be a subspace of V. If B is a basis for W, prove that:

$$W^\perp = \left\{ \vec{v} \in V | \langle \vec{v} | \vec{w} \rangle = 0_{\mathbb{C}} \text{ for all } \vec{w} \in B \right\},$$

that is, it is necessary and sufficient that we check if $\vec{v}$ is orthogonal to every member of a basis for W.

21. Prove that the Gram-Schmidt Algorithm is still valid on a complex inner product space V.

22. Prove the Dimension Theorem for Orthogonal Complements: If V is a ***finite-dimensional*** complex inner product space, and W is a subspace of V, then: $dim(W) + dim(W^\perp) = dim(V)$.

23. Let V be a ***finite dimensional*** complex inner product space, and let W be any subspace of V. Prove that:

$$(W^\perp)^\perp = W.$$

8.4 Complex Linear Transformations and The Adjoint

We shall continue with the task of generalizing the concepts of linear transformations, eigentheory and diagonalizability of vector spaces over arbitrary fields, and over the complex field in particular, when it is appropriate to do so.

Complex Linear Transformations

The concept of a linear transformation is easily extendible to vector spaces over an arbitrary field:

Definition: Suppose $(V, \oplus_V, \odot_V)$ and $(W, \oplus_W, \odot_W)$ are both vector spaces over the same field $(\mathbf{F}, +, \cdot)$. A ***linear transformation:***

$$T : V \to W,$$

is a ***function*** that assigns a unique vector $\vec{w} \in W$ to every vector $\vec{v} \in V$, such that for all $\vec{u}$ and $\vec{v} \in V$ and all scalars $c \in \mathbf{F}$, the function T satisfies:

$$T(\vec{u} \oplus_V \vec{v}) = T(\vec{u}) \oplus_W T(\vec{v}), \text{ and}$$
$$T(c \odot_V \vec{u}) = c \odot_W T(\vec{u}).$$

As before, we call V the ***domain*** of T and W the ***codomain*** of T. If $T : V \to V$, we call T a ***linear operator***. The two properties above are again known as ***additivity*** and ***homogeneity***. When the operations are clear, we will simply write these properties as:

$$T(\vec{u} + \vec{v}) = T(\vec{u}) + T(\vec{v}), \text{ and } T(c \cdot \vec{u}) = c \cdot T(\vec{u}).$$

In the case of complex linear transformations, the ones that are easiest to understand are those with domain $\mathbb{C}^n$ and codomain $\mathbb{C}^m$, for some positive integers n and m. In this case, the additivity and homogeneity properties again allow us to construct the ***standard matrix*** for T, namely the $m \times n$ complex matrix $[T]$, with ***columns:***

$$[T] = \left[\ T(\vec{e}_1) \,|\, T(\vec{e}_2) \,|\, \cdots \,|\, T(\vec{e}_n) \right].$$

The action of T on a vector $\vec{v} \in \mathbb{C}^n$ is given by ***matrix multiplication:***

$$[T(\vec{v})] = [T] \cdot [\vec{v}] \in \mathbb{C}^m,$$

where $[\vec{v}]$ is the $n \times 1$ coordinate matrix of $\vec{v}$ with respect to the standard basis of $\mathbb{C}^n$, and similarly, $[T(\vec{v})]$ is the $m \times 1$ coordinate matrix of $T(\vec{v})$ with respect to the standard basis of $\mathbb{C}^m$.

Once again, linear transformations ***preserve subspaces*** of the domain and the codomain: if $T : V \to W$, $U \trianglelefteq V$, and $Z \trianglelefteq W$, then:

$$T(U) \trianglelefteq W \text{ and } T^{-1}(Z) \trianglelefteq V.$$

In particular, we can define the ***kernel*** of T and the ***range*** of T as the subspaces:

$$ker(T) = T^{-1}\left(\left\{\vec{\mathbf{0}}_W\right\}\right) \trianglelefteq V \text{ and } range(T) = T(V) \trianglelefteq W.$$

We define:

$$nullity(T) = dim(ker(T)) \text{ and } rank(T) = dim(range(T)).$$

A linear transformation $T : V \to W$ is ***one-to-one*** ***if and only if*** $ker(T) = \{\vec{\mathbf{0}}_V\}$, and T is ***onto*** ***if and only if*** $range(T) = W$.

The Dimension Theorem will again tell us that for $T : V \to W$:

$$rank(T) + nullity(T) = dim(V).$$

Thus if $dim(V) < dim(W)$, then T cannot be onto, and if $dim(V) > dim(W)$, then T cannot be one-to-one.

An ***isomorphism*** $T : V \to W$ of vector spaces is a linear transformation that is both one-to-one and onto. Every complex space of dimension n is isomorphic to $\mathbb{C}^n$, and two finite-dimensional vector spaces V and W are isomorphic to each other ***if and only if*** $dim(V) = dim(W)$. More explicitly, if $dim(V) = n$ and $B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a basis for V, then the assignment:

$$T(\vec{v}_i) = \vec{e}_i,$$

for $i = 1..n$, extends by linearity to an isomorphism $T : V \to \mathbb{C}^n$.

Example: Suppose we have a linear transformation $T : \mathbb{C}^4 \to \mathbb{C}^3$. Then T is automatically ***not*** one-to-one. Suppose T is given by:

$$[T] = \begin{bmatrix} -i & 2 & 3+i & 9-5i \\ 1+2i & -4+2i & -3i & -10+8i \\ 3-i & 2+6i & 1+2i & 17+13i \end{bmatrix}, \text{ with rref: } \begin{bmatrix} 1 & 2i & 0 & 3+5i \\ 0 & 0 & 1 & 1-i \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Indeed, we find two free variables, and thus the kernel of T is:

$$ker(T) = Span(\{\langle -2i, 1, 0, 0\rangle, \langle -3-5i, 0, -1+i, 1\rangle\}).$$

and $nullity(T) = 2$. The leading 1's are in columns 1 and 3, so the range of T is:

$$range(T) = Span(\{\langle -i, 1+2i, 3-i\rangle, \langle 3+i, -3i, 1+2i\rangle\}).$$

Thus $rank(T) = 2 < 3$, so T is ***not*** onto, either. We verify that:

$$rank(T) + nullity(T) = 2 + 2 = 4,$$

as it should be, according to the Dimension Theorem.□

Eigentheory of Complex Linear Transformations

We can define the ***characteristic polynomial*** $p(\lambda)$ of an $n \times n$ complex matrix A, as before, as:

$$p(\lambda) = det(\lambda \boldsymbol{I}_n - A)$$

We obtain a polynomial of degree n with complex coefficients, and thanks to the Fundamental Theorem of Algebra, we can now say with confidence that $p(\lambda)$ has exactly n complex ***roots***, counting multiplicities. These roots are again called the ***eigenvalues*** of A, and a non-zero vector $\vec{v} \in \mathbb{C}^n$ is called an ***eigenvector*** for A with respect to λ, if:

$$A\vec{v} = \lambda\vec{v}.$$

If $T : \mathbb{C}^n \to \mathbb{C}^n$ is a linear operator, then $[T]$ is an $n \times n$ matrix so we can find its eigenvalues and eigenvectors. More generally, if V is a (possibly infinite dimensional) complex vector space and $T : V \to V$ is a linear transformation, we say that λ is an eigenvalue for T with associated (non-zero) eigenvector $\vec{v}$ if:

$$T(\vec{v}) = \lambda\vec{v}.$$

The set of all eigenvectors associated to an eigenvalue λ, together with the zero vector, form a subspace of V called the ***eigenspace*** $Eig(A, \lambda)$ or $Eig(T, \lambda)$. Thus:

$$Eig(T, \lambda) = \{\vec{v} \in V | T(\vec{v}) = \lambda\vec{v}\}.$$

Example: Let $A = \begin{bmatrix} 3+8i & 4-6i \\ 6+4i & -7i \end{bmatrix}$. Then:

$$\begin{aligned} p(\lambda) &= \begin{vmatrix} \lambda-3-8i & -4+6i \\ -6-4i & \lambda+7i \end{vmatrix} \\ &= (\lambda-3-8i)(\lambda+7i) - (-4+6i)(-6-4i) \\ &= \lambda^2 - (3+i)\lambda + 8 - i. \end{aligned}$$

We can find the roots of this quadratic polynomial using our method in Section 8.1. The discriminant is:

$$b^2 - 4ac = (3+i)^2 - 4(8-i) = -24 + 10i.$$

The length of this complex number is:

$$\|b^2 - 4ac\| = \sqrt{(-24)^2 + 10^2} = \sqrt{576 + 100} = 26$$

and thus:

$$b^2 - 4ac = 26\left(-\frac{12}{13} + \frac{5}{13}i\right),$$

which is now in polar form, with $\cos(\theta) = -\frac{12}{13}$ and $\sin(\theta) = \frac{5}{13}$, i.e. $\pi/2 < \theta < \pi$, and thus $\pi/4 < \theta/2 < \pi/2$. In other words, $\theta/2$ is in the ***first quadrant***. Thus, using the Half-Angle Formulas, we get:

$$\cos(\theta/2) = \sqrt{\frac{1 + \left(-\frac{12}{13}\right)}{2}} = \frac{1}{\sqrt{26}}, \quad \text{and}$$

$$\sin(\theta/2) = \sqrt{\frac{1 - \left(-\frac{12}{13}\right)}{2}} = \frac{5}{\sqrt{26}}.$$

Thus, we get the principal square root:

$$\sqrt{b^2 - 4ac} = \sqrt{26}\left(\frac{1}{\sqrt{26}} + \frac{5}{\sqrt{26}}i\right) = 1 + 5i,$$

and finally, our eigenvalues are:

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{3 + i \pm (1+5i)}{2} = 1 - 2i \text{ or } 2 + 3i.$$

To find the eigenvectors for each eigenvalue, we need to find the ***kernel*** of the matrix $\lambda \boldsymbol{I_2} - A$ (or $A - \lambda \boldsymbol{I_2}$) for the two values of λ.

For $\lambda = 1 - 2i$, we get:

$$(1-2i)\boldsymbol{I_2} - A = \begin{bmatrix} -2-10i & -4+6i \\ -6-4i & 1+5i \end{bmatrix}, \text{ with rref: } \begin{bmatrix} 1 & -\frac{1}{2}-\frac{1}{2}i \\ 0 & 0 \end{bmatrix},$$

so $Eig(A, 1-2i)$ is 1-dimensional, with basis, say $\{\langle 1+i, 2\rangle\}$.
For $\lambda = 2+3i$, we get:

$$(2+3i)\boldsymbol{I_2} - A = \begin{bmatrix} -1-5i & -4+6i \\ -6-4i & 2+10i \end{bmatrix}, \text{ with rref: } \begin{bmatrix} 1 & -1-i \\ 0 & 0 \end{bmatrix},$$

so $Eig(A, 2+3i)$ is 1-dimensional again, with basis, say $\{\langle 1+i, 1\rangle\}$. □

The Spectrum of an Operator

The highlight of this Chapter is Section 8.6, where we present the ***Spectral Theorems***. Their collective name comes from the following:

> ***Definition:*** Let $T : V \to V$ be a linear operator on a (possibly infinite-dimensional) complex vector space V. The ***spectrum*** of T, denoted ***Spec(T)*** is the set of all eigenvalues of T, thus:
> $$Spec(T) = \left\{\lambda \in \mathbb{C} \,|\, T(\vec{v}) = \lambda\vec{v} \text{ for some } \textbf{\textit{non-zero}} \text{ vector } \vec{v} \in V\right\}.$$
> If $V = \mathbb{C}^{\boldsymbol{n}}$, and $A = [T]$, then we write:
> $$Spec(T) = Spec(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\},$$
> namely, the set of ***distinct eigenvalues*** of T, where $1 \leq k \leq n$.

Example: Let $A = \begin{bmatrix} 3+8i & 4-6i \\ 6+4i & -7i \end{bmatrix}$, from our previous Example. For this matrix:

$$Spec(A) = \{1-2i, 2+3i\}. \;\square$$

The Adjoint of a Matrix

We now introduce an important operation on complex matrices and linear transformations: we will perform an extra "twist" on the transpose:

> ***Definition:*** Let A be an $n \times n$ complex matrix. We define the ***adjoint*** of A, written as A^*, as:
> $$A^* = \overline{A^\top}.$$
> that is, we take the ***complex conjugate*** of each entry of $A^\top$. Similarly, if $T : \mathbb{C}^{\boldsymbol{n}} \to \mathbb{C}^{\boldsymbol{n}}$ is a linear operator with standard matrix $[T]$, we denote by T^*, the ***adjoint*** of T, as the linear operator on $\mathbb{C}^{\boldsymbol{n}}$ such that $[T^*] = [T]^*$.

This is also called the ***Hermitian adjoint*** or the ***Hermitian transpose*** of A. Notice that if A is a pure real matrix, then A^* is again $A^\top$. The adjoint is the matrix analog of ***complex conjugation***. In the area of ***Quantum Mechanics***, the adjoint is sometimes written as $A^\dagger$ or A^H, although it is usually in the context of operators on an infinite-dimensional complex vector space.

Example: Let $A = \begin{bmatrix} 7-4i & 2+i \\ 5 & 3i \end{bmatrix}$. Then:

$$A^* = \overline{\begin{bmatrix} 7-4i & 5 \\ 2+i & 3i \end{bmatrix}} = \begin{bmatrix} 7+4i & 5 \\ 2-i & -3i \end{bmatrix}. \square$$

Due to its connection with the transpose operation as well as the complex conjugation, the adjoint also shares properties that are possessed by these two operations:

Theorem — Properties of the Adjoint:
Let A and B be $n \times n$ complex matrices and $c \in \mathbb{C}$. Then the following properties hold:

1. ***The Double Adjoint Property:*** $(A^*)^* = A$.
2. ***The Additivity Property:*** $(A+B)^* = A^* + B^*$.
3. ***The Conjugate-Homogeneity Property:*** $(c \cdot A)^* = \bar{c} \cdot A^*$.
4. ***The Anti-Commutativity Property:*** $(A \cdot B)^* = B^* \cdot A^*$.

The proofs of these properties are left as Exercises. The adjoint of A also has an interesting and useful property in connection with the complex inner product:

Theorem: Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be any linear operator. Then, for all $\vec{v}, \vec{w} \in \mathbb{C}^n$:

$$\langle T(\vec{v}) | \vec{w} \rangle = \langle \vec{v} | T^*(\vec{w}) \rangle.$$

Proof: Recall that in Section 7.6, the dot product can be viewed as a ***matrix product:***

$$\vec{v} \circ \vec{w} = [\vec{w}]^\top \cdot [\vec{v}],$$

where $[\vec{v}]$ and $[\vec{w}]$ are column matrices corresponding to the real vectors $\vec{v}$ and $\vec{w}$. In the complex case:

$$\langle \vec{v} | \vec{w} \rangle = \vec{v} \circ \overline{\vec{w}} = \overline{[\vec{w}]}^\top \cdot [\vec{v}] = [\vec{w}]^* \cdot [\vec{v}].$$

Thus:

$$\begin{aligned} \langle T(\vec{v}) | \vec{w} \rangle &= [\vec{w}]^* \cdot [T(\vec{v})] \\ &= [\vec{w}]^* \cdot [T] \cdot [\vec{v}], \text{ and} \\ \langle \vec{v} | T^*(\vec{w}) \rangle &= [T^*(\vec{w})]^* \cdot [\vec{v}] \\ &= ([T^*] \cdot [\vec{w}])^* \cdot [\vec{v}] \\ &= ([T]^* \cdot [\vec{w}])^* \cdot [\vec{v}] \\ &= [\vec{w}]^* \cdot ([T]^*)^* \cdot [\vec{v}] \\ &= [\vec{w}]^* \cdot [T] \cdot [\vec{v}]. \end{aligned}$$

Hence, the two sides are equal. ■

Informally, we say that if we ***slide*** an operator from the left side to the right side of the inner product, it becomes its adjoint:

$$\langle T(\vec{v})|\vec{w}\rangle = \langle\vec{v}|T^*(\vec{w})\rangle,$$

that is, T on the left becomes T^* on the right.

Similarly, in terms of matrices, we have:

$$\left\langle A \cdot [\vec{v}] \,|\, [\vec{w}] \right\rangle = \left\langle [\vec{v}] \,|\, A^* \cdot [\vec{w}] \right\rangle.$$

Diagonalization

Since we are now allowed to have imaginary eigenvalues and complex eigenvectors from $\mathbb{C}^n$, we can expand the definition of diagonalizability and allow more matrices with real entries to be diagonalized.

Let A be an $n \times n$ complex matrix. We say that A is ***diagonalizable over*** $\mathbb{C}$ if we can find an ***invertible*** complex matrix C such that:

$$C^{-1}AC = D,$$

where $D = Diag\left(\alpha_1, \alpha_2, \ldots, \alpha_n\right)$ is a diagonal matrix, or equivalently:

$$AC = CD \quad \text{or} \quad A = CDC^{-1}.$$

We also say that C ***diagonalizes*** A over $\mathbb{C}$.

Let A be an $n \times n$ matrix. Then, A is ***diagonalizable*** ***if and only if*** we can find a set of n ***linearly independent eigenvectors*** for A, say $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$. If this is the case, then the diagonalizing matrix C is again the matrix whose ***columns*** are $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$, and the diagonal matrix D contains the corresponding ***eigenvalues*** along the main ***diagonal***. This tells us that if all the eigenvalues of A were real numbers, and if A cannot be diagonalized over the field of real numbers, then A cannot be diagonalized either over the field of complex numbers. However, if a ***real*** matrix had imaginary eigenvalues, we can now attempt to find a complete set of linearly independent ***complex*** eigenvectors.

Let $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ be an ordered set of eigenvectors for an $n \times n$ matrix A, and suppose that the corresponding eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ for these eigenvectors are all ***distinct***. Then, S is again ***linearly independent***. Thus, if A has a total of m distinct eigenvalues, we can find ***at least m*** linearly independent eigenvectors for A.

Furthermore, suppose the characteristic polynomial $p(\lambda)$ of A factors as:

$$p(\lambda) = (\lambda - \lambda_1)^{n_1} \cdot (\lambda - \lambda_2)^{n_2} \cdot \cdots \cdot (\lambda - \lambda_k)^{n_k},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are distinct and $n_1 + n_2 + \cdots + n_k = n$. As before, we call the exponent n_i the ***algebraic multiplicity*** of λ_i, and we call:

$$dim(Eig(A, \lambda_i))$$

(the dimension of the eigenspace over $\mathbb{C}$) the ***geometric multiplicity*** of λ_i.

For any eigenvalue λ_i of an $n \times n$ matrix A, the ***geometric multiplicity*** of λ_i is ***at most*** equal to the ***algebraic multiplicity*** of λ_i. Thus A is diagonalizable ***if and only if*** for every eigenvalue λ_i of A, the geometric multiplicity of λ_i is ***exactly equal*** to its algebraic multiplicity. Thus, if A has n ***distinct eigenvalues***, then A is certainly diagonalizable.

Example: Let $A = \begin{bmatrix} 3+8i & 4-6i \\ 6+4i & -7i \end{bmatrix}$.

We saw above that the two eigenvalues of A are distinct, and thus A is diagonalizable. In particular, we found that:

$$Eig(A, 1-2i) = Span(\{\langle\, 1+i, 2\,\rangle\}), \quad \text{and}$$
$$Eig(A, 2+3i) = Span(\{\langle\, 1+i, 1\,\rangle\}).$$

Thus, we have:

$$D = \begin{bmatrix} 1-2i & 0 \\ 0 & 2+3i \end{bmatrix}, \quad \text{and}$$

$$C = \begin{bmatrix} 1+i & 1+i \\ 2 & 1 \end{bmatrix}, \quad \text{with } C^{-1} = \begin{bmatrix} -\frac{1}{2}+\frac{1}{2}i & 1 \\ 1-i & -1 \end{bmatrix}$$

Indeed, we can check that:

$$\begin{aligned} CDC^{-1} &= \begin{bmatrix} 1+i & 1+i \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 1-2i & 0 \\ 0 & 2+3i \end{bmatrix}\begin{bmatrix} -\frac{1}{2}+\frac{1}{2}i & 1 \\ 1-i & -1 \end{bmatrix} \\ &= \begin{bmatrix} 3-i & -1+5i \\ 2-4i & 2+3i \end{bmatrix}\begin{bmatrix} -\frac{1}{2}+\frac{1}{2}i & 1 \\ 1-i & -1 \end{bmatrix} = \begin{bmatrix} 3+8i & 4-6i \\ 6+4i & -7i \end{bmatrix} \\ &= A. \square \end{aligned}$$

8.4 Section Summary

We can generalize the concepts of:

- a ***linear transformation*** from one complex vector space to another;
- the ***matrix*** of a linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$;
- the ***kernel***, ***range***, ***nullity*** and ***rank*** of a complex linear transformation;
- ***one-to-one*** linear transformations, ***onto*** linear transformations, and ***isomorphisms***;
- the ***characteristic polynomial*** and ***equation*** of an $n \times n$ complex matrix;
- the ***eigenvalues*** and associated ***eigenvectors*** of an $n \times n$ complex matrix;
- the ***diagonalizability*** of an $n \times n$ complex matrix.

Let $T : V \to V$ be a linear operator on a (possibly infinite-dimensional) complex vector space V. The ***spectrum*** of T, denoted ***Spec(T)*** is the set of all eigenvalues of T, thus:

$$Spec(T) = \left\{\lambda \in \mathbb{C} \mid T(\vec{v}) = \lambda\vec{v} \text{ for some } \textbf{\textit{non-zero}} \text{ vector } \vec{v} \in V\right\}.$$

If $V = \mathbb{C}^n$, and $A = [T]$, then we write: $Spec(T) = Spec(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$, namely, the set of ***distinct eigenvalues*** of T, where $1 \leq k \leq n$.

Let A be an $n \times n$ complex matrix. We define the ***adjoint*** of A, written as A^*, as: $A^* = \overline{A^\top}$. That is, we take the complex conjugate of each entry of $A^\top$. Similarly, if $T : \mathbb{C}^n \to \mathbb{C}^n$ is a linear operator with standard matrix $[T]$, we denote by T^* the linear operator on $\mathbb{C}^n$ such that $[T^*] = [T]^*$. The adjoint operation enjoys similar properties as the transpose, with the notable exception that $(w \cdot A)^* = \overline{w} \cdot A^*$.

8.4 Exercises

For Exercises (1) to (9): Find a basis for the (a) kernel, and (b) range of the linear transformation T using its rref; decide if T is (c) one-to-one and/or (d) onto, and (e) if T is an isomorphism, find its inverse:

1. $T : \mathbb{C}^2 \to \mathbb{C}^2$; $[T] = \begin{bmatrix} 3-i & 3-11i \\ 5+2i & 16-11i \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 2-3i \\ 0 & 0 \end{bmatrix}$

2. $T : \mathbb{C}^2 \to \mathbb{C}^2$; $[T] = \begin{bmatrix} 2+i & 3 \\ 2i & 4+i \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

3. $T : \mathbb{C}^2 \to \mathbb{C}^3$; $[T] = \begin{bmatrix} -2i & 2-6i \\ 2+i & 5i \\ 1-3i & 9 \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$

4. $T : \mathbb{C}^3 \to \mathbb{C}^2$; $[T] = \begin{bmatrix} 3-i & 3-11i & 2i \\ 4+3i & 17-6i & 3 \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 2-3i & 0 \\ 0 & 0 & 1 \end{bmatrix}$

5. $T : \mathbb{C}^2 \to \mathbb{C}^3$; $[T] = \begin{bmatrix} 2+i & -2+4i \\ 3 & 6i \\ 3-5i & 10+6i \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 2i \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$

6. $T : \mathbb{C}^3 \to \mathbb{C}^2$; $[T] = \begin{bmatrix} 2 & -i & 3i \\ 3i & 2 & -2 \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 0 & 4i \\ 0 & 1 & 5 \end{bmatrix}$

7. $T : \mathbb{C}^3 \to \mathbb{C}^3$; $[T] = \begin{bmatrix} 3+2i & 13 & 5 \\ 2-i & 4-7i & 3i \\ 3i & 6+9i & -2 \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 3-2i & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$

8. $T : \mathbb{C}^3 \to \mathbb{C}^3$; $[T] = \begin{bmatrix} 1-i & 2i & 4+8i \\ 3+i & 2-i & 3+i \\ 3i & 5 & 9-5i \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 0 & 2i \\ 0 & 1 & 3-i \\ 0 & 0 & 0 \end{bmatrix}$

9. $T : \mathbb{C}^3 \to \mathbb{C}^3$; $[T] = \begin{bmatrix} 1-i & i & -i \\ 2i & 2-i & i \\ 3 & -1 & 1+i \end{bmatrix}$, with rref $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

For Exercises (10) and (11): For the given matrices A, find the (a) Hermitian adjoint, (b) characteristic polynomial, (c) spectrum, and (d) a basis for each eigenspace of the matrix; (e) decide if A is diagonalizable, and if so, find a diagonal matrix D and an invertible matrix C such that $CDC^{-1} = A$.

10. $A = \begin{bmatrix} -43-30i & 4+97i \\ -30+12i & 51+31i \end{bmatrix}$

11. $A = \begin{bmatrix} 3-7i & -9+9i \\ -1-7i & -3+11i \end{bmatrix}$

12. We saw in Chapter 6 that the rotation matrices:

$$[rot_\theta] = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

do not have any real eigenvalues unless $\theta = n\pi$ for some integer n.

a. Find the (imaginary) eigenvalues of this matrix when $\theta \neq n\pi$.

b. Find a basis for the eigenspace of each eigenvalue.

c. Show that these matrices are diagonalizable over $\mathbb{C}$.

13. Consider $\mathbb{C}$ as a 2-dimensional vector space over $\mathbb{R}$. Define the functions:

$$Re : \mathbb{C} \to \mathbb{R} \quad \text{and} \quad Im : \mathbb{C} \to \mathbb{R} \quad \text{via:}$$
$$Re(x+iy) = x \quad \text{and} \quad Im(x+iy) = y$$

Show that Re and Im are linear transformations of ***real*** vector spaces. These symbols stand for the *Real* part of $x+iy$ and the *Imaginary* part of $x+iy$, respectively.

14. Suppose that $T : V \to W$ is a one-to-one linear transformation of complex vector spaces, and suppose also that W is an inner product space under $\langle | \rangle_W$. Prove that we can induce an inner product on V, denoted $\langle | \rangle_V$, via:

$$\langle \vec{z} | \vec{w} \rangle_V = \langle T(\vec{z}) | T(\vec{w}) \rangle_W.$$

For Exercises (15) to (18): Let A and B be $n \times n$ complex matrices and $w \in \mathbb{C}$. Show that the following properties hold:

15. ***The Double Adjoint Property:*** $(A^*)^* = A$.

16. ***The Additivity Property:*** $(A+B)^* = A^* + B^*$

17. ***The Conjugate-Homogeneity Property:*** $(w \cdot A)^* = \overline{w} \cdot A^*$

18. ***The Anti-Commutativity Property:*** $(A \bullet B)^* = B^* \bullet A^*$

19. Prove that if $T : \mathbb{C}^n \to \mathbb{C}^n$ is any linear operator, then for all $\vec{v}, \vec{w} \in \mathbb{C}^n$:

$$\left\langle T^*(\vec{v}) \,|\, \vec{w} \right\rangle = \left\langle \vec{v} \,|\, T(\vec{w}) \right\rangle.$$

20. Let V be the complex vector space of ***all polynomials*** $p(z)$ in the complex variable z, with complex coefficients. Show that:

$$T : V \to V, \text{ given by}$$
$$T(p(z)) = z \bullet p(z),$$

is a linear transformation, but T does not have any eigenvalues or eigenvectors, whether real or complex. In other words, $Spec(T) = \emptyset$.

21. Generalize the previous Exercise: Suppose that $q(z)$ is a ***fixed*** non-constant polynomial, and V is the space of ***all polynomials*** in z, as before. Show that:

$$T : V \to V, \text{ given by}$$
$$T(p(z)) = q(z) \bullet p(z),$$

is a linear transformation, but T does not have any eigenvalues or eigenvectors, whether real or complex.

22. Let S be the complex vector space of ***all sequences*** of complex numbers:

$$\{z_1, z_2, z_3, \ldots\}$$

under addition of sequences and the natural scalar multiplication:

$$w \odot \{z_1, z_2, z_3, \ldots\} = \{w \bullet z_1, w \bullet z_2, w \bullet z_3, \ldots\}.$$

Let T be the ***deleting*** linear operator, given by:

$$T : S \to S, \quad \text{where:}$$
$$T(\{z_1, z_2, z_3, \ldots\}) = \{z_2, z_3, z_4, \ldots\}.$$

This means we remove the first term in the sequence.

Prove that T is indeed a linear transformation, and that furthermore:

$$Spec(T) = \mathbb{C}.$$

This means that for ***any*** complex number λ, there exists a non-zero sequence $\{z_1, z_2, z_3, \ldots\}$ such that:

$$T(\{z_1, z_2, z_3, \ldots\}) = \lambda \bullet \{z_1, z_2, z_3, \ldots\}, \text{ that is:}$$
$$\{z_2, z_3, z_4, \ldots\} = \lambda \bullet \{z_1, z_2, z_3, \ldots\}.$$

Hint: think of a common type of sequence that you see in Precalculus, but generalize it to complex numbers. Note that we are not requiring these sequences to “converge” in any way; they are simply infinite ***ordered lists*** of complex numbers.

8.5 Normal Matrices

We will now look at a special set of matrices that are diagonalizable in a certain way, as we shall see in the next Section, where we see the Spectral Theorems. They are collectively called ***normal matrices***, but within this group, there are more specific subcategories, such as Hermitian, Skew-Hermitian and unitary matrices. Let us start with the first two kinds of these special matrices:

Hermitian and Skew-Hermitian Matrices

We will now extend the definition of symmetric and skew-symmetric matrices to complex matrices:

> ***Definition:*** We say that an $n \times n$ complex matrix A is ***Hermitian*** if:
> $$A^* = A,$$
> and similarly, A is ***Skew-Hermitian*** if:
> $$A^* = -A.$$
> For linear operators $T : \mathbb{C}^n \to \mathbb{C}^n$, these conditions translate to $T^* = T$ and $T^* = -T$, respectively.

It is easy to see that if A is a pure real matrix, then A is Hermitian precisely when it is symmetric, and A is Skew-Hermitian precisely when it is skew-symmetric. In the same way that we can visually check if a matrix is symmetric or skew-symmetric, we have the following Theorem, whose proof is left as an Exercise:

> ***Theorem:*** Let A be an $n \times n$ complex matrix. Let us refer to the entries $a_{j,k}$ and $a_{k,j}$, where $j \neq k$, as ***off-diagonal pairs***, and the entries $a_{i,i}$ as the diagonal entries, as before. Then:
> 1. A is ***Hermitian*** ***if and only if*** the diagonal entries of A are all ***pure real***, and the off-diagonal pairs are ***complex-conjugate pairs***.
> 2. A is ***Skew-Hermitian*** ***if and only if*** the diagonal entries of A are all ***pure imaginary***, and the off-diagonal pairs are ***negative complex-conjugate pairs***.

Example: The matrix B is Hermitian, and C is Skew-Hermitian:

$$B = \begin{bmatrix} 3 & 5-2i & 7+3i \\ 5+2i & 6 & i/2 \\ 7-3i & -i/2 & -4 \end{bmatrix}; \quad C = \begin{bmatrix} 7i & 7+2i & -6 \\ -7+2i & -i/3 & 3-4i \\ 6 & -3-4i & 5i \end{bmatrix}. \square$$

Notice that in the Skew-Hermitian matrix C, the off-diagonal pairs have the same imaginary part but the real parts are opposite in sign. These are also called ***real-conjugate pairs***.$\square$

Recall that real symmetric and skew-symmetric matrices in fact form ***subspaces*** of the space of all $n \times n$ real matrices. Unfortunately, their complex analogs do not form a subspace of $Mat(\mathbb{C}, n, n)$ over $\mathbb{C}$: Hermitian and Skew-Hermitian $n \times n$ matrices are closed under addition, but ***not*** under scalar multiplication ***by a complex number***. However, we can say the following:

Theorem: The set of all $n \times n$ Hermitian matrices, denoted ***Herm(n)***, is a ***real vector space***. This means that if A and B are $n \times n$ ***Hermitian*** matrices and r is a ***pure real*** number, then $A + B$ and $r \cdot A$ are again ***Hermitian***.

Furthermore, $Herm(n)$ is closed under the ***transpose*** and ***adjoint*** operations: $A^\top$ and A^* are again ***Hermitian***.

Similarly, the set of all $n \times n$ Skew-Hermitian matrices, denoted ***SkewHerm(n)***, is also a ***real vector space***. This means that if C and D are $n \times n$ ***Skew-Hermitian*** matrices and r is a ***pure real*** number, then $C + D$ and $r \cdot C$ are again ***Skew-Hermitian***.

Furthermore, $SkewHerm(n)$ is closed under the ***transpose*** and ***adjoint*** operations: $C^\top$ and C^* are again ***Skew-Hermitian***.

The proofs are left as Exercises. Notice that $(a + bi)A$ is not necessarily Hermitian if $b \neq 0$.

Let us pause for a minute and compare some properties we have seen so far: A complex number z is pure real if and only if $z = \bar{z}$, and z is pure imaginary if and only if $z = -\bar{z}$. Similarly, A is Hermitian if and only if $A = A^*$, and A is Skew-Hermitian if and only if $A = -A^*$. Thus, we can say, philosophically, that:

Hermitian matrices are analogous to pure real numbers.

Skew-Hermitian matrices are analogous to pure imaginary numbers.

Because of the connection between the adjoint and the inner product, we can describe these special linear transformations as follows:

Theorem: Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation. Then:

1. T is ***Hermitian*** ***if and only if*** for all $\vec{v}, \vec{w} \in \mathbb{C}^n$:
$$\langle T(\vec{v}) | \vec{w} \rangle = \langle \vec{v} | T(\vec{w}) \rangle.$$
2. T is ***Skew-Hermitian*** ***if and only if*** for all $\vec{v}, \vec{w} \in \mathbb{C}^n$:
$$\langle T(\vec{v}) | \vec{w} \rangle = -\langle \vec{v} | T(\vec{w}) \rangle.$$

Proof: For any linear operator T, we have:

$$\begin{aligned} \langle T(\vec{v}) | \vec{w} \rangle &= [\vec{w}]^* \cdot [T(\vec{v})] \\ &= [\vec{w}]^* \cdot [T] \cdot [\vec{v}]. \end{aligned}$$

On the other hand:

$$\begin{aligned} \langle \vec{v} | T(\vec{w}) \rangle &= [T(\vec{w})]^* \cdot [\vec{v}] \\ &= ([T] \cdot [\vec{w}])^* \cdot [\vec{v}] \\ &= [\vec{w}]^* \cdot [T]^* \cdot [\vec{v}]. \end{aligned}$$

Let us prove Part 1. If T is Hermitian, then $T = T^*$, or $[T] = [T]^*$, so the two indicated inner products are indeed equal by our computations above. For the converse, suppose the two indicated inner products are equal for ***all*** $\vec{v}, \vec{w} \in \mathbb{C}^n$. Subtracting, we get:

$$0 = [\vec{w}]^* \cdot ([T] - [T]^*) \cdot [\vec{v}] = \langle ([T] - [T]^*)\vec{v} | \vec{w} \rangle$$

for all $\vec{v},\ \vec{w} \in \mathbb{C}^n$. In particular, it is true if $\vec{w} = ([T]-[T]^*)\vec{v}$. But this says that the ***length*** of $([T]-[T]^*)\vec{v}$ is zero for ***all*** $\vec{v} \in \mathbb{C}^n$. In other words, $([T]-[T]^*)\vec{v} = \vec{0}$ for all $\vec{v} \in \mathbb{C}^n$. But this is possible ***if and only if*** $[T]-[T]^*$ is the zero linear transformation, i.e. $[T] = [T]^*$. The proof for Part 2 only requires a slight modification of this proof due to the $(-)$ sign.■

Note that the real analog of this Theorem says that the standard matrix $[T]$ of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is ***symmetric*** if and only if:

$$T(\vec{v}) \circ \vec{w} = \vec{v} \circ T(\vec{w})$$

for all $\vec{v},\ \vec{w} \in \mathbb{R}^n$. We informally say that T ***slides back and forth*** the dot product.

The Theorem can also be used to ***define*** Hermitian and Skew-Hermitian operators on ***infinite dimensional*** complex inner product spaces V. We say that $T : V \to V$ is ***Hermitian*** ***if and only if*** for all $\vec{v},\ \vec{w} \in V$:

$$\langle T(\vec{v}) | \vec{w} \rangle = \langle \vec{v} | T(\vec{w}) \rangle,$$

and similarly, T is ***Skew-Hermitian*** ***if and only if:***

$$\langle T(\vec{v}) | \vec{w} \rangle = -\langle \vec{v} | T(\vec{w}) \rangle.$$

The determinant and eigenvalues of Hermitian and Skew-Hermitian matrices have special properties:

Theorem: Let A be an $n \times n$ ***Hermitian*** matrix. Then: the ***determinant*** and ***eigenvalues*** of A are all ***pure real numbers***.

Let B be an $n \times n$ ***Skew-Hermitian*** matrix. Then: the ***determinant*** of B is a ***pure real number*** if n is ***even***, and a ***pure imaginary number*** if n is ***odd***. However, the ***eigenvalues*** of B are ***pure imaginary numbers***, regardless of whether n is odd or even.

Proof: Recall that for any square matrix A (whether real or complex):

$$det(A) = det(A^\top).$$

Similarly, from the general definition of the determinant and the additivity and multiplicativity of the conjugate operation, we can see that:

$$det\left(\overline{A}\right) = \overline{det(A)}.$$

Now, if A is Hermitian, then $A = A^* = \overline{A^\top}$. Thus:

$$det(A) = det\left(\overline{A^\top}\right) = \overline{det(A^\top)} = \overline{det(A)}.$$

Thus, $det(A)$ is a pure real number by the Test for Pure Real Numbers.

Now, let λ be an eigenvalue of A, and T the linear operator with matrix A. By the previous Theorem:

$$\langle T(\vec{v}) | \vec{w} \rangle = \langle \vec{v} | T(\vec{w}) \rangle$$

for all vectors $\vec{v},\ \vec{w} \in \mathbb{C}^n$. But now, let $\vec{v}$ be an eigenvector associated to λ, and let $\vec{w} = \vec{v}$. We get:

$$\langle T(\vec{v}) | \vec{v} \rangle = \langle \vec{v} | T(\vec{v}) \rangle, \text{ and thus:}$$
$$\langle \lambda\vec{v} | \vec{v} \rangle = \langle \vec{v} | \lambda\vec{v} \rangle, \text{ or}$$
$$\lambda\langle \vec{v} | \vec{v} \rangle = \overline{\lambda}\langle \vec{v} | \vec{v} \rangle.$$

Note that we used the ***left-homogeneity*** and ***right-conjugate-homogeneity*** properties of the inner product. But since an eigenvector $\vec{v}$ is a ***non-zero*** vector, $\langle \vec{v}, \vec{v} \rangle > 0$, and thus we can divide it out of both sides of the equation to get $\lambda = \bar{\lambda}$. Thus λ is pure real. The proofs for the statements regarding Skew-Hermitian matrices are similar and left as Exercises.■

We remark that this Theorem directly proves that the eigenvalues of a real ***symmetric*** matrix are ***pure real numbers***, as we stated in Chapter 7, since they are also Hermitian matrices.

Example: Consider:

$$A = \begin{bmatrix} 1 & 2-3i \\ 2+3i & 5 \end{bmatrix}.$$

By our criteria above, A is ***Hermitian***. Its characteristic polynomial is:

$$\begin{aligned} p(\lambda) &= \begin{vmatrix} \lambda-1 & -2+3i \\ -2-3i & \lambda-5 \end{vmatrix} \\ &= (\lambda-1)(\lambda-5) - (-2+3i)(-2-3i) \\ &= \lambda^2 - 6\lambda + 5 - (4+9) \\ &= \lambda^2 - 6\lambda - 8. \end{aligned}$$

Using the (ordinary) quadratic formula, we get as our eigenvalues $\lambda = 3 \pm \sqrt{17}$, which are, as expected, pure real.□

Unitary Matrices

Recall that in Chapter 6, we called a square matrix A ***orthogonal*** if $A \cdot A^\top = \boldsymbol{I_n} = A^\top \cdot A$. The following matrix type is the complex analogue:

Definition: We say that an $n \times n$ complex matrix A is ***unitary*** if:

$$A^* \cdot A = \boldsymbol{I_n} = A \cdot A^*,$$

that is, A is ***invertible***, and its ***adjoint*** is also its ***inverse***. Equivalently, a linear operator T is unitary if $T^* \circ T = \boldsymbol{I}_{\mathbb{C}^n} = T \circ T^*$.

In the same way that the rows and columns of an orthogonal matrix form orthonormal sets, the rows and columns of unitary matrices form orthonormal sets, but under our new complex Euclidean inner product. We leave the proof of this generalization as an Exercise:

Theorem: The following conditions are equivalent regarding an $n \times n$ complex matrix A:

1. A is ***unitary***.
2. The ***rows*** of A form an ***orthonormal*** set of vectors from $\mathbb{C}^n$ under the complex Euclidean inner product.
3. The ***columns*** of A form an ***orthonormal*** set of vectors from $\mathbb{C}^n$ under the complex Euclidean inner product.

Example: Consider the matrix:

$$A = \begin{bmatrix} \frac{1}{3} - \frac{2}{3}i & \frac{2}{3}i \\ \frac{2}{3}i & \frac{1}{3} + \frac{2}{3}i \end{bmatrix}.$$

If we form the matrix product $A \cdot A^*$, we get:

$$\begin{aligned} &A \cdot A^* \\ &= \begin{bmatrix} \frac{1}{3} - \frac{2}{3}i & \frac{2}{3}i \\ \frac{2}{3}i & \frac{1}{3} + \frac{2}{3}i \end{bmatrix} \begin{bmatrix} \frac{1}{3} + \frac{2}{3}i & -\frac{2}{3}i \\ -\frac{2}{3}i & \frac{1}{3} - \frac{2}{3}i \end{bmatrix} \\ &= \begin{bmatrix} \left(\frac{1}{3} - \frac{2}{3}i\right)\left(\frac{1}{3} + \frac{2}{3}i\right) + \frac{2}{3}i\left(-\frac{2}{3}i\right) & \left(\frac{1}{3} - \frac{2}{3}i\right)\left(-\frac{2}{3}i\right) + \frac{2}{3}i\left(\frac{1}{3} - \frac{2}{3}i\right) \\ \frac{2}{3}i\left(\frac{1}{3} + \frac{2}{3}i\right) + \left(\frac{1}{3} + \frac{2}{3}i\right)\left(-\frac{2}{3}i\right) & \frac{2}{3}i\left(-\frac{2}{3}i\right) + \left(\frac{1}{3} + \frac{2}{3}i\right)\left(\frac{1}{3} - \frac{2}{3}i\right) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{9} + \frac{4}{9} + \frac{4}{9} & 0 \\ 0 & \frac{4}{9} + \frac{1}{9} + \frac{4}{9} \end{bmatrix} \\ &= \boldsymbol{I}_2. \end{aligned}$$

Thus, A is unitary.□

Just like Hermitian and Skew-Hermitian matrices, we can characterize unitary matrices using the inner product:

> ***Theorem:*** Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation. Then $[T]$ is ***unitary*** ***if and only if*** for all $\vec{v}, \vec{w} \in \mathbb{C}^n$:
>
> $$\langle \vec{v} | \vec{w} \rangle = \langle T(\vec{v}) | T(\vec{w}) \rangle.$$

The proof is very similar to the analogous Theorem on Hermitian matrices, and is left as an Exercise. Similarly, the determinant and eigenvalues of unitary matrices have special properties:

> ***Theorem:*** Let A be an $n \times n$ ***unitary*** matrix. Then: the ***complex norm*** of $det(A)$ is 1, and ***all*** the ***eigenvalues*** λ_i of A also have ***complex norm*** 1, that is:
>
> $$\|det(A)\| = 1, \quad \text{and} \quad \|\lambda_i\| = 1 \text{ for all } i = 1 \ldots n.$$

This Theorem justifies the root word ***unit*** in ***unitary***.

Because of the ***multiplicative*** way by which we defined unitary matrices, i.e. $A^* \cdot A = \boldsymbol{I}_n = A \cdot A^*$, we should not expect these matrices to form a subspace of $Mat(\mathbb{C}, n, n)$. In particular, the sum of two unitary matrices is not necessarily unitary. However, they do enjoy some ***multiplicative*** properties instead:

Theorem: Let A and B be ***unitary*** $n \times n$ matrices. Then:
1. $A \cdot B$ is also ***unitary***.
2. A is invertible and $A^{-1} = A^*$ is also ***unitary***.
3. If $\alpha = \cos(\theta) + \sin(\theta)i$ is a ***unit complex scalar*** for some real number θ, then $\alpha \cdot A$ is also ***unitary***.

Again, the proofs are left as Exercises. All the properties above lead us to say, philosophically, that:

Unitary matrices are analogous to complex numbers of unit length.

Example: The unitary matrix:

$$A = \begin{bmatrix} \frac{1}{3} - \frac{2}{3}i & \frac{2}{3}i \\ \frac{2}{3}i & \frac{1}{3} + \frac{2}{3}i \end{bmatrix}$$

from our previous Example, has determinant:

$$\left(\frac{1}{3} - \frac{2}{3}i\right)\left(\frac{1}{3} + \frac{2}{3}i\right) - \left(\frac{2}{3}i\right)\left(\frac{2}{3}i\right) = 1.$$

Its characteristic polynomial is:

$$p(\lambda) = \left(\lambda - \frac{1}{3} + \frac{2}{3}i\right)\left(\lambda - \frac{1}{3} - \frac{2}{3}i\right) - \left(\frac{2}{3}i\right)\left(\frac{2}{3}i\right) = \lambda^2 - \frac{2}{3}\lambda + 1.$$

Using the (ordinary) quadratic formula, its eigenvalues are:

$$\lambda = \frac{1}{3} \pm \frac{2}{3}i,$$

and both eigenvalues indeed have length:

$$\|\lambda\| = \sqrt{\frac{1}{9} + \frac{8}{9}} = 1. \square$$

Normal Matrices

We are now ready to assemble our special matrices above under the umbrella of a general category:

Definition: We say that an $n \times n$ complex matrix A is ***normal*** if:

$$A^* \cdot A = A \cdot A^*,$$

that is, A ***commutes*** with its adjoint A^*. Similarly, a linear operator T on $\mathbb{C}^n$ is normal if $T^* \circ T = T \circ T^*$

The three special matrix types that we saw above are all normal, because:

If A is ***Hermitian***, then $A = A^*$, hence $A^* \cdot A = A^2 = A \cdot A^*$.

If A is ***Skew-Hermitian***, then $A^* = -A$, hence $A^* \cdot A = -A^2 = A \cdot A^*$.

If A is ***unitary***, then $A^* \cdot A = \boldsymbol{I_n} = A \cdot A^*$.

In particular, among the ***real*** $n \times n$ matrices, ***symmetric***, ***skew-symmetric*** and ***orthogonal*** matrices are also normal, since these are the real analogs of the complex matrix types above.

However, we must caution that there are normal matrices that are ***not*** one of these special types!

Example: The matrix:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

is pure real, and we can visually check that it is ***not*** symmetric, skew-symmetric, or orthogonal. However, multiplying A by $A^* = A^\top$, we get:

$$A \bullet A^\top = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}.$$

Thus $A \bullet A^* = A^* \bullet A$ and A is normal.□

As with the other special transformations, we can use the inner product to characterize normal operators:

Theorem: Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation. Then T is ***normal*** ***if and only if*** for all $\vec{v}, \vec{w} \in \mathbb{C}^n$:

$$\langle T(\vec{v}) | T(\vec{w}) \rangle = \langle T^*(\vec{v}) | T^*(\vec{w}) \rangle.$$

The proof is left as an Exercise.

Although Hermitian, Skew-Hermitian and unitary matrices are important, there are matrix types that do not fall into any of these categories, but are easily seen to be normal. For instance, take the simplest kind of square matrix:

Theorem: An $n \times n$ complex ***diagonal*** matrix D is ***normal***.

Proof: If $D = Diag(d_1, d_2, \ldots, d_n)$ where the diagonal entries are complex numbers, then:

$$D^* = Diag\left(\overline{d_1}, \overline{d_2}, \ldots, \overline{d_n}\right).$$

Thus:

$$\begin{aligned} D \bullet D^* &= Diag\left(d_1\overline{d_1}, d_2\overline{d_2}, \ldots, d_n\overline{d_n}\right) \\ &= Diag\left(\|d_1\|^2, \|d_2\|^2, \ldots, \|d_n\|^2\right) \\ &= Diag\left(\overline{d_1}d_1, \overline{d_2}d_2, \ldots, \overline{d_n}d_n\right) \\ &= D^* \bullet D. \blacksquare \end{aligned}$$

Notice that a diagonal matrix with at least one ***non-real*** entry is ***not*** Hermitian nor Skew-Hermitian. However, a ***real*** diagonal matrix is symmetric, hence obviously normal.

The next simplest kind of matrices after diagonal matrices are the triangular matrices. Ironically, it turns out that normal upper-triangular matrices do not give us anything new:

Theorem: If the $n \times n$ complex matrix A is ***upper (or lower) triangular*** and ***normal***, then A is actually ***diagonal***.

Proof: We will prove this Theorem in the case when A is upper triangular using Induction on n, and leave the lower triangular case as an Exercise.

Obviously a 1×1 matrix is already diagonal, so there is nothing to prove. For the Induction Hypothesis, assume that if an $(k-1) \times (k-1)$ matrix B is both upper triangular and normal, then B is actually diagonal. Now for the Inductive Step: let A be an $k \times k$ upper triangular complex matrix which is also normal. We must show that A is diagonal. Let us explicitly write the entries of A and A^* :

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ 0 & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{k,k} \end{bmatrix}, \quad \text{and} \quad A^* = \begin{bmatrix} \overline{a_{1,1}} & 0 & \cdots & 0 \\ \overline{a_{1,2}} & \overline{a_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1,k}} & \overline{a_{2,k}} & \cdots & \overline{a_{k,k}} \end{bmatrix}.$$

We can easily see that the upper-left entry of $A \bullet A^*$ is:

$$a_{1,1} \bullet \overline{a_{1,1}} + a_{1,2} \bullet \overline{a_{1,2}} + \cdots + a_{1,k} \bullet \overline{a_{1,k}} = \|a_{1,1}\|^2 + \|a_{1,2}\|^2 + \cdots + \|a_{1,k}\|^2.$$

On the other hand if we look at $A^* \bullet A$, we get:

$$A^* \bullet A = \begin{bmatrix} \overline{a_{1,1}} & 0 & \cdots & 0 \\ \overline{a_{1,2}} & \overline{a_{2,2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1,k}} & \overline{a_{2,k}} & \cdots & \overline{a_{k,k}} \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,k} \\ 0 & a_{2,2} & \cdots & a_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{k,k} \end{bmatrix}$$

and the upper-left entry of $A^* \bullet A$ is just $a_{1,1} \bullet \overline{a_{1,1}} = \|a_{1,1}\|^2$. But if $A \bullet A^* = A^* \bullet A$, then we must have:

$$\|a_{1,1}\|^2 + \|a_{1,2}\|^2 + \cdots + \|a_{1,k}\|^2 = \|a_{1,1}\|^2.$$

In other words:

$$\|a_{1,2}\|^2 + \cdots + \|a_{1,k}\|^2 = 0.$$

Since every term in this sum is a non-negative real number, this is possible only if:

$$a_{1,2} = a_{1,3} = \cdots = a_{1,k} = 0.$$

Thus, their conjugates are likewise 0, so in block form:

$$A = \begin{bmatrix} a_{1,1} & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & B \end{bmatrix}$$

where the $(k-1) \times (k-1)$ sub-matrix B is both upper triangular and normal. By the Induction Hypothesis, B is diagonal, and thus A is also diagonal.■

8.5 Key Concepts

We summarize below the special matrix types and analogous operators that generalize from real matrices to complex matrices.

Real Matrices A and Linear Transformations $T : \mathbb{R}^n \to \mathbb{R}^n$ (below, $\vec{v}$, $\vec{w} \in \mathbb{R}^n$, and $[T] = A$)	Complex Matrices A and Linear Transformations $T : \mathbb{C}^n \to \mathbb{C}^n$ (below, $\vec{v}$, $\vec{w} \in \mathbb{C}^n$, and $[T] = A$)
Transpose: $A^\top$	***Adjoint:*** $A^* = \overline{A^\top}$ Action: $[T^*(\vec{v})] = [T]^*[\vec{v}]$
Symmetric: $A = A^\top$ $T(\vec{v}) \circ \vec{w} = \vec{v} \circ T(\vec{w})$	***Hermitian:*** $A = A^*$ $\langle T(\vec{v}) \mid \vec{w} \rangle = \langle \vec{v} \mid T(\vec{w}) \rangle$
Skew-Symmetric: $A = -A^\top$ $T(\vec{v}) \circ \vec{w} = -\vec{v} \circ T(\vec{w})$	***Skew-Hermitian:*** $A = -A^*$ $\langle T(\vec{v}) \mid \vec{w} \rangle = -\langle \vec{v} \mid T(\vec{w}) \rangle$
Orthogonal: $A \cdot A^\top = \boldsymbol{I_n}$ $\vec{v} \circ \vec{w} = T(\vec{v}) \circ T(\vec{w})$	***Unitary:*** $A \cdot A^* = \boldsymbol{I_n}$ $\langle \vec{v} \mid \vec{w} \rangle = \langle T(\vec{v}) \mid T(\vec{w}) \rangle$

An $n \times n$ complex matrix A is ***normal*** if $A \cdot A^* = A^* \cdot A$.

All matrix types shown above are examples of normal matrices. In addition, diagonal complex matrices are also normal. However, ***not all*** normal matrices fall into one of these special categories.

Hermitian and Skew-Hermitian matrices are ***closed*** under addition, multiplication by a pure real number, the transpose operation and the adjoint operation.

In contrast, unitary matrices have ***multiplicative*** properties.

Eigenvalues of special normal matrices: If λ is an eigenvalue of a normal matrix A, then:

1. λ is a ***pure real number*** if A is ***Hermitian***.
2. λ is a ***pure imaginary number*** if A is ***Skew-Hermitian***.
3. $\|\lambda\| = 1$ if A is ***unitary***.

8.5 Exercises

For Exercises (1) to (15): For each matrix A: (a) Determine all the adjectives which apply to A, among the choices: (i) Hermitian (ii) Skew-Hermitian (iii) unitary (iv) normal (v) none of these. Next, if A ***is*** normal: (b) find $det(A)$ and verify that it possesses the property that is specific to that type of matrix if A belongs to a special type (i.e. $det(A)$ is pure real if A is Hermitian, etc.), (c) find $charpoly(A)$, and (d) $Spec(A)$ and verify that $Spec(A)$ possesses the property that is specific to that type of matrix, again if A belongs to a special type.

1. $A = \begin{bmatrix} 5 & 2i \\ -2i & 2 \end{bmatrix}$

2. $A = \begin{bmatrix} 5i & 2 \\ -2 & 2i \end{bmatrix}$

3. $A = \begin{bmatrix} 7 & 2 \\ 2 & 4 \end{bmatrix}$

4. $A = \begin{bmatrix} 7 & -2 \\ 2 & 4 \end{bmatrix}$

5. $A = \begin{bmatrix} 3 & 6i \\ -6i & -2 \end{bmatrix}$

6. $A = \frac{1}{3}\begin{bmatrix} 2-i & 2i \\ 2i & 2+i \end{bmatrix}$

7. $A = \begin{bmatrix} 3 & -2 \\ 2 & 3 \end{bmatrix}$

8. $A = \begin{bmatrix} 15i & -30 \\ 30 & -10i \end{bmatrix}$

9. $A = \frac{1}{\sqrt{2}}\begin{bmatrix} i & -i \\ -1 & -1 \end{bmatrix}$

10. $A = \begin{bmatrix} 1 & -i & 0 \\ i & -1 & 1+i \\ 0 & 1-i & 1 \end{bmatrix}$

11. $A = \begin{bmatrix} i & 1+i & -1 \\ -1+i & -i & 1 \\ 1 & -1 & i \end{bmatrix}$

12. $A = \begin{bmatrix} 1 & -i & 1+i \\ i & -1 & i \\ 1-i & -i & -1 \end{bmatrix}$

13. $A = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{i}{\sqrt{2}} & \frac{-i}{\sqrt{2}} & 0 \\ 0 & 0 & -i \end{bmatrix}$

14. $A = \begin{bmatrix} 5 & -2 & 1 \\ 2 & 5 & -2 \\ -1 & 2 & 5 \end{bmatrix}$

15. $A = \begin{bmatrix} 1 & 3 & -2 \\ -3 & 1 & -6 \\ 2 & 6 & 1 \end{bmatrix}$

16. Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation. Prove that T is ***normal*** ***if and only if*** for all $\vec{v}$, $\vec{w} \in \mathbb{C}^n : \langle T(\vec{v}) | T(\vec{w}) \rangle = \langle T^*(\vec{v}) | T^*(\vec{w}) \rangle$.

17. Prove that the following conditions are equivalent regarding an $n \times n$ complex matrix A:

    a. A is ***unitary***.

    b. The ***rows*** of A form an ***orthonormal*** set of vectors from $\mathbb{C}^n$ under the complex Euclidean inner product.

    c. The ***columns*** of A form an ***orthonormal*** set of vectors from $\mathbb{C}^n$ under the complex Euclidean inner product.

    Hint: Mimic the proof of the analogous Theorem for orthogonal matrices in Chapter 7.

18. Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation. Prove that $[T]$ is ***unitary*** ***if and only if*** for all $\vec{v}$, $\vec{w} \in \mathbb{C}^n : \langle \vec{v} | \vec{w} \rangle = \langle T(\vec{v}) | T(\vec{w}) \rangle$.

19. Let A be a ***unitary*** matrix. Prove that:

    a. the ***complex norm*** of $det(A)$ is 1, that is, $\| det(A) \| = 1$.

    b. all the ***eigenvalues*** λ of A also have ***complex norm*** 1, that is, for every eigenvalue λ for A, $\| \lambda \| = 1$.

20. Let A and B be ***unitary*** $n \times n$ matrices. Prove that:

a. $A \bullet B$ is also ***unitary***.

b. A is ***invertible*** and $A^{-1} = A^*$ is also unitary.

c. If $\alpha = \cos(\theta) + \sin(\theta)i$ is a ***unit*** complex scalar for some real number θ, then $\alpha \bullet A$ is also unitary.

21. Prove that T is a ***unitary*** operator on $\mathbb{C}^n$ ***if and only if*** for every ***orthonormal basis*** $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ for $\mathbb{C}^n$, the set $B' = \{T(\vec{u}_1), T(\vec{u}_2), \ldots, T(\vec{u}_n)\}$ is also an ***orthonormal basis***.

22. Prove that T is a ***unitary*** operator on $\mathbb{C}^n$ ***if and only if*** $\|\vec{v}\| = \|T(\vec{v})\|$ for all $\vec{v} \in \mathbb{C}^n$. Again, this is similar to a Theorem from Chapter 7.

23. Let A be an $n \times n$ complex matrix. Prove that:

a. A is ***Hermitian*** ***if and only if*** the diagonal entries of A are all pure real, and the off-diagonal pairs are complex-conjugate pairs.

b. A is ***Skew-Hermitian*** ***if and only if*** the diagonal entries of A are all pure imaginary, and the off-diagonal pairs are real-conjugate pairs.

24. Let A and B be $n \times n$ ***Hermitian*** matrices, and $r \in \mathbb{R}$. Prove that:

a. $A + B$ is again Hermitian.

b. $r \bullet A$ is again Hermitian.

c. $A^\top$ and A^* are again Hermitian.

25. Let A and B be $n \times n$ ***Skew-Hermitian*** matrices, and $r \in \mathbb{R}$. Prove that:

a. $A + B$ is again Skew-Hermitian.

b. $r \bullet A$ is again Skew-Hermitian.

c. $A^\top$ and A^* are again Skew-Hermitian.

26. Let B be an $n \times n$ ***Skew-Hermitian*** matrix. Prove that:

a. the ***determinant*** of B is a ***pure real number*** if n is ***even***, and a ***pure imaginary number*** if n is ***odd***.

b. the ***eigenvalues*** of B are always ***pure imaginary numbers***.

27. Let $T : \mathbb{C}^n \to \mathbb{C}^n$ be a linear transformation. Prove that T is ***Skew-Hermitian*** ***if and only if*** for all $\vec{v}, \vec{w} \in \mathbb{C}^n : \langle T(\vec{v}) | \vec{w} \rangle = -\langle \vec{v} | T(\vec{w}) \rangle$.

28. Prove that if A is any $n \times n$ complex matrix, then $A \bullet A^*$ is always ***Hermitian***.

29. Prove that if A is any $n \times n$ complex matrix, then $i \bullet A \bullet A^*$ is always ***Skew-Hermitian***.

30. Prove that if A ***Hermitian***, then $i \bullet A$ is ***Skew-Hermitian***.

31. Prove that if A is ***Skew-Hermitian***, then $i \bullet A$ is ***Hermitian***.

32. If A is ***Hermitian***, is its ***conjugate*** $\overline{A}$ also necessarily Hermitian?

33. If A is ***Skew-Hermitian***, is its ***conjugate*** $\overline{A}$ also necessarily Skew-Hermitian?

34. Prove that if A and B are both $n \times n$ ***Hermitian*** matrices ***and*** $AB = BA$, then AB is also ***Hermitian***. In other words, the product of two ***commuting*** Hermitian matrices is again Hermitian.

35. Prove that the only $n \times n$ matrix that are ***both Hermitian*** and ***Skew-Hermitian*** is $\mathbf{0}_{n \times n}$.

36. Prove that if A is ***any*** square matrix, then $A + A^*$ is always ***Hermitian***, and $A - A^*$ is always ***Skew-Hermitian***.

37. Prove that every square complex matrix A can be expressed ***uniquely*** as $A = B + C$, where B is a ***Hermitian*** matrix and C is a ***Skew-Hermitian*** matrix. Hint: Use the two previous Exercises for the uniqueness and existence properties, respectively.

38. In Chapter 7, we saw the following Theorem relating the sets of symmetric and skew-symmetric matrices: Consider the two subspaces of $Mat(n,n)$:

$$V = Sym(n) = \{A \in Mat(n,n) \,|\, A^\top = A\}, \quad \text{and}$$
$$W = SkewSym(n) = \{B \in Mat(n,n) \,|\, B^\top = -B\},$$

namely, the subspaces of ***symmetric*** and ***skew-symmetric*** matrices, respectively. Then: $V^\perp = W$.

The complex analog of this Theorem is as follows: Consider $Mat(\mathbb{C},n,n)$ as a ***real*** vector space, that is, matrices with complex entries are added in the usual way, but we allow only scalar multiplication by a ***real*** number. Denote by:

$$V = Herm(n) = \{A \in Mat(\mathbb{C},n,n) \,|\, A^* = A\}, \quad \text{and}$$
$$W = SkewHerm(n) = \{B \in Mat(\mathbb{C},n,n) \,|\, B^* = -B\},$$

namely, the ***real*** vector subspace of all complex ***Hermitian*** $n \times n$ matrices, and the real vector subspace of all ***Skew-Hermitian*** $n \times n$ matrices. Our goal is to prove that $V^\perp = W$ with respect to a natural inner product $\langle A \,|\, B \rangle$ on $Mat(\mathbb{C},n,n)$.

a. Define a bilinear form on $Mat(\mathbb{C},n,n)$ by:

$$\langle A \,|\, B \rangle = \sum_{\text{all subscripts } j,k} [Re(a_{j,k}) \cdot Re(b_{j,k}) + Im(a_{j,k}) \cdot Im(b_{j,k})],$$

where Re and Im are the functions: $Re(x + iy) = x$ and $Im(x + iy) = y$, where $x, y \in \mathbb{R}$.

Prove that this bilinear form is a ***(real) inner product*** on $Mat(\mathbb{C},n,n)$.

b. Find an orthonormal basis for $Herm(n)$ under the inner product in (a). What is the dimension of $Herm(n)$?

c. Find an orthonormal basis for $SkewHerm(n)$ under the inner product in (a). What is the dimension of $SkewHerm(n)$?

d. Show that $dim(Mat(\mathbb{C},n,n)) = dim(Herm(n)) + dim(SkewHerm(n))$.

Reminder: these are all regarded as ***real*** vector spaces.

e. Prove that every member of $Herm(n)$ is ***orthogonal*** to every member of $SkewHerm(n)$ under the inner product of (a). Hint/Reminder: you only have to do it for every pair of basis vectors.

f. Prove that $Herm(n)$ and $SkewHerm(n)$ are ***orthogonal complements*** of each other under the inner product of (a).

39. Mimic the proof in the final Theorem to prove that if the $n \times n$ complex matrix A is ***lower triangular*** and ***normal***, then A is actually ***diagonal***.

8.6 The Spectral Theorems

Recall from Section 6.3 that an $n \times n$ complex matrix A is ***diagonalizable*** ***if and only if*** the ***algebraic multiplicity*** of each ***eigenvalue*** of A is equal to its ***geometric multiplicity***. Consequently, there is a complete set of linearly independent eigenvectors for A from $\mathbb{C}^n$ that acts as a basis for $\mathbb{C}^n$. The Spectral Theorems tell us precisely not just ***when***, but also ***how*** a certain type of matrix is diagonalizable, as well as the nature of its eigenvalues or ***spectrum***.

Also recall that we say A is ***similar*** to B if there exists an invertible matrix C such that:

$$B = C^{-1}AC,$$

where all three are $n \times n$ matrices (but this time, they can have ***complex*** entries). Thus we can say that A is ***diagonalizable*** ***if and only if*** A is ***similar*** to a ***diagonal*** matrix D, and we say that C ***diagonalizes*** A. The Spectral Theorems require a special kind of a diagonalizing matrix, namely, one that is ***unitary***, so let us first make the following:

Definitions: We say that two $n \times n$ complex matrices A and B are ***unitarily equivalent*** if there exists an $n \times n$ ***unitary*** matrix U such that:

$$B = U^{-1}AU = U^*AU.$$

We say that A is ***unitarily diagonalizable*** if there exists an $n \times n$ ***unitary*** matrix U and a ***diagonal*** matrix D (possibly with complex entries) such that:

$$D = U^{-1}AU = U^*AU.$$

In other words, A is ***unitarily equivalent*** to a ***diagonal*** matrix D.

In Chapter 6, we stated without proof that a (real) symmetric matrix A is diagonalizable using an orthogonal matrix C. But since an orthogonal matrix is the real analog of a unitary matrix, this Theorem says that a symmetric matrix is unitarily diagonalizable. This is in fact one of the Spectral Theorems, and we will be seeing it later.

The following Theorem is analogous to the first Theorem in Section 6.3:

Theorem: Let A be an $n \times n$ complex matrices. Then: A is ***unitarily diagonalizable*** ***if and only if*** there is an ***orthonormal basis*** for $\mathbb{C}^n$ consisting of ***eigenvectors*** of A.

Proof: ($\Rightarrow$) Suppose A is unitarily diagonalizable. Let U be a diagonalizing unitary matrix, and D a diagonal matrix, such that:

$$D = U^*AU, \quad \text{or in other words:}$$

$$UD = AU.$$

If $U = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_n]$ and $D = Diag(d_1, d_2, \ldots, d_n)$, then this last equation tells us that:

$$[d_1\vec{u}_1 | d_2\vec{u}_2 | \ldots | d_n\vec{u}_n] = [A\vec{u}_1 | A\vec{u}_2 | \ldots | A\vec{u}_n].$$

But this says that each $\vec{u}_i$ is an ***eigenvector*** of A with associated eigenvalue d_i. Since the n columns of U form an ***orthonormal set***, we have found an ***orthonormal basis*** $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ for $\mathbb{C}^n$.

($\Leftarrow$) For the converse, we essentially reverse the construction above. Suppose $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is an orthonormal basis for $\mathbb{C}^n$ consisting of eigenvectors of A. Assemble these vectors into the columns of a matrix $U = [\vec{u}_1 | \vec{u}_2 | \ldots | \vec{u}_n]$, and the corresponding eigenvalues into $Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. By construction, U is ***unitary***, and $UD = AU$ as above. Thus $D = U^*AU$, and A is ***unitarily diagonalizable***.■

As the name implies, the property of being unitarily equivalent is an ***equivalence relation:***

Theorem: Let A, B and C be $n \times n$ complex matrices. Define the relation $A \backsim B$ ***if and only if*** A is ***unitarily equivalent*** to B. Then the following properties hold:

1. ***Symmetry:*** $A \backsim A$.
2. ***Reflexivity:*** If $A \backsim B$ then $B \backsim A$.
3. ***Transitivity:*** If $A \backsim B$ and $B \backsim C$ then $A \backsim C$.

Thus, unitary equivalence is an ***equivalence relation***.

Proof: These properties follow easily because $\boldsymbol{I}_n$ is unitary, the inverse of a unitary matrix is also unitary, and the product of two unitary matrices is unitary.■

The property of being ***normal*** is also ***preserved*** by ***unitary equivalence:***

Theorem: If an $n \times n$ complex matrix A is ***normal***, then U^*AU is also ***normal*** for all ***unitary*** matrices U.

Proof: Suppose A is normal and U is unitary, and both are $n \times n$ matrices. Let $B = U^*AU$. Then $B^* = U^*A^*(U^*)^* = U^*A^*U$. Since U is unitary, we have $U \cdot U^* = \boldsymbol{I}_n = U^* \cdot U$. Using the Associative Property, we get:

$$\begin{aligned} B^* \cdot B &= U^*A^*U \cdot U^*AU \\ &= U^*A^*\boldsymbol{I}_nAU \\ &= U^*A^*AU, \quad \text{and} \\ B \cdot B^* &= U^*AU \cdot U^*A^*U \\ &= U^*A\boldsymbol{I}_nA^*U \\ &= U^*AA^*U. \end{aligned}$$

Since $A^* \cdot A = A \cdot A^*$, we get $B^* \cdot B = B \cdot B^*$.■

The Main Spectral Theorem

We are now ready to state the main Spectral Theorem, from which the others follow:

Theorem — The Spectral Theorem for Normal Matrices:

Let A be an $n \times n$ complex matrix. Then: ***A is unitarily diagonalizable if and only if A is normal***, that is, $A \cdot A^* = A^* \cdot A$.

This Theorem is extremely powerful because all we need to determine if A is unitarily diagonalizable is to perform two matrix products and see if they are equal! Once the matrix passes this test, we can go through the arduous task of finding D and U. We certainly do not want to go through all this effort if it were impossible to unitarily diagonalize A. We also know that the following special matrices are all normal: (real) symmetric, Hermitian, skew-symmetric, skew-Hermitian and unitary matrices.

To prove this Theorem, we need the following ***Lemma*** by ***Issai Schur*** (1875-1941). A Lemma is a Theorem that is used to prove an even more important Theorem. Schur was born in Mogilyov, Belarus, but lived most of his life in Germany. Being Jewish, he was forced to resign from his professorial post in the Prussian Academy in 1938 during the Holocaust. He died in poverty in Tel Aviv, but he is remembered for the following:

> ***Theorem — Schur's Lemma:***
>
> Let A be an $n \times n$ complex matrix. Then there exists a ***unitary*** matrix U and an ***upper triangular*** complex matrix B so that:
>
> $$A = UBU^*, \quad \text{or equivalently,} \quad B = U^*AU.$$
>
> We call the factorization $A = UBU^*$ the ***Schur Decomposition*** of A.

Proof: The idea is to use Induction on the dimension n of A. If $n = 1$, then A is already upper triangular, so there is nothing to prove.

For the Induction Hypothesis, let us assume that if C is a $(k-1) \times (k-1)$ complex matrix, then there exists a $(k-1) \times (k-1)$ unitary matrix X such that $X^{-1}CX = X^*CX$ is an upper triangular complex matrix.

Now, let A be an $k \times k$ complex matrix. We must find a $k \times k$ unitary matrix, say Z, such that $Z^{-1}AZ = Z^*AZ$ is upper triangular.

Let λ be any eigenvalue of A, and let $\vec{u}_1$ be a ***unit*** eigenvector of A associated to λ. We can extend $\{\vec{u}_1\}$ to a basis $B = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_k\}$ for all of $\mathbb{C}^k$. By applying the Gram-Schmidt algorithm, we may as well assume that B is an ***orthonormal basis***.

If we assemble these vectors into the ***columns*** of a matrix U:

$$U = [\,[\vec{u}_1]\,|\,[\vec{u}_2]\,|\cdots|\,[\vec{u}_k]\,],$$

then U is ***unitary***. Its adjoint U^* has as its ***rows:***

$$U^* = \begin{bmatrix} [\vec{u}_1]^* \\ [\vec{u}_2]^* \\ \vdots \\ [\vec{u}_k]^* \end{bmatrix}.$$

But now:

$$\begin{aligned} U^*AU &= U^*A[\,[\vec{u}_1]\,|\,[\vec{u}_2]\,|\cdots|\,[\vec{u}_k]\,] \\ &= [\,U^*A\vec{u}_1\,|\,U^*A\vec{u}_2\,|\cdots|\,U^*A\vec{u}_k\,] \\ &= [\,U^*\lambda\vec{u}_1\,|\,U^*A\vec{u}_2\,|\cdots|\,U^*A\vec{u}_k\,]. \end{aligned}$$

Let us focus on the ***first column*** only of this matrix. We get:

$$U^* \lambda \vec{u}_1 = \begin{bmatrix} [\vec{u}_1]^* \\ [\vec{u}_2]^* \\ \vdots \\ [\vec{u}_k]^* \end{bmatrix} \lambda \vec{u}_1 = \begin{bmatrix} \lambda[\vec{u}_1]^* \bullet [\vec{u}_1] \\ \lambda[\vec{u}_2]^* \bullet [\vec{u}_1] \\ \vdots \\ \lambda[\vec{u}_k]^* \bullet [\vec{u}_1] \end{bmatrix} = \begin{bmatrix} \lambda \bullet 1 \\ \lambda \bullet 0 \\ \vdots \\ \lambda \bullet 0 \end{bmatrix} = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

by the ***orthonormality*** property of B. Thus, U^*AU has the form:

$$U^*AU = \begin{bmatrix} \lambda & \vec{R}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & C \end{bmatrix},$$

where $\vec{R}_{1\times(k-1)}$ consists of the rest of the $k-1$ entries on the first row (which are irrelevant). But C is now a $(k-1)\times(k-1)$ complex matrix, so by the Induction Hypothesis, there exists a $(k-1)\times(k-1)$ ***unitary*** matrix X such that $X^{-1}CX = X^*CX$ is an ***upper triangular*** complex matrix. Let us "enlarge" X to the $k\times k$ matrix:

$$Y = \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X \end{bmatrix}.$$

Since X is unitary, we get:

$$\begin{aligned} YY^* &= \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X \end{bmatrix} \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X^* \end{bmatrix} \\ &= \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & XX^* \end{bmatrix} = \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & I_{k-1} \end{bmatrix} = \boldsymbol{I}_k, \end{aligned}$$

so Y is also ***unitary***. Since the product of two unitary matrices is also unitary, $Z = UY$ is ***unitary***. But now:

$$\begin{aligned} Z^* \bullet A \bullet Z &= (UY)^* \bullet A \bullet (UY) \\ &= Y^* \bullet U^* \bullet A \bullet U \bullet Y. \end{aligned}$$

By substituting Y^*, U^*AU and Y above, we get:

$$\begin{aligned} Z^* \bullet A \bullet Z &= \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X^* \end{bmatrix} \begin{bmatrix} \lambda & \vec{R}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & C \end{bmatrix} \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X \end{bmatrix} \\ &= \begin{bmatrix} \lambda & \vec{R}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X^*C \end{bmatrix} \begin{bmatrix} 1 & \vec{\mathbf{0}}_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X \end{bmatrix} \\ &= \begin{bmatrix} \lambda & \vec{R}'_{1\times(k-1)} \\ \vec{\mathbf{0}}_{(k-1)\times 1} & X^*CX \end{bmatrix}. \end{aligned}$$

Since X^*CX is an upper triangular complex matrix and the rest of column 1 below λ consists of zeroes, our final matrix is also ***upper triangular*** (note that the rest of the entries in row 1, denoted $\vec{R}'_{1\times(k-1)}$, are irrelevant as before). This completes the proof of Schur's Lemma.■

Proof of the Main Spectral Theorem:

We will now show that: for every $n \times n$ complex matrix A:

A is unitarily diagonalizable ***if and only if*** ***A is normal***.

($\Rightarrow$) Suppose A is ***unitarily diagonalizable***. We must show that A is ***normal***. Let U be a diagonalizing unitary matrix, and D a diagonal matrix, such that:

$$D = U^*AU.$$

But this equation says that A is ***unitarily equivalent*** to a ***diagonal*** matrix. Since diagonal matrices are ***normal***, and normality is ***preserved*** by unitary equivalence, A must also be ***normal***.

($\Leftarrow$) For the converse, suppose now that A is ***normal***. We must show that A is ***unitarily diagonalizable***. By ***Schur's Lemma***, we can find an upper triangular matrix B and a unitary matrix U such that:

$$B = U^*AU.$$

Again, this says that A is ***unitarily equivalent*** to B, but since A is ***normal*** so is B. However, the only upper triangular matrices that are also normal are the ***diagonal*** matrices. Thus B is actually diagonal, and so A is ***unitarily diagonalizable***.■

Notice that the proof above is quite short, thanks to Schur's Lemma. This echoes a recurring theme in mathematics:

To prove a Theorem about something ***special*** (in this case, the Spectral Theorem is only about ***normal*** matrices) you might try to prove a Theorem which is more ***general*** (in this case, Schur's Lemma is about ***all $n \times n$ complex matrices***) that can be used to easily prove the special case.

Orthogonality of Distinct Eigenspaces

In Chapter 7, we were able to prove that two eigenvectors belonging to two distinct eigenspaces of a ***symmetric*** matrix are in fact orthogonal under the ordinary dot product. It turns out that this is again a special case of a more general phenomenon, and it is an easy consequence of the Main Spectral Theorem:

Theorem: If $\vec{v}$ and $\vec{w}$ are vectors from two ***distinct*** eigenspaces of a ***normal*** $n \times n$ matrix A, then $\langle \vec{v} | \vec{w} \rangle = 0$.

Proof: If A is normal, then A is unitarily diagonalizable by the Main Spectral Theorem. From our first Theorem, we saw that we can find an ***orthonormal basis*** $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ for $\mathbb{C}^n$ consisting of ***eigenvectors*** of A.

Thus, if we have two distinct eigenvalues λ_1 and λ_2, then, without loss of generality, we can write:

$$Eig(A, \lambda_1) = Span(\{\vec{u}_{j_1}, \ldots, \vec{u}_{j_2}\}), \text{ and}$$

$$Eig(A, \lambda_2) = Span(\{\vec{u}_{j_3}, \ldots, \vec{u}_{j_4}\}),$$

where $j_2 < j_3$. Since the members of these two indicated sets of vectors are pairwise ***orthogonal,*** so are any two vectors in their ***Spans***. Thus two vectors from distinct eigenspaces are ***orthogonal***.■

The Spectral Theorems for Special Families

We can now state the Spectral Theorems for the special types of normal matrices. Let us begin with our main goal, which is to tie up the loose end from Chapter 7:

> ***Theorem — The Spectral Theorem for Real Symmetric Matrices:***
> An $n \times n$ ***real symmetric*** matrix A is ***orthogonally diagonalizable***, and all of the ***eigenvalues*** in the diagonal matrix D are ***pure real***.

Proof: We already know that a real ***symmetric*** matrix is ***Hermitian***, and from the previous section, all of the ***eigenvalues*** of a Hermitian matrix are ***pure real***. Now, the main Spectral Theorem only says that A is ***diagonalizable*** by a ***unitary*** matrix U, that is, U could have ***imaginary*** entries.

But since all the entries of A are real, and the eigenvalues λ are also real, then every $\lambda \boldsymbol{I}_n - A$ is also real. Thus, the ***rref*** of $\lambda \boldsymbol{I}_n - A$ is also ***real***, so we **can** find a basis for each eigenspace that consists of vectors in $\mathbb{R}^n$ and not just $\mathbb{C}^n$. By the (real) Gram-Schmidt Algorithm, we can again construct an ***orthonormal basis*** for $\mathbb{R}^n$ consisting of eigenvectors of A. Thus, U can be constructed not just to be unitary, but ***orthogonal***.■

Other Spectral Theorems follow directly from the main one, and our knowledge of the eigenvalues of each matrix type that we saw in the previous Section:

> ***Theorem — The Spectral Theorem for Hermitian Matrices:***
> An $n \times n$ ***Hermitian*** matrix A is unitarily diagonalizable, and all of the eigenvalues in the diagonal matrix D are ***pure real***.

> ***Theorem — The Spectral Theorem for Skew-Hermitian Matrices:***
> An $n \times n$ ***Skew-Hermitian*** matrix A is unitarily diagonalizable, and all of the eigenvalues in the diagonal matrix D are ***pure imaginary***.

> ***Theorem — The Spectral Theorem for Unitary Matrices:***
> An $n \times n$ ***unitary*** matrix A is unitarily diagonalizable, and all of the eigenvalues in the diagonal matrix D are complex numbers of ***unit length***.

The Unitary Diagonalization Algorithm

Let us describe the steps to unitarily diagonalize an $n \times n$ normal matrix A:

1. Make sure that A is normal, that is, $A \bullet A^* = A^* \bullet A$.
2. Find the characteristic polynomial $p(\lambda) = det(\lambda \boldsymbol{I}_n - A)$.
3. Find the distinct eigenvalues of A from the characteristic polynomial, that is:
$$Spec(A) = \{\lambda_1, \lambda_2, \ldots, \lambda_k\}$$
4. For each eigenspace $Eig(A, \lambda_j)$, find a basis for this eigenspace by finding the nullspace of $\lambda_j \boldsymbol{I}_n - A$ using the ***Gauss-Jordan Algorithm*** with complex numbers.
5. For each eigenspace $Eig(A, \lambda_j)$, convert the basis from the previous step into an ***orthonormal basis*** using the ***Gram-Schmidt Algorithm***.
6. Assemble all the unit eigenvectors that you obtained for all the k eigenspaces into the ***columns*** of the $n \times n$ ***unitary*** matrix:
$$U = [\vec{u}_1 | \vec{u}_2 | \cdots | \vec{u}_n],$$
and assemble the $n \times n$ ***diagonal*** matrix:
$$D = Diag(\lambda_1, \ldots, \lambda_k),$$
repeating the eigenvalue λ_i by its algebraic ***multiplicity***. In other words, a double root should appear twice, and so on.
7. We obtain the Schur decomposition: $A = UDU^* = UDU^{-1}$.

Example: Let us apply the algorithm above to:

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix},$$

that we already know is normal from the previous Section. We get:

$$p(\lambda) = \begin{vmatrix} \lambda - 1 & -1 & 0 \\ 0 & \lambda - 1 & -1 \\ -1 & 0 & \lambda - 1 \end{vmatrix}$$
$$= (\lambda - 1)^3 - 1 = \lambda^3 - 3\lambda^2 + 3\lambda - 2 = (\lambda - 2)(\lambda^2 - \lambda + 1)$$

We note that in this case, we were lucky that $p(\lambda)$ had an ***integer*** root. In general, this will not be the case for a 3×3 normal matrix. Using the quadratic formula, we get:

$$Spec(A) = \left\{2,\ \frac{1}{2} + \frac{\sqrt{3}}{2}i,\ \frac{1}{2} - \frac{\sqrt{3}}{2}i\right\}.$$

Notice that we get 3 distinct eigenvalues, so each eigenspace is only 1-dimensional. Still, there is much work to do. We find the matrices $\lambda_j \boldsymbol{I}_3 - A$ for each eigenvalue above, to get the three matrices:

$$\begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \\ -1 & 0 & 1 \end{bmatrix}, \quad \text{with rref} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix};$$

$$\begin{bmatrix} -\frac{1}{2}+\frac{\sqrt{3}}{2}i & -1 & 0 \\ 0 & -\frac{1}{2}+\frac{\sqrt{3}}{2}i & -1 \\ -1 & 0 & -\frac{1}{2}+\frac{\sqrt{3}}{2}i \end{bmatrix}, \quad \text{with rref} \begin{bmatrix} 1 & 0 & \frac{1}{2}-\frac{\sqrt{3}}{2}i \\ 0 & 1 & \frac{1}{2}+\frac{\sqrt{3}}{2}i \\ 0 & 0 & 0 \end{bmatrix}; \text{ and}$$

$$\begin{bmatrix} -\frac{1}{2}-\frac{\sqrt{3}}{2}i & -1 & 0 \\ 0 & -\frac{1}{2}-\frac{\sqrt{3}}{2}i & -1 \\ -1 & 0 & -\frac{1}{2}-\frac{\sqrt{3}}{2}i \end{bmatrix}, \quad \text{with rref} \begin{bmatrix} 1 & 0 & \frac{1}{2}+\frac{\sqrt{3}}{2}i \\ 0 & 1 & \frac{1}{2}-\frac{\sqrt{3}}{2}i \\ 0 & 0 & 0 \end{bmatrix}.$$

We thus get a basis for each 1-dimensional eigenspace:

$$\begin{aligned} \vec{v}_1 &= \langle 1,1,1 \rangle, \\ \vec{v}_2 &= \tfrac{1}{2}\left\langle -1+\sqrt{3}\,i, -1-\sqrt{3}\,i, 2 \right\rangle, \text{ and} \\ \vec{v}_3 &= \tfrac{1}{2}\left\langle -1-\sqrt{3}\,i, -1+\sqrt{3}\,i, 2 \right\rangle. \end{aligned}$$

One can quickly verify that:

$$\langle \vec{v}_1 | \vec{v}_2 \rangle = \langle \vec{v}_1 | \vec{v}_3 \rangle = \langle \vec{v}_2 | \vec{v}_3 \rangle = 0,$$

hence we have an ***orthogonal set***. Each of these vectors, by coincidence, has length $\sqrt{3}$, so we just divide each vector by $\sqrt{3}$ before we assemble them into the columns of our unitary matrix U. We obtain:

$$U = \begin{bmatrix} \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{3}}+\frac{1}{2}i & -\frac{1}{2\sqrt{3}}-\frac{1}{2}i \\ \frac{1}{\sqrt{3}} & -\frac{1}{2\sqrt{3}}-\frac{1}{2}i & -\frac{1}{2\sqrt{3}}+\frac{1}{2}i \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}, \text{ and } D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2}+\frac{\sqrt{3}}{2}i & 0 \\ 0 & 0 & \frac{1}{2}-\frac{\sqrt{3}}{2}i \end{bmatrix},$$

where $A = UDU^*$. □

8.6 Section Summary

We say that two $n \times n$ complex matrices A and B are ***unitarily equivalent*** if there exists an $n \times n$ ***unitary*** matrix U such that: $B = U^{-1}AU = U^{*}AU$.

A is ***unitarily diagonalizable*** if A is ***unitarily equivalent*** to a ***diagonal*** matrix D.

Let A be an $n \times n$ complex matrices. Then: A is ***unitarily diagonalizable*** ***if and only if*** there is an ***orthonormal basis*** for $\mathbb{C}^n$ consisting of ***eigenvectors*** of A.

Unitary equivalence is an ***equivalence relation:*** it is ***reflexive***, ***symmetric***, and ***transitive***.

If an $n \times n$ complex matrix A is ***normal***, then $U^{*}AU$ is also ***normal*** for all ***unitary*** matrices U. In other words, the property of being normal is ***preserved*** under unitary equivalence.

If $\vec{v}$ and $\vec{w}$ are from ***distinct*** eigenspaces of a ***normal*** $n \times n$ matrix A, then $\langle \vec{v} | \vec{w} \rangle = 0$. Thus, distinct eigenspaces are ***orthogonal*** to each other.

The Spectral Theorem for Normal Matrices:

Let A be an $n \times n$ complex matrix. Then: ***A is unitarily diagonalizable*** ***if and only if*** ***A is normal***, that is, $A \cdot A^{*} = A^{*} \cdot A$.

The Spectral Theorem for Real Symmetric Matrices:

An $n \times n$ ***real symmetric*** matrix A is ***orthogonally diagonalizable***, and the eigenvalues in the diagonal matrix D are ***pure real***.

The Spectral Theorem for Hermitian Matrices:

An $n \times n$ ***Hermitian matrix*** A is ***unitarily diagonalizable***, and the eigenvalues in the diagonal matrix D are ***pure real***.

The Spectral Theorem for Skew-Hermitian Matrices:

An $n \times n$ ***Skew-Hermitian matrix*** A is ***unitarily diagonalizable***, and all of the eigenvalues in the diagonal matrix D are ***pure imaginary***.

The Spectral Theorem for Unitary Matrices:

An $n \times n$ ***unitary matrix*** A is ***unitarily diagonalizable***, and all of the eigenvalues in the diagonal matrix D are complex numbers of ***unit length***.

Schur's Lemma: Let A be an $n \times n$ complex matrix. Then there exists a ***unitary*** matrix U and an ***upper triangular*** complex matrix B so that $A = UBU^{*}$, or equivalently, $B = U^{*}AU$. We call the factorization $A = UBU^{*}$ the ***Schur Decomposition*** of A.

We unitarily diagonalize a normal matrix by finding an orthonormal basis for each eigenspace of A, using a combination of the Gauss-Jordan Algorithm, followed by the Gram-Schmidt Algorithm if the eigenspace is more than 1-dimensional. The unitary matrix U has for its columns the basis vectors for the eigenspaces, and the diagonal matrix D has the corresponding eigenvalues along the diagonal.

8.6 Exercises

For Exercises (1) to (15): Unitarily diagonalize, if possible, the matrix A in the corresponding Exercise from Section 7.5. In other words, find a diagonal matrix D and a unitary matrix U such that $D = U^*AU$, or explain why this is not possible.

16. ***The Pauli matrices:*** The matrices:

$$\sigma_1 = \sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \text{and } \sigma_3 = \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

are known as the Pauli matrices, named after the physicist ***Wolfgang Pauli***. They are important in the field of ***Quantum Mechanics***.

a. Identify the types of normal matrices that these belong to.

b. Unitarily diagonalize σ_1 and σ_2 (notice that σ_3 is already diagonal).

c. Show that $\sigma_x^2 = \sigma_y^2 = \sigma_z^2 = \boldsymbol{I}_2$.

d. Show that $\sigma_x\sigma_y = \sigma_z$.

e. Show that $\sigma_z\sigma_y = -i \cdot \sigma_x$.

f. Based on the previous two parts, guess and prove similar formulas for the four other products $\sigma_a\sigma_b$ where a and b are distinct members of $\{x, y, z\}$.

17. Prove that $A = \begin{bmatrix} 3 & 7 \\ 0 & -2 \end{bmatrix}$ is diagonalizable, but ***not*** unitarily diagonalizable. Hint: there are no computations needed. Use Proof by Contradiction and a Theorem from the previous Section.

18. Give an example of a normal matrix, with at least ***one non-real entry***, that is ***not*** diagonal, Hermitian, skew-Hermitian or unitary.

19. Prove directly that eigenvectors from distinct eigenspaces of a Hermitian matrix are orthogonal. Hint: mimic the proof in Chapter 7 for symmetric matrices by considering $\langle T(\vec{u}) | \vec{v} \rangle$ and $\langle \vec{u} | T(\vec{v}) \rangle$, where $\vec{u}$ is an eigenvector for λ_1 and $\vec{v}$ is an eigenvector for λ_2, and $\lambda_1 \neq \lambda_2$.

20. Repeat Exercise 19 for Skew-Hermitian matrices.

21. Repeat Exercise 19 for unitary matrices. Further hint: recall that for eigenvalues of a unitary matrix: $\lambda_1 \cdot \overline{\lambda_1} = 1 = \lambda_2 \cdot \overline{\lambda_2}$.

22. Find all complex numbers $a + bi$ such that $\begin{bmatrix} 2 & a+bi \\ i & -i \end{bmatrix}$ is a normal matrix.

23. Repeat the previous Exercise for $\begin{bmatrix} a+bi & 2 \\ 1 & i \end{bmatrix}$.

24. Apply the Unitary Diagonalization Algorithm to the orthogonal rotation matrix:

$$rot_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

for any real number θ. In other words, find D and U such that $rot_\theta = UDU^*$.

8.7 Simultaneous Diagonalization

We know that an $n \times n$ matrix A can be diagonalized ***if and only if*** there is a basis $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors for A. These vectors are assembled into the columns of an invertible matrix, $C = [\vec{v}_1\, \vec{v}_2 \ldots \vec{v}_n]$, and we get:

$$C^{-1}AC = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n) = D,$$

a diagonal matrix containing the eigenvalues of A along the diagonal. In this Section, we will look at the situation where A and B are diagonalizable $n \times n$ matrices, and we ask if we can find a ***single*** invertible matrix C where ***both*** $C^{-1}AC$ and $C^{-1}BC$ are diagonal, although possibly with different eigenvalues. This is a pretty tall order for A and B, and there is no clear condition as to when this is even possible. We will therefore make the following:

> ***Definition:*** Let A and B be $n \times n$ matrices. We say that A and B are ***simultaneously diagonalizable*** if there exists an ***invertible*** matrix C such that:
>
> $$C^{-1}AC = D_1, \quad \textbf{\textit{and}} \quad C^{-1}BC = D_2,$$
>
> where D_1 and D_2 are both ***diagonal*** matrices.

Now, recall that all $n \times n$ diagonal matrices ***commute***. So, suppose that we were lucky, and A and B are known to be simultaneously diagonalizable. In this case, with the notation above:

$$(C^{-1}AC)(C^{-1}BC) = D_1 \cdot D_2 = D_2 \cdot D_1 = (C^{-1}BC)(C^{-1}AC)$$

But by the Associative Property of Matrix Multiplication, we get:

$$(C^{-1}AC)(C^{-1}BC) = C^{-1}A(C \cdot C^{-1})BC = C^{-1}ABC, \text{ and similarly,}$$
$$(C^{-1}BC)(C^{-1}AC) = C^{-1}B(C \cdot C^{-1})AC = C^{-1}BAC.$$

Thus, we can conclude that $AB = BA$, or in other words, A and B also ***commute***.

The main Theorem for this Section is that the ***converse*** is also true: if $AB = BA$, and both A and B are diagonalizable, then A and B are simultaneously diagonalizable. This is far from obvious. To prepare ourselves to prove this Theorem, we need to understand the following construction involving direct sums:

Diagonalizability of Direct Sums

Recall that in the Exercises of Sections 2.8, 2.9 and 5.3, we created the ***direct sum*** of two or more matrices, and investigated their properties. We will now see when these matrices are diagonalizable:

> ***Theorem:*** Let E_1 be an $n \times n$ matrix and let E_2 be an $m \times m$ matrix. Let us form the ***direct sum*** of these two matrices, defined as the $(n+m) \times (n+m)$ matrix:
>
> $$E = E_1 \oplus E_2 = \begin{bmatrix} E_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & E_2 \end{bmatrix}.$$
>
> Then: E is ***diagonalizable*** ***if and only if*** both E_1 and E_2 are ***diagonalizable***.

As before, we say that E is in ***block diagonal form*** with blocks E_1 and E_2.

Proof: This Theorem almost sounds "obvious," but actually it is not. The converse, though is indeed obvious:

($\Leftarrow$) Suppose E_1 and E_2 are both diagonalizable. Thus, there exist C_1, an invertible $n \times n$ matrix, and C_2, invertible $m \times m$ matrix, such that:

$$C_1^{-1} E_1 C_1 = D_1, \quad \text{and} \quad C_2^{-1} E_2 C_2 = D_2,$$

where both D_1 and D_2 are diagonal matrices. Then, the matrix:

$$C = C_1 \oplus C_2 = \begin{bmatrix} C_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & C_2 \end{bmatrix} \text{ is invertible, with inverse: } C^{-1} = C_1^{-1} \oplus C_2^{-1} = \begin{bmatrix} C_1^{-1} & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & C_2^{-1} \end{bmatrix}$$

This can be verified by checking that $C \cdot C^{-1} = \boldsymbol{I}_{n+m}$. But then, using the Exercises in Section 2.8:

$$\begin{aligned} C^{-1}EC &= \begin{bmatrix} C_1^{-1} & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & C_2^{-1} \end{bmatrix} \begin{bmatrix} E_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & E_2 \end{bmatrix} \begin{bmatrix} C_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & C_2 \end{bmatrix} \\ &= \begin{bmatrix} C_1^{-1}E_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & C_2^{-1}E_2 \end{bmatrix} \begin{bmatrix} C_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & C_2 \end{bmatrix} \\ &= \begin{bmatrix} C_1^{-1}E_1C_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & C_2^{-1}E_2C_2 \end{bmatrix} = \begin{bmatrix} D_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & D_2 \end{bmatrix} = D, \end{aligned}$$

where D is diagonal. Thus E is diagonalizable.

($\Rightarrow$) Now, suppose that E as constructed above is diagonalizable. We must show that the blocks E_1 and E_2 are diagonalizable. Since E is diagonalizable, there exists an $(n+m) \times (n+m)$ invertible matrix C such that $C^{-1}EC = D$, where D contains $\lambda_1, \lambda_2, \ldots, \lambda_{n+m}$, the eigenvalues of E, along the diagonal.

The key idea here is that since C is ***invertible:***

$$rank(C) = n + m.$$

Now, let us partition C into $n + m$ columns but with ***two*** rows of vectors:

$$C = \begin{bmatrix} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_{n+m} \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n+m} \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_{n+m} \end{bmatrix}$$

where every $\vec{v}_i \in \mathbb{R}^{\boldsymbol{n}}$ and every $\vec{w}_i \in \mathbb{R}^{\boldsymbol{m}}$. As usual, we can rewrite the diagonalization equation as $EC = C \cdot Diag(\lambda_1, \lambda_2, \ldots, \lambda_{n+m})$, or in partitioned form:

$$\begin{bmatrix} E_1 & \mathbf{0}_{n \times m} \\ \mathbf{0}_{m \times n} & E_2 \end{bmatrix} \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n+m} \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_{n+m} \end{bmatrix} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n+m} \\ \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_{n+m} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \lambda_{n+m} \end{bmatrix}.$$

Performing the block multiplications on each side, we conclude that:

$$E_1\vec{v}_i = \lambda_i\vec{v}_i, \quad \text{and}$$
$$E_2\vec{w}_i = \lambda_i\vec{w}_i, \quad \text{for all } i = 1..n+m.$$

This means that every $\vec{v}_i$ is an eigenvector for E_1, and every $\vec{w}_i$ is an eigenvector for E_2.

Now, let us separate the top and bottom parts of C into two matrices:

$$C_1 = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_{n+m} \end{bmatrix}, \quad \text{and}$$
$$C_2 = \begin{bmatrix} \vec{w}_1 & \vec{w}_2 & \cdots & \vec{w}_{n+m} \end{bmatrix},$$

where C_1 is $n \times (n+m)$ and C_2 is $m \times (n+m)$. We mentioned that the key idea here is that $rank(C) = n+m$. We will now exploit that fact to show that C_1 and C_2 contain a full set of eigenvectors for E_1 and E_2, respectively, within its columns.

Since both $n < n+m$ and $m < n+m$, we get:

$$rank(C_1) \leq n \quad \text{and} \quad rank(C_2) \leq m.$$

But notice that n is the number of ***rows*** of C_1 and likewise m is the number of ***rows*** of C_2. Thus, if $rank(C_1) < n$, then C_1 will have fewer than n linearly independent rows. Similarly, if $rank(C_2) < m$, then C_2 will have fewer than m linearly independent rows. Thus, if ***either*** situation occurs, than C will have ***fewer*** than $n+m$ linearly independent rows. But this is impossible because $rank(C) = n+m$. Thus we can conclude that

$$rank(C_1) = n \quad \text{and} \quad rank(C_2) = m.$$

This means that C_1 contains n ***linearly independent columns***, and similarly C_2 contains m ***linearly independent columns***. Since we have a full set of eigenvectors for both E_1 and E_2, they are both diagonalizable.■

By Induction, we can generalize the previous Theorem into the following:

Theorem: Let $E_1, E_2, \ldots, E_k$ be square matrices, not necessarily of the same size. Then: the matrix E formed as the direct sum of these k matrices:

$$E = E_1 \oplus E_2 \oplus \cdots \oplus E_k = \begin{bmatrix} E_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & E_k \end{bmatrix}$$

(where each zero matrix is of the appropriate size) is diagonalizable ***if and only if*** every block E_i is diagonalizable.

The Simultaneous Diagonalizability Theorem

Now we are ready to state the Main Theorem regarding simultaneous diagonalization:

Theorem — The Simultaneous Diagonalizability Theorem:
If A and B are both ***diagonalizable*** $n \times n$ matrices, then A and B are ***simultaneously diagonalizable*** ***if and only if*** $AB = BA$.

Proof: We already showed in the Introduction that the forward implication is true. Conversely, suppose we are given that A and B are both diagonalizable and that $AB = BA$. We will show that A and B are simultaneously diagonalizable by dividing our proof into two cases, based on the geometric multiplicities of our matrices:

Case 1: Suppose that each eigenspace of either A or B has geometric multiplicity 1, that is, the characteristic polynomial factors into n distinct linear factors. Since A has n distinct eigenvalues, we know that A is diagonalizable. Suppose A has this quality, λ is one of the eigenvalues of A, and $\vec{v}$ is an associated eigenvector for A with respect to λ. Thus, $A\vec{v} = \lambda\vec{v}$.

Now, let us see what we can say about $B\vec{v}$. By the Associative Property of Matrix Multiplication:

$$A(B\vec{v}) = (AB)\vec{v} = (BA)\vec{v} = B(A\vec{v}) = B(\lambda\vec{v}) = \lambda(B\vec{v}).$$

Notice it was crucial that $AB = BA$. This equation tells us:

$$A(B\vec{v}) = \lambda(B\vec{v}).$$

Thus, $B\vec{v} \in Eig(A,\lambda)$. It is possible, though, that $B\vec{v} = \vec{\mathbf{0}}_n$. But in any case, since $Eig(A,\lambda)$ is 1-dimensional, this means that $B\vec{v}$ is ***parallel*** to $\vec{v}$, that is:

$$B\vec{v} = k\vec{v}, \text{ for some } k \in \mathbb{R} \text{ (where } k \textbf{\textit{ could}} \text{ be 0).}$$

Since $\vec{v}$ is an eigenvector, it is a ***non-zero*** vector, so this equation says that $\vec{v}$ is likewise an eigenvector for B, but for some (possibly different) eigenvalue k. Thus, every eigenvector for A is also an eigenvector for B, and so ***any*** matrix C that diagonalizes A will also diagonalize B.

Case 2: Suppose that A and B both have an eigenspace with geometric multiplicity bigger then 1. We know that A is diagonalizable, so suppose C is an invertible matrix such that

$$C^{-1}AC = D,$$

where $D = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$. By rearranging the columns of C if necessary, we may assume that eigenvalues in D appear in ***monotonic increasing order:***

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

Thus, any ***repeated*** eigenvalues will all appear ***consecutively***. Let us denote our ***distinct*** eigenvalues as:

$$\lambda_1^{/} < \lambda_2^{/} < \cdots < \lambda_k^{/},$$

where all we know is that $\lambda_1 = \lambda_1^{/}$ and $\lambda_n = \lambda_k^{/}$. At this point, let us also state that $dim\left(Eig\left(A,\lambda_i^{/}\right)\right) = n_i$, where $n_1 + n_2 + \cdots + n_k = n$. Now that we have our notation in order, let us continue.

Although we know that $C^{-1}AC$ is diagonal, we know nothing about $C^{-1}BC$, so let us give it a name, say, E:

$$C^{-1}BC = E.$$

But now, solving for A and B, we get:

$$A = CDC^{-1} \text{ and } B = CEC^{-1}, \text{ so:}$$

$$AB = (CDC^{-1})(CEC^{-1}) = CDEC^{-1}, \text{ and}$$

$$BA = (CEC^{-1})(CDC^{-1}) = CEDC^{-1}.$$

Since $AB = BA$, we can therefore conclude that $DE = ED$.

Let us now investigate the implications of this equation. Since D is diagonal, we obtain DE by multiplying row i of E by λ_i, and we obtain ED by multiplying column j of E by λ_j. If we write $E = [e_{i,j}]$ as usual, in order to satisfy $DE = ED$, we must have for all $i, j = 1 \ldots n$:

$$\begin{aligned} [DE]_{i,j} &= [ED]_{i,j} && \Rightarrow \\ \lambda_i e_{i,j} &= \lambda_j e_{i,j} && \Rightarrow \\ (\lambda_i - \lambda_j) e_{i,j} &= 0 && \Rightarrow \\ \text{either } \lambda_i = \lambda_j & \text{ or } e_{i,j} = 0. \end{aligned}$$

Since repeated eigenvalues are grouped together, this tells us that E is in ***block diagonal form:***

$$E = \begin{bmatrix} E_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & E_k \end{bmatrix}.$$

Each block E_i corresponds to $Eig\left(A, \lambda_i^{/}\right)$, and has dimension $n_i \times n_i$, and each $\mathbf{0}$ represents a zero matrix of an appropriate size. Correspondingly, we can also express the diagonal matrix D also in block diagonal form:

$$D = \begin{bmatrix} \Lambda_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Lambda_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \Lambda_k \end{bmatrix}$$

where $\Lambda_i = \lambda_i^{/} \boldsymbol{I}_{n_i}$. We note that it is appropriate that we use the symbol Λ, which is the capital version of λ.

Now, recall from the section on Similarity that ***diagonalizability*** is an ***invariant*** under ***similarity***. Thus, since B is diagonalizable, E is also diagonalizable. But according to our previous Theorem on direct sums, this implies that every E_i is diagonalizable. Thus for every E_i, we can find an invertible matrix G_i with the same size as E_i, such that $G_i^{-1} E_i G_i = D_i$, a diagonal matrix. However, since $\Lambda_i = \lambda_i^{/} \boldsymbol{I}_{n_i}$, we get:

$$G_i^{-1} \Lambda_i G_i = G_i^{-1} \lambda_i^{/} \boldsymbol{I}_{n_i} G_i = \lambda_i^{/} \boldsymbol{I}_{n_i} = \Lambda_i$$

Thus, if we let $G = G_1 \oplus G_2 \oplus \cdots \oplus G_k$, then:

$$\begin{aligned} G^{-1} E G &= \begin{bmatrix} G_1^{-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G_2^{-1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_k^{-1} \end{bmatrix} \begin{bmatrix} E_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & E_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & E_k \end{bmatrix} \begin{bmatrix} G_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_k \end{bmatrix} \\ &= \begin{bmatrix} G_1^{-1} E_1 G_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G_2^{-1} E_2 G_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & G_k^{-1} E_k G_k \end{bmatrix} = \begin{bmatrix} D_1 & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & D_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & D_k \end{bmatrix} = D^{/}, \quad \text{and} \\ G^{-1} D G &= D, \end{aligned}$$

with both D' and D diagonal. Thus, if we let $H = CG$, then:

$$H^{-1}AH = (CG)^{-1}A(CG) = G^{-1}C^{-1}ACG = G^{-1}DG = D, \text{ and}$$
$$H^{-1}BH = (CG)^{-1}B(CG) = G^{-1}C^{-1}BCG = G^{-1}EG = D'.$$

A and B are therefore simultaneously diagonalizable, with diagonalizing matrix H for both.■

Notice that the main idea in Case 1 was that every eigenvector of A is also an eigenvector of B, if each eigenspace of A is one dimensional and $AB = BA$. Nowhere in Case 1 did we use the fact that B is also diagonalizable. But in this case, if $S = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ is a complete set of independent eigenvectors for A, then S is also a complete set of eigenvectors for B. Thus, B is ***automatically*** diagonalizable as well, and we can rephrase Case 1 into the following simpler version:

Theorem — The Simple Simultaneous Diagonalizability Theorem:
If A and B are both $n \times n$ matrices, A has n ***distinct*** eigenvalues and $AB = BA$, then A and B are ***simultaneously diagonalizable***, even if we do not know beforehand that B is diagonalizable.

If we are in Case 2 of the Proof, we need two matrices, C and G, to obtain our final matrix $H = CG$. Let us give these matrices special names:

Definition: In the notation of Case 2 of the Main Theorem, we call C the ***main factor***, G the ***secondary factor***, and $H = CG$ the ***simultaneous diagonalizing matrix*** for A and B.

There are many computational steps in the Proof of Main Theorem, including finding the characteristic polynomials, finding eigenvalues (finding the roots of the characteristic polynomial, which may be cubic or larger), finding a basis for each eigenspace of both matrices, multiplying matrices, assembling the diagonalizing matrix C or $H = CG$, and finding inverses. If the student has sufficiently mastered all these processes, it would be appropriate to use technology to perform these steps in the interest of saving time and effort. In our examples, we will freely use technology to assist our computations. We encourage the reader to experiment with technology and verify that our computations are correct.

Example: Consider the two matrices:

$$A = \begin{bmatrix} 22 & 42 & -9 \\ -17 & -33 & 7 \\ -30 & -60 & 13 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 19 & 30 & -5 \\ -15 & -26 & 5 \\ -30 & -60 & 14 \end{bmatrix}.$$

First, let us check that:

$$AB = \begin{bmatrix} 58 & 108 & -26 \\ -38 & -72 & 18 \\ -60 & -120 & 32 \end{bmatrix} = BA.$$

Thus, the matrices commute. Notice that neither matrix is symmetric, so there is no guarantee that they are both diagonalizable. Using some technology, we find that:

$$p_A(\lambda) = \lambda^3 - 2\lambda^2 - 5\lambda + 6 = (\lambda+2)(\lambda-1)(\lambda-3), \text{ and}$$
$$p_B(\lambda) = \lambda^3 - 7\lambda^2 + 8\lambda + 16 = (\lambda+1)(\lambda-4)^2.$$

We see that the eigenvalues of A all have geometric multiplicity 1, and so we are in the Simple Case of the Theorem. All we need is a basis for each eigenspace of A, which again we can find using some technology:

$$\begin{aligned} Eig(A,-2) &= Span(\{\langle -1, 1, 2\rangle\}), \\ Eig(A,1) &= Span(\{\langle -2, 1, 0\rangle\}), \text{ and} \\ Eig(A,3) &= Span(\{\langle -3, 2, 3\rangle\}). \end{aligned}$$

Thus, we can let:

$$C = \begin{bmatrix} -1 & -2 & -3 \\ 1 & 1 & 2 \\ 2 & 0 & 3 \end{bmatrix}, \text{ with } C^{-1} = \begin{bmatrix} 3 & 6 & -1 \\ 1 & 3 & -1 \\ -2 & -4 & 1 \end{bmatrix}.$$

We verify that:

$$C^{-1}AC = \begin{bmatrix} 3 & 6 & -1 \\ 1 & 3 & -1 \\ -2 & -4 & 1 \end{bmatrix}\begin{bmatrix} 22 & 42 & -9 \\ -17 & -33 & 7 \\ -30 & -60 & 13 \end{bmatrix}\begin{bmatrix} -1 & -2 & -3 \\ 1 & 1 & 2 \\ 2 & 0 & 3 \end{bmatrix} = \begin{bmatrix} -2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \text{ and}$$
$$C^{-1}BC = \begin{bmatrix} 3 & 6 & -1 \\ 1 & 3 & -1 \\ -2 & -4 & 1 \end{bmatrix}\begin{bmatrix} 19 & 30 & -5 \\ -15 & -26 & 5 \\ -30 & -60 & 14 \end{bmatrix}\begin{bmatrix} -1 & -2 & -3 \\ 1 & 1 & 2 \\ 2 & 0 & 3 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

Thus, we have simultaneously diagonalized A and B. □

Example: Consider the two matrices:

$$A = \begin{bmatrix} 7 & 4 & -16 \\ 6 & 9 & -24 \\ 2 & 2 & -5 \end{bmatrix}, \text{ and } B = \begin{bmatrix} 19 & -10 & -10 \\ 20 & -11 & -10 \\ 10 & -5 & -6 \end{bmatrix}.$$

Again, we first check that the two matrices commute:

$$AB = \begin{bmatrix} 53 & -34 & -14 \\ 54 & -39 & -6 \\ 28 & -17 & -10 \end{bmatrix} = BA.$$

Their characteristics polynomials are:

$$p_A(\lambda) = \lambda^3 - 11\lambda^2 + 39\lambda - 45 = (\lambda-3)^2(\lambda-5), \text{ and}$$
$$p_B(\lambda) = \lambda^3 - 2\lambda^2 - 7\lambda - 4 = (\lambda+1)^2(\lambda-4).$$

Since both matrices have a repeated eigenvalue, it is not obvious that they are diagonalizable. Again, with the use of some technology, we get:

$$\begin{aligned} Eig(A,3) &= Span(\{\langle -1, 1, 0\rangle, \langle 4, 0, 1\rangle\}), \\ Eig(A,5) &= Span(\{\langle 2, 3, 1\rangle\}), \\ Eig(B,-1) &= Span(\{\langle 1, 2, 0\rangle, \langle 1, 0, 2\rangle\}), \quad \text{and} \\ Eig(B,4) &= Span(\{\langle 2, 2, 1\rangle\}). \end{aligned}$$

Thus, both matrices are diagonalizable. This time, we are in Case 2 of the Proof because both matrices have a 2-dimensional eigenspace. In our lists above, we deliberately arranged the eigenvalues of both matrices in increasing order, as prescribed in the Proof. Choosing the eigenvectors of A in the order stated, we assemble the ***main factor:***

$$C = \begin{bmatrix} -1 & 4 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 1 \end{bmatrix}, \quad \text{with inverse} \quad C^{-1} = \begin{bmatrix} -3 & -2 & 12 \\ -1 & -1 & 5 \\ 1 & 1 & -4 \end{bmatrix}.$$

We check that:

$$C^{-1}AC = \begin{bmatrix} -3 & -2 & 12 \\ -1 & -1 & 5 \\ 1 & 1 & -4 \end{bmatrix}\begin{bmatrix} 7 & 4 & -16 \\ 6 & 9 & -24 \\ 2 & 2 & -5 \end{bmatrix}\begin{bmatrix} -1 & 4 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \quad \text{but}$$

$$C^{-1}BC = \begin{bmatrix} -3 & -2 & 12 \\ -1 & -1 & 5 \\ 1 & 1 & -4 \end{bmatrix}\begin{bmatrix} 19 & -10 & -10 \\ 20 & -11 & -10 \\ 10 & -5 & -6 \end{bmatrix}\begin{bmatrix} -1 & 4 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -31 & 70 & 0 \\ -15 & 34 & 0 \\ 0 & 0 & -1 \end{bmatrix}.$$

As predicted in the Proof, $C^{-1}BC$ is not necessarily diagonal, but it is in ***block diagonal form***. We need only to diagonalize the block:

$$E_1 = \begin{bmatrix} -31 & 70 \\ -15 & 34 \end{bmatrix},$$

since the other block $E_2 = [-1]$ is already diagonal. The characteristic polynomial of E_1 is:

$$p_{E_1}(\lambda) = (\lambda + 31)(\lambda - 34) - (-15)(70) = \lambda^2 - 3\lambda - 4 = (\lambda + 1)(\lambda - 4)$$

Notice that the eigenvalue $\lambda = -1$ already appears in E_2. This should be the case because $Eig(B,-1)$ is 2-dimensional. Continuing with the algorithm, we find the eigenvectors:

$$\begin{aligned} Eig(E_1,-1) &= Span(\{\langle 7, 3\rangle\}), \quad \text{and} \\ Eig(E_1,4) &= Span(\{\langle 2, 1\rangle\}). \end{aligned}$$

Thus, we get the diagonalizing matrix for E_1:

$$G_1 = \begin{bmatrix} 7 & 2 \\ 3 & 1 \end{bmatrix}, \quad \text{with inverse} \quad G_1^{-1} = \begin{bmatrix} 1 & -2 \\ -3 & 7 \end{bmatrix}.$$

Although E_2 is already diagonal, we still need $G_2 = [1]$ to "trivially" diagonalize E_2. The next step of the Proof says that we assemble the ***secondary factor*** as the direct sum:

$$G = G_1 \oplus G_2 = \begin{bmatrix} 7 & 2 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \text{ with inverse:}$$

$$G^{-1} = G_1^{-1} \oplus G_2^{-1} = \begin{bmatrix} 1 & -2 & 0 \\ -3 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

From this, we can construct the ***simultaneous diagonalizing matrix:***

$$H = CG = \begin{bmatrix} -1 & 4 & 2 \\ 1 & 0 & 3 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} 7 & 2 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 2 & 2 \\ 7 & 2 & 3 \\ 3 & 1 & 1 \end{bmatrix}, \quad \text{with inverse:}$$

$$H^{-1} = G^{-1}C^{-1} = \begin{bmatrix} 1 & -2 & 0 \\ -3 & 7 & 0 \\ 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -3 & -2 & 12 \\ -1 & -1 & 5 \\ 1 & 1 & -4 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & -4 \end{bmatrix}.$$

Finally, we can check that:

$$H^{-1}AH = \begin{bmatrix} -1 & 0 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & -4 \end{bmatrix}\begin{bmatrix} 7 & 4 & -16 \\ 6 & 9 & -24 \\ 2 & 2 & -5 \end{bmatrix}\begin{bmatrix} 5 & 2 & 2 \\ 7 & 2 & 3 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix} = D_1, \text{ and}$$

$$H^{-1}BH = \begin{bmatrix} -1 & 0 & 2 \\ 2 & -1 & -1 \\ 1 & 1 & -4 \end{bmatrix}\begin{bmatrix} 19 & -10 & -10 \\ 20 & -11 & -10 \\ 10 & -5 & -6 \end{bmatrix}\begin{bmatrix} 5 & 2 & 2 \\ 7 & 2 & 3 \\ 3 & 1 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & -1 \end{bmatrix} = D_2.$$

Thus, both matrices have been diagonalized. Note, however, that the Proof only required that the eigenvalues of A appear in increasing order in D_1, but did not guarantee this ordering in D_2. □

The Symmetric Case

In the special case that A and B are ***symmetric*** matrices, then we know from the Spectral Theorem that they are automatically ***diagonalizable***. Furthermore, for each matrix, we can choose our diagonalizing matrix Q to be an ***orthogonal*** matrix, that is, $Q^{-1} = Q^\top$. It is therefore natural to wonder if we can simultaneously orthogonally diagonalize A and B, that is, can we find an orthogonal matrix Q such that $Q^\top AQ = D_1$ and $Q^\top BQ = D_2$, where D_1 and D_2 are both diagonal matrices. The following should not be a surprise:

Theorem — The Simultaneous Orthogonal Diagonalizability Theorem for Symmetric Matrices:

If A and B are both ***symmetric*** $n \times n$ matrices, then they are ***simultaneously orthogonally diagonalizable*** ***if and only if*** $AB = BA$.

The Proof is an easy modification of that for our Main Theorem, and will be outlined in the Exercises. The ideas behind it will be seen in the Example below.

Example: Consider the two matrices:

$$A = \begin{bmatrix} 0 & 2 & 0 & -2 \\ 2 & -1 & -2 & 1 \\ 0 & -2 & -2 & -2 \\ -2 & 1 & -2 & -1 \end{bmatrix}, \quad \text{and} \quad B = \begin{bmatrix} -15 & -8 & -16 & 40 \\ -8 & 17 & -32 & -8 \\ -16 & -32 & -31 & -16 \\ 40 & -8 & -16 & -15 \end{bmatrix}.$$

Both are obviously symmetric, hence orthogonally diagonalizable, but we still have to check that:

$$AB = \begin{bmatrix} -96 & 50 & -32 & 14 \\ 50 & 23 & 46 & 105 \\ -32 & 46 & 158 & 78 \\ 14 & 105 & 78 & -41 \end{bmatrix} = BA.$$

Their characteristics polynomials are:

$$p_A(\lambda) = \lambda^4 + 4\lambda^3 - 12\lambda^2 - 32\lambda + 64 = (\lambda+4)^2(\lambda-2)^2, \quad \text{and}$$
$$p_B(\lambda) = \lambda^4 + 44\lambda^3 - 3146\lambda^2 - 79860\lambda + 3294225 = (\lambda+55)^2(\lambda-33)^2.$$

Since we are certain that both matrices are diagonalizable, we only need to find eigenspaces for one of them to construct the main factor C. Although they both have two 2-dimensional eigenspaces, A has smaller eigenvalues, so we find:

$$Eig(A,-4) = Span(\{\langle -1, 2, 2, 0\rangle, \langle 1, -1, 0, 1\rangle\}), \quad \text{and}$$
$$Eig(A,2) = Span(\{\langle -2, -2, 1, 0\rangle, \langle -2, -1, 0, 1\rangle\}).$$

We apply the ***Gram-Schmidt Algorithm*** to our basis for each eigenspace. For $Eig(A,-4)$:

$$\vec{v}_1 = \langle -1, 2, 2, 0\rangle, \quad \text{and}$$
$$\vec{v}_2 = \langle 1, -1, 0, 1\rangle - \frac{\langle 1, -1, 0, 1\rangle \circ \langle -1, 2, 2, 0\rangle}{\langle -1, 2, 2, 0\rangle \circ \langle -1, 2, 2, 0\rangle}\langle -1, 2, 2, 0\rangle$$
$$= \langle 1, -1, 0, 1\rangle - \frac{-3}{9}\langle -1, 2, 2, 0\rangle = \frac{1}{3}\langle 2, -1, 2, 3\rangle,$$

so we shall use $\vec{v}_2 = \langle 2, -1, 2, 3\rangle$. For $Eig(A,2)$:

$$\vec{v}_3 = \langle -2, -2, 1, 0\rangle, \quad \text{and}$$
$$\vec{v}_4 = \langle -2, -1, 0, 1\rangle - \frac{\langle -2, -1, 0, 1\rangle \circ \langle -2, -2, 1, 0\rangle}{\langle -2, -2, 1, 0\rangle \circ \langle -2, -2, 1, 0\rangle}\langle -2, -2, 1, 0\rangle$$
$$= \langle -2, -1, 0, 1\rangle - \frac{6}{9}\langle -2, -2, 1, 0\rangle = \frac{1}{3}\langle -2, 1, -2, 3\rangle,$$

so we shall use $\vec{v}_4 = \langle -2, 1, -2, 3\rangle$. A quick check shows that:

$$S = \{\langle -1, 2, 2, 0\rangle, \langle 2, -1, 2, 3\rangle, \langle -2, -2, 1, 0\rangle, \langle -2, 1, -2, 3\rangle\}$$

is an ***orthogonal*** set. These vectors have lengths 3, $3\sqrt{2}$, 3, and $3\sqrt{2}$, respectively.

Thus, our ***main factor*** will be:

$$C = \begin{bmatrix} \frac{-1}{3} & \frac{2}{3\sqrt{2}} & \frac{-2}{3} & \frac{-2}{3\sqrt{2}} \\ \frac{2}{3} & \frac{-1}{3\sqrt{2}} & \frac{-2}{3} & \frac{1}{3\sqrt{2}} \\ \frac{2}{3} & \frac{2}{3\sqrt{2}} & \frac{1}{3} & \frac{-2}{3\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix}, \quad \text{with} \quad C^{-1} = C^{\top} = \begin{bmatrix} \frac{-1}{3} & \frac{2}{3} & \frac{2}{3} & 0 \\ \frac{2}{3\sqrt{2}} & \frac{-1}{3\sqrt{2}} & \frac{2}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{-2}{3} & \frac{-2}{3} & \frac{1}{3} & 0 \\ \frac{-2}{3\sqrt{2}} & \frac{1}{3\sqrt{2}} & \frac{-2}{3\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

We check that:

$$C^{\top}AC = \begin{bmatrix} -4 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \quad \text{but} \quad C^{\top}BC = \begin{bmatrix} -\frac{77}{3} & -\frac{88\sqrt{2}}{3} & 0 & 0 \\ -\frac{88\sqrt{2}}{3} & \frac{11}{3} & 0 & 0 \\ 0 & 0 & \frac{35}{3} & -\frac{80\sqrt{2}}{3} \\ 0 & 0 & -\frac{80\sqrt{2}}{3} & -\frac{101}{3} \end{bmatrix}.$$

As expected, we were not guaranteed that $Q^{\top}BQ$ is diagonal, but it is ***symmetric*** because B is symmetric. Therefore, ***each block*** in this matrix is likewise ***symmetric***. We will orthogonally diagonalize each 2×2 block:

$$E_1 = \begin{bmatrix} -\frac{77}{3} & -\frac{88\sqrt{2}}{3} \\ -\frac{88\sqrt{2}}{3} & \frac{11}{3} \end{bmatrix}, \quad \text{with eigenspaces:}$$

$$Eig(E_1, 33) = Span\left(\left\{\left\langle \sqrt{2}, -2 \right\rangle\right\}\right) = Span\left(\left\{\left\langle \tfrac{1}{\sqrt{3}}, \tfrac{-2}{\sqrt{6}} \right\rangle\right\}\right), \text{ and}$$

$$Eig(E_1, -55) = Span\left(\left\{\left\langle \sqrt{2}, 1 \right\rangle\right\}\right) = Span\left(\left\{\left\langle \tfrac{2}{\sqrt{6}}, \tfrac{1}{\sqrt{3}} \right\rangle\right\}\right);$$

$$E_2 = \begin{bmatrix} \frac{35}{3} & -\frac{80\sqrt{2}}{3} \\ -\frac{80\sqrt{2}}{3} & -\frac{101}{3} \end{bmatrix}, \quad \text{with eigenspaces:}$$

$$Eig(E_1, 33) = Span\left(\left\{\left\langle 5\sqrt{2}, -4 \right\rangle\right\}\right) = Span\left(\left\{\left\langle \tfrac{5}{\sqrt{33}}, \tfrac{-4}{\sqrt{66}} \right\rangle\right\}\right), \text{ and}$$

$$Eig(E_1, -55) = Span\left(\left\{\left\langle 2\sqrt{2}, 5 \right\rangle\right\}\right) = Span\left(\left\{\left\langle \tfrac{4}{\sqrt{66}}, \tfrac{5}{\sqrt{33}} \right\rangle\right\}\right).$$

Notice that the eigenspaces are 1-dimensional, and we found bases containing a unit vector. Thus, the blocks are ***orthogonally diagonalizable*** using:

$$G_1 = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}, \quad \text{with inverse} \quad G_1^{-1} = G_1^{\top} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix}, \quad \text{and}$$

$$G_2 = \begin{bmatrix} \frac{5}{\sqrt{33}} & \frac{4}{\sqrt{66}} \\ \frac{-4}{\sqrt{66}} & \frac{5}{\sqrt{33}} \end{bmatrix}, \quad \text{with inverse} \quad G_2^{-1} = G_2^{\top} = \begin{bmatrix} \frac{5}{\sqrt{33}} & \frac{-4}{\sqrt{66}} \\ \frac{4}{\sqrt{66}} & \frac{5}{\sqrt{33}} \end{bmatrix}.$$

We assemble our ***secondary matrix*** (which is also ***orthogonal***), $G = G_1 \oplus G_2$, with inverse $G^{-1} = G^\top = G_1^\top \oplus G_2^\top$:

$$G = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 & 0 \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & \frac{5}{\sqrt{33}} & \frac{4}{\sqrt{66}} \\ 0 & 0 & \frac{-4}{\sqrt{66}} & \frac{5}{\sqrt{33}} \end{bmatrix} \text{ and } G^{-1} = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} & 0 & 0 \\ \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & \frac{5}{\sqrt{33}} & \frac{-4}{\sqrt{66}} \\ 0 & 0 & \frac{4}{\sqrt{66}} & \frac{5}{\sqrt{33}} \end{bmatrix}.$$

Our ***simultaneous orthogonal diagonalizing matrix*** Q is thus:

$$Q = CG = \begin{bmatrix} \frac{-1}{3} & \frac{2}{3\sqrt{2}} & \frac{-2}{3} & \frac{-2}{3\sqrt{2}} \\ \frac{2}{3} & \frac{-1}{3\sqrt{2}} & \frac{-2}{3} & \frac{1}{3\sqrt{2}} \\ \frac{2}{3} & \frac{2}{3\sqrt{2}} & \frac{1}{3} & \frac{-2}{3\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}} & 0 & 0 \\ \frac{-2}{\sqrt{6}} & \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & 0 & \frac{5}{\sqrt{33}} & \frac{4}{\sqrt{66}} \\ 0 & 0 & \frac{-4}{\sqrt{66}} & \frac{5}{\sqrt{33}} \end{bmatrix}$$

$$= \begin{bmatrix} \frac{-1}{\sqrt{3}} & 0 & \frac{-2}{\sqrt{33}} & \frac{-6}{\sqrt{66}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-4}{\sqrt{33}} & \frac{-1}{\sqrt{66}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{3}{\sqrt{33}} & \frac{-2}{\sqrt{66}} \\ \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{33}} & \frac{5}{\sqrt{66}} \end{bmatrix}, \text{ with inverse:}$$

$$Q^{-1} = Q^\top = \begin{bmatrix} \frac{-1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & \frac{-1}{\sqrt{3}} \\ 0 & \frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{6}} \\ \frac{-2}{\sqrt{33}} & \frac{-4}{\sqrt{33}} & \frac{3}{\sqrt{33}} & \frac{-2}{\sqrt{33}} \\ \frac{-6}{\sqrt{66}} & \frac{-1}{\sqrt{66}} & \frac{-2}{\sqrt{66}} & \frac{5}{\sqrt{66}} \end{bmatrix}$$

Finally, we verify that:

$$Q^\top A Q = \begin{bmatrix} -4 & 0 & 0 & 0 \\ 0 & -4 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, \text{ and } Q^\top A Q = \begin{bmatrix} 33 & 0 & 0 & 0 \\ 0 & -55 & 0 & 0 \\ 0 & 0 & 33 & 0 \\ 0 & 0 & 0 & -55 \end{bmatrix}.$$

Again, only the eigenvalues of A were guaranteed to appear in increasing order, but not the eigenvalues of B. □

In closing, we know that there is a one-to-one correspondence between ***operators*** T on $\mathbb{R}^n$, and $n \times n$ matrices $A = [T]$, where matrix multiplication corresponds to the ***composition*** of transformations. We also know that if $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ is any ***basis*** for $\mathbb{R}^n$, then Section 6.4 tells us that:

$$[T]_S = [S]^{-1}[T][S],$$

where $[S] = [\vec{w}_1\, \vec{w}_2 \ldots \vec{w}_n]$, as usual. In other words, $[S]$ is the ***change of basis matrix*** that transforms $[T]$ to $[T]_S$.

Let us now rephrase our Theorems in the language of operators:

Theorem — The Simultaneous Diagonalizability Theorem for Operators:
Suppose that T_1 and T_2 are operators on $\mathbb{R}^n$.

If one of the operators, say T_1, has n distinct eigenvalues, and $T_1 \circ T_2 = T_2 \circ T_1$, then the set of eigenvectors $B = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ for T_1, made by choosing one eigenvector from each 1-dimensional eigenspace of T_1, is also a set of eigenvectors for T_2.

More generally, suppose there exists a basis $S_1 = \{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors for T_1, and another basis $S_2 = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors for T_2, where the eigenspaces may have a dimension bigger than 1.

Then: $T_1 \circ T_2 = T_2 \circ T_1$, that is, the operators commute, ***if and only if*** there exists a basis $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ for $\mathbb{R}^n$ consisting of eigenvectors for **both** T_1 and T_2. In this case, $T_1(\vec{w}_i) = \lambda_i \vec{w}_i$ and $T_2(\vec{w}_i) = \mu_i \vec{w}_i$, for some eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ for T_1, and eigenvalues $\mu_1, \mu_2, \ldots, \mu_n$ for T_2, and for all $i = 1 \ldots n$. Thus:

$$[T_1]_S = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n), \quad \text{and}$$

$$[T_2]_S = Diag(\mu_1, \mu_2, \ldots, \mu_n).$$

Moreover, if S_1 and S_2 can both be chosen to be ***orthogonal*** bases for $\mathbb{R}^n$, then another orthogonal basis $S = \{\vec{w}_1, \vec{w}_2, \ldots, \vec{w}_n\}$ exists consisting of eigenvectors for **both** T_1 and T_2.

8.7 Section Summary

Let A and B be $n \times n$ matrices. We say that A and B are ***simultaneously diagonalizable*** if there exists an ***invertible*** matrix C such that $C^{-1}AC = D_1$, and $C^{-1}BC = D_2$, where D_1 and D_2 are both ***diagonal*** matrices.

Let E_1, E_2, …, E_k be square matrices, not necessarily of the same size. Then: the matrix $E = E_1 \oplus E_2 \oplus \cdots \oplus E_k$ is ***diagonalizable*** ***if and only if*** every block E_i is ***diagonalizable***.

The Simultaneous Diagonalizability Theorem: If A and B are both ***diagonalizable*** $n \times n$ matrices, then A and B are ***simultaneously diagonalizable*** ***if and only if*** $AB = BA$.

If A and B are both $n \times n$ matrices, A has n distinct eigenvalues and $AB = BA$, then A and B are ***simultaneously diagonalizable***, even if we do not know beforehand that B is diagonalizable.

If A and B are both ***symmetric*** $n \times n$ matrices, then they are ***simultaneously orthogonally diagonalizable*** ***if and only if*** $AB = BA$.

These Theorems can also be stated in the language of operators, as shown above.

8.7 Exercises

For Exercises (1) to (14): The goal is to simultaneously diagonalize A and B in each of the following problems using the same invertible matrix C. Verify first that $AB = BA$. Find the characteristic polynomial of both matrices and find all the eigenvalues. If one of the matrices has only 1-dimensional eigenspaces, diagonalize both matrices using the eigenvectors of that matrix, as in Case 1 of the proof. If both matrices have an eigenspace which is at least 2-dimensional, use Case 2 of the proof to find the diagonalizing matrix $H = CG$ (and for the sake of uniformity, use the eigenvectors of A to construct C, and arrange them in order of increasing eigenvalues as is done in the proof). If both A and B are symmetric, construct C to be an orthogonal matrix, as in the last Example. ***We strongly recommend the use of technology to assist in the computations.*** We note, though, that some packages do not use the Gauss-Jordan Algorithm to find a basis for the nullspace, and consequently eigenvectors, and so the simultaneous diagonalizing matrix that you obtain may be different from that in the Answer Key. However, the final diagonal matrices (i.e. containing the eigenvalues) should be the same, up to ordering.

1. $A = \begin{bmatrix} 21 & 32 & -16 \\ -14 & -20 & 11 \\ -4 & -2 & 3 \end{bmatrix}, \ B = \begin{bmatrix} 7 & 12 & -6 \\ -9 & -14 & 6 \\ -9 & -12 & 4 \end{bmatrix}$

2. $A = \begin{bmatrix} -9 & -8 & 16 \\ -12 & -5 & 16 \\ -12 & -8 & 19 \end{bmatrix}, \quad B = \begin{bmatrix} -12 & -10 & 20 \\ -23 & -7 & 28 \\ -19 & -10 & 27 \end{bmatrix}$

3. $A = \begin{bmatrix} -23 & 20 & 40 \\ -10 & 7 & 20 \\ -10 & 10 & 17 \end{bmatrix}, \quad B = \begin{bmatrix} -4 & 9 & 9 \\ -3 & 8 & 3 \\ -3 & 3 & 8 \end{bmatrix}$

4. $A = \begin{bmatrix} 11 & -8 & 8 \\ 4 & -1 & 4 \\ -6 & 6 & -3 \end{bmatrix}, \quad B = \begin{bmatrix} 9 & 14 & 28 \\ 7 & 2 & 14 \\ -7 & -7 & -19 \end{bmatrix}$

5. $A = \begin{bmatrix} -2 & 10 & -20 \\ 10 & 13 & 10 \\ -20 & 10 & -2 \end{bmatrix}, \ B = \begin{bmatrix} 10 & 4 & 19 \\ 4 & 43 & 4 \\ 19 & 4 & 10 \end{bmatrix}$

6. $A = \begin{bmatrix} 2 & 2 & 4 \\ 2 & -1 & 2 \\ 4 & 2 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 41 & 16 & -4 \\ 16 & -19 & 16 \\ -4 & 16 & 41 \end{bmatrix}$

7. $A = \begin{bmatrix} 3 & -2 & -1 \\ -2 & 0 & -2 \\ -1 & -2 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 5 \\ 0 & 6 & 0 \\ 5 & 0 & 1 \end{bmatrix}$

8. $A = \begin{bmatrix} -16 & -10 & 13 & 38 \\ -20 & -6 & 15 & 40 \\ 44 & 10 & -35 & -88 \\ -29 & -10 & 23 & 61 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & 6 & -5 & -16 \\ 12 & 5 & -9 & -24 \\ -28 & -6 & 24 & 56 \\ 16 & 6 & -13 & -33 \end{bmatrix}$

9. $A = \begin{bmatrix} 58 & 8 & -12 & -16 \\ 8 & 47 & -10 & 27 \\ -12 & -10 & 4 & -2 \\ -16 & 27 & -2 & 23 \end{bmatrix}, \quad B = \begin{bmatrix} 4 & -10 & 0 & -10 \\ -10 & 9 & 10 & 5 \\ 0 & 10 & 14 & -10 \\ -10 & 5 & -10 & 9 \end{bmatrix}$

10. $A = \begin{bmatrix} -15 & 10 & 10 & -4 \\ -14 & 12 & 10 & 2 \\ -3 & 0 & 2 & -6 \\ 7 & -5 & -5 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 7 & 0 & 0 & 8 \\ 8 & -1 & 0 & 8 \\ -4 & 4 & 3 & 0 \\ -4 & 0 & 0 & -5 \end{bmatrix}$

11. $A = \begin{bmatrix} 4 & 1 & -3 & -2 \\ 1 & 4 & 3 & 2 \\ -3 & 3 & -4 & -6 \\ -2 & 2 & -6 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & -1 & 0 & -1 \\ -1 & -1 & 1 & 0 \\ 0 & 1 & -2 & -1 \\ -1 & 0 & -1 & -1 \end{bmatrix}$

12. $A = \begin{bmatrix} -9 & 9 & 12 & 3 & 15 \\ 9 & -29 & 8 & -23 & -15 \\ 12 & 8 & -56 & -4 & 0 \\ 3 & -23 & -4 & -41 & 15 \\ 15 & -15 & 0 & 15 & -45 \end{bmatrix}, \quad B = \begin{bmatrix} 11 & 9 & 12 & 3 & 15 \\ 9 & -9 & 8 & -23 & -15 \\ 12 & 8 & -16 & 16 & -20 \\ 3 & -23 & 16 & -1 & -5 \\ 15 & -15 & -20 & -5 & -5 \end{bmatrix}$

13. $A = \begin{bmatrix} 25 & 10 & 5 & 5 & 10 \\ 10 & 4 & 2 & 2 & 4 \\ 5 & 2 & 1 & 1 & 2 \\ 5 & 2 & 1 & 1 & 2 \\ 10 & 4 & 2 & 2 & 4 \end{bmatrix}, \quad B = \begin{bmatrix} -2 & 2 & 1 & 1 & 2 \\ 2 & -2 & -1 & -1 & -2 \\ 1 & -1 & -4 & 3 & -1 \\ 1 & -1 & 3 & -4 & -1 \\ 2 & -2 & -1 & -1 & -2 \end{bmatrix}$

14. $A = \begin{bmatrix} 54 & 162 & 13 & -22 & 64 & -61 \\ 0 & -22 & 7 & 24 & -12 & 9 \\ -30 & -80 & -10 & 0 & -30 & 30 \\ -13 & -40 & -3 & 7 & -16 & 15 \\ -45 & -116 & -17 & -2 & -43 & 43 \\ -2 & -42 & 11 & 40 & -22 & 17 \end{bmatrix}$, $B = \begin{bmatrix} 56 & 168 & 16 & -16 & 64 & -64 \\ -9 & -41 & -2 & 6 & -12 & 18 \\ -27 & -74 & -8 & 6 & -30 & 27 \\ -14 & -42 & -4 & 4 & -16 & 16 \\ -40 & -106 & -12 & 8 & -44 & 38 \\ -17 & -72 & -4 & 10 & -22 & 31 \end{bmatrix}$

15. Prove the following subtle relaxation of our Theorem on symmetric matrices: Suppose that A is a ***symmetric*** matrix, B is a ***diagonalizable*** matrix, and $AB = BA$. Prove that B is also ***symmetric***.

16. In Exercise 43 of Section 3.4, we defined the ***centralizer*** of a matrix A as the set:

$$\textbf{\textit{Centralizer}}(A) = \{B \in Mat(n) \mid AB = BA\}.$$

In other words, $Centralizer(A)$ consists of all the matrices B that ***commute*** with A. In that Exercise, we saw that $Centralizer(A)$ is always a subspace of the space of all $n \times n$ matrices, and it is an infinite set containing all multiples of $\boldsymbol{I}_n$.

a. Let B_1 and B_2 be from $Centralizer(A)$. Prove that B_1B_2 and B_2B_1 are also in $Centralizer(A)$.

Now, suppose that A has ***distinct eigenvalues***.

b. Prove that $B_1B_2 = B_2B_1$. Hint: review the proof of Case 1 of our main theorem in this Section.

c. If B is an invertible matrix from $Centralizer(A)$, prove that B^{-1} is also in $Centralizer(A)$.

This Exercise shows that the invertible matrices in $Centralizer(A)$ form a ***commutative*** or ***abelian group*** under matrix multiplication.

17. The goal of this Exercise is to mimic the Proof of the Simultaneous Diagonalizability Theorem to prove the Simultaneous Orthogonal Diagonalizability Theorem for Symmetric Matrices. Assume that both A and B are $n \times n$ ***symmetric*** matrices.

a. What does the Spectral Theorem tell us about ***symmetric*** matrices?

b. What do we know about the ***product*** of two $n \times n$ ***orthogonal*** matrices?

c. What do we know about the ***direct sum*** of two ***orthogonal*** matrices (not necessarily of the same size)?

d. Show that if A has n ***distinct*** eigenvalues, then the matrix C in Case 1 of the Proof can be chosen to be an ***orthogonal*** matrix Q. Explain why Q also diagonalizes B if $AB = BA$. This takes care of Case 1 of the Proof.

e. Now, suppose A has eigenspaces of dimension 2 or more. Explain how to produce an ***orthogonal*** basis for each eigenspace.

f. Explain how to produce an ***orthogonal*** matrix Q (the ***main factor***) and a diagonal matrix D such that $Q^\top AQ = D$, and the diagonal entries in D are in increasing order.

g. Explain why $Q^\top BQ$ is again in ***block diagonal form*** and is also a ***symmetric*** matrix. What can we also say about each block of $Q^\top BQ$?

h. Explain how to produce an ***orthogonal*** set of eigenvectors for each block of $Q^\top BQ$.

i. Explain how to obtain the ***secondary factor*** G, where G is also an ***orthogonal*** matrix.

j. Explain how to obtain the ***orthogonal diagonalizing matrix*** H, such that $H^\top AH$ and $H^\top BH$ are both diagonal.

A Summary of Chapter 8

The set of all complex numbers: $\mathbb{C} = \{a + bi | a, b \in \mathbb{R}\}$, is constructed using the imaginary unit $i = \sqrt{-1}$. $\mathbb{R}$ and $\mathbb{C}$ are examples of a ***field***, a set upon which we define an addition and a multiplication, such that the 11 field axioms are satisfied.

Complex Euclidean n-Space, $\mathbb{C}^n = \{\langle z_1, z_2, \ldots, z_n\rangle | z_i \in \mathbb{C}\}$, is the basic example of a vector space over $\mathbb{C}$. Other examples of vector spaces over $\mathbb{C}$ are $\mathbb{P}^n(\mathbb{C})$ and $Mat(\mathbb{C}, m, n)$.

The following terms can be defined for vector spaces over $\mathbb{C}$:

the ***Span*** of a set of vectors; the ***linear combinations*** of a set of vectors; linear ***dependence*** or ***independence*** of a set of vectors; a ***basis*** for a vector space; the ***dimension*** of a vector space; a ***subspace*** of a vector space; ***matrix arithmetic:*** addition, subtraction, multiplication, finding determinants and inverses of square matrices, when they exist; a ***linear transformation*** from one complex vector space to another; the ***matrix*** of a linear transformation from $\mathbb{C}^n$ to $\mathbb{C}^m$; the ***kernel***, ***range***, ***nullity*** and ***rank*** of a complex linear transformation; ***one-to-one*** linear transformations, ***onto*** linear transformations, and ***isomorphisms***; the ***characteristic polynomial*** of an $n \times n$ complex matrix; the ***eigenvalues*** and associated ***eigenvectors*** of an $n \times n$ complex matrix; the ***diagonalizability*** of an $n \times n$ complex matrix.

Let $\vec{z} = \langle z_1, z_2, \ldots, z_n\rangle$, $\vec{w} = \langle w_1, w_2, \ldots, w_n\rangle \in \mathbb{C}^n$. We define their ***Complex Euclidean inner product***, or simply their ***inner product***, by: $\langle\vec{z}|\vec{w}\rangle = \vec{z} \circ \overline{\vec{w}} = z_1\overline{w_1} + z_2\overline{w_2} + \cdots + z_n\overline{w_n}$.

Let $\vec{z}, \vec{w}, \vec{u} \in \mathbb{C}^n$, and $k \in \mathbb{C}$. Under the complex inner product, the following properties are true:

1. ***The Hermitian-Symmetry Property:*** $\langle\vec{z}|\vec{w}\rangle = \overline{\langle\vec{w}|\vec{z}\rangle}$
2. ***The Left Homogeneity Property:*** $\langle k \cdot \vec{z}|\vec{w}\rangle = k \cdot \langle\vec{z}|\vec{w}\rangle$
3. ***The Left Additivity Property:*** $\langle\vec{u} + \vec{v}|\vec{w}\rangle = \langle\vec{u}|\vec{w}\rangle + \langle\vec{v}|\vec{w}\rangle$
4. ***The Positivity Property:*** If $\vec{z} \neq \vec{0}_{\mathbb{C}^n}$, then $\langle\vec{z}|\vec{z}\rangle > 0$.

A complex vector space V is a ***complex inner product space*** under a bilinear form $\langle|\rangle$ if the above four axioms are satisfied by $\langle|\rangle$. The following properties also hold in a complex inner product space:

1. ***The Right Additivity Property:*** $\langle\vec{z}|\vec{w} + \vec{u}\rangle = \langle\vec{z}|\vec{w}\rangle + \langle\vec{z}|\vec{u}\rangle$
2. ***The Right Conjugate-Homogeneity Property:*** $\langle\vec{z}|k \cdot \vec{w}\rangle = \overline{k} \cdot \langle\vec{z}|\vec{w}\rangle$
3. ***The Inner Product with the Zero Vector Property:*** $\left\langle\vec{z}|\vec{\mathbf{0}}_V\right\rangle = 0_{\mathbb{C}} = \left\langle\vec{\mathbf{0}}_V|\vec{z}\right\rangle$.

We can define the following concepts for complex inner product spaces:

- the ***length*** of a vector: $\|\vec{z}\| = \sqrt{\langle\vec{z}|\vec{z}\rangle}$.
- the ***distance*** between two vectors: $d(\vec{z}, \vec{w}) = \|\vec{z} - \vec{w}\|$.
- orthogonality: $\vec{z}$ is ***orthogonal*** to $\vec{w}$ ***if and only if*** $\langle\vec{z}|\vec{w}\rangle = 0$.
- ***orthogonal*** and ***orthonormal sets*** of vectors.
- the applicability of the ***Gram-Schmidt Algorithm***.
- the ***orthogonal complement*** $W^\perp$ of a subspace W of V.

Let $T : V \to V$ be a linear operator on a (possibly infinite-dimensional) complex vector space V. The ***spectrum*** of T, denoted ***Spec(T)*** is the set of all eigenvalues of T.

Let A be an $n \times n$ complex matrix. We define the ***adjoint*** of A as: $A^* = \overline{A^\top}$. If $T : \mathbb{C}^n \to \mathbb{C}^n$ is a linear operator with standard matrix $[T]$, then T^* is the linear operator on $\mathbb{C}^n$ such that $[T^*] = [T]^*$.

A is ***Hermitian*** if $A = A^*$, or analogously: $\langle T(\vec{v})|\vec{w}\rangle = \langle\vec{v}|T(\vec{w})\rangle$ for all $\vec{v}, \vec{w} \in \mathbb{C}^n$.

A is ***Skew-Hermitian*** if $A = -A^*$, or analogously: $\langle T(\vec{v}), \vec{w} \rangle = -\langle \vec{v} | T(\vec{w}) \rangle$ for all $\vec{v}, \vec{w} \in \mathbb{C}^n$.

A is ***unitary*** if $A \cdot A^* = I_n$, or analogously: $\left\langle \vec{v} | \vec{w} \right\rangle = \left\langle T(\vec{v}) | T(\vec{w}) \right\rangle$ for all $\vec{v}, \vec{w} \in \mathbb{C}^n$.

A is ***normal*** if $A \cdot A^* = A^* \cdot A$. All matrix types shown above are normal matrices, as are diagonal complex matrices. However, some normal matrices do not fall into any of these four categories.

Hermitian and Skew-Hermitian matrices are ***closed*** under addition, multiplication by a pure real number, the transpose operation and the adjoint operation. However, ***unitary*** matrices have ***multiplicative*** properties.

Eigenvalues of special normal matrices: If λ is an eigenvalue of a normal matrix A, then: (1) λ is a ***pure real number*** if A is ***Hermitian***. (2) λ is a ***pure imaginary number*** if A is ***Skew-Hermitian***. (3) $\|\lambda\| = 1$ if A is ***unitary***.

We say that two $n \times n$ complex matrices A and B are ***unitarily equivalent*** if there exists an $n \times n$ ***unitary*** matrix U such that: $B = U^{-1}AU = U^*AU$.

A is ***unitarily diagonalizable*** if A is ***unitarily equivalent*** to a ***diagonal*** matrix D.

Let A be an $n \times n$ complex matrices. Then: A is ***unitarily diagonalizable*** ***if and only if*** there is an ***orthonormal basis*** for $\mathbb{C}^n$ consisting of ***eigenvectors*** of A.

Unitary equivalence is an ***equivalence relation:*** it is ***reflexive***, ***symmetric***, and ***transitive***.

If an $n \times n$ complex matrix A is ***normal***, then U^*AU is also ***normal*** for all ***unitary*** matrices U. In other words, the property of being normal is ***preserved*** under unitary equivalence.

If $\vec{v}$ and $\vec{w}$ are from ***distinct*** eigenspaces of a ***normal*** $n \times n$ matrix A, then $\langle \vec{v} | \vec{w} \rangle = 0$. Thus, distinct eigenspaces are ***orthogonal*** to each other.

Schur's Lemma: Let A be an $n \times n$ complex matrix. Then there exists a ***unitary*** matrix U and an ***upper triangular*** complex matrix B so that $A = UBU^*$, or equivalently, $B = U^*AU$. We call the factorization $A = UBU^*$ the ***Schur Decomposition*** of A.

The Spectral Theorem for Normal Matrices: Let A be an $n \times n$ complex matrix. Then: ***A is unitarily diagonalizable*** ***if and only if*** ***A is normal***, that is, $A \cdot A^* = A^* \cdot A$. More specific versions can be stated for the families of symmetric, Hermitian, Skew-Hermitian, and unitary matrices.

We ***unitarily diagonalize*** a normal matrix by finding an orthonormal basis for each eigenspace of A, using a combination of the Gauss-Jordan Algorithm, followed by the Gram-Schmidt Algorithm if the eigenspace is more than 1-dimensional. The unitary matrix U has for its columns the basis vectors for the eigenspaces, and the diagonal matrix D has the corresponding eigenvalues along the diagonal.

Let A and B be $n \times n$ matrices. We say that A and B are ***simultaneously diagonalizable*** if there exists an ***invertible*** matrix C such that $C^{-1}AC = D_1$, and $C^{-1}BC = D_2$, where D_1 and D_2 are both ***diagonal*** matrices.

Let $E_1, E_2, \ldots, E_k$ be square matrices, not necessarily of the same size. Then: the matrix $E = E_1 \oplus E_2 \oplus \cdots \oplus E_k$ is ***diagonalizable*** ***if and only if*** every block E_i is ***diagonalizable***.

The Simultaneous Diagonalizability Theorem: If A and B are both ***diagonalizable*** $n \times n$ matrices, then A and B are ***simultaneously diagonalizable*** ***if and only if*** $AB = BA$.

If A and B are both $n \times n$ matrices, A has n distinct eigenvalues and $AB = BA$, then A and B are ***simultaneously diagonalizable***, even if we do not know beforehand that B is diagonalizable.

If A and B are both ***symmetric*** $n \times n$ matrices, then they are ***simultaneously orthogonally diagonalizable*** ***if and only if*** $AB = BA$.

Chapter 9

The Big Picture:

The Fundamental Theorem of Linear Algebra and Applications

Linear Algebra is a powerful tool in science and engineering. We will see in this Chapter how systems of linear equations naturally occur in fields such as chemistry and physics, particularly in balancing chemical equations and studying the current flowing through resistors in simple circuits.

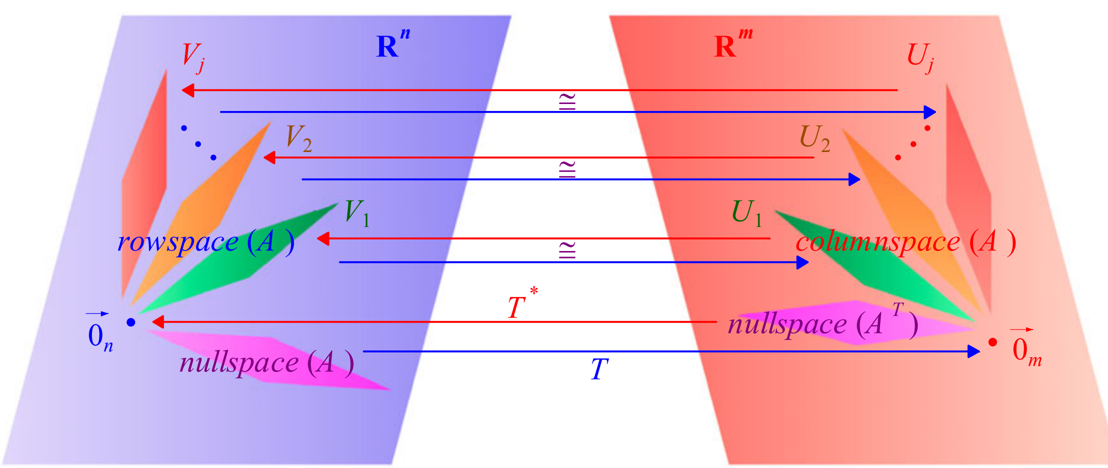
Eigentheory can be used to find closed formulas for the terms of a recursive sequence. Not surprisingly, some differential equations can be solved using a similar idea.

The main goal of this Chapter is to present ***The Fundamental Theorem of Linear Algebra*** and its fraternal twin ***The Singular Value Decomposition***. The latter has incredible applications in modern data processing, particularly by taming the high-volume demands of the Internet.

In order to fully appreciate these constructions, though, we need to discuss the theory of Quadratic Forms, particularly that of Positive Definite and Semi-Definite forms and matrices. We will see that matrices which are of the form $A \cdot A^{\top}$ are not just symmetric, but that all of its eigenvalues are non-negative real numbers.

As a bonus, we apply this theory to a more pedestrian application, which is to rotate a general quadratic equation in two variables so that we can recognize its graph as one of the conic sections. We will also see an alternative to solving the least-squares problem.

The Portrait of Linear Algebra that we have painted is by no means a complete one, nor is it the only possible way to capture the soul of its subject. The Fundamental Theorem of Linear Algebra, however, pulls together the major concepts that we have encountered on our journey: rowspaces, columnspaces, nullspaces, isomorphisms, eigenspaces and orthogonality. It is a fitting climax to our story.



The Fundamental Theorem of Linear Algebra

9.1 Balancing Chemical Equations

A ***chemical reaction*** is a process by which substances called ***reactants***, which are either ***atoms*** (also known as ***elements***) or ***molecules*** (also known as ***compounds***), break some of their bonds and recombine with each other in order to form new substances called ***products***.

We show below a short list of atoms which appear in common chemical reactions. Some atoms such as Helium (He) and Krypton (Kr) are not included on this list because they are ***inert***, that is, they do not naturally bind with other elements to form compounds. The periodic table of the elements, which is easily found on the Web, shows the complete list of atoms and their chemical symbols.

Atomic Number	Name	Symbol
1	Hydrogen	H
3	Lithium	Li
4	Beryllium	Be
5	Boron	B
6	Carbon	C
7	Nitrogen	N
8	Oxygen	O
9	Fluorine	F
11	Sodium	Na
12	Magnesium	Mg
13	Aluminium	Al
14	Silicon	Si
15	Phosphorus	P
16	Sulfur	S

Atomic Number	Name	Symbol
17	Chlorine	Cl
19	Potassium	K
20	Calcium	Ca
21	Scandium	Sc
22	Titanium	Ti
23	Vanadium	V
24	Chromium	Cr
25	Manganese	Mn
26	Iron	Fe
27	Cobalt	Co
28	Nickel	Ni
29	Copper	Cu
30	Zinc	Zn
31	Gallium	Ga

Some Common Elements, Their Atomic Numbers and Symbols

Molecules are written using the chemical symbols for each element with ***subscripts*** denoting the number of atoms of that element appearing in the molecule. Chemists have their own system of nomenclature, that is, the naming compounds and parts of compounds. We will not get into this nomenclature, as it is not relevant to our goal.

A ***chemical equation*** symbolically describes the reaction, where the ***reactants*** are written on the left side, separated by "+" signs, and the ***products*** are written on the right side, again separated by "+" signs, if there are more than one of them. A right arrow "→" separates the reactants from the products, and indicates that a chemical reaction has taken place. An equation is called ***balanced*** if exactly the same number of atoms of each element is found on each side of the equation, otherwise the equation is called ***unbalanced***.

The goal of this application is to show how to balance an unbalanced chemical equation by solving a homogeneous system of linear equations.

Example: One of the simplest chemical reactions that we can easily understand is the combination of ***hydrogen*** and ***oxygen*** gases to form water:

$$\mathrm{H_2 + O_2 \rightarrow H_2O}.$$

Notice that there are two oxygen atoms in an oxygen molecule, but only one oxygen atom is in a water molecule. This equation is therefore ***unbalanced***. Since we need two oxygen atoms on the right, we will put a coefficient of 2 on the left of our water molecule. Thus we get:

$$\mathrm{H_2 + O_2 \rightarrow 2H_2O}.$$

However, we now have the side-effect of having ***four*** hydrogen atoms on the product side. But we can easily fix that on the reactant side:

$$\mathrm{2H_2 + O_2 \rightarrow 2H_2O}.$$

This equation is now ***balanced,*** since the number of atoms of each element are the same on both sides of the equation.□

Obviously, reactions can be very complicated and involve several elements and subscripts. It is also possible that the same element appears in several reactants and/or several products. Thus we need a systematic way by which we can find the correct coefficients of each substance involved in the reaction. Linear Algebra can help us balance chemical reactions by solving a system of linear equations involving the unknown coefficients of each substance involved.

Example: Let us demonstrate how to use a system of equations on our example above. We will simply put an unknown coefficient in front of each substance:

$$x_1 \cdot \mathrm{H_2} + x_2 \cdot \mathrm{O_2} \rightarrow x_3 \cdot \mathrm{H_2O}.$$

Now, each element will give us an equation involving our coefficients. We simply have to require that the total number of that element on each side of the equation is the same, keeping the subscripts into account:

$$\begin{aligned} 2x_1 + 0x_2 &= 2x_3 && \text{for Hydrogen, and} \\ 0x_1 + 2x_2 &= x_3 && \text{for Oxygen.} \end{aligned}$$

Notice that we get the homogeneous, underdetermined system of linear equations:

$$2x_1 + 0x_2 - 2x_3 = 0, \text{ and}$$
$$0x_1 + 2x_2 - x_3 = 0.$$

There will thus be an infinite number of solutions to this system, but we want the solution with integer values for all the coefficients, which are as small as possible. In practice, we should find the rref of this system, but our system is simple enough that we can see that:

$$x_1 = x_3 = 2x_2,$$

gives us all solutions, with one free variable x_3. But to get a solution with integer values, we use $x_2 = 1$, and thus $x_1 = x_3 = 2$, giving us exactly the balanced equation we saw above.□

Let us see more complicated reactions:

Example: Let us consider the reaction where ***ethane*** and ***oxygen gas*** (what animals inhale) combine to create ***carbon dioxide gas*** (what animals exhale) and ***steam*** (water):

$$C_2H_6 + O_2 \to CO_2 + H_2O.$$

There are two substances in the reactant side and two also on the product side. There are three elements involved: Carbon (C), Hydrogen (H) and Oxygen (O). Let us put a coefficient beside each substance:

$$x_1 \cdot C_2H_6 + x_2 \cdot O_2 \to x_3 \cdot CO_2 + x_4 \cdot H_2O.$$

Let us count the number of atoms of each element, excluding the coefficients which do not contribute to that element:

$$\begin{aligned} 2x_1 &= x_3 && \text{for Carbon,} \\ 6x_1 &= 2x_4 && \text{for Hydrogen, and} \\ 2x_2 &= 2x_3 + x_4 && \text{for Oxygen.} \end{aligned}$$

Converting this to a homogeneous system of equations, we get:

$$\begin{array}{rrrrl} 2x_1 & & - \; x_3 & & = 0 \\ 6x_1 & & & - \; 2x_4 & = 0 \\ & 2x_2 & - \; 2x_3 & - \; x_4 & = 0 \end{array}$$

The coefficient matrix and its rref are:

$$\begin{bmatrix} 2 & 0 & -1 & 0 \\ 6 & 0 & 0 & -2 \\ 0 & 2 & -2 & -1 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 0 & -\frac{1}{3} \\ 0 & 1 & 0 & -\frac{7}{6} \\ 0 & 0 & 1 & -\frac{2}{3} \end{bmatrix}.$$

We thus get one free variable, x_4, and a basis for our nullspace is:

$$\langle 1/3, 7/6, 2/3, 1 \rangle \quad \text{or} \quad \langle 2, 7, 4, 6 \rangle,$$

by clearing denominators. Thus, we get our coefficients and our balanced equation:

$$2C_2H_6 + 7O_2 \rightarrow 4CO_2 + 6H_2O,$$

which can be verified to be numerically correct.□

Some groups of atoms within a reactant can stay together during the reaction. We show in the table below some common examples:

Name of Group	Symbol for the Group
hydroxide	OH
cyanide	CN
nitrate	NO_3
phosphate	PO_4
sulphate	SO_4

Some Common Molecular Groups

When groups appear in a reaction, we can sometimes replace these groups with a new symbol, like Y or Z, and balance the equation using these group symbols. However, we warn that this technique should ***not*** be used if a group is the only source of an element which appears in ***another*** product that does not contain this group.

Example: Consider the reaction:

$$Ba(NO_3)_2 + Al_2(SO_4)_3 \rightarrow BaSO_4 + Al(NO_3)_3.$$

We will keep barium (Ba) and aluminum (Al), but replace the nitrate and sulphate groups with Y and Z respectively. Let us set up our coefficients and new group symbols:

$$x_1 \cdot BaY_2 + x_2 \cdot Al_2Z_3 \rightarrow x_3 \cdot BaZ + x_4 \cdot AlY_3.$$

Our system of equations is:

$$\begin{aligned} x_1 &= x_3 && \text{for Barium,} \\ 2x_2 &= x_4 && \text{for Aluminum} \\ 2x_1 &= 3x_4 && \text{for Nitrate (Y)} \\ 3x_2 &= x_3 && \text{for Sulphate (Z).} \end{aligned}$$

Thus our homogenous system is:

$$\begin{array}{ccccccccc} x_1 & & & - & x_3 & & & = & 0 \\ & & 2x_2 & & & - & x_4 & = & 0 \\ 2x_1 & & & & & - & 3x_4 & = & 0 \\ & & 3x_2 & - & x_3 & & & = & 0 \end{array}.$$

The coefficient matrix and its rref are:

$$\begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 2 & 0 & -1 \\ 2 & 0 & 0 & -3 \\ 0 & 3 & -1 & 0 \end{bmatrix} \text{ with rref } \begin{bmatrix} 1 & 0 & 0 & -\frac{3}{2} \\ 0 & 1 & 0 & -\frac{1}{2} \\ 0 & 0 & 1 & -\frac{3}{2} \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

We thus get one free variable, x_4, and a basis for our nullspace is:

$$\langle 3/2, 1/2, 3/2, 1 \rangle \quad \text{or} \quad \langle 3, 1, 3, 2 \rangle,$$

by clearing denominators. Thus, we get our coefficients and our balanced equation:

$$3Ba(NO_3)_2 + Al_2(SO_4)_3 \rightarrow 3BaSO_4 + 2Al(NO_3)_3,$$

which can be verified to be correct.□

Notice that in this Example, we had four linear equations in four variables, and it was not clear that we would get a dependent system. However, since we knew for certain that this reaction does indeed occur, then there has to be a nontrivial solution, that is, we must have a free variable, and therefore a row of zeroes must appear in the rref.

9.1 Exercises

Balance the following chemical equations using a homogeneous system of equations:

1. $Fe + O_2 \rightarrow Fe_2O_3$

   (iron and oxygen combine to form iron oxide or rust)

2. $Cl_2O_5 + H_2O \rightarrow HClO_3$

   (dichloride pentoxide and water combine to form chloric acid)

3. $V_2O_5 + H_2 \rightarrow V_2O_3 + H_2O$

   (vanadium pentoxide and hydrogen gas combine to form vanadium trioxide and water)

4. $NH_3 + O_2 \rightarrow NO + H_2O$

   (ammonia and oxygen combine to form nitric oxide and water)

5. $CaO + P_4O_{10} \rightarrow Ca_3(PO_4)_2$

   (calcium oxide and tetraphosphorus decoxide combine to form calcium phosphate)

6. $C_3H_8 + O_2 \rightarrow CO_2 + H_2O$

   (propane and oxygen combine to form carbon dioxide and water)

7. $I_2 + Na_2S_2O_3 \rightarrow NaI + Na_2S_4O_6$

   (iodide and sodium thiosulphate combine to form sodium iodide and sodium tetrathionate)

8. $(NH_4)_2CO_3 \rightarrow NH_3 + CO_2 + H_2O$

(ammonium carbonate decomposes into ammonia, carbon dioxide and water)

9. $C_4H_{10} + O_2 \rightarrow CO_2 + H_2O$

(butane and oxygen combine to produce carbon dioxide and water)

10. $CH_3OH + O_2 \rightarrow CO_2 + H_2O$

(methanol and oxygen combine to produce carbon dioxide and water)

Some parts of the following reactions can be replaced with groups:

11. $Pb + PbO_2 + H_2SO_4 \rightarrow PbSO_4 + H_2O$

(free lead, lead dioxide and sulfuric acid combine to form lead sulfate and water)

12. $Ca(OH)_2 + H_3PO_4 \rightarrow Ca_3(PO_4)_2 + H_2O$

(calcium hydroxide and phosphoric acid combine to form calcium phosphate and water; there is only one group in this reaction)

13. $Fe_2O_3 + H_2SO_4 \rightarrow Fe_2(SO_4)_3 + H_2O$

(iron oxide and sulfuric acid combine to form iron sulfite and water)

14. $Zn(OH)_2 + H_3PO_4 \rightarrow Zn_3(PO_4)_2 + H_2O$

(zinc hydroxide and phosphoric acid combine to form zinc phosphate and water)

There are two groups in each of the following reactions:

15. $H_3PO_4 + NaCN \rightarrow HCN + Na_3PO_4$

(phosphoric acid and sodium cyanide combine to form hydrogen cyanide and sodium phosphate)

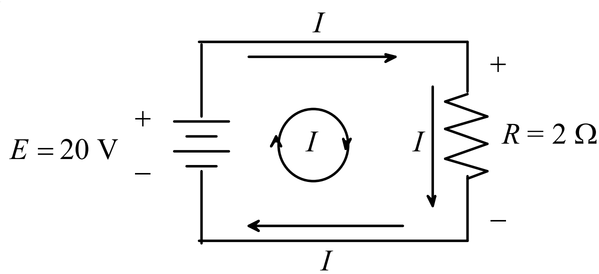
16. $Fe_2(SO_4)_3 + KSCN \rightarrow K_3Fe(SCN)_6 + K_2SO_4$

(iron sulphate combines with potassium thiocyanate to form alum and potassium sulfate)

9.2 Basic Circuit Analysis

In this Section, we will see how systems of linear equations naturally appear in the study of simple electrical circuits. We will look at circuits which only involve a constant ***voltage source*** (such as a battery), ***resistors***, and ***wires*** of negligible resistance. We will also limit our analysis to ***planar circuits***, that is, circuits where wires do not have to cross each other except at a ***connecting node***, and so the circuit can be assembled on a plane.

The simplest kind of electric circuit involves a single battery, a single resistor, and connecting wires. A battery is often denoted by the symbol ***E***, which stands for ***electromotive force*** or ***emf***. Its ***voltage***, which is assumed to be constant, is measured in ***volts***, abbreviated as **V**. Resistors are denoted by the symbol ***R***, and we use the unit of ***ohms*** (with symbol **Ω**) to measure resistance. For example, we could have:



A Simple Electrical Circuit

The battery, wires and resistor form a single ***loop*** or ***mesh***. When the circuit is complete, electrons flow from the negative end of the battery to the positive end, and go around in the circuit, creating a ***current***, denoted by the symbol ***I***. Due to the orientation of the battery, the current above goes ***clockwise***. The current goes through the resistor, which releases the energy as heat. Current is measured in the unit of ***amperes***, with symbol **A**.

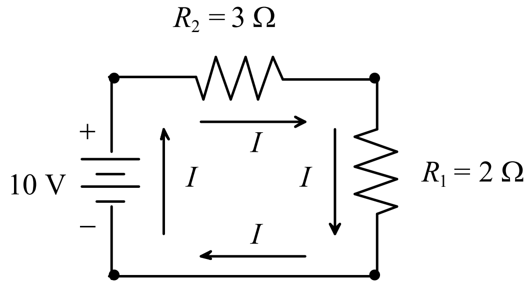
The quantities E, R and I are related by ***Ohm's Law***:

$$E = IR.$$

Thus, in our circuit above:

$$I = \frac{E}{R} = \frac{10\text{ V}}{2\ \Omega} = 5\text{ A}.$$

The analysis of course becomes more complicated if we add more components to our circuit. Components are joined together at ***nodes***, often by soldering, or through the built-in connections in a circuit board. For example, we can attach a 3 Ω resistor to our circuit in essentially two different ways. The first way, shown below, is called a ***series*** configuration:



Two Resistors in Series

Notice that our new circuit still consists of a single loop, but the current I now goes through two resistors R_1 and R_2. We should therefore expect a different value for I. In order to find it, we will need a couple of ideas. First, Ohm's Law can be applied to each resistor. We can think of the current as a river which goes through two waterfalls. As we traverse the circuit in a clockwise manner, we experience a ***potential drop*** through each resistor, which we denote by V_1 and V_2 respectively. According to Ohm's Law:

$$V_1 = IR_1 = 2I \quad \text{and} \quad V_2 = IR_2 = 3I.$$

However, as we go through the battery, we experience a ***potential rise***, in the same way that water can be lifted up from a well using a bucket. To tie these three potential changes together, we need the following important law:

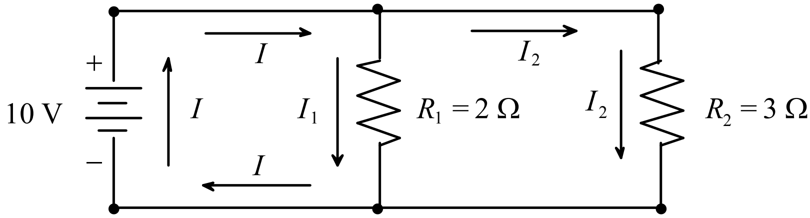
Kirchoff's Voltage Law or KVL: If we go around a mesh, starting and ending at the same node, the sum of the potential changes must be ***zero***.

We agree that a potential rise will have a positive sign, while a potential drop will have a negative sign. Thus, according to KVL, our new current I satisfies:

$$10 - 3I - 2I = 0,$$

so $I = 10/5 = 2$ A. Notice that this new current is ***smaller*** than our previous current of 5 A.

The other way that we can attach another resistor to our first circuit is called a ***parallel*** configuration:



Two Resistors in Parallel

Notice that our new circuit now consists of ***two*** loops, and so we expect a ***different*** current to flow through R_1 and R_2. Finding them individually is a direct consequence of Ohm's Law. Since the top and bottom nodes of each resistor are essentially the same node as the top and bottom of the battery (assuming wires have negligible resistance), Ohm's Law says that:

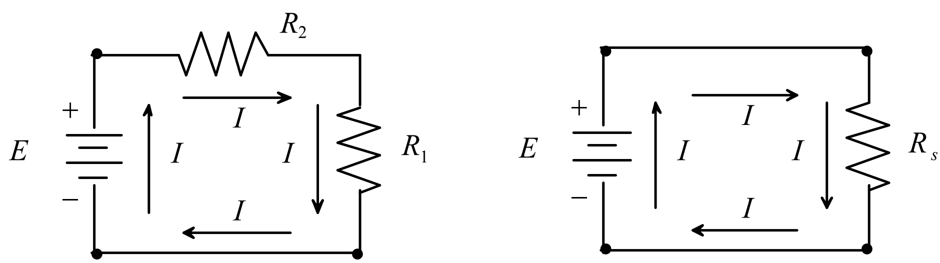
$$10 = 2I_1, \text{ and}$$
$$10 = 3I_2.$$

Thus, $I_1 = 10/2 = 5$ A, the same current in our original circuit, but $I_2 = 10/3 = 3\frac{1}{3}$ A. If we want to know the current I flowing through the battery, though, we need the following twin to KVL:

> ***Kirchoff's Current Law or KCL:*** The sum of the currents ***entering*** any node must equal the sum of the currents ***leaving*** that node.

Thus, $I = I_1 + I_2 = 5 + 3\frac{1}{3} = 8\frac{1}{3}$ A. Notice that this current is now ***greater*** than the old current of 5 A going through the battery.

Having two resistors and a single battery in one circuit is not a very complicated system. In fact, we can replace the two resistors with a single ***equivalent*** resistor. Let us see how to do it for two resistors in series:



Two Resistors in Series and An Equivalent Resistor R_s

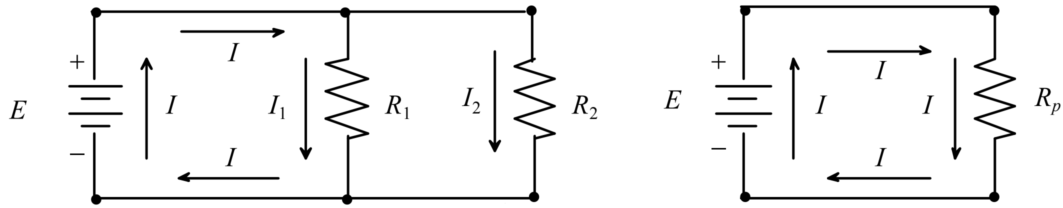
Applying KVL to the first circuit, we get $E - IR_2 - IR_1 = 0$, and so:

$$I = \frac{E}{R_1 + R_2}.$$

The two circuits are equivalent if the ***same current*** flows through the battery. But from the equivalent circuit, we have $I = E/R_s$. Thus:

$$R_s = R_1 + R_2.$$

Now, for two resistors in parallel:



Two Resistors in Parallel and An Equivalent Resistor R_p

We saw above that $E = I_1R_1$ and $E = I_2R_2$, and thus:

$$I = I_1 + I_2 = \frac{E}{R_1} + \frac{E}{R_2} = E\left(\frac{1}{R_1} + \frac{1}{R_2}\right)$$

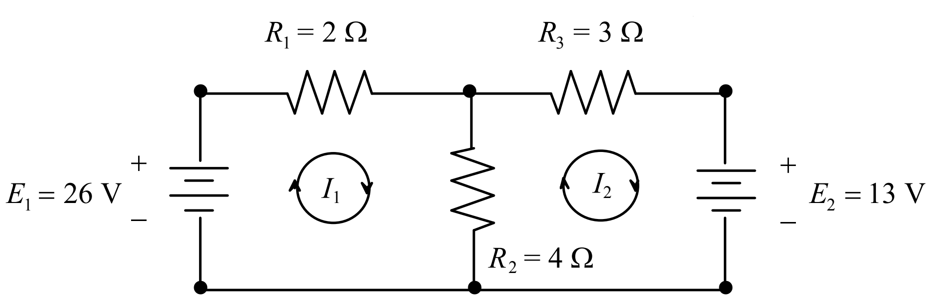
As before, we have $I = E/R_p$, and so to get equal currents, we need:

$$\frac{1}{R_p} = \frac{1}{R_1} + \frac{1}{R_2}.$$

Mesh Analysis

To find the currents going through the elements in more complicated planar circuits involving only batteries and resistors, we can use what is called the ***Superposition Principle***. We will assign a current I_k to each mesh, which by convention will be ***clockwise***. We then traverse the mesh clockwise and apply KVL to write a linear equation for this mesh. However, if a resistor is straddling between two meshes, we ***subtract*** the current of the neighboring mesh from I_k, because it will be going in the opposite direction relative to I_k. Keep in mind also that when we traverse a battery, the emf is taken to be positive if we go from − to + as we go clockwise, and negative if we go from + to −. We will usually obtain an ***invertible*** square system which we can then solve using the Gauss-Jordan Algorithm. This makes sense because once we complete the circuit, there should be exactly one value for the current flowing through each component. Once we have the value of each I_k, we return to the circuit and find the current through components straddling two meshes. We illustrate with the following:

Example: Suppose we have the following circuit:



For the 1st mesh, the battery goes from – to + as we go clockwise, R_1 only experiences I_1, but R_2 is straddling the 2nd mesh. Thus, our 1st equation is:

$$26 - 2I_1 - 4(I_1 - I_2) = 0.$$

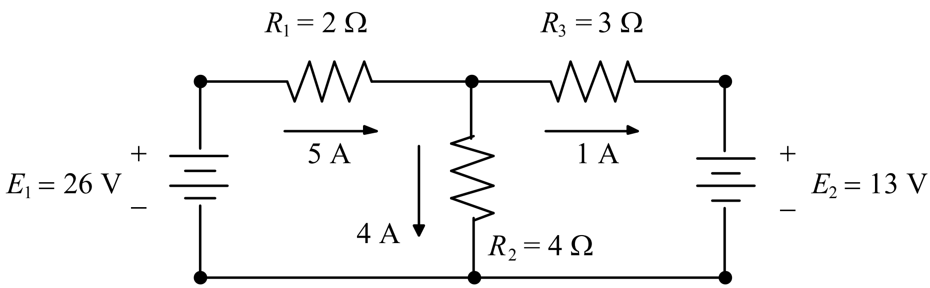
For the 2nd mesh, the battery goes from + to – as we go clockwise, R_2 is straddling the 1st mesh, and R_3 only experiences I_2. Thus, our 2nd equation is:

$$-13 - 4(I_2 - I_1) - 3I_2 = 0.$$

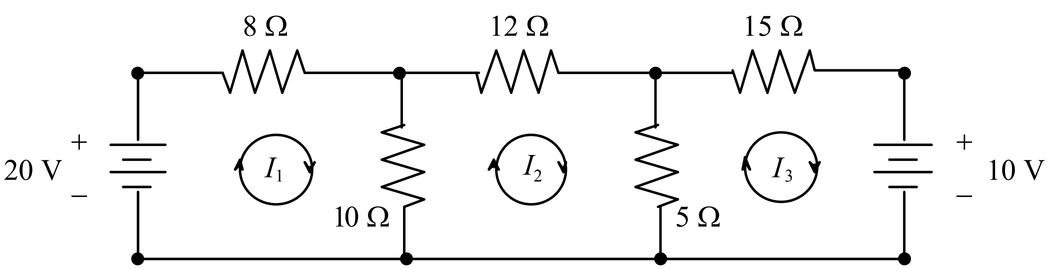
We obtain the system:

$$6I_1 - 4I_2 = 26, \text{ and}$$
$$-4I_1 + 7I_2 = -13.$$

This is only a 2 × 2 system, and we easily find that $I_1 = 5$ A. and $I_2 = 1$ A. Since $I_1 > I_2$, a current of $5 - 1 = 4$ A flows ***downwards*** through R_2. We show the currents through the three resistors below:



Example: Let us analyze the following circuit, which has three meshes:



The three resistors at the top of the circuit only experience their respective mesh currents, but the 10 Ω resistor is straddling between I_1 and I_2, while the 5 Ω resistor is straddling between I_2 and I_3. Thus, we get the following three equations using KVL:

$$20 - 8I_1 - 10(I_1 - I_2) = 0,$$
$$-10(I_2 - I_1) - 12I_2 - 5(I_2 - I_3) = 0, \text{ and}$$
$$-10 - 5(I_3 - I_2) - 15I_3 = 0.$$

Notice that the coefficient of each resistor is ***negative***, and that the mesh current always goes ***first*** in every resistor that is straddling two meshes. Now we convert our equations to standard form:

$$
\begin{aligned}
18I_1 \quad - \quad 10I_2 \quad & \quad = \quad 20 \\
-10I_1 \quad + \quad 27I_2 \quad - \quad 5I_3 & \quad = \quad 0 \\
-5I_2 \quad + \quad 20I_3 & \quad = \quad -10
\end{aligned}
$$

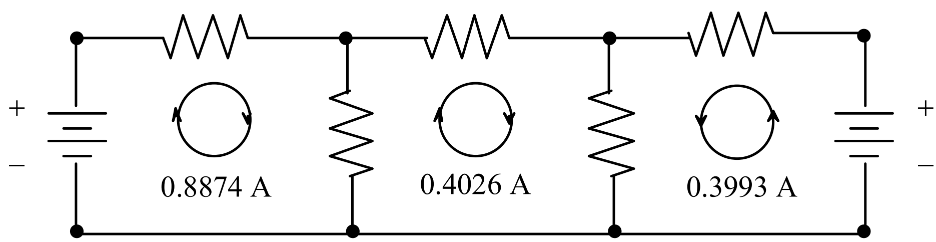
We wrote the three equations above so that the coefficient of the respective mesh current is ***positive***: $18I_1$, $27I_2$ and $15I_3$. Notice that these coefficients are precisely the ***sum*** of the resistors involved in each mesh:

$$
\begin{aligned}
18 &= 8 + 10 \\
27 &= 10 + 12 + 5 \\
20 &= 5 + 15
\end{aligned}
$$

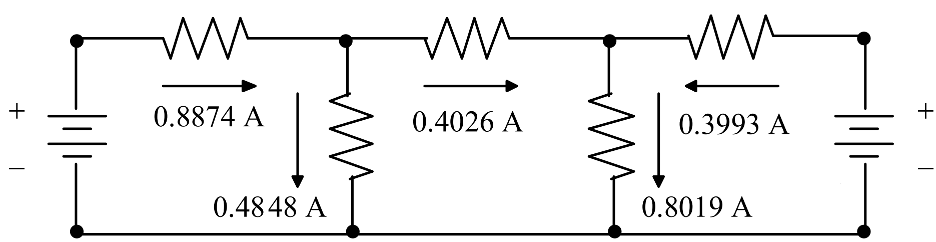
The other currents have a negative coefficient, corresponding to the resistors that straddle other meshes. This is a good way to check that we set up our system correctly. Now, the (approximate) solutions to our system are:

$$
\begin{aligned}
I_1 &\approx 0.8874 \text{ A}, \\
I_2 &\approx 0.4026 \text{ A}, \text{ and} \\
I_3 &\approx -0.3993 \text{ A}.
\end{aligned}
$$

Notice that I_3 has a ***negative*** sign. This means that the assumed direction of counterclockwise is ***wrong***. Thus, the correct mesh currents are:

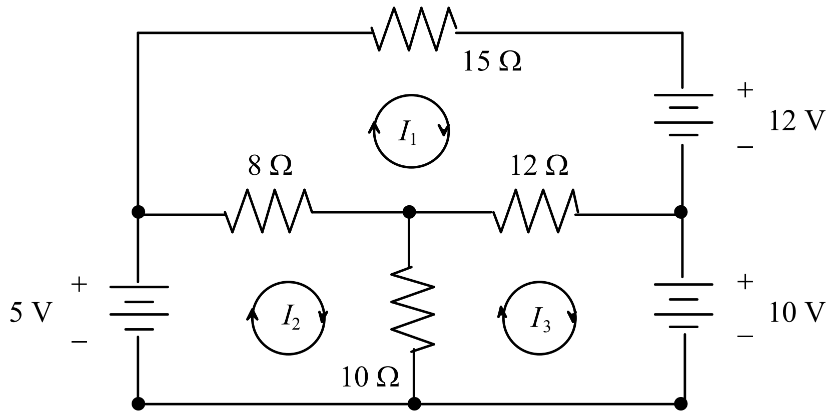


From here, we find the currents through the straddling resistors, along with their correct orientations:



Notice that I_1 and I_2 are opposite currents with respect to the 10 Ω resistor, with I_1 stronger, and so the current through this resistor is $0.8874 - 0.4026 = 0.4848$ A. However, I_2 and I_3 are going in the same direction through the 5 Ω resistor, and so the current through this resistor is $0.4026 + 0.3993 = 0.8019$ A.$\square$

Example: For our final example, let us analyze the circuit below:



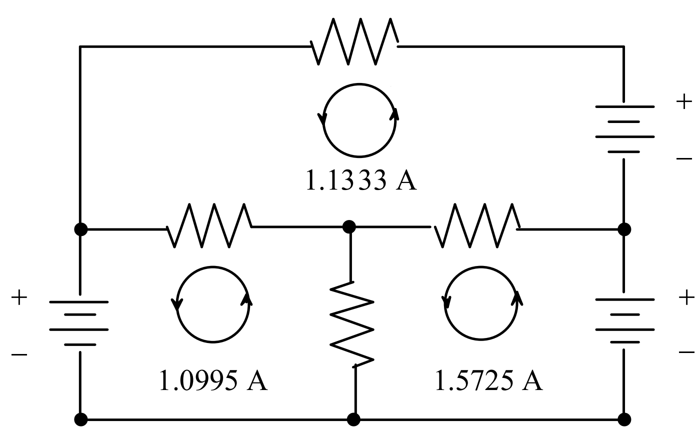
Although this circuit also has three meshes, it is different from the previous example because each mesh is adjacent to the other two meshes. Thus, the 8 Ω resistor is straddling I_1 and I_2, the 12 Ω resistor is straddling I_1 and I_3, and the 10 Ω resistor is straddling I_2 and I_3. Each of our three equations will therefore involve all three mesh currents. We get the system:

$$\begin{array}{rcrcrcr} 35I_1 & - & 8I_2 & - & 12I_3 & = & -12 \\ -8I_1 & + & 18I_2 & - & 10I_3 & = & 5 \\ -12I_1 & - & 10I_2 & + & 22I_3 & = & -10 \end{array}$$

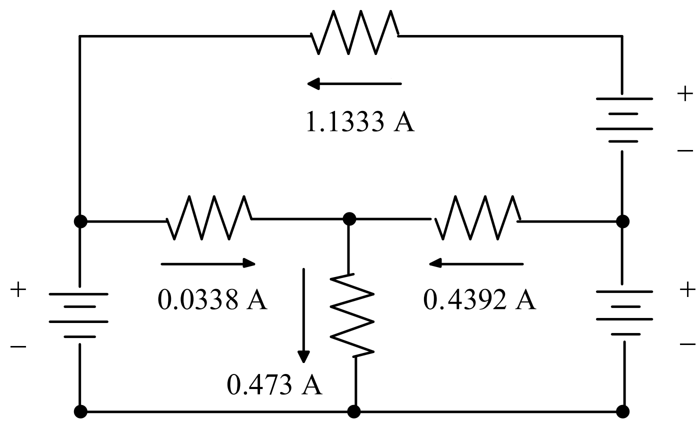
Observe further that the coefficient matrix is ***symmetric***. The diagonal entries are the sums of the resistors in each mesh, appearing with a ***positive*** coefficient as previously mentioned, and the off-diagonal entries are the resistors that straddle an adjacent mesh, appearing with a ***negative*** coefficient. This makes sense because if each mesh were treated separately, ignoring the others, the resistors are connected ***in series***, and so the effective resistance is simply the sum. The emf's are positive when we go from – to + as we go clockwise, and negative otherwise. Now, solving the system, we get:

$$I_1 \approx -1.1333 \text{ A}, \; I_2 \approx -1.0995 \text{ A}, \text{ and } I_3 \approx -1.5725 \text{ A}.$$

Thus, all of the mesh currents are actually going ***counterclockwise***:



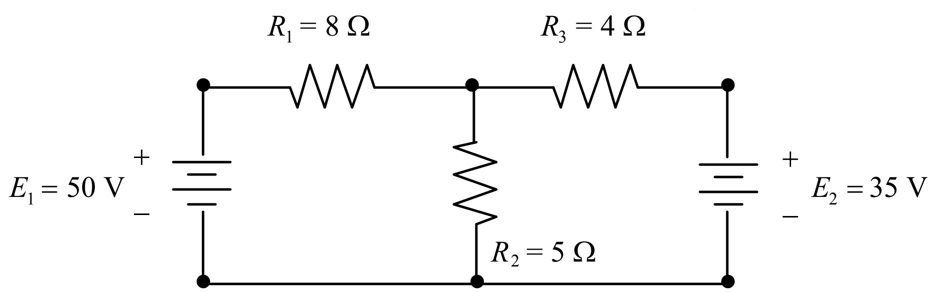
Finally, we solve for the current through each resistor by subtracting the smaller mesh current from the larger. The direction of the net current is the same as that of the stronger current. We observe that none of the mesh currents reinforce each other, as opposed to the previous example, so the final currents are:



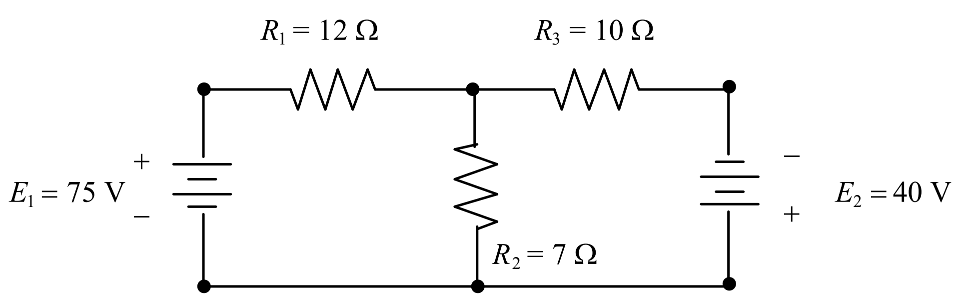
9.2 Exercises

Find the current in each resistor of the following circuits:

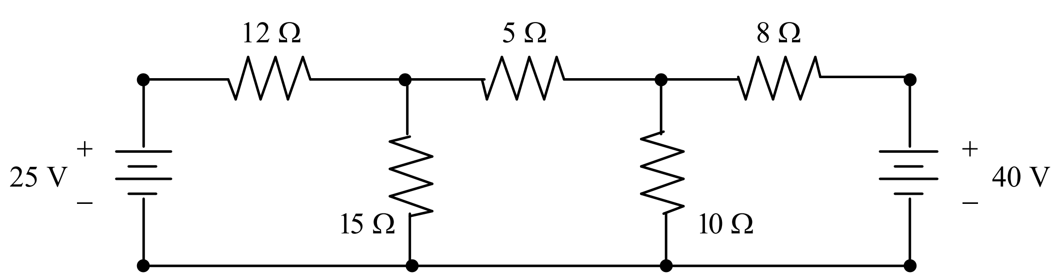
1.



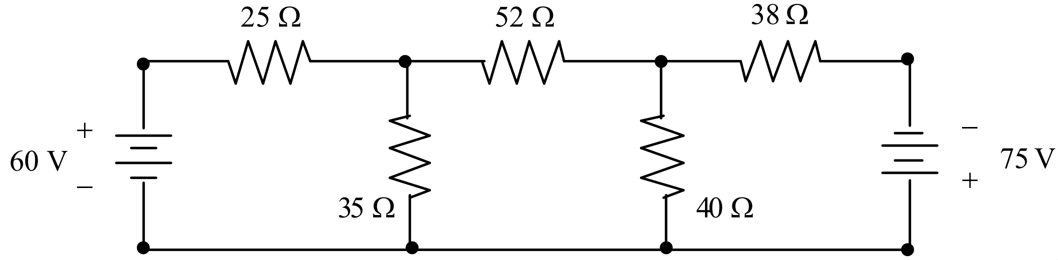
2.



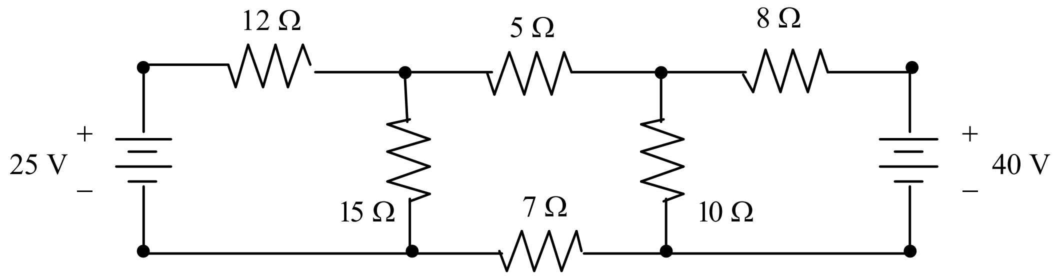
3.



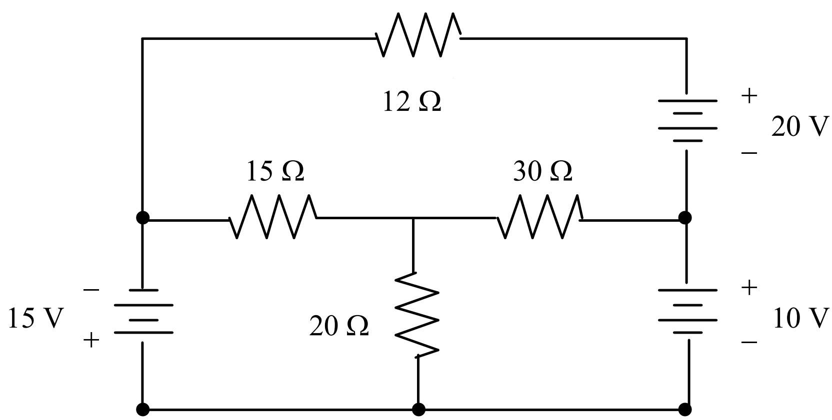
4.


25 Ω
52 Ω
38 Ω
60 V
+
–
35 Ω
40 Ω
–
+
75 V

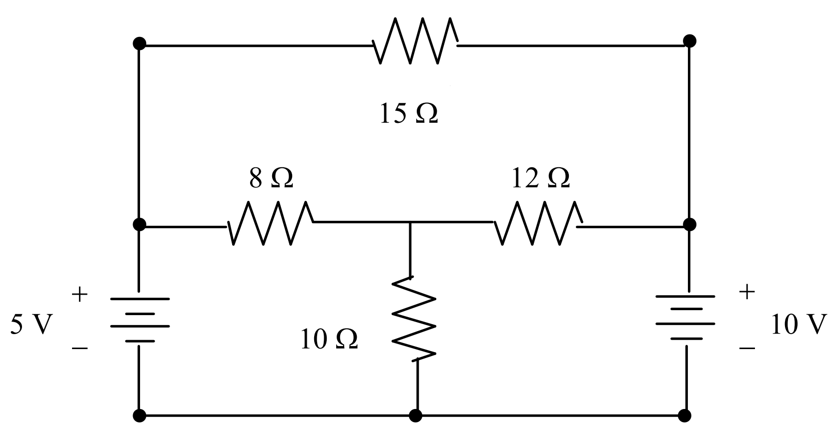
5.


12 Ω
5 Ω
8 Ω
25 V
+
–
15 Ω
7 Ω
10 Ω
+
–
40 V

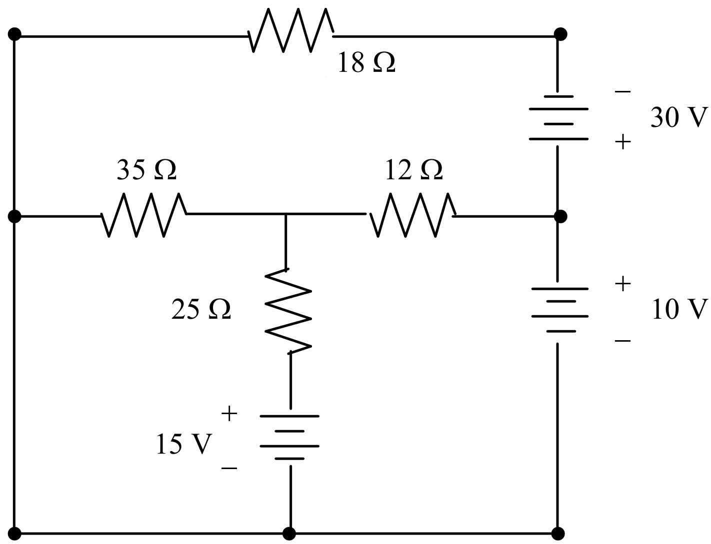
6.


12 Ω
+
20 V
–
15 Ω
30 Ω
15 V
–
+
20 Ω
+
10 V
–

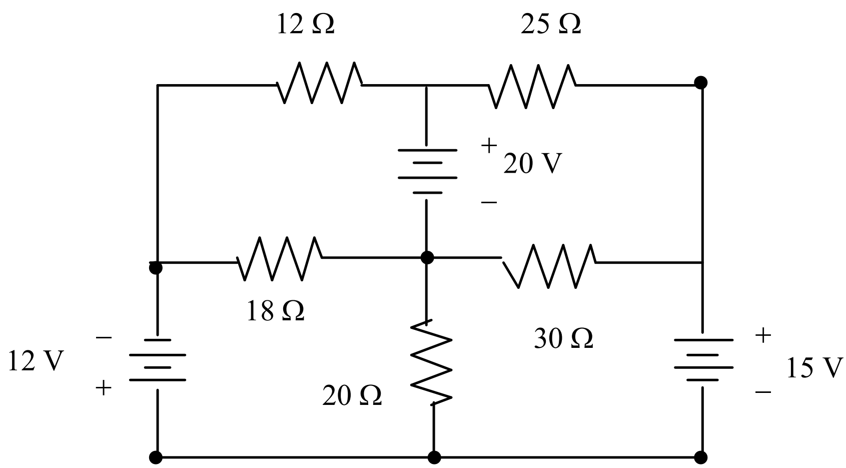
7.


15 Ω
8 Ω
12 Ω
5 V
+
–
10 Ω
+
10 V
–

8.



9.



9.3 Recurrence Relations

When we study sequences in Precalculus or Calculus, we usually mention ***recursive sequences***, that is, sequences where we specify the first or first few terms of the sequence, and a formula or algorithm to generate the next term of the sequence based on the previous term or terms. The most famous of these recursive sequences is the Fibonacci sequence:

Definition: The ***Fibonacci sequence*** $\{F_n\}_{n=0}^{\infty}$ is the recursive sequence given by:

$$\begin{aligned} F_0 &= 0, \\ F_1 &= 1, \text{ and} \\ F_{n+1} &= F_n + F_{n-1} \text{ for all } n \geq 1. \end{aligned}$$

The final equation says that in order to get the next term of the sequence $\left(\text{which is } F_{n+1}\right)$, we need to add together the two previous terms $\left(F_n \text{ and } F_{n-1}\right)$. Thus, the first few terms of the Fibonacci sequence are:

$$0,\ 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, \ldots .$$

This sequence is so important, there is a journal dedicated to these numbers and their applications, called ***The Fibonacci Quarterly***.

Our goal is to find a ***closed formula*** for F_n, that is, a formula that explicitly gives F_n as a function of n, without having to go through all the terms before F_n. In other words, we want a formula:

$$F_n = F(n) = \text{ some algebraic function of } n.$$

Such a formula will tell us, for example, that $F_{20} = F(20) = 6765$.

The only clue we have for such a formula is the ***recurrence relation:***

$$F_{n+1} = F_n + F_{n-1}.$$

Thus:

$$\begin{aligned} F_2 &= F_1 + F_0, \text{ and} \\ F_3 &= F_2 + F_1. \end{aligned}$$

However, notice that the equations above simply involve ***linear combinations*** of F_0, F_1 and F_2. Thus, let us write our equations in matrix form as:

$$\begin{bmatrix} F_1 \\ F_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_1 + F_0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} F_0 \\ F_1 \end{bmatrix}.$$

Notice that F_1 appears ***both*** in the column matrix on the left side and the column matrix on the right side. Similarly, we can write:

$$\begin{bmatrix} F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \end{bmatrix}.$$

If we substitute the first matrix equation into the second, we get:

$$\begin{bmatrix} F_2 \\ F_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} F_0 \\ F_1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^2 \bullet \begin{bmatrix} F_0 \\ F_1 \end{bmatrix}.$$

Proceeding by Induction, we can show that:

$$\begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^n \bullet \begin{bmatrix} F_0 \\ F_1 \end{bmatrix}.$$

But we know how to find large powers of a matrix if we can ***diagonalize*** it. Thus, let:

$$A = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}.$$

We shall now diagonalize A. First, its characteristic polynomial is:

$$p(\lambda) = \begin{vmatrix} \lambda & -1 \\ -1 & \lambda - 1 \end{vmatrix} = \lambda(\lambda - 1) - 1 = \lambda^2 - \lambda - 1.$$

The quadratic formula yields our eigenvalues:

$$\lambda = \frac{1 \pm \sqrt{1 - 4(-1)}}{2} = \frac{1 \pm \sqrt{5}}{2}$$

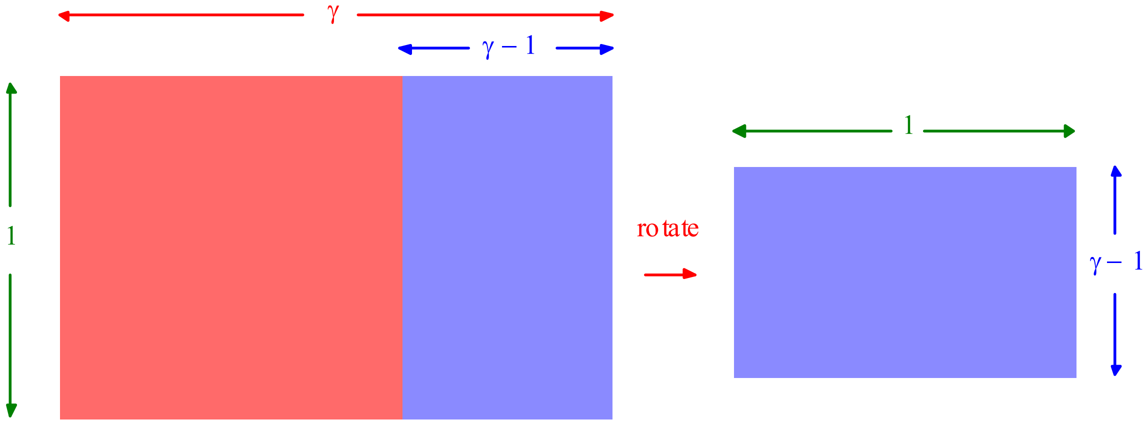
The number:

$$\gamma = \frac{1 + \sqrt{5}}{2} \approx 1.618$$

is a famous number, known as the ***Golden Ratio***. We often write the other root as:

$$\bar{\gamma} = \frac{1 - \sqrt{5}}{2} \approx -0.618.$$

The Golden ratio is associated to a special geometric figure: a ***Golden Rectangle*** is a rectangle whose sides are in the proportion of $1 : \gamma$. If we were to cut off a square using the shorter side of such a rectangle, we would be left with a smaller rectangle that is also a Golden Rectangle:



A Large Golden Rectangle, and a Smaller Golden Rectangle
Obtained by Cutting Off a Square from the Larger Rectangle

We have rotated the smaller Golden Rectangle so that we can see more convincingly that its proportions are the same as the larger rectangle. We will prove this algebraically below.

Before we proceed further, let us look at some interesting properties of γ and $\overline{\gamma}$. Since they are both roots of $p(\lambda) = \lambda^2 - \lambda - 1$, they each satisfy this equation, thus:

$$\gamma^2 = \gamma + 1, \quad \text{and} \quad (\overline{\gamma})^2 = \overline{\gamma} + 1.$$

Furthermore, we have:

$$\begin{aligned} \gamma \cdot \overline{\gamma} &= \frac{1+\sqrt{5}}{2} \cdot \frac{1-\sqrt{5}}{2} = \frac{1-5}{4} = -1, \\ \gamma + \overline{\gamma} &= \frac{1+\sqrt{5}}{2} + \frac{1-\sqrt{5}}{2} = 1, \quad \text{and} \\ \gamma - \overline{\gamma} &= \frac{1+\sqrt{5}-1+\sqrt{5}}{2} = \sqrt{5}. \end{aligned}$$

If we use the first two equations to solve for $\overline{\gamma}$, we get:

$$\overline{\gamma} = -\frac{1}{\gamma} = 1 - \gamma$$

This last equation shows that the smaller rectangle obtained by cutting off a square from one side of the larger Golden Rectangle indeed has the same proportions:

$$\frac{\gamma - 1}{1} = \frac{1}{\gamma} \quad \text{or} \quad (\gamma - 1) : 1 = 1 : \gamma.$$

We will be using the equations above to simplify our computations. Next, we find the eigenvectors: For $\lambda = \gamma$, we solve the system:

$$\begin{bmatrix} \gamma & -1 \\ -1 & \gamma - 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

This should be a dependent system. Indeed, if we divide the first row by γ and add it to the second row, the entries will be:

$$-1 + \frac{\gamma}{\gamma} = 0, \quad \text{and} \quad \gamma - 1 - \frac{1}{\gamma} = \gamma - 1 + \overline{\gamma} = 1 - 1 = 0,$$

thus producing zeroes on the entire second row. Hence, we only need to solve:

$$\gamma x - y = 0.$$

Although y should be our free variable, we will instead make x our free variable in order to get the simpler general solution:

$$y = \gamma x.$$

Thus $Eig(A, \gamma)$ has basis $\{\langle 1, \gamma \rangle\}$.

Similarly, to find the eigenvectors for $\overline{\gamma}$, we must solve the system:

$$\begin{bmatrix} \overline{\gamma} & -1 \\ -1 & \overline{\gamma} - 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

We will leave the details as an easy exercise, but the outcome is that $\{\langle 1, \overline{\gamma} \rangle\}$ is a basis for $Eig(A, \overline{\gamma})$.

Thus, $D = Diag(\gamma, \overline{\gamma})$, and our diagonalizing matrix C is:

$$C = \begin{bmatrix} 1 & 1 \\ \gamma & \overline{\gamma} \end{bmatrix}, \quad \text{with inverse}$$

$$C^{-1} = \frac{1}{\overline{\gamma} - \gamma} \begin{bmatrix} \overline{\gamma} & -1 \\ -\gamma & 1 \end{bmatrix} = -\frac{1}{\sqrt{5}} \begin{bmatrix} \overline{\gamma} & -1 \\ -\gamma & 1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} -\overline{\gamma} & 1 \\ \gamma & -1 \end{bmatrix}.$$

Now, we are ready to diagonalize:

$$A = CDC^{-1}$$

$$\begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ \gamma & \overline{\gamma} \end{bmatrix} \cdot \begin{bmatrix} \gamma & 0 \\ 0 & \overline{\gamma} \end{bmatrix} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} -\overline{\gamma} & 1 \\ \gamma & -1 \end{bmatrix}.$$

From this, we get:

$$\begin{aligned} A^n &= CD^nC^{-1} \\ &= \begin{bmatrix} 1 & 1 \\ \gamma & \overline{\gamma} \end{bmatrix} \cdot \begin{bmatrix} \gamma & 0 \\ 0 & \overline{\gamma} \end{bmatrix}^n \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} -\overline{\gamma} & 1 \\ \gamma & -1 \end{bmatrix} \\ &= \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 1 \\ \gamma & \overline{\gamma} \end{bmatrix} \cdot \begin{bmatrix} \gamma^n & 0 \\ 0 & \overline{\gamma}^n \end{bmatrix} \cdot \begin{bmatrix} -\overline{\gamma} & 1 \\ \gamma & -1 \end{bmatrix} \end{aligned}$$

$$= \frac{1}{\sqrt{5}} \begin{bmatrix} \gamma^n & \bar{\gamma}^n \\ \gamma^{n+1} & \bar{\gamma}^{n+1} \end{bmatrix} \cdot \begin{bmatrix} -\bar{\gamma} & 1 \\ \gamma & -1 \end{bmatrix}$$

$$= \frac{1}{\sqrt{5}} \begin{bmatrix} -\gamma^n \bar{\gamma} + \gamma \bar{\gamma}^n & \gamma^n - \bar{\gamma}^n \\ -\gamma^{n+1} \bar{\gamma} + \gamma \bar{\gamma}^{n+1} & \gamma^{n+1} - \bar{\gamma}^{n+1} \end{bmatrix}$$

$$= \frac{1}{\sqrt{5}} \begin{bmatrix} \gamma^{n-1} - \bar{\gamma}^{n-1} & \gamma^n - \bar{\gamma}^n \\ \gamma^n - \bar{\gamma}^n & \gamma^{n+1} - \bar{\gamma}^{n+1} \end{bmatrix}.$$

In the last step, we made use of the property that $\gamma \cdot \bar{\gamma} = -1$. Now, let us use A^n to find our closed formula:

$$\begin{bmatrix} F_n \\ F_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix}^n \cdot \begin{bmatrix} F_0 \\ F_1 \end{bmatrix}$$

$$= \frac{1}{\sqrt{5}} \begin{bmatrix} \gamma^{n-1} - \bar{\gamma}^{n-1} & \gamma^n - \bar{\gamma}^n \\ \gamma^n - \bar{\gamma}^n & \gamma^{n+1} - \bar{\gamma}^{n+1} \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} \gamma^n - \bar{\gamma}^n \\ \gamma^{n+1} - \bar{\gamma}^{n+1} \end{bmatrix}.$$

Finally, we get the formula:

$$F_n = \frac{1}{\sqrt{5}} (\gamma^n - \bar{\gamma}^n).$$

This formula is elegant in and of itself, but remember that $\bar{\gamma} \approx -0.618$, thus if n is a large number $\bar{\gamma}^{n+1} \approx 0$. For instance, $\bar{\gamma}^{20} \approx 6.6034 \times 10^{-5}$ so it is almost negligible. Thus, if n is large:

$$F_n \approx \frac{\gamma^n}{\sqrt{5}},$$

and therefore F_n is practically an ***exponential*** function with base γ. For instance:

$$F_{20} \approx \frac{(1.61803)^{20}}{\sqrt{5}} \approx 6764.7,$$

which agrees with the actual value of 6765.

Linear Homogeneous Recurrence Relations

The Fibonacci sequence is a particular example of a recursive sequence $\{a_n\}_{n=0}^{\infty}$ that can be defined using a ***linear homogeneous recurrence relation with constant coefficients***. First, we specify the initial d terms of the sequence, namely:

$$a_0, a_1, a_2, \ldots, a_{d-1},$$

which are also called the ***seeds*** of the sequence (because the sequence grows from them). We call d the ***order*** of the recurrence relation. The next term of the sequence is computed using a ***linear combination with constant coefficients*** of the previous d terms.

Thus, we can write the general equation for a_n in the form:

$$a_n = c_1 a_{n-1} + c_2 a_{n-2} + \cdots + c_d a_{n-d}.$$

for all $n \geq d$, where $c_1, c_2, \ldots, c_d$ are constant coefficients. Notice that in the equation above, we write the terms in ***descending*** order. This equation is ***homogeneous*** because every term is a term in the sequence, with a coefficient beside it. We will show by example how to solve such relations.

Example: Let us define a sequence recursively by:

$$a_0 = 5, \;\; a_1 = 7, \text{ and}$$

$$a_n = 2a_{n-1} + 3a_{n-2}, \text{ for all } n > 1.$$

This is a linear homogeneous recurrence relation of order 2. The first few terms of this sequence are: 5, 7, 29, 79, 245, 727, …

We will simply mimic the method that we used to derive a closed-formula for the Fibonacci numbers in order to find an analogous closed-formula for a_n. According to the relation, $a_2 = 2a_1 + 3a_0$. Thus we can write:

$$\begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ 2a_1 + 3a_0 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}.$$

By induction, as before:

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 3 & 2 \end{bmatrix}^n \begin{bmatrix} a_0 \\ a_1 \end{bmatrix}.$$

We must diagonalize the matrix above, whose characteristic polynomial is:

$$p(\lambda) = \begin{vmatrix} \lambda & -1 \\ -3 & \lambda - 2 \end{vmatrix} = \lambda(\lambda - 2) - 3 = \lambda^2 - 2\lambda - 3 = (\lambda + 1)(\lambda - 3),$$

so our eigenvalues are fortunately integers, and more importantly, ***distinct***. We find that:

$$Eig(A, -1) = Span(\{\langle 1, -1 \rangle\}), \text{ and}$$

$$Eig(A, 3) = Span(\{\langle 1, 3 \rangle\}).$$

Thus, $D = Diag(-1, 3)$, and our diagonalizing matrix is:

$$C = \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix}, \text{ with inverse: } C^{-1} = \frac{1}{4} \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix}.$$

We can now compute:

$$\begin{aligned}
A^n &= CD^nC^{-1} \\
&= \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & 3 \end{bmatrix}^n \frac{1}{4} \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{1}{4} \begin{bmatrix} 1 & 1 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} (-1)^n & 0 \\ 0 & 3^n \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{1}{4} \begin{bmatrix} (-1)^n & 3^n \\ (-1)^{n+1} & 3^{n+1} \end{bmatrix} \begin{bmatrix} 3 & -1 \\ 1 & 1 \end{bmatrix} \\
&= \frac{1}{4} \begin{bmatrix} 3(-1)^n + 3^n & (-1)^{n+1} + 3^n \\ 3(-1)^{n+1} + 3^{n+1} & (-1)^{n+2} + 3^{n+1} \end{bmatrix}.
\end{aligned}$$

Now, for our closed formula:

$$\begin{aligned}
\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ 3 & 2 \end{bmatrix}^n \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \\
&= \frac{1}{4} \begin{bmatrix} 3(-1)^n + 3^n & (-1)^{n+1} + 3^n \\ 3(-1)^{n+1} + 3^{n+1} & (-1)^{n+2} + 3^{n+1} \end{bmatrix} \begin{bmatrix} 5 \\ 7 \end{bmatrix} \\
&= \begin{bmatrix} \dfrac{15(-1)^n + 5 \cdot 3^n - 7(-1)^n + 7 \cdot 3^n}{4} \\ \dfrac{15(-1)^{n+1} + 5 \cdot 3^{n+1} - 7(-1)^{n+1} + 7 \cdot 3^{n+1}}{4} \end{bmatrix} \\
&= \begin{bmatrix} 2(-1)^n + 3 \cdot 3^n \\ 2(-1)^{n+1} + 3 \cdot 3^{n+1} \end{bmatrix}.
\end{aligned}$$

The first entry gives us:

$$a_n = 2(-1)^n + 3 \cdot 3^n$$

By testing $n = 5,$ we find that:

$$a_5 = 2(-1)^5 + 3 \cdot 3^5 = -2 + 729 = 727,$$

which agrees with the value that we obtained for a_5 in the list at the beginning of this example.□

If we compare the results of our two examples above:

$$F_n = \frac{1}{\sqrt{5}}(\gamma^n - \overline{\gamma}^n), \text{ and}$$

$$a_n = 2(-1)^n + 3 \cdot 3^n,$$

we could make an educated guess that if the matrix A associated with the linear recurrence relation of order 2 has ***distinct*** eigenvalues λ_1 and $\lambda_2,$ then the closed-formula for a_n has the form:

$$a_n = d_1\lambda_1^n + d_2\lambda_2^n,$$

for some coefficients d_1 and d_2. That would indeed be correct in general. Unfortunately, if the eigenvalues are ***repeated***, this method does not work. However, this situation has been studied in detail, and so we leave it to a more advanced course.

9.3 Exercises

For the following recursive sequences: (a) Find the next 4 terms in the sequence; (b) Find a closed formula $a_n = a(n)$ for the terms; (c) Check that your formula is correct using your data in (a).

1. $a_0 = 4$, $a_1 = 5$, and $a_n = 3a_{n-1} - 2a_{n-2}$.
2. $a_0 = 3$, $a_1 = 1$, and $a_n = -3a_{n-1} + 4a_{n-2}$.
3. $a_0 = 2$, $a_1 = 7$, and $a_n = 3a_{n-1} + 4a_{n-2}$.
4. $a_0 = 4$, $a_1 = 5$, and $a_n = 6a_{n-1} + 7a_{n-2}$.
5. $a_0 = 1$, $a_1 = 3$, and $a_n = 2a_{n-1} + 8a_{n-2}$.
6. $a_0 = 2$, $a_1 = 5$, and $a_n = 4a_{n-1} + 7a_{n-2}$.
7. $a_0 = 4$, $a_1 = 7$, and $a_n = 8a_{n-1} + 3a_{n-2}$.

9.4 Introduction to Quadratic Forms

Quadratic forms have many applications and appear in many disguises. The Calculus student already has seen these in the study of ***conic sections*** (ellipses, hyperbolas and parabolas) and their 3-dimensional counterparts called ***quadric surfaces*** (ellipsoids, hyperboloids, paraboloids, cones and hyperbolic paraboloids). We also saw them in Chapter 7 when we constructed ***weighted dot products*** for $\mathbb{R}^n$.

Definition: A ***quadratic form*** Q in the variables $x_1, x_2, \ldots, x_n$ is a function of the form:

$$Q(x_1, x_2, \ldots, x_n) = \sum_{\substack{i,j=1 \\ (i \le j)}}^{n} a_{i,j} x_i x_j,$$

for some coefficients $a_{i,j} \in \mathbb{R}$. We also call this a ***homogeneous 2nd degree polynomial***. Notice that there are no linear terms nor a constant term.

If $i < j$, and $a_{i,j} \neq 0$, we call the term $a_{i,j} x_i x_j$ a ***mixed term*** or ***cross term***, because two distinct variables appear in the same term.

A quadratic form is ***diagonal*** if it has no mixed terms, that is, $a_{i,j} = 0$ whenever $i < j$. Thus, we can write it in the simpler form:

$$Q(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} a_i \cdot x_i^2 = a_1 x_1^2 + a_2 x_2^2 + \cdots + a_n x_n^2.$$

Examples: A quadratic form in one variable x has the form:

$$Q_1(x) = ax^2.$$

Thus, it can be visualized as a ***parabola*** where $y = Q_1(x) = ax^2$, for example:

$$Q_1(x) = -2x^2.$$

A form in only one variable, of course, is not very interesting.

A quadratic form in two variables, say x and y, has the form:

$$Q_2(x,y) = ax^2 + bxy + cy^2.$$

It is also called a ***binary*** quadratic form. For example, we can have:

$$Q_2(x,y) = 11x^2 + 24xy + 4y^2.$$

This quadratic form is ***not*** diagonal. However, we could have:

$$Q_3(x,y) = 4x^2 + 7y^2,$$

which is a ***diagonal*** quadratic form. □

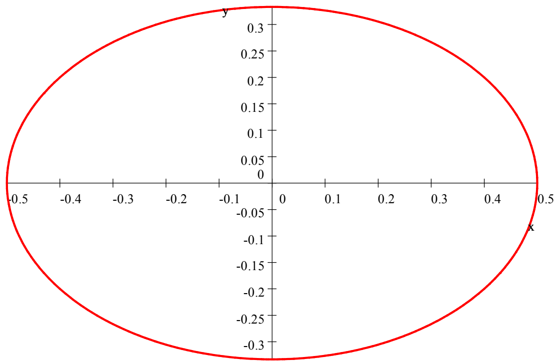
Conic Sections and Quadric Surfaces

Quadratic forms appear naturally when we study conic sections. For example, if we set the quadratic form $Q_3(x,y)$ equal to 1, we will get an ellipse:

$$4x^2 + 9y^2 = 1, \text{ or equivalently:}$$

$$\frac{x^2}{(1/2)^2} + \frac{x^2}{(1/3)^2} = 1.$$

This is a horizontal ellipse with vertices at $(\pm 1/2, 0)$ and covertices $(0, \pm 1/3)$.

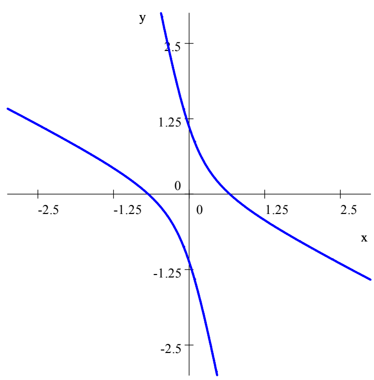


The Ellipse $4x^2 + 9y^2 = 1$

The quadratic form Q_2, though, is not diagonal. If we set it equal to 5, we will get:

$$11x^2 + 24xy + 4y^2 = 5.$$

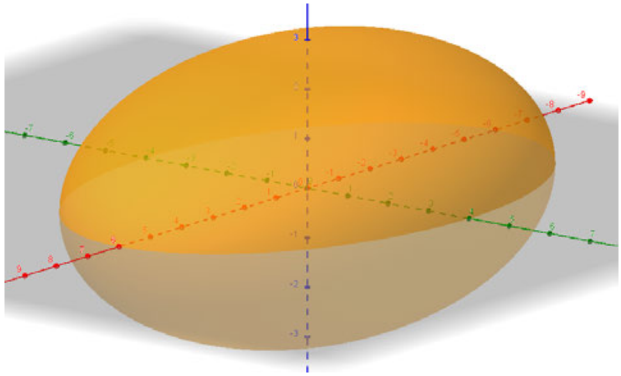
It is not obvious what its graph looks like, but we used a Computer Algebra System to draw it below:



The Hyperbola $11x^2 + 24xy + 4y^2 = 5$

It certainly looks like a hyperbola, but even if it were, its axes of symmetry are neither horizontal nor vertical. We will see in a little bit how to graph it without the use of technology.

A quadratic form in three variables, say x, y and z, would naturally represent a quadric surface when we set it equal to a constant as we saw above. For example, we show below the graph of an ellipsoid and some of its cross-sections.



The Ellipsoid $4x^2 + 9y^2 + 16z^2 = 144$

Obviously, an ellipsoid is the 3-dimensional analog of an ellipse.

The Matrix Form of a Quadratic Form

In order to prove properties of quadratic forms or perform computations on them in an efficient way, it is often more convenient to rewrite them using square matrices.

Definition: Let $\vec{x} = \langle x_1, x_2, \ldots, x_n \rangle \in \mathbb{R}^n$. Denote by $[\vec{x}]$ the ***column matrix*** with entries from $\vec{x}$, as usual. The quadratic form:

$$Q(x_1, x_2, \ldots, x_n) = \sum_{\substack{i,j=1 \\ (i \leq j)}}^{n} a_{i,j} x_i x_j,$$

can be written as a ***matrix product:***

$$Q(x_1, x_2, \ldots, x_n) = [\vec{x}]^\top [Q][\vec{x}],$$

where the $n \times n$ matrix $[Q]$, known as ***the matrix of*** Q, is defined by:

$$[Q]_{i,j} = \begin{cases} a_{i,i} & \text{if } i = j \\ a_{i,j}/2 & \text{if } i < j \\ a_{j,i}/2 & \text{if } i > j \end{cases}, \text{ and thus:}$$

$$[Q] = \begin{bmatrix} a_{1,1} & a_{1,2}/2 & \cdots & a_{1,k}/2 \\ a_{1,2}/2 & a_{2,2} & \cdots & a_{2,k}/2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,k}/2 & a_{2,k}/2 & \cdots & a_{k,k} \end{bmatrix}.$$

The reason for this cumbersome definition is that we only have the term $a_{i,j}x_ix_j$ where $i \leq j$. For example, we have an x_1x_3 term but not an x_3x_1 term. The “1/2” factors disappear when we expand the product and combine both terms. It is obvious that $[Q]$ is ***symmetric***, and Q is diagonal ***if and only if*** $[Q]$ is a diagonal matrix.

Examples: Let us rewrite our first Examples. The definition says that a quadratic form in two variables has the form:

$$\begin{aligned} Q_2(x,y) &= ax^2 + bxy + cy^2 \\ &= \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}. \end{aligned}$$

If we expand this product step-by-step, we obtain:

$$\begin{aligned} &\begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} ax + by/2 \\ bx/2 + cy \end{bmatrix} \\ &= (ax + by/2)x + (bx/2 + cy)y \\ &= ax^2 + byx/2 + bxy/2 + cy^2 \\ &= ax^2 + bxy + cy^2, \end{aligned}$$

as expected. For $Q_2(x,y) = 11x^2 + 24xy + 4y^2$, we have:

$$[Q_2] = \begin{bmatrix} 11 & 12 \\ 12 & 4 \end{bmatrix}.$$

For our diagonal example, $Q_3(x,y) = 4x^2 + 9y^2$, we obtain:

$$[Q_3] = \begin{bmatrix} 4 & 0 \\ 0 & 9 \end{bmatrix}.$$

This is a diagonal matrix, as expected. □

Diagonalizing Quadratic Forms

It should come as no surprise that the matrix form of a quadratic form unleashes the full power of the Spectral Theorem for Symmetric Matrices:

> ***Theorem:*** Let $Q(x_1, x_2, \ldots, x_n)$ be a quadratic form. Then: there exists a change of variables:
>
> $$[\vec{y}] = U^{\top}[\vec{x}], \text{ or } [\vec{x}] = U[\vec{y}],$$
>
> where U is an ***orthogonal matrix***, and $y_1, y_2, \ldots, y_n$ are new variables, such that Q can be rewritten as the ***equivalent*** quadratic form:
>
> $$Q(y_1, y_2, \ldots, y_n) = [\vec{y}]^{\top} D [\vec{y}],$$
>
> and D is a ***diagonal*** matrix. In other words, it is always possible to ***diagonalize*** Q.
> In some instances, we can require D to contain the eigenvalues of $[Q]$ in ***ascending order***:
>
> $$D = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n), \quad \text{where } \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

Proof: By the Spectral Theorem, we can diagonalize $[Q]$ using an ***orthogonal*** matrix U :

$$[Q] = UDU^{-1} = UDU^{\top},$$

since $U^{-1} = U^{\top}$ by the orthogonal property. Since we are free to arrange the eigenvalues in any order we please, we can assume that they are in ascending order in D, if it is so desired. In any case, using the associative property of multiplication and properties of the transpose, we have:

$$\begin{aligned} [\vec{x}]^{\top}[Q][\vec{x}] &= [\vec{x}]^{\top}(UDU^{\top})[\vec{x}] \\ &= ([\vec{x}]^{\top}U)D(U^{\top}[\vec{x}]) \\ &= (U^{\top}[\vec{x}])^{\top}D(U^{\top}[\vec{x}]) \\ &= [\vec{y}]^{\top}D[\vec{y}], \end{aligned}$$

where $[\vec{y}] = U^{\top}[\vec{x}]$. Thus, we can rewrite the quadratic form diagonally as:

$$Q(y_1, y_2, \ldots, y_n) = [\vec{y}]^{\top}D[\vec{y}]. \blacksquare$$

Example: Suppose we are given:

$$Q(x, y, z) = 24x^2 - 16xy + 16xz - y^2 - 34yz - z^2.$$

Then:

$$[Q] = \begin{bmatrix} 24 & -8 & 8 \\ -8 & -1 & -17 \\ 8 & -17 & -1 \end{bmatrix}.$$

Using a Computer Algebra System, we find that its eigenvalues are $\lambda = -18,\ 8$ and 32, and so each eigenspace is 1-dimensional. We further find that:

$$\begin{aligned} Eig([Q], -18) &= Span(\{\langle 0, 1, 1\rangle\}), \\ Eig([Q], 8) &= Span(\{\langle 1, 1, -1\rangle\}), \text{ and} \\ Eig([Q], 32) &= Span(\{\langle 2, -1, 1\rangle\}). \end{aligned}$$

Converting these to unit vectors, we can form the orthogonal matrix U and the diagonal matrix D:

$$U = \begin{bmatrix} 0 & 1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & -1/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}, \text{ and } D = \begin{bmatrix} -18 & 0 & 0 \\ 0 & 8 & 0 \\ 0 & 0 & 32 \end{bmatrix}.$$

If we denote the new variables $X,\ Y$ and $Z,$ then our old variables are related to our new ones by:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 & 1/\sqrt{3} & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & -1/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{3} & 1/\sqrt{6} \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}.$$

In other words:

$$\begin{aligned} x &= & & Y/\sqrt{3} &+& 2Z/\sqrt{6} \\ y &= X/\sqrt{2} &+& Y/\sqrt{3} &-& Z/\sqrt{6} \\ z &= X/\sqrt{2} &-& Y/\sqrt{3} &+& Z/\sqrt{6} \end{aligned} .$$

With respect to the new variables, the quadratic form becomes:

$$Q(X, Y, Z) = -18X^2 + 8Y^2 + 32Z^2. \square$$

The Discriminant of a Binary Quadratic Form

In the classic quadratic formula, the discriminant tells us if we have two distinct real roots, a repeated real root, or two distinct non-real roots. In the case of a binary quadratic form, we define the discriminant using essentially the same formula. We will use this in the next Section to help us graph conic sections.

Definition/Theorem — The Invariance of the Discriminant:
Consider the binary quadratic form:

$$Q(x,y) = ax^2 + bxy + cy^2,$$

where $a, b, c \in \mathbb{R}$. We define the ***discriminant*** of Q to be the number:

$$\Delta = b^2 - 4ac.$$

Then: the discriminant is an ***invariant*** of Q under an ***orthogonal*** change of variables. This means the following: Suppose that U is any 2×2 ***orthogonal matrix***, with associated change of variables:

$$\begin{bmatrix} x \\ y \end{bmatrix} = U \begin{bmatrix} X \\ Y \end{bmatrix}.$$

If we substitute x and y into our quadratic above, we obtain a new quadratic form:

$$Q'(X, Y) = a'X^2 + b'XY + c'Y^2,$$

for some new coefficients a', b' and c'. However, the new discriminant $\Delta' = (b')^2 - 4(a')(c')$ of $Q'(X, Y)$ is ***exactly the same*** as the discriminant of $Q(x,y)$:

$$(b')^2 - 4(a')(c') = \Delta = b^2 - 4ac.$$

Proof: Recall that any 2×2 orthogonal matrix has the form:

$$U = \begin{bmatrix} \cos(\theta) & \mp\sin(\theta) \\ \pm\sin(\theta) & \cos(\theta) \end{bmatrix},$$

where the choice of sign only determines if U is proper or improper. Thus, we get the change of variables:

$$x = \cos(\theta)X \mp \sin(\theta)Y, \text{ and}$$

$$y = \pm\sin(\theta)X + \cos(\theta)Y.$$

Substituting into Q, we get:

$$\begin{aligned}
&ax^2 + bxy + cy^2 \\
&= a(\cos(\theta)X \mp \sin(\theta)Y)^2 + b(\cos(\theta)X \mp \sin(\theta)Y)(\pm\sin(\theta)X + \cos(\theta)Y) \\
&\quad + c(\pm\sin(\theta)X + \cos(\theta)Y)^2 \\
&= \quad a(\cos^2(\theta)X^2 \mp 2cos(\theta)\sin(\theta)XY + \sin^2(\theta)Y^2) \\
&\quad + b(\pm\cos(\theta)\sin(\theta)X^2 - \sin^2(\theta)XY + \cos^2(\theta)XY \mp \sin(\theta)\cos(\theta)Y^2) \\
&\quad + c(\sin^2(\theta)X^2 \pm 2sin(\theta)\cos(\theta)XY + \cos^2(\theta)Y^2).
\end{aligned}$$

Distributing a, b, and c, and collecting coefficients of X^2, XY and Y^2, we get:

$$\begin{aligned}
a' &= a \cdot \cos^2(\theta) \pm b \cdot \cos(\theta)\sin(\theta) + c \cdot \sin^2(\theta) \\
&= \frac{a+c}{2} + \left[\pm\frac{b}{2}\sin(2\theta) + \frac{a-c}{2}\cos(2\theta)\right] \\
b' &= \mp 2a \cdot \cos(\theta)\sin(\theta) + b \cdot [\cos^2(\theta) - \sin^2(\theta)] \pm 2c \cdot \sin(\theta)\cos(\theta) \\
&= b \cdot \cos(2\theta) \pm (c-a) \cdot \sin(2\theta) \\
c' &= a \cdot \sin^2(\theta) \mp b \cdot \cos(\theta)\sin(\theta) + c \cdot \cos^2(\theta) \\
&= \frac{a+c}{2} - \left[\pm\frac{b}{2}\sin(2\theta) + \frac{a-c}{2}\cos(2\theta)\right],
\end{aligned}$$

where we used the identities:

$$\cos^2(\theta) = \frac{1+\cos(2\theta)}{2}, \quad \sin^2(\theta) = \frac{1-\cos(2\theta)}{2}, \quad \text{and} \quad \sin(2\theta) = 2\sin(\theta)\cos(\theta).$$

Notice that the terms in square brackets found in a' and c' are ***identical***, and so a' and c' have the form:

$$\begin{aligned}
a' &= \alpha + \beta, \text{ and } c' = \alpha - \beta, \text{ hence} \\
4a'c' &= 4\alpha^2 - 4\beta^2.
\end{aligned}$$

Finally, we substitute these into Δ' to get:

$$\begin{aligned}
\Delta' &= (b')^2 - 4(a')(c') \\
&= [b \cdot \cos(2\theta) \pm (c-a) \cdot \sin(2\theta)]^2 - 4\alpha^2 + 4\beta^2 \\
&= b^2 \cdot \cos^2(2\theta) \pm 2b(c-a) \cdot \cos(2\theta)\sin(2\theta) + (c-a)^2\sin^2(2\theta) \\
&\quad - 4\left[\frac{a+c}{2}\right]^2 + 4\left[\pm\frac{b}{2}\sin(2\theta) + \frac{a-c}{2}\cos(2\theta)\right]^2 \\
&= b^2 \cdot \cos^2(2\theta) \pm 2b(c-a) \cdot \cos(2\theta)\sin(2\theta) + (c-a)^2\sin^2(2\theta) \\
&\quad - (a+c)^2 + b^2\sin^2(2\theta) \pm 2b(a-c)\cos(2\theta)\sin(2\theta) + (a-c)^2\cos^2(2\theta) \\
&= b^2 \cdot \cos^2(2\theta) + b^2\sin^2(2\theta) \\
&\quad + (c-a)^2\sin^2(2\theta) + (a-c)^2\cos^2(2\theta) - (a+c)^2 \\
&= b^2 + (a-c)^2 - (a+c)^2 \\
&= b^2 - 4ac. \blacksquare
\end{aligned}$$

We remark that a binary quadratic form is essentially what is referred to as a ***homogenization*** of the usual quadratic equation:

$$ax^2 + bx + c = 0.$$

To make all three terms have total degree 2, we multiply each term by an appropriate power of a new variable y :

$$ax^2 + bxy + cy^2 = 0.$$

We obtain the original quadratic by plugging in $y = 1$.

The Complex Case

We must be careful in defining a quadratic form with complex coefficients and in complex variables. Suppose that $z_1, z_2, \dots, z_n \in \mathbb{C}$. Then we define a quadratic form:

$$Q(z_1, z_2, \dots, z_n) = \sum_{i,j=1}^{n} a_{i,j}\overline{z_i}z_j,$$

where we require that the coefficients $a_{i,j}$ satisfy:

$$\begin{aligned} a_{i,i} &\in \mathbb{R} && \text{for all } i = 1..n, \text{ and} \\ a_{i,j} &= a_{j,i}^* && \text{for all } i, j = 1..n. \end{aligned}$$

Notice that we are forcing the appearance of both $\overline{z_i}z_j$ and $\overline{z_j}z_i$ in this definition, where $i \neq j$. But because of the restrictions on the coefficients, the matrix:

$$[Q] = [a_{i,j}]$$

is an ***Hermitian matrix***, i.e. $[Q]^* = [Q]$, where $[Q]^*$ is the adjoint of $[Q]$. This is of course the natural extension because $[Q]$ is ***symmetric*** in the real case.

By the Spectral Theorem for Hermitian Matrices, $[Q]$ is once more diagonalizable, but this time, using a ***unitary*** matrix U, that is, with $UU^* = \boldsymbol{I}_n = U^*U$:

$$[Q] = UDU^{-1} = UDU^*,$$

Once again, the eigenvalues of $[Q]$, and hence the diagonal entries of D, are all ***real*** numbers. If we let $[\vec{z}\,]$ represent the column matrix with entries z_i, then we get:

$$\begin{aligned} & Q(z_1, z_2, \dots, z_n) \\ &= [\vec{z}\,]^*[Q][\vec{z}\,] \\ &= [\vec{z}\,]^*(UDU^*)[\vec{z}\,] \\ &= ([\vec{z}\,]^*U)D(U^*[\vec{z}\,]) \\ &= [\vec{w}\,]^*D[\vec{w}\,], \end{aligned}$$

where $[\vec{w}\,] = U^*[\vec{z}\,]$ represents a unitary change of variables.

Example: Suppose we have:

$$\begin{aligned} Q(z_1, z_2) &= 3\overline{z_1}z_1 + (2-i)\overline{z_1}z_2 + (2+i)\overline{z_2}z_1 - \overline{z_2}z_2 \\ &= 3\|z_1\|^2 + (2-i)\overline{z_2}z_2 + (2+i)z_1\overline{z_2} - \|z_2\|^2. \end{aligned}$$

The matrix of Q is:

$$[Q] = \begin{bmatrix} 3 & 2-i \\ 2+i & -1 \end{bmatrix},$$

and indeed $[Q]$ is Hermitian. Its characteristic polynomial is:

$$\begin{aligned} p(\lambda) &= (\lambda-3)(\lambda+1) - (2-i)(2+i) \\ &= \lambda^2 - 2\lambda - 8 = (\lambda-4)(\lambda+2). \end{aligned}$$

Its eigenspaces are:

$$\begin{aligned} Eig([Q],-2) &= Span(\{\vec{u}\}) = Span(\{\langle -1, 2+i\rangle\}), \text{ and} \\ Eig([Q],4) &= Span(\{\vec{v}\}) = Span(\{\langle 2-i, 1\rangle\}). \end{aligned}$$

Recall that to check if two complex vectors are orthogonal, we take the complex conjugate of one of the vectors before we take their dot product, or equivalently:

$$\langle \vec{u} | \vec{v} \rangle = [\vec{v}]^*[\vec{u}] = (2+i)(-1) + 1(2+i) = 0,$$

hence the eigenspaces are orthogonal, as they should be. Both of the basis vectors have length $\sqrt{6}$, so we obtain the unitary matrix:

$$U = \frac{1}{\sqrt{6}}\begin{bmatrix} -1 & 2-i \\ 2+i & 1 \end{bmatrix}.$$

Thus, under the change of variables:

$$\begin{aligned} z_1 &= (-w_1 + (2-i)w_2)/\sqrt{6}, \text{ and} \\ z_1 &= ((2+i)w_1 + w_2)/\sqrt{6}, \end{aligned}$$

the quadratic form can be written in diagonal form as:

$$Q(w_1, w_2) = -2\|w_1\|^2 + 4\|w_2\|^2.$$

In particular, the value of the quadratic form is always ***real***.$\square$

9.4 Exercises

For Exercises (1) to (9): For each of the following quadratic forms in x, y (or x, y, z), find new variables X, Y (or X, Y, Z) and the corresponding change of variables, such that the quadratic form is diagonal with respect to the new variables, and find this equivalent quadratic form.

1. $Q(x, y) = 8x^2 - 4xy + 5y^2$
2. $Q(x, y) = -x^2 + 4xy - 4y^2$
3. $Q(x, y) = 3x^2 + 12xy - 2y^2$
4. $Q(x, y) = 13x^2 - 12xy + 8y^2$
5. $Q(x, y, z) = 7x^2 + 7y^2 + 7z^2 - 2xz$
6. $Q(x, y, z) = 4x^2 + 5y^2 + 4z^2 + 2xy + 4xz + 2yz$
7. $Q(x, y, z) = x^2 - 2xy + 2y^2 - 2yz + z^2$
8. $Q(x, y, z) = x^2 + xy + xz + yz$
9. $Q(x, y, z) = x^2 + y^2 + z^2 - 2xy - 2xz - 2yz$

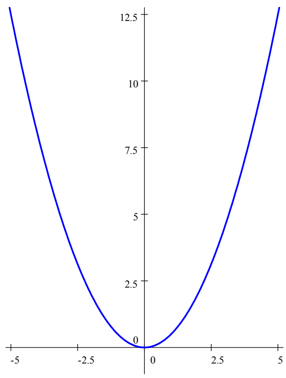
For Exercises (10) to (13): Find the discriminant of the binary quadratic forms in Exercises (1) through (4), respectively. Check that the discriminant of the equivalent diagonal quadratic form is the same.

For Exercises (14) to (20): For the quadratic form in the complex variables z_1 and z_2, find new complex variables w_1 and w_2 and the corresponding change of variables, such that the quadratic form is diagonal with respect to w_1 and w_2, and find this equivalent quadratic form.

14. $Q(z_1, z_2) = 2\overline{z_1}z_1 + (1 + 3i)\overline{z_1}z_2 + (1 - 3i)\overline{z_2}z_1 + 5\overline{z_2}z_2$
15. $Q(z_1, z_2) = \overline{z_1}z_1 + (2 + 2i)\overline{z_1}z_2 + (2 - 2i)\overline{z_2}z_1 - \overline{z_2}z_2$
16. $Q(z_1, z_2) = 7\overline{z_1}z_1 + (3 + 4i)\overline{z_1}z_2 + (3 - 4i)\overline{z_2}z_1 + 7\overline{z_2}z_2$
17. $Q(z_1, z_2) = 6\overline{z_1}z_1 + (7 - 24i)\overline{z_1}z_2 + (7 + 24i)\overline{z_2}z_1 + 6\overline{z_2}z_2$
18. $Q(z_1, z_2) = 2\overline{z_1}z_1 + (1 + 3i)\overline{z_1}z_2 + (1 - 3i)\overline{z_2}z_1 + 5\overline{z_2}z_2$
19. $Q(z_1, z_2) = 2\overline{z_1}z_1 + (3 - i)\overline{z_1}z_2 + (3 + i)\overline{z_2}z_1 + 5\overline{z_2}z_2$
20. $Q(z_1, z_2) = -5\overline{z_1}z_1 + (4 - 3i)\overline{z_1}z_2 + (4 + 3i)\overline{z_2}z_1 - 5\overline{z_2}z_2$

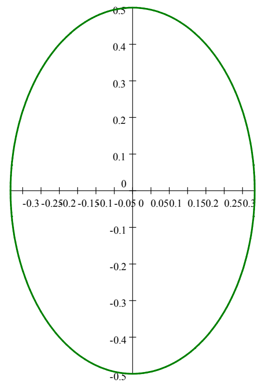
9.5 Rotations of Conics

We will now use the theory of quadratic forms in order to draw the graph of a quadratic equation that contains a mixed term xy. Let us first recall the three basic conic sections and the standard forms of their equations:



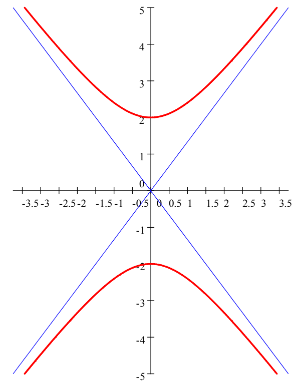
The Parabola

$$y = ax^2$$



The Ellipse

$$\frac{y^2}{a^2} + \frac{x^2}{b^2} = 1$$



The Hyperbola

$$\frac{y^2}{a^2} - \frac{x^2}{b^2} = 1$$

Of course, it is also possible for x and y to exchange roles to change the orientation of the major axes of these conic sections. A ***general quadratic equation in x and y*** has the form:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0.$$

It has as its graph a conic section or one of its ***degenerate*** forms:

- A parabola can degenerate into ***a line*** or ***two parallel lines***.
- An ellipse can degenerate (in this case, keep shrinking) into ***a single point*** or ***no graph*** whatsoever.
- A hyperbola can degenerate (in this case, open wider and wider) into ***two intersecting lines***.

In order to identify the conic section that it represents and sketch its graph by hand, however, we must bring it to an equivalent ***diagonal*** form. In the traditional method found in many Precalculus books, we would compute an angle of rotation and a change of variables that will eliminate the mixed term. Instead, our method below will produce a rotation matrix using ***eigenvectors***, without trigonometry!

Theorem: Consider the general quadratic equation:

$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \text{ with } b \neq 0, \text{ and}$$

$$Q(x, y) = ax^2 + bxy + cy^2,$$

its associated quadratic form. Then, there exists a change of variables by a ***rotation matrix*** rot_θ :

$$\begin{bmatrix} x \\ y \end{bmatrix} = rot_\theta \begin{bmatrix} X \\ Y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} X \\ Y \end{bmatrix},$$

so that the equation that we obtain by substituting in:

$$x = \cos(\theta)X - \sin(\theta)Y, \text{ and}$$
$$y = \sin(\theta)X + \cos(\theta)Y,$$

has no mixed term $b'XY$. In other words, the resulting quadratic has the form:

$$a'X^2 + c'Y^2 + d'X + e'Y + f = 0$$

(the constant term will not change). The rotation matrix rot_θ can be chosen so that $\theta \in (0, \pi/2)$.

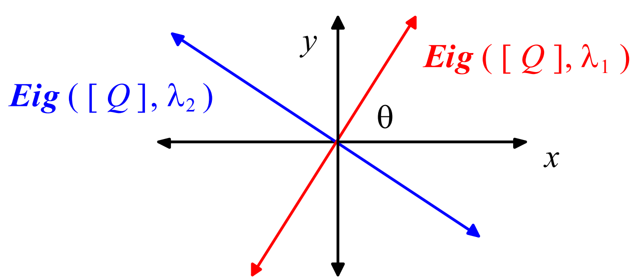
Proof: Let us form the matrix of Q:

$$[Q] = \begin{bmatrix} a & b/2 \\ b/2 & c \end{bmatrix}.$$

The characteristic polynomial of $[Q]$ is:

$$\begin{aligned} p(\lambda) &= (\lambda - a)(\lambda - c) - b^2/4 \\ &= \lambda^2 - (a+c)\lambda - (b^2 - 4ac)/4, \quad \text{so} \\ \lambda &= \frac{a + c \pm \sqrt{(a+c)^2 + (b^2 - 4ac)}}{2} \\ &= \frac{a + c \pm \sqrt{a^2 + 2ac + c^2 + b^2 - 4ac}}{2} \\ &= \frac{a + c \pm \sqrt{(a-c)^2 + b^2}}{2}. \end{aligned}$$

Since we assumed that $b \neq 0$, the radical above is non-zero, so we get two ***distinct*** eigenvalues. Thus we will get two orthogonal 1-dimensional eigenspaces:



We have chosen λ_1 to be the eigenvalue whose eigenspace passes through the 1st quadrant. In this case, the angle $\theta \in (0, \pi/2)$ made by $Eig([Q], \lambda_1)$ with respect to the x-axis gives us the eigenvector $\vec{v}_1 = \langle \cos(\theta), \sin(\theta) \rangle$. We need not explicitly compute θ, because a ***unit vector*** automatically has this form. Correspondingly, we get the eigenvector $\vec{v}_2 = \langle -\sin(\theta), \cos(\theta) \rangle$ for λ_2, so that it is in the 2nd quadrant. Note that $\vec{v}_2$ is 90^0 counterclockwise from $\vec{v}_1$.

Incidentally, it is impossible for θ to be 0 or $\pi/2$, otherwise the original quadratic will not contain a mixed term. Thus, from our previous Theorem and its proof, our change of basis matrix U is:

$$U = rot_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

and our change of variables is:

$$\begin{bmatrix} X \\ Y \end{bmatrix} = U^\top \begin{bmatrix} x \\ y \end{bmatrix}, \text{ and thus: } \begin{bmatrix} x \\ y \end{bmatrix} = U \begin{bmatrix} X \\ Y \end{bmatrix}.$$

Under this change of variables, Q can be written as:

$$Q(X, Y) = \lambda_1 X^2 + \lambda_2 Y^2,$$

and therefore there is no XY term. If our original quadratic contains a linear (degree 1) term, we can now bring the quadratic equation into one of the three standard forms for conic sections so that it can be graphed.■

Example: Let us see how to apply this diagonalization to graph our suspected hyperbola:

$$11x^2 + 24xy + 4y^2 = 5$$

that we saw in the previous Section. The matrix of the associated quadratic form is:

$$[Q] = \begin{bmatrix} 11 & 12 \\ 12 & 4 \end{bmatrix}.$$

We can easily find its characteristic polynomial:

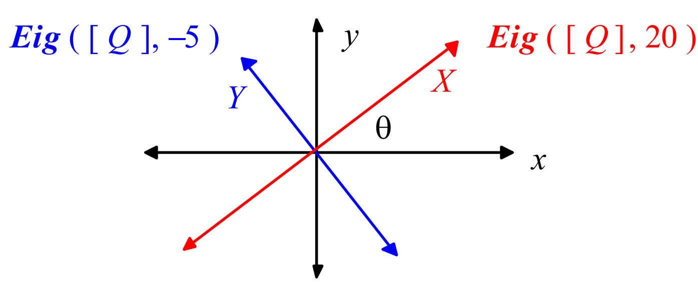
$$\begin{aligned} p(\lambda) &= (\lambda - 11)(\lambda - 4) - 144 \\ &= \lambda^2 - 15\lambda - 100 = (\lambda - 20)(\lambda + 5). \end{aligned}$$

The eigenspaces for $[Q]$ are:

$$Eig([Q], 20) = Span(\{\langle 4/5, 3/5 \rangle\}), \text{ and}$$

$$Eig([Q], -5) = Span(\{\langle -3/5, 4/5 \rangle\}).$$

These two Spans are clearly orthogonal, and we show them below:



These will be the new X and Y axes, respectively. With $\lambda_1 = 20$ and $\lambda_2 = -5$, we obtain:

$$rot_\theta = \begin{bmatrix} 4/5 & -3/5 \\ 3/5 & 4/5 \end{bmatrix} \text{ and } D = \begin{bmatrix} 20 & 0 \\ 0 & -5 \end{bmatrix}.$$

The corresponding change of variables is:

$$x = \frac{4}{5}X - \frac{3}{5}Y, \text{ and}$$
$$y = \frac{3}{5}X + \frac{4}{5}Y.$$

Substituting these into our original quadratic, we get:

$$\begin{aligned}
5 &= 11x^2 + 24xy + 4y^2 \\
&= 11\left(\tfrac{4}{5}X - \tfrac{3}{5}Y\right)^2 + 24\left(\tfrac{4}{5}X - \tfrac{3}{5}Y\right)\left(\tfrac{3}{5}X + \tfrac{4}{5}Y\right) + 4\left(\tfrac{3}{5}X + \tfrac{4}{5}Y\right)^2 \\
&= 11\left(\tfrac{16}{25}X^2 - \tfrac{24}{25}XY + \tfrac{9}{25}Y^2\right) + 24\left(\tfrac{12}{25}X^2 + \tfrac{7}{25}XY - \tfrac{12}{25}Y^2\right) \\
&\quad + 4\left(\tfrac{9}{25}X^2 + \tfrac{24}{25}XY + \tfrac{16}{25}Y^2\right) \\
&= \frac{176}{25}X^2 - \frac{264}{25}XY + \frac{99}{25}Y^2 + \frac{288}{25}X^2 + \frac{168}{25}XY - \frac{288}{25}Y^2 \\
&\quad + \frac{36}{25}X^2 + \frac{96}{25}XY + \frac{64}{25}Y^2 \\
&= 20X^2 - 5Y^2.
\end{aligned}$$

This is correct because our eigenvalues are 20 and -5. Scaling by 5, we get the standard equation:

$$1 = 4X^2 - Y^2 = \frac{X^2}{(1/2)^2} - \frac{Y^2}{1},$$

so indeed our original equation represents a rotated hyperbola.

Let us put some finishing touches: The vertices of the hyperbola are at $(X, Y) = (\pm 1/2, 0)$, and the asymptotes are $Y = \pm 2X$. Let us find their coordinates and equations in terms of x and y. If $X = \pm 1/2$ and $Y = 0$, we get:

$$x = \pm\frac{2}{5} \text{ and } y = \pm\frac{3}{10}.$$

so the vertices are at $(\pm 0.4, \pm 0.6)$. Notice that this is consistent with the slope of the major axis. Now for the asymptotes: the equations for the ***inverse*** change of coordinates are:

$$X = \cos(\theta)x + \sin(\theta)y = \frac{4}{5}x + \frac{3}{5}y, \text{ and}$$
$$Y = -\sin(\theta)X + \cos(\theta)Y = -\frac{3}{5}x + \frac{4}{5}y,$$

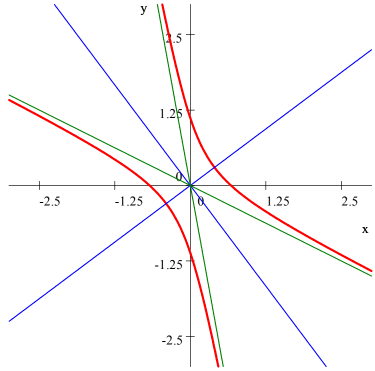
so if we substitute these into $X = \pm 2Y$, we get:

$$-\frac{3}{5}x + \frac{4}{5}y = 2\left(\frac{4}{5}x + \frac{3}{5}y\right), \text{ and}$$
$$-\frac{3}{5}x + \frac{4}{5}y = -2\left(\frac{4}{5}x + \frac{3}{5}y\right).$$

Simplifying, we get:

$$y = -\frac{11}{2}x \text{ and } y = -\frac{1}{2}x.$$

We redraw our original graph with this new information:



The Hyperbola $11x^2 + 24xy + 4y^2 = 5$
with its asymptotes and axes of symmetry

If we were to graph this by hand, we would draw the new X and Y axes using the eigenvectors, mark off units in the same scale as our original, then ignore the x and y axes. We can draw $4X^2 - Y^2 = 1$ on the new X and Y axes using the standard steps from Precalculus: plot the vertices, sketch the asymptotes, plot four more points using symmetry, then connect the dots keeping the asymptotes in mind.□

The Discriminant of a General Quadratic

Now, since the discriminant of a quadratic form is invariant, we can use the standard equations of the three types of conic sections to develop a litmus test for any general quadratic. We will first eliminate denominators, if need be, to simplify the discriminant:

- For a parabola $y = ax^2$, or $ax^2 - y = 0$; $\Delta = 0^2 - 4 \bullet a \bullet 0 = 0$.
- For an ellipse $a^2x^2 + b^2y^2 - a^2b^2 = 0$; $\Delta = 0^2 - 4a^2b^2 < 0$.
- For a hyperbola $a^2x^2 - b^2y^2 - a^2b^2 = 0$; $\Delta = 0^2 + 4a^2b^2 > 0$.

Thus, we can conclude:

Theorem: If the ***discriminant***, Δ, of a general quadratic equation is:
a) ***positive***, then its graph is a ***hyperbola*** or two intersecting lines.
b) ***zero***, then its graph is a ***parabola***, or a line, or two parallel lines.
c) ***negative***, then its graph is an ***ellipse***, a point, or it has no graph.

Example: Consider the quadratic:

$$4x^2 + 12xy + 9y^2 - 4x - 6y - 35 = 0.$$

Its discriminant is:

$$\Delta = 12^2 - 4 \bullet 4 \bullet 9 = 144 - 144 = 0,$$

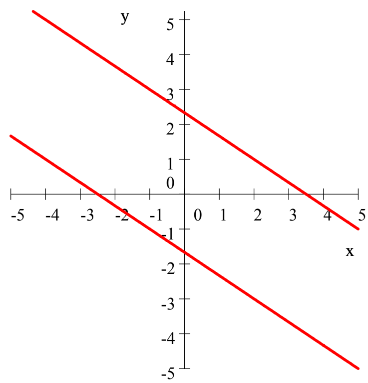
and therefore its graph is either a parabola, a line, or two parallel lines. Instead of rotating this quadratic as we did in the previous example, let us point out that the left side can be factored by grouping, obtaining:

$$(2x+3y)^2 - 2(2x+3y) - 35 = (2x+3y-7)(2x+3y+5) = 0.$$

Thus, we get two parallel lines:

$$2x+3y = 7 \text{ and } 2x+3y = -5.$$

We show the graph below:



The Degenerate Quadratic: $4x^2 + 12xy + 9y^2 - 4x - 6y - 35 = 0$

As a final result, we saw in the proof about the invariance of the discriminant from the previous Section that the coefficient of the new mixed term is given by:

$$b' = b \cdot \cos(2\theta) \pm (c-a) \cdot \sin(2\theta).$$

If the change of variables were accomplished by a rotation matrix (i.e. a ***proper*** orthogonal matrix), we would choose the ***positive*** sign. Thus we get a well-known result:

Theorem: To eliminate the mixed term bxy, $b \neq 0$, under a change of variables via a rotation matrix rot_θ, the angle θ must satisfy the equation:

$$\cot(2\theta) = \frac{a-c}{b}.$$

Example: Let us look again at our previous quadratic:

$$4x^2 + 12xy + 9y^2 - 4x - 6y - 35 = 0.$$

It was certainly not obvious that the left side factors nicely, so let us go through the process of rotation to eliminate $12xy$. By our Theorem above:

$$\cot(2\theta) = \frac{a-c}{b} = \frac{4-9}{12} = \frac{-5}{12}.$$

By completing the triangle, we get $\cos(2\theta) = -5/13$. Now, by the half angle formula:

$$\cos(\theta) = \sqrt{\frac{1+\cos(2\theta)}{2}} = \sqrt{\frac{1-5/13}{2}} = \frac{2}{\sqrt{13}}, \text{ and so: } \sin(\theta) = \frac{3}{\sqrt{13}},$$

so that θ is in the 1st quadrant. We get the change of variables:

$$x = \frac{2}{\sqrt{13}}X - \frac{3}{\sqrt{13}}Y, \quad \text{and}$$

$$y = \frac{3}{\sqrt{13}}X + \frac{2}{\sqrt{13}}Y.$$

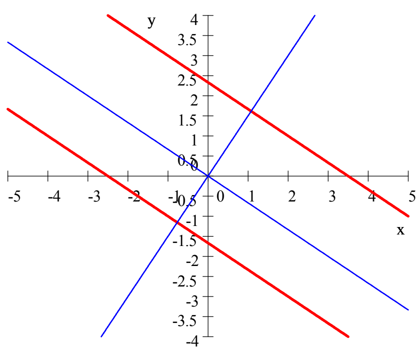
Substituting these, we get:

$$\begin{aligned}
0 &= 4x^2 + 12xy + 9y^2 - 4x - 6y - 35 \\
&= 4\left(\frac{2}{\sqrt{13}}X - \frac{3}{\sqrt{13}}Y\right)^2 + 12\left(\frac{2}{\sqrt{13}}X - \frac{3}{\sqrt{13}}Y\right)\left(\frac{3}{\sqrt{13}}X + \frac{2}{\sqrt{13}}Y\right) + \\
&\quad 9\left(\frac{3}{\sqrt{13}}X + \frac{2}{\sqrt{13}}Y\right)^2 - 4\left(\frac{2}{\sqrt{13}}X - \frac{3}{\sqrt{13}}Y\right) - 6\left(\frac{3}{\sqrt{13}}X + \frac{2}{\sqrt{13}}Y\right) - 35 \\
&= \frac{16}{13}X^2 - \frac{48}{13}XY + \frac{36}{13}Y^2 + \frac{72}{13}X^2 - \frac{60}{13}XY - \frac{72}{13}Y^2 + \\
&\quad \frac{81}{13}X^2 + \frac{108}{13}XY + \frac{36}{13}Y^2 - \frac{8}{13}\sqrt{13}\,X + \frac{12}{13}\sqrt{13}\,Y - \frac{18}{13}\sqrt{13}\,X - \frac{12}{13}\sqrt{13}\,Y - 35 \\
&= 13X^2 - 2\sqrt{13}\,X - 35.
\end{aligned}$$

Solving this (ordinary) quadratic equation, we get:

$$\begin{aligned}
X &= \frac{2\sqrt{13} \pm \sqrt{52 - 4(13)(-35)}}{26} \\
&= \frac{2\sqrt{13} \pm 12\sqrt{13}}{26} = \frac{7}{\sqrt{13}} \quad \text{or} \quad \frac{-5}{\sqrt{13}} \approx 1.9415 \ \text{and} \ -1.3868.
\end{aligned}$$

If we redraw our previous graph with the new X and Y axes, we can confirm that these X coordinates are reasonable:



The Degenerate Quadratic: $4x^2 + 12xy + 9y^2 - 4x - 6y - 35 = 0$

However, we also have reverse equations:

$$X = \frac{2}{\sqrt{13}}x + \frac{3}{\sqrt{13}}y, \quad \text{and} \quad Y = \frac{3}{\sqrt{13}}x - \frac{2}{\sqrt{13}}y.$$

and so we get:

$$X = \frac{7}{\sqrt{13}} \text{ or } \frac{-5}{\sqrt{13}} \Rightarrow$$

$$\frac{2}{\sqrt{13}}x + \frac{3}{\sqrt{13}}y = \frac{7}{\sqrt{13}} \text{ or } \frac{-5}{\sqrt{13}} \Rightarrow$$

$$2x + 3y = 7 \text{ or } -5,$$

which are indeed the equations of the two lines that we found.□

9.5 Exercises

For the following quadratic equations in x and y: find new variables X and Y and the corresponding change of variables such that the resulting equivalent quadratic in X and Y has no XY term. Sketch the quadratic, showing the new X and Y axes. In addition, provide the following information in the ***original*** x, y coordinates:

- If the graph is a parabola, specify the vertex and the axis of symmetry.
- If the graph is an ellipse, specify the center, vertices (endpoints of the major or longer axis) and covertices (endpoints of the minor or shorter axis).
- If the graph is a hyperbola, specify the center and vertices, and give the equations of the asymptotes.
- If the graph is a line or two lines, give their equations in the form $ax + by = c$.

Note: It is sometimes possible to "scale down" the rotated equation (in X and Y) and eliminate common factors. Before proceeding, check the Answer Key to see if your rotated equation is correct and simplified.

Furthermore, in Exercises 8 through 12, you will need to "complete the square" after changing variables in order to bring the quadratic into standard form.

1. $52x^2 - 72xy + 73y^2 = 900$
2. $2x^2 + 8xy + 8y^2 + 2\sqrt{5}x - \sqrt{5}y = 0$
3. $175x^2 + 1230xy - 481y^2 = 13600$
4. $181x^2 + 192xy + 261y^2 = 2925$
5. $32x^2 - 48xy + 18y^2 + 45x + 60y = 0$
6. $9x^2 - 30xy + 25y^2 + 9x - 15y = 10$
7. $6831x^2 - 6960xy - 71y^2 = 74529$
8. $16x^2 + 24xy + 9y^2 - 150x - 50y = -325$
9. $85x^2 - 96xy + 45y^2 + 124\sqrt{13}x - 48\sqrt{13}y = -559$
10. $96x^2 + 116xy + 9y^2 + 476\sqrt{5}x + 298\sqrt{5}y = -3445$
11. $801x^2 + 600xy + 1396y^2 - 8346x - 13000y = -34645$
12. $2375x^2 + 8880xy - 6431y^2 - 9230x - 21788y = 198406$

9.6 Positive Definite Quadratic Forms and Matrices

We have seen that there is a one-to-one correspondence between quadratic forms and symmetric matrices: every quadratic form Q in n variables can be written in terms of a symmetric $n \times n$ matrix $[Q] = A$, and conversely every symmetric $n \times n$ matrix A represents a quadratic form:

$$Q(x_1, x_2, \ldots, x_n) = Q(\vec{x}) = [\vec{x}]^\top [Q][\vec{x}] = [\vec{x}]^\top A[\vec{x}].$$

We also saw that due to their symmetry, every quadratic form can be ***diagonalized***, so that after a change of variables $[\vec{y}] = U[\vec{x}]$, where U is orthogonal, we can write:

$$Q(y_1, y_2, \ldots, y_n) = Q(\vec{y}) = \sum_{i=1}^{n} \lambda_i y_i^2,$$

where we can order the real eigenvalues of $[Q] : \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$. Now, since each eigenvalue is real, and a real number can be positive, negative, or zero, we can analogously classify each quadratic form using the following criteria:

Definition: A ***quadratic form*** Q can be classified into exactly one of the following types:

1. Q is ***positive definite*** if all of its eigenvalues are ***positive***.
2. Q is ***negative definite*** if all of its eigenvalues are ***negative***.
3. Q is ***positive semi-definite*** if all of its eigenvalues are ***positive or zero***.
4. Q is ***negative semi-definite*** if all of its eigenvalues are ***negative or zero***.
5. Q is ***indefinite*** if it has at least one positive eigenvalue and at least one negative eigenvalue. In other words, it is not one of the four types above.

Analogously, we say that a ***symmetric matrix*** A is ***positive definite*** if the associated quadratic form $[\vec{x}]^\top A[\vec{x}]$ is positive definite, and so on. We will refer to this as the ***definiteness type*** of a quadratic form Q or a symmetric matrix A.

Notice that under the definitions above, a positive definite matrix is also positive semi-definite. This is consistent with the convention that a positive number is also a non-negative number:

$$\text{If } x > 0, \quad \text{then } x \geq 0 \text{ as well.}$$

Thus, we use the ordinary inequality symbols as notation for these matrix types:

Type of Symmetric Matrix	***Symbol***
A is positive definite	$A > 0$
A is negative definite	$A < 0$
A is positive semi-definite	$A \geq 0$
A is negative semi-definite	$A \leq 0$

Analogously, we will write $Q > 0$, etc., for the corresponding quadratic form. It is worth noting that the definiteness type is equivalent to the nature of the ***values*** that the quadratic form assumes, further justifying the use of this notation:

Theorem: Let $Q(x_1, x_2, \ldots, x_n)$ be a quadratic form. Then:

1. $Q > 0$ ***if and only if*** for ***all non-zero*** $\vec{x} \in \mathbb{R}^n$: $[\vec{x}]^\top[Q][\vec{x}] > 0$.
2. $Q < 0$ ***if and only if*** for ***all non-zero*** $\vec{x} \in \mathbb{R}^n$: $[\vec{x}]^\top[Q][\vec{x}] < 0$.
3. $Q \geq 0$ ***if and only if*** for ***all*** $\vec{x} \in \mathbb{R}^n$: $[\vec{x}]^\top[Q][\vec{x}] \geq 0$.
4. $Q \leq 0$ ***if and only if*** for ***all*** $\vec{x} \in \mathbb{R}^n$: $[\vec{x}]^\top[Q][\vec{x}] \leq 0$.

Proof: Every quadratic form can be diagonalized so that under an orthogonal change of variables $[\vec{y}] = U[\vec{x}]$, we can write:

$$Q(y_1, y_2, \ldots, y_n) = \sum_{i=1}^{n} \lambda_i y_i^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2.$$

Each of the four equivalences now follow. For example, if $Q > 0$, then all the eigenvalues are positive. If $\vec{x}$ is not the zero-vector, then neither is $U[\vec{x}]$ since U is invertible. Thus at least one of the y_i is non-zero, hence $y_i^2 > 0$ and the value of Q is positive. Conversely, suppose for all non-zero $\vec{x} \in \mathbb{R}^n$: $[\vec{x}]^\top[Q][\vec{x}] > 0$. We must show that all of the eigenvalues are positive. So suppose that one of the eigenvalues were zero, say $\lambda_1 = 0$. Since U is invertible, we can find a non-zero vector $\vec{x}$ so that $U[\vec{x}] = \langle 1, 0, \ldots, 0 \rangle = \vec{y}$. But then $Q(\vec{y}) = \lambda_1 = 0$, yielding a contradiction. We can use the same idea to show that none of the eigenvalues can be negative either. Thus, all eigenvalues are positive. The other equivalences are proven using the same format.■

Closure Properties of Definite Matrices

This new symbol for symmetric matrices allows us to extend some (but not all) of the properties of order among the real numbers:

Theorem: Let A and B be symmetric $n \times n$ matrices. If $A > 0$ and $B > 0$, then $A + B > 0$.

Proof: Let $\vec{x} \in \mathbb{R}^n$ be any non-zero vector. Since A and B are both positive:

$$[\vec{x}]^\top A[\vec{x}] > 0 \quad \text{and} \quad [\vec{x}]^\top B[\vec{x}] > 0.$$

Thus:

$$[\vec{x}]^\top A[\vec{x}] + [\vec{x}]^\top B[\vec{x}] > 0, \quad \text{or}$$

$$[\vec{x}]^\top (A + B)[\vec{x}] > 0.$$

Therefore, $A + B$ is likewise positive.■

Unfortunately, the process of matrix multiplication is more sensitive than addition. However, scalar multiplication gives us an easy result:

Theorem: Let A and B be symmetric matrices. If $A > 0$, then for all $k \in \mathbb{R}$:

$$\text{If } k > 0, \text{ then } kA > 0, \text{ and}$$

$$\text{if } k < 0, \text{ then } kA < 0.$$

In particular, $-A < 0$. Analogously, if $B < 0$, then for all $k \in \mathbb{R}$:

$$\text{If } k > 0, \text{ then } kA < 0, \text{ and}$$

$$\text{if } k < 0, \text{ then } kA > 0.$$

In particular, $-B > 0$. Analogous statements can be made for both semi-definite types.

Proof: Using the same idea as the proof of the previous Theorem, for all $\vec{x} \in \mathbb{R}^{\boldsymbol{n}}$:

$$[\vec{x}\,]^\top A[\vec{x}\,] > 0 \text{ and thus if } k > 0, \text{ then:}$$

$$[\vec{x}\,]^\top (kA)[\vec{x}\,] > 0,$$

and so $kA > 0$. Similarly, if $k < 0$, $kA < 0$ also. A similar idea works if $B < 0$. ■

Sylvester's Criterion

Apart from explicitly computing all the eigenvalues of a symmetric matrix in order to determine its definiteness, the following criterion also gives us a way to determine if a symmetric matrix is positive definite by simply computing certain determinants:

Theorem — Sylvester's Criterion: Let A be an $n \times n$ ***symmetric*** matrix. For every k, define the submatrix:

$$A^{(k)} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}.$$

In other words, $A^{(k)}$ is the $k \times k$ submatrix extracted from the upper left hand corner of A. Then:

A is ***positive definite*** ***if and only if*** $det(A^{(k)}) > 0$ for all $k = 1 \ldots n$.

The proof of this Theorem is rather long and involved, so we will omit it; proofs are easily found on the Internet. We mention this Theorem only because we use it briefly to compute the Cholesky Decomposition later.

Matrices of the form $B^\top B$

We have seen before that if $A = B^\top B$, then $A^\top = B^\top(B^\top)^\top = A$, so A is symmetric. However, the following Theorem says that a matrix of this form can be determined with certainty to be of one of the types above:

> ***Theorem:*** Let B be an $m \times n$ matrix (not necessarily square), and let $A = B^\top B$.
> Then: $A \geq 0$. Furthermore, if $rank(B) = n$, then $A > 0$.
> Moreover, $rank(A) = rank(B)$, and $nullspace(A) = nullspace(B)$.

Proof: Since B is $m \times n$, $A = B^\top B$ is $n \times n$. Thus, let $\vec{x} \in \mathbb{R}^n$. Then:

$$\begin{aligned}[\vec{x}]^\top A[\vec{x}] &= [\vec{x}]^\top (B^\top B)[\vec{x}] = [\vec{x}]^\top B^\top B[\vec{x}] \\ &= (B[\vec{x}])^\top B[\vec{x}] = B[\vec{x}] \circ B[\vec{x}].\end{aligned}$$

Thus, by the non-degeneracy of the dot product, $B[\mathbf{x}] \circ B[\mathbf{x}] \geq 0$, so $A \geq 0$.
Now, suppose $rank(B) = k$, with $k \leq min(m,n)$. By the Dimension Theorem, $nullity(B) = n - k$. So suppose $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_{n-k}\}$ is a basis for $nullspace(B)$. Thus $B\vec{v}_i = \vec{0}_m$ for all i. But then $B^\top B\vec{v}_i = \vec{0}_n$, and so $nullspace(B) \subseteq nullspace(A)$. Thus $nullity(A) \geq n - k$. However, $B^\top B$ is $n \times n$, and so again by the Dimension Theorem, this means that $rank(B^\top B) \leq k$.
Now, by the Minimizing Theorem of Chapter 1, we can find a subset $C = \{\vec{c}_{i_1}, \vec{c}_{i_2}, \ldots, \vec{c}_{i_k}\}$ of k original columns from B that is a basis for $colspace(B)$. But C is also a basis for $rowspace(B^\top)$. However, we showed in the Section on the Significance of the Rowspace in Chapter 4 that a linear transformation T is ***one-to-one*** when it is restricted to the rowspace of $[T]$. Thus, the linear transformation represented by $B^\top$ is one-to-one on $Span(C)$, and so:

$$\{B^\top\vec{c}_{i_1}, B^\top\vec{c}_{i_2}, \ldots, B^\top\vec{c}_{i_k}\} \subseteq colspace(B^\top B) = colspace(A)$$

is still linearly independent. This proves that $rank(A) = k = rank(B)$, and so again by the Dimension Theorem, $nullity(A) = n - k$. Thus $nullspace(B) = nullspace(A)$. Note however that unless $m = n$, A and B have different number of rows, and so in general we cannot say that $colspace(B) = colspace(A)$, even though they have the same dimension.
Now, if $rank(B) = n$, then B represents a one-to-one linear transformation $T_B : \mathbb{R}^n \to \mathbb{R}^m$, and thus $B[\vec{x}] \neq \vec{0}_m$, so $B[\vec{x}] \circ B[\vec{x}] > 0$, whenever $\vec{x}$ is not $\vec{0}_n$. Thus $A > 0$ in this case.■

Example: Let $B = \begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix}$.

Notice that B is not symmetric, but it is invertible, so $rank(B) = 2$. However:

$$A = B^\top B = \begin{bmatrix} 3 & 1 \\ -2 & 0 \end{bmatrix}\begin{bmatrix} 3 & -2 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 10 & -6 \\ -6 & 4 \end{bmatrix}$$

is indeed symmetric. The characteristic polynomial of A is:

$$\begin{aligned}p(\lambda) &= (\lambda - 10)(\lambda - 4) - 36 \\ &= \lambda^2 - 14\lambda + 76\end{aligned}$$

so its eigenvalues are $7 + 3\sqrt{5} \approx 13.71$ and $7 - 3\sqrt{5} \approx 0.29$. Thus A is positive definite.□

The Cholesky Decomposition

There is an interesting and powerful converse to our previous Theorem:

> ***Theorem — The Cholesky Decomposition:*** Let A be a ***symmetric, positive definite*** $n \times n$ matrix. Then there exists a ***unique, lower triangular*** $n \times n$ matrix L such that the entries on the diagonal of L are ***positive***, and:
>
> $$A = L \cdot L^{\top} = U^{\top} \cdot U,$$
>
> where $U = L^{\top}$ is ***upper triangular***.

Proof: There are at least two ways to explicitly compute the Cholesky Decomposition of a symmetric, positive definite $n \times n$ matrix A. We will present the ***Cholesky-Crout Algorithm***, that computes L entry by entry, starting with $L_{1,1}$ and proceeding by columns to the right.

The idea is to simply assume the decomposition exists, and then show that we can uniquely solve for the entries. Let us illustrate with the 3×3 case. We want to satisfy the equation:

$$A = L \cdot L^{\top}, \text{ or:}$$

$$\begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{21} & A_{22} & A_{32} \\ A_{31} & A_{32} & A_{33} \end{bmatrix} = \begin{bmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{bmatrix} \begin{bmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{bmatrix}$$

$$= \begin{bmatrix} L_{11}^2 & L_{11}L_{21} & L_{11}L_{31} \\ L_{11}L_{21} & L_{21}^2 + L_{22}^2 & L_{21}L_{31} + L_{22}L_{32} \\ L_{11}L_{31} & L_{21}L_{31} + L_{22}L_{32} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{bmatrix}.$$

By Sylvester's Criterion, $A_{11} > 0$, so by comparing entries:

$$L_{11} = \sqrt{A_{11}}, \qquad \text{and we also get:}$$

$$L_{21} = A_{21}/L_{11} = A_{21}/\sqrt{A_{11}}, \text{ and}$$

$$L_{31} = A_{31}/L_{11} = A_{31}/\sqrt{A_{11}}.$$

Now, for the 2nd column, we must first solve:

$$\begin{aligned} A_{22} &= L_{21}^2 + L_{22}^2 \\ &= A_{21}^2/A_{11} + L_{22}^2, \text{ or} \end{aligned}$$

$$\begin{aligned} L_{22}^2 &= A_{22} - A_{21}^2/A_{11} \\ &= (A_{11}A_{22} - A_{21}^2)/A_{11}. \end{aligned}$$

But again, we notice that $A_{11}A_{22} - A_{21}^2$ is $det(A^{(2)})$, so it is again positive (we defined $A^{(2)}$ in Sylvester's Criterion). Since A_{11} is also positive, we now get:

$$L_{22} = \sqrt{det(A^{(2)})/A_{11}}.$$

Now we must solve for L_{32} from:

$$\begin{aligned} A_{32} &= L_{21}L_{31} + L_{22}L_{32}, \text{ or} \\ L_{32} &= (A_{32} - L_{21}L_{31})/L_{22} = (A_{32} - A_{21}A_{31}/A_{11})/L_{22}. \end{aligned}$$

Finally, we must solve for L_{33} from:

$$\begin{aligned} A_{33} &= L_{31}^2 + L_{32}^2 + L_{33}^2 \\ &= A_{31}^2/A_{11} + (A_{32} - A_{21}A_{31}/A_{11})^2/((A_{11}A_{22} - A_{21}^2)/A_{11}) + L_{33}^2 \\ &= A_{31}^2/A_{11} + (A_{11}A_{32} - A_{21}A_{31})^2/(A_{11}[A_{11}A_{22} - A_{21}^2]) + L_{33}^2, \text{ or} \\ L_{33}^2 &= A_{33} - A_{31}^2/A_{11} - (A_{11}A_{32} - A_{21}A_{31})^2/(A_{11}[A_{11}A_{22} - A_{21}^2]) \\ &= \frac{A_{11}A_{33}[A_{11}A_{22} - A_{21}^2] - A_{31}^2[A_{11}A_{22} - A_{21}^2] - (A_{11}A_{32} - A_{21}A_{31})^2}{A_{11}[A_{11}A_{22} - A_{21}^2]}. \end{aligned}$$

Expanding and then simplifying the numerator above yields:

$$\begin{aligned} &A_{11}(A_{11}A_{22}A_{33} - A_{21}^2A_{33} - A_{31}^2A_{22} - A_{11}A_{32}^2 + 2A_{32}A_{21}A_{31}) \\ &= A_{11}det(A), \end{aligned}$$

and so we get:

$$L_{33}^2 = det(A)/det(A^{(2)}).$$

In the $n \times n$ case, we have the general formulas:

$$L_{ii} = \sqrt{A_{ii} - \sum_{j=1}^{i-1} L_{i,j}^2} = \sqrt{\frac{det(A^{(i)})}{det(A^{(i-1)})}}, \text{ and}$$

$$L_{ij} = \frac{1}{L_{jj}}\left(A_{ij} - \sum_{k=1}^{j-1} L_{ik}L_{jk}\right),$$

where $A^{(0)} = 1$, and $A^{(i)}$ is defined in Sylvester's Criterion.■

Incidentally, a positive semi-definite matrix also has a Cholesky Decomposition, however, it is ***not*** unique, and its construction is via a limiting process (in other words, it is not explicit as we saw above). We can thus put together our two previous Theorems:

Theorem — Equivalence of Cholesky Decomposition:

Let A be a symmetric $n \times n$ matrix. Then:

A is ***positive definite*** ***if and only if*** there exists a ***unique, invertible, lower triangular*** $n \times n$ matrix L such that $A = L \cdot L^\top$.

A is ***positive semi-definite*** ***if and only if*** there exists a ***lower triangular*** $n \times n$ matrix L such that $A = L \cdot L^\top$.

Example: For the sake of brevity, let us do a 2×2 example. We saw that:

$$A = \begin{bmatrix} 10 & -6 \\ -6 & 4 \end{bmatrix}$$

is positive definite. According to the algorithm above:

$$L_{11} = \sqrt{A_{11}} = \sqrt{10},$$
$$L_{21} = \frac{A_{21}}{L_{11}} = \frac{-6}{\sqrt{10}}, \text{ and}$$
$$L_{22} = \sqrt{\frac{det(A)}{A_{11}}} = \sqrt{\frac{40-36}{10}} = \frac{2}{\sqrt{10}}.$$

Thus:

$$L = \frac{1}{\sqrt{10}}\begin{bmatrix} 10 & 0 \\ -6 & 2 \end{bmatrix}, \text{ and } L^{\top} = \frac{1}{\sqrt{10}}\begin{bmatrix} 10 & -6 \\ 0 & 2 \end{bmatrix}.$$

We easily check that:

$$\begin{aligned} LL^{\top} &= \frac{1}{10}\begin{bmatrix} 10 & 0 \\ -6 & 2 \end{bmatrix}\begin{bmatrix} 10 & -6 \\ 0 & 2 \end{bmatrix} \\ &= \frac{1}{10}\begin{bmatrix} 100 & -60 \\ -60 & 40 \end{bmatrix} = \begin{bmatrix} 10 & -6 \\ -6 & 4 \end{bmatrix} = A. \square \end{aligned}$$

Q Defines a Bilinear Form

We saw in Chapter 6 that an inner product $\langle \vec{u} | \vec{v} \rangle$ on a vector space V is a bilinear form, that is, it is symmetric, additive and homogeneous. Furthermore, we required that $\langle \vec{u} | \vec{u} \rangle > 0$ for all non-zero vectors $\vec{u}$. We now see that quadratic forms are related to these bilinear forms:

> ***Definition/Theorem:*** A ***quadratic form*** $Q(x_1, x_2, \ldots, x_n)$ defines a ***bilinear form*** on $\mathbb{R}^n$, that is, a real-valued function $\langle \vec{u} | \vec{v} \rangle_Q$, given by:
>
> $$\langle \vec{u} | \vec{v} \rangle_Q = [\vec{v}]^{\top}[Q][\vec{u}].$$
>
> This means that the function $\langle \vec{u} | \vec{v} \rangle_Q$ is ***symmetric, additive*** and ***homogeneous***.
> Q is ***positive definite*** ***if and only if*** $\langle \vec{u} | \vec{v} \rangle_Q$ defines an ***inner-product*** on $\mathbb{R}^n$.

Proof: To prove symmetry, we compute:

$$\begin{aligned} \langle \vec{v} | \vec{u} \rangle_Q &= [\vec{u}]^{\top}[Q][\vec{v}] \\ &= ([\vec{v}]^{\top}[Q]^{\top}[\vec{u}])^{\top} \\ &= ([\vec{v}]^{\top}[Q][\vec{u}])^{\top} \\ &= \langle \vec{u} | \vec{v} \rangle_Q^{\top} = \langle \vec{u} | \vec{v} \rangle_Q. \end{aligned}$$

using the symmetry of Q and the properties of the transpose operator. Note that we regarded the number $\langle \vec{u} | \vec{v} \rangle_Q$ as a 1×1 matrix in the final step, that is, a real number. Additivity and homogeneity will be left as straightforward Exercises. Finally, we know that $Q > 0$ ***if and only if*** $[\vec{u}]^{\top}[Q][\vec{u}] > 0$ for all non-zero $\vec{u}$, and this is equivalent to $\langle \vec{u} | \vec{u} \rangle_Q > 0$. ■

Example: In Chapter 7, we showed that if A is an ***invertible*** $n \times n$ matrix, then the bilinear form on $\mathbb{R}^{\boldsymbol{n}}$ defined by:

$$\langle \vec{u} | \vec{v} \rangle = A[\vec{u}\,] \circ A[\vec{v}\,]$$

is actually an ***inner-product*** on $\mathbb{R}^{\boldsymbol{n}}$. We can now prove this in a slightly different way. Using the matrix form of the dot product, we get:

$$\begin{aligned} A[\vec{u}\,] \circ A[\vec{v}\,] &= (A[\vec{v}\,])^{\top}(A[\vec{u}\,]) \\ &= [\vec{v}\,]^{\top} A^{\top} A[\vec{u}\,], \end{aligned}$$

and thus if we let $[Q] = A^{\top}A$, then $[Q]$ is ***positive definite*** since $rank(A) = n$, and thus $[Q]$ defines an inner product on $\mathbb{R}^{\boldsymbol{n}}$.$\square$

The Complex Case

We defined a complex quadratic form Q such that $[Q]$ is a ***Hermitian*** matrix. By the Spectral Theorem, we can still define its definite type in exactly the same way as the real case, and equivalently for Hermitian matrices, but substituting the adjoint for the transpose. Thus, for example, we can state:

> ***Theorem:*** The following are equivalent for a ***Hermitian*** matrix $[Q]$:
>
> 1. Q is a ***positive definite*** complex quadratic form.
> 2. All of the eigenvalues of $[Q]$ are ***positive***.
> 3. For ***all*** non-zero $\vec{z} \in \mathbb{C}^n$: $[\vec{z}\,]^*[Q][\vec{z}\,] > 0$.

We can write similar equivalences for the other three definiteness types.
The closure properties and Sylvester's Criterion are still valid. We can rephrase the Cholesky Decomposition as:

> ***Theorem — Cholesky Decomposition:*** Let A be a ***Hermitian, positive definite*** $n \times n$ matrix. Then there exists a ***unique, lower triangular*** $n \times n$ matrix L such that the entries on the diagonal of L are ***positive***, and:
>
> $$A = L \bullet L^* = U^* \bullet U,$$
>
> where $U = L^*$ is ***upper triangular***.

Applications

Positive semi-definite matrices appear as ***covariance matrices*** in Statistics. A random variable $X \in \mathbb{R}$ has a mean value or ***expected value***, denoted by μ:

$$\mu = E(X).$$

If X is a discrete random variable (for example, the number of dots appearing in a throw of a die), then μ is simply the sum of the outcomes divided by the number (or count) of the outcomes (i.e. how many outcomes are possible). For a continuous random variable, its mean is defined to be the integral of its density function with respect to its probability measure. The ***variance*** σ^2 of a random variable is defined by:

$$\sigma^2 = var(X) = E\big((X-\mu)^2\big).$$

The covariance matrix generalizes the concept of a variance when we deal with several random variables at the same time, say $X_1, X_2, \ldots, X_n \in \mathbb{R}$. We assemble them in a ***column matrix:***

$$X = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}.$$

We can define the expected value $E(X)$ of X simply to be the column matrix whose entries are $E(X_i)$. From this, we can define the covariance matrix:

$$\Sigma = cov(X) = E((X-E(X))(X-E(X))^\top).$$

Because Σ has the form $B^\top B$, we know that it is symmetric and positive semi-definite. Conversely, it can be shown that any positive semi-definite matrix is the covariance matrix for some column matrix X consisting of random variables.

The entry $\Sigma_{i,j}$ is a measure of the linear coupling between the random variables X_i and X_j. But because Σ can be diagonalized, there is a change of variables $Y = U^\top X$ such that:

$$\Sigma' = cov(Y)$$

is a diagonal matrix. This means that two distinct random variables Y_i and Y_j are ***uncorrelated***, or essentially independent of each other.

We are sometimes interested in computing the ***square root*** of a matrix A. There are, however, at least two possible conventions or definitions for this. The first one is the generalization of the square root of a number: B is a square root of A if:

$$A = B^2.$$

Unfortunately, such a square root need not be unique, in the same way that there are two real numbers x that solve the equation $4 = x^2$. But if A happens to be ***positive semi-definite***, then there is a ***unique*** positive semi-definite matrix B such that $A = B^2$. This is analogous to the unique positive square root of a positive real number. It is easily computed using our diagonalization:

$$A = CDC^{-1},$$

where $D = Diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$, and every $\lambda_i \geq 0$. Thus we can define:

$$\sqrt{D} = D^{1/2} = Diag\big(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n}\big),$$

and from this, we can construct:

$$B = \sqrt{A} = C\sqrt{D}\,C^{-1}.$$

We easily see that $B^2 = A$ and B is positive semi-definite as well.

The second definition of a square root involves the adjoint or transpose: B is a square root of A if $A = B^*B$. The ***Cholesky Decomposition*** $A = L \bullet L^* = U^* \bullet U$ obviously yields a square root of A under this definition.

9.6 Exercises

For Exercises (1) to (9): Determine the definiteness type of the quadratic form Q in the corresponding Exercise in Section 9.4.

For Exercises (10) to (16): Determine the definiteness type of the quadratic form Q in Exercises 14 through 20, respectively, in Section 9.4.

For Exercises (17) to (23): Use the Cholesky-Crout Algorithm to find the Cholesky Decomposition $A = L \cdot L^\top$ (or $L \cdot L^*$, in the complex case) of the following positive definite matrices A:

17. $\begin{bmatrix} 3 & -5 \\ -5 & 12 \end{bmatrix}$

18. $\begin{bmatrix} 3 & 4 \\ 4 & 7 \end{bmatrix}$

19. $\begin{bmatrix} 4 & 1+3i \\ 1-3i & 7 \end{bmatrix}$

20. $\begin{bmatrix} 5 & 3-2i \\ 3+2i & 4 \end{bmatrix}$

21. $\begin{bmatrix} 5 & 0 & 1 \\ 0 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix}$

22. $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

23. $\begin{bmatrix} 5 & 0 & 1 \\ 0 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix}$

For Exercises (24) to (30): Find $B = \sqrt{A}$ for the following positive semi-definite matrices A, that is, find a positive semi-definite matrix B such that $A = B^2$.

24. $\begin{bmatrix} 9 & -3 \\ -3 & 1 \end{bmatrix}$

25. $\begin{bmatrix} 4 & 2 \\ 2 & 1 \end{bmatrix}$

26. $\begin{bmatrix} 10 & 6 \\ 6 & 5 \end{bmatrix}$

27. $\begin{bmatrix} 6 & 2 \\ 2 & 3 \end{bmatrix}$

28. $\begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}$

29. $\begin{bmatrix} 5 & 1 & 1 \\ 1 & 5 & 1 \\ 1 & 1 & 5 \end{bmatrix}$

30. $\begin{bmatrix} 5 & 0 & 1 \\ 0 & 5 & 0 \\ 1 & 0 & 5 \end{bmatrix}$

31. Let $Q(x_1, x_2, \ldots, x_n)$ be a quadratic form in n variables. Show that the bilinear form:

$$\langle \vec{u} | \vec{v} \rangle_Q = [\vec{v}]^\top [Q][\vec{u}]$$

is both ***additive*** and ***homogeneous***.

32. Use induction to prove the formula for the entries $L_{i,j}$ of L in the Cholesky-Crout algorithm for general $n \times n$ positive-definite matrices.

33. Prove that if A is a positive semi-definite matrix, then there exists a ***unique*** positive semi-definite matrix B such that $A = B^2$. Hint: use the ideas at the end of the Section.

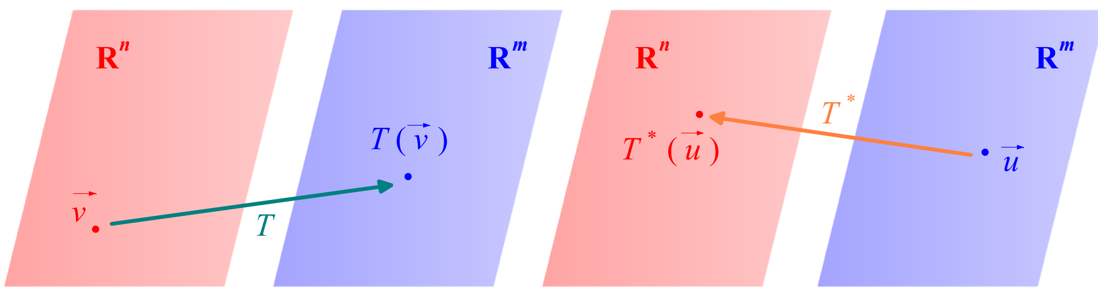
9.7 The Fundamental Theorem of Linear Algebra

We have arrived at the climax of our journey. The Fundamental Theorem of Linear Algebra neatly ties together several key concepts: the rowspace, columnspace and nullspace of any matrix A, as well as its transpose, and the concepts of eigenvectors, isomorphism, orthogonality, and definiteness.

Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is ***any*** linear transformation, with $m \times n$ standard matrix $[T] = A$. We know that the $n \times m$ transpose matrix $A^\top$ represents another linear transformation, that we will call T^*, the ***adjoint transformation of T:***

$$T^* : \mathbb{R}^m \to \mathbb{R}^n, \quad \text{with} \quad [T^*] = A^\top.$$

It is therefore natural to ask: is there a ***geometric connection*** between T and T^*?



T and its Adjoint T^*

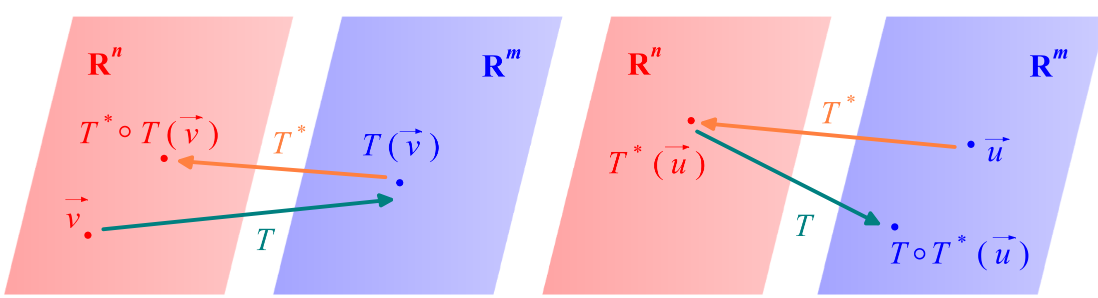
The key is to ***compose*** the two transformations in both orders:

$$T^* \circ T : \mathbb{R}^n \to \mathbb{R}^n, \quad \text{and}$$
$$T \circ T^* : \mathbb{R}^m \to \mathbb{R}^m,$$

with standard matrices $A^\top A$ and $AA^\top$, respectively. We have seen that:

$$(A^\top A)^\top = A^\top (A^\top)^\top = A^\top A.$$

Thus $T^* \circ T$, and analogously $T \circ T^*$, are both ***symmetric***. Note that both of these are ***operators***, but they may be operating on ***different*** Euclidean spaces, in general, if $n \neq m$. Moreover, there is absolutely no reason for $T^* \circ T$ or $T \circ T^*$ to be the identity operator, and so after we complete each round trip, in general we will end up at a ***different*** vector from where we started:



The Compositions $T^* \circ T$ and $T \circ T^*$

But since $A^\top A$ and $AA^\top$ are ***positive semi-definite symmetric matrices***, every eigenvalue λ of either product is either ***zero*** (corresponding to the ***nullspace***) or ***positive***.

Now, suppose that $\vec{v} \in \mathbb{R}^n$ is an eigenvector for $T^* \circ T$ with eigenvalue $\lambda \in \mathbb{R}$, that is:

$$(T^* \circ T)(\vec{v}) = \lambda\vec{v}, \quad \text{or as a matrix product:} \quad (A^\top A)[\vec{v}] = \lambda[\vec{v}].$$

Let us next multiply both sides on the left by A and apply the associative properties of matrix and scalar multiplication, to get:

$$A((A^\top A)[\vec{v}]) = A(\lambda[\vec{v}]), \quad \text{and by regrouping:}$$

$$(AA^\top)(A[\vec{v}]) = \lambda(A[\vec{v}]).$$

This equation seems to be saying that λ is also an eigenvalue for $AA^\top$ with eigenvector $A[\vec{v}]$, but it is possible that $A[\vec{v}] = \vec{\mathbf{0}}_m$. If so, then $\vec{v} \in \mathit{nullspace}(A)$. But then $(A^\top A)[\vec{v}] = \vec{\mathbf{0}}_n$, and since $\vec{v}$ is a non-zero vector, this means that $\lambda = 0$.

On the other hand, suppose $\lambda > 0$. Since $\vec{v} \neq \vec{\mathbf{0}}_n$, $\lambda[\vec{v}] \neq \vec{\mathbf{0}}_n$ either, by the ***Zero-Factors Theorem***. But since $(A^\top A)[\vec{v}] = \lambda[\vec{v}]$, this tells us that $A[\vec{v}]$ cannot be $\vec{\mathbf{0}}_m$, otherwise:

$$A^\top(A[\vec{v}]) = A^\top\vec{\mathbf{0}}_m = \vec{\mathbf{0}}_n,$$

and so $\lambda[\vec{v}] = \vec{\mathbf{0}}_n$, a contradiction. Thus, if $\lambda > 0$, we are guaranteed that λ is indeed an ***eigenvalue*** for $AA^\top$, and $A[\vec{v}]$ is an ***eigenvector*** for $AA^\top$ corresponding to λ.

Similarly, if $\vec{u} \in \mathbb{R}^m$ is an eigenvector for $T \circ T^*$ with eigenvalue $\lambda > 0$, then λ is also an eigenvalue for $A^\top A$, and $A^\top[\vec{u}] \in \mathbb{R}^n$ is a eigenvector for $A^\top A$ corresponding to λ. Thus, the ***positive eigenvalues*** of $A^\top A$ are ***exactly the same*** as those of $AA^\top$.

Now, let us put it all together: suppose that $k = \mathit{rank}(A) = \mathit{rank}(A^\top) = \mathit{rank}(A^\top A)$, and $\lambda_1 > \lambda_2 > \cdots > \lambda_j$ are the distinct ***positive*** eigenvalues of $A^\top A$ and $AA^\top$. By the Spectral Theorem, there exist orthonormal bases for $\mathbb{R}^n$ and $\mathbb{R}^m$:

$$B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k, \vec{v}_{k+1}, \ldots, \vec{v}_n\} \subseteq \mathbb{R}^n, \quad \text{and}$$

$$B' = \{\vec{u}_1, \vec{u}_2, .., \vec{u}_k, \vec{u}_{k+1}, \ldots, \vec{u}_m\} \subseteq \mathbb{R}^m,$$

consisting of eigenvectors for $A^\top A$ and $AA^\top$ respectively. We may assume that the eigenvectors are ordered so that $\{\vec{v}_{k+1}, \ldots, \vec{v}_n\}$ is a basis for $\mathit{nullspace}(A)$, if $k < n$, and similarly $\{\vec{u}_{k+1}, \ldots, \vec{u}_m\}$ is a basis for $\mathit{nullspace}(A^\top)$, if $k < m$. But recall that the orthogonal complement of the nullspace of a matrix is its rowspace, and thus $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_k\}$ is a basis for $\mathit{rowspace}(A)$, and $\{\vec{u}_1, \vec{u}_2, .., \vec{u}_k\}$ is a basis for $\mathit{rowspace}(A^\top) = \mathit{colspace}(A)$. Thus, ***all four fundamental matrix spaces*** of A are involved. For this reason, our result is known as the following:

Theorem — The Fundamental Theorem of Linear Algebra:

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with $m \times n$ standard matrix $[T] = A$. The $n \times m$ transpose matrix $A^\top$ represents the ***adjoint transformation*** $T^* : \mathbb{R}^m \to \mathbb{R}^n$. Then: both compositions:

$$T^* \circ T : \mathbb{R}^n \to \mathbb{R}^n, \quad \text{and} \quad T \circ T^* : \mathbb{R}^m \to \mathbb{R}^m,$$

are ***symmetric*** and ***positive semi-definite*** operators with standard matrices $A^\top A$ and $AA^\top$, respectively. Thus, $\lambda \geq 0$ for all of their eigenvalues.

Moreover, all the ***positive*** eigenvalues of $A^{\top}A$ and $AA^{\top}$ are exactly the ***same***. Suppose we list these distinct eigenvalues in decreasing order as $\lambda_1 > \lambda_2 > \cdots > \lambda_k$, and let:

$$V_i = Eig(A^{\top}A, \lambda_i) \quad \text{and} \quad U_i = Eig(AA^{\top}, \lambda_i),$$

for all $i = 1 \ldots k$. Then, for all $i \neq j$, $i, j = 1 \ldots k$:

$$V_i \cap V_j = \vec{0}_n \quad \text{and} \quad U_i \cap U_j = \vec{0}_m.$$

Furthermore, for every eigenvalue $\lambda_i > 0$:

$$T : V_i \to U_i, \quad \text{and} \quad T^* : U_i \to V_i$$

are ***isomorphisms of eigenspaces*** (although they are not necessarily inverses of each other).

If $rank(T) = r = rank(T^*)$, there exist ***orthogonal bases:***

$$B = \{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r, \vec{v}_{r+1}, \ldots, \vec{v}_n\}, \text{ and}$$
$$B' = \{\vec{u}_1, \vec{u}_2, .., \vec{u}_r, \vec{u}_{r+1}, \ldots, \vec{u}_m\},$$

for $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, consisting of eigenvectors for $A^{\top}A$ and $AA^{\top}$ respectively. The first r vectors in B and B' are eigenvectors corresponding to the ***positive*** eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$, in that order, counting multiplicities.

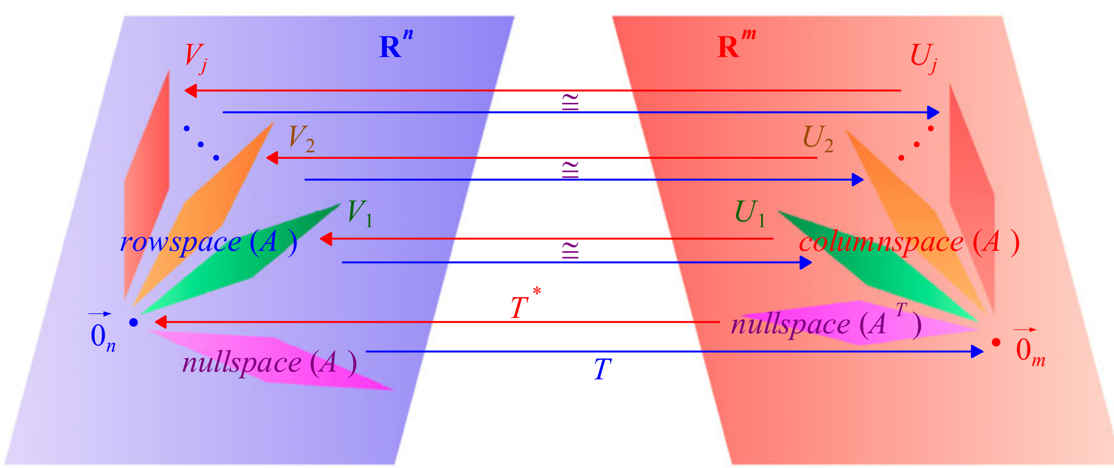
The set $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}$ is a ***basis*** for $rowspace(A)$, the set $\{\vec{u}_1, \vec{u}_2, .., \vec{u}_r\}$ is a ***basis*** for $colspace(A)$, and:

$$T : Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\}) \to Span(\{\vec{u}_1, \vec{u}_2, .., \vec{u}_r\}), \text{ and}$$
$$T^* : Span(\{\vec{u}_1, \vec{u}_2, .., \vec{u}_r\}) \to Span(\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_r\})$$

are ***isomorphisms*** between the rowspace and the columnspace (again, not necessarily inverses of each other).

By the Dimension Theorem: $nullity(T) = n - r$, and $nullity(T^*) = m - r$.

If $n > r$, then $\{\vec{v}_{r+1}, \ldots, \vec{v}_n\}$ is a ***basis*** for $ker(T) = nullspace(A)$, otherwise T is one-to-one. Similarly, if $m > r$, then $\{\vec{u}_{r+1}, \ldots, \vec{u}_m\}$ is a ***basis*** for $ker(T^*) = nullspace(A^{\top})$, otherwise T^* is one-to-one.



The Fundamental Theorem of Linear Algebra

We remark that in the language of Chapter 4, the Fundamental Theorem says that the rowspace and columnspace can be expressed as ***direct sums***:

$$rowspace(A) = V_1 \oplus V_2 \oplus \cdots \oplus V_k, \quad \text{and}$$
$$colspace(A) = U_1 \oplus U_2 \oplus \cdots \oplus U_k.$$

This means that every $\vec{v} \in rowspace(A)$ can be expressed ***uniquely*** as a sum $\vec{v} = \vec{v}_1 + \vec{v}_2 + \cdots + \vec{v}_k$, where every $\vec{v}_i \in V_i = Eig(A^\top A, \lambda_i)$, and analogously for the members of $colspace(A)$.

Example: Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ with standard matrix:

$$[T] = A = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}.$$

The two rows are linearly independent, so $rank(T) = 2$, and thus $nullity(T) = 1$. On the other hand, $nullity(T^*) = 0$, so T^* is one-to-one. Let us compute the products:

$$AA^\top = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 2 & 6 \end{bmatrix}, \quad \text{and}$$

$$A^\top A = \begin{bmatrix} -1 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 5 & 3 \\ 0 & 3 & 2 \end{bmatrix}.$$

Both are, as expected, ***symmetric*** matrices. Their characteristic polynomials are:

$$p_1(\lambda) = \lambda^2 - 9\lambda + 14 = (\lambda - 2)(\lambda - 7), \text{ and}$$
$$p_2(\lambda) = \lambda^3 - 9\lambda^2 + 14\lambda = \lambda(\lambda - 2)(\lambda - 7).$$

As predicted, they have common positive eigenvalues of $\lambda = 2$ and 7.

The eigenspaces for $AA^\top$ are:

$$Eig(AA^\top, 2) = Span(\{\langle 2, -1\rangle\}), \quad \text{and}$$
$$Eig(AA^\top, 7) = Span(\{\langle 1, 2\rangle\}).$$

The eigenspaces for $A^\top A$ are:

$$Eig(A^\top A, 2) = Span(\{\langle -3, 0, 1\rangle\}),$$
$$Eig(A^\top A, 7) = Span(\{\langle 1, 5, 3\rangle\}), \quad \text{and}$$
$$Eig(A^\top A, 0) = Span(\{\langle 1, -2, 3\rangle\}).$$

A quick check of dot products will verify that:

$$B = \{\langle 2, -1\rangle, \langle 1, 2\rangle\}, \quad \text{and}$$
$$B' = \{\langle -3, 0, 1\rangle, \langle 1, 5, 3\rangle, \langle 1, -2, 3\rangle\}$$

are orthogonal bases for $\mathbb{R}^2$ and $\mathbb{R}^3$, respectively. We can check that:

$$T(\langle -3, 0, 1\rangle) = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}\begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} = \langle 4, -2\rangle = 2\langle -2, 1\rangle,$$

$$T(\langle 1, 5, 3\rangle) = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix} = \langle 7, 14\rangle = 7\langle 1, 2\rangle, \quad \text{and}$$

$$T(\langle 1, -2, 3\rangle) = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix} = \langle 0, 0\rangle.$$

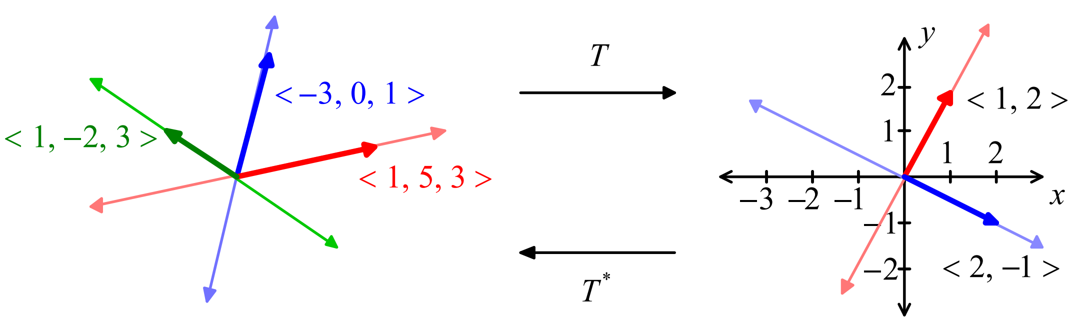
Similarly, we see that:

$$T^*(\langle 2, -1\rangle) = \begin{bmatrix} -1 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 2 \\ -1 \end{bmatrix} = \langle -3, 0, 1\rangle, \quad \text{and}$$

$$T^*(\langle 1, 2\rangle) = \begin{bmatrix} -1 & 1 \\ 1 & 2 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \end{bmatrix} = \langle 1, 5, 3\rangle.$$

Geometrically, this means that the lines $Span(\{\langle -3, 0, 1\rangle\})$ and $Span(\{\langle 2, -1\rangle\})$ are mapped into each other, and every round-trip stretches the original vector by $\lambda = 2$:

$$\langle -3, 0, 1\rangle \overset{T}{\to} \langle 4, -2\rangle \overset{T^*}{\to} 2\langle -3, 0, 1\rangle \overset{T}{\to} 2\langle 4, -2\rangle \overset{T^*}{\to} \cdots$$



The Eigenspaces of $T^* \circ T$ in $\mathbb{R}^3$ and of $T \circ T^*$ in $\mathbb{R}^2$

Similarly, the lines $Span(\{\langle 1, 5, 3\rangle\})$ and $Span(\{\langle 1, 2\rangle\})$ are also mapped into each other, with every round-trip stretching the original vector by $\lambda = 7$:

$$\langle 1, 5, 3\rangle \overset{T}{\to} \langle 7, 14\rangle \overset{T^*}{\to} 7\langle 1, 5, 3\rangle \overset{T}{\to} 7\langle 7, 14\rangle \overset{T^*}{\to} \cdots$$

The nullspace $Span(\{\langle 1, -2, 3\rangle\})$ is sent by T to $\vec{\mathbf{0}}_2$, whereas T^* is one-to-one. $\square$

The Operator Case

If $T : \mathbb{R}^n \to \mathbb{R}^n$ is an operator, then the eigenspaces of $T^* \circ T$ and $T \circ T^*$ are both in $\mathbb{R}^n$. However, there is no guarantee that they will be the same subspaces of $\mathbb{R}^n$ for the same λ. In general, they are not.

Example: Let $T : \mathbb{R}^2 \to \mathbb{R}^2$ with standard matrix:

$$[T] = A = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}.$$

Then we have:

$$A^\top A = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} = \begin{bmatrix} 13 & 8 \\ 8 & 5 \end{bmatrix}, \text{ and}$$

$$AA^\top = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}\begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 5 & 8 \\ 8 & 13 \end{bmatrix}.$$

They share the characteristic polynomial:

$$p_1(\lambda) = \lambda^2 - 18\lambda + 1 = p_2(\lambda).$$

Thus, they are both invertible, and have common positive eigenvalues of $\lambda = 9 \pm 4\sqrt{5}$. Surprisingly, the eigenvectors of these two matrices involve the Golden Ratio $\gamma = \left(1 + \sqrt{5}\right)/2 \approx 1.618$ and its conjugate $\overline{\gamma} = \left(1 - \sqrt{5}\right)/2 \approx -0.618$. We saw these numbers when we studied the Fibonacci Numbers and Recurrence Relations. They satisfy the equations:

$$\gamma + \overline{\gamma} = 1, \;\; \gamma\overline{\gamma} = -1, \;\text{ and}$$
$$\gamma^2 - \gamma - 1 = 0 = \overline{\gamma}^2 - \overline{\gamma} - 1.$$

We can rewrite our eigenvalues as:

$$9 \pm 4\sqrt{5} = 9 + 8\left(\frac{1 \pm \sqrt{5}}{2}\right) - 4 = 8\gamma + 5 \approx 17.94 \;\text{ and }\; 8\overline{\gamma} + 5 \approx 0.06$$

Now for their eigenspaces:

$$Eig(A^\top A, 8\gamma + 5) = Span(\{\langle \gamma, 1 \rangle\}),$$
$$Eig(A^\top A, 8\overline{\gamma} + 5) = Span(\{\langle \overline{\gamma}, 1 \rangle\}),$$
$$Eig(AA^\top, 8\gamma + 5) = Span(\{\langle 1, \gamma \rangle\}), \;\text{ and}$$
$$Eig(AA^\top, 8\overline{\gamma} + 5) = Span(\{\langle -1, -\overline{\gamma} \rangle\}).$$

We remark that indeed we could have chosen $\langle 1, \overline{\gamma} \rangle$ as a basis for the last eigenspace, but we chose $\langle -1, -\overline{\gamma} \rangle$ for reasons that will become clearer when we discuss the Singular Value Decomposition or SVD.

Now, we can check that:

$$\langle \gamma, 1 \rangle \circ \langle \overline{\gamma}, 1 \rangle = \gamma\overline{\gamma} + 1 = -1 + 1 = 0 = \langle 1, \gamma \rangle \circ \langle -1, -\overline{\gamma} \rangle,$$

so the pairs of eigenspaces are indeed orthogonal. Now we can see where each eigenvector goes to under the appropriate operator:

$$T(\langle \gamma, 1 \rangle) = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} \gamma \\ 1 \end{bmatrix} = \begin{bmatrix} 2\gamma + 1 \\ 3\gamma + 2 \end{bmatrix} = (2\gamma + 1)\langle 1, \gamma \rangle,$$

$$T(\langle \overline{\gamma}, 1 \rangle) = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \begin{bmatrix} \overline{\gamma} \\ 1 \end{bmatrix} = \begin{bmatrix} 2\overline{\gamma} + 1 \\ 3\overline{\gamma} + 2 \end{bmatrix} = (-2\overline{\gamma} - 1)\langle -1, -\overline{\gamma} \rangle,$$

$$T^*(\langle 1, \gamma \rangle) = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ \gamma \end{bmatrix} = \begin{bmatrix} 3\gamma + 2 \\ 2\gamma + 1 \end{bmatrix} = (2\gamma + 1)\langle \gamma, 1 \rangle,$$

$$T^*(\langle -1, -\overline{\gamma} \rangle) = \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ -\overline{\gamma} \end{bmatrix} = \begin{bmatrix} -3\overline{\gamma} - 2 \\ -2\overline{\gamma} - 1 \end{bmatrix} = (-2\overline{\gamma} - 1)\langle \overline{\gamma}, 1 \rangle.$$

Observe that:

$$(2\gamma + 1)^2 = 4\gamma^2 + 4\gamma + 1 = 4(\gamma + 1) + 4\gamma + 1 = 8\gamma + 5 = \lambda_1,$$

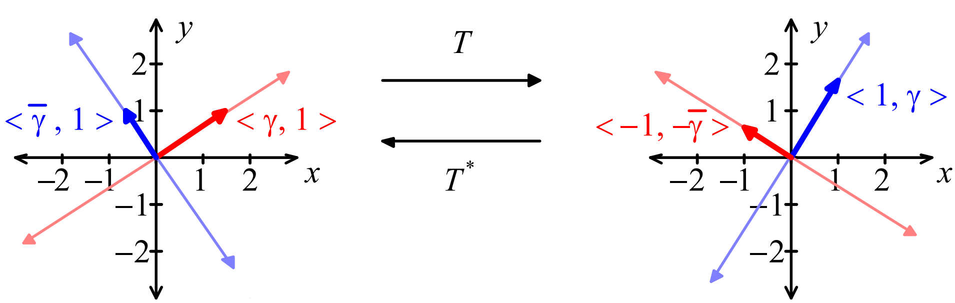
and similarly, $(-2\overline{\gamma} - 1)^2 = 8\overline{\gamma} + 5 = \lambda_2$. Furthermore, both $2\gamma + 1 \approx 4.24$ and $-2\overline{\gamma} - 1 \approx 0.24$ are ***positive***. This will be relevant when we discuss the SVD. Thus we get:

$$\langle \gamma, 1 \rangle \overset{T}{\to} (2\gamma + 1)\langle 1, \gamma \rangle \overset{T^*}{\to} (8\gamma + 5)\langle \gamma, 1 \rangle \overset{T}{\to} (34\gamma + 21)\langle 1, \gamma \rangle \overset{T^*}{\to} \cdots$$

and similarly:

$$\langle \overline{\gamma}, 1 \rangle \overset{T^*}{\to} (2\overline{\gamma} + 1)\langle 1, \overline{\gamma} \rangle \overset{T}{\to} (8\overline{\gamma} + 5)\langle \overline{\gamma}, 1 \rangle \overset{T^*}{\to} (34\overline{\gamma} + 21)\langle 1, \overline{\gamma} \rangle \overset{T}{\to} \cdots$$

It should not be a surprise that ***Fibonacci Numbers*** are starting to appear as coefficients. Finally, we show the two eigenspaces below:



The Eigenspaces of $T^* \circ T$ and $T \circ T^*$. □

The Complex Case

Our previous Theorem generalizes naturally when $T : \mathbb{C}^n \to \mathbb{C}^m$. In this case, $[T] = A$ may have imaginary entries, and so we form $A^* = \overline{A^\top}$, the ***Hermitian adjoint*** of A. As seen in Chapter 8, the matrices A^*A and AA^* are ***Hermitian***, and thus their eigenvalues are once again pure real numbers. By our theory on Quadratic Forms, both matrices are again positive semi-definite, and so all eigenvalues are positive or zero. However, we now obtain orthonormal bases for $\mathbb{C}^n$ and $\mathbb{C}^m$ under the usual complex inner-product $\langle \vec{u} | \vec{v} \rangle = [\boldsymbol{v}]^* \cdot [\boldsymbol{u}]$.

Once again, A^*A and AA^* have exactly the same positive eigenvalues, and $T : Eig(A^*A, \lambda) \to Eig(AA^*, \lambda)$ and $T^* : Eig(AA^*, \lambda) \to Eig(A^*A, \lambda)$ are ***isomorphisms*** if $\lambda > 0$.

Example: Let $T : \mathbb{C}^2 \to \mathbb{C}^3$ with standard matrix:

$$[T] = A = \begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix}.$$

The columns are obviously linearly independent, and so $rank(T) = 2$. Thus $nullity(T) = 0$, and $nullity(T^*) = 1$.

The adjoint operator $T^* : \mathbb{C}^3 \to \mathbb{C}^2$ has standard matrix:

$$[T^*] = A^* = \begin{bmatrix} 1 & 0 & -i \\ -i & -1 & 0 \end{bmatrix}.$$

(recall that we take the ***conjugate*** of the transpose). Let us form the two products:

$$A^*A = \begin{bmatrix} 1 & 0 & -i \\ -i & -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix} = \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix}, \text{ and}$$

$$AA^* = \begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & -i \\ -i & -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & -i & -i \\ i & 1 & 0 \\ i & 0 & 1 \end{bmatrix}$$

Observe that these matrix products are not symmetric, but they are ***Hermitian***. The characteristic polynomials of these two matrix products are:

$$\begin{aligned} charpoly(A^*A, \lambda) &= det\left(\begin{bmatrix} \lambda - 2 & -i \\ i & \lambda - 2 \end{bmatrix}\right) \\ &= (\lambda - 2)^2 + i^2 \\ &= \lambda^2 - 4\lambda + 3 = (\lambda - 1)(\lambda - 3), \text{ and} \end{aligned}$$

$$
\begin{aligned}
charpoly(AA^*, \lambda) &= det\left(\begin{bmatrix} \lambda - 2 & i & i \\ -i & \lambda - 1 & 0 \\ -i & 0 & \lambda - 1 \end{bmatrix}\right) \\
&= (\lambda - 2)(\lambda - 1)^2 + 0 + 0 + i^2(\lambda - 1) + 0 + i^2(\lambda - 1) \\
&= (\lambda - 2)(\lambda - 1)^2 - 2(\lambda - 1) \\
&= (\lambda - 1)[(\lambda - 2)(\lambda - 1) - 2] \\
&= (\lambda - 1)(\lambda^2 - 3\lambda) = \lambda(\lambda - 1)(\lambda - 3).
\end{aligned}
$$

Thus, the eigenvalues are pure real and non-negative as expected, and both compositions share the common eigenvalues of $\lambda = 1$ and $\lambda = 3$. We can find the eigenspaces for AA^*:

$$
\begin{aligned}
Eig(AA^*, 1) &= Span(\{\langle 0, -1, 1\rangle\}), \\
Eig(AA^*, 3) &= Span(\{\langle 2, i, i\rangle\}), \text{ and} \\
Eig(AA^*, 0) &= Span(\{\langle -1, i, i\rangle\}).
\end{aligned}
$$

We remark that it is only by coincidence that we are able to find an eigenvector for $\lambda = 1$ with pure real components. Similarly the eigenspaces for A^*A are:

$$
\begin{aligned}
Eig(A^*A, 1) &= Span(\{\langle -i, 1\rangle\}), \text{ and} \\
Eig(A^*A, 3) &= Span(\{\langle 1, -i\rangle\}).
\end{aligned}
$$

We can easily check that:

$$
\begin{aligned}
B &= \{\langle 0, -1, 1\rangle, \langle 2, i, i\rangle, \langle -1, i, i\rangle\}, \text{ and} \\
B^{/} &= \{\langle -i, 1\rangle, \langle 1, -i\rangle\}
\end{aligned}
$$

form ***orthogonal bases***, respectively, for $\mathbb{C}^3$ and $\mathbb{C}^2$ (recall that we need to apply the ***conjugate*** to the 2nd vector), so for example:

$$
\langle\langle 2, i, i\rangle | \langle -1, i, i\rangle\rangle = \langle 2, i, i\rangle \circ \overline{\langle -1, i, i\rangle} = \langle 2, i, i\rangle \circ \langle -1, -i, -i\rangle = -2 - i^2 - i^2 = 0.
$$

Similarly, we can verify that:

$$
T^*(\langle 0, -1, 1\rangle) = \begin{bmatrix} 1 & 0 & -i \\ -i & -1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ -1 \\ 1 \end{bmatrix} = \langle -i, 1\rangle,
$$

$$
T^*(\langle 2, i, i\rangle) = \begin{bmatrix} 1 & 0 & -i \\ -i & -1 & 0 \end{bmatrix}\begin{bmatrix} 2 \\ i \\ i \end{bmatrix} = \langle 3, -3i\rangle = 3\langle 1, -i\rangle, \text{ and}
$$

$$
T^*(\langle -1, i, i\rangle) = \begin{bmatrix} 1 & 0 & -i \\ -i & -1 & 0 \end{bmatrix}\begin{bmatrix} -1 \\ i \\ i \end{bmatrix} = \langle 0, 0\rangle.
$$

Thus, they are indeed sent to the corresponding eigenspaces when $\lambda > 0$. Likewise:

$$T(\langle -i, 1\rangle) = \begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix}\begin{bmatrix} -i \\ 1 \end{bmatrix} = \langle 0, -1, 1\rangle, \text{ and}$$

$$T(\langle 1, -i\rangle) = \begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix}\begin{bmatrix} 1 \\ -i \end{bmatrix} = \langle 2, i, i\rangle.$$

We can thus see that $Eig(AA^*, 1)$ and $Eig(A^*A, 1)$ are mapped into each other, and similarly that $Eig(AA^*, 3)$ and $Eig(A^*A, 3)$ are mapped into each other.

Unfortunately, we cannot directly visualize the action of complex transformations. □

9.7 Exercises

For the following matrices A, the standard matrix for some linear transformation T: (a) Find a basis of each eigenspace of $A^\top A$ and $AA^\top$ (or A^*A and AA^*, in the complex case); (b) Show explicitly that the image under T of each eigenvector that you found in (a) for $A^\top A$ is an eigenvector for $AA^\top$, and similarly that the image under T^* of each eigenvector for $AA^\top$ is an eigenvector for $A^\top A$.

1. $\begin{bmatrix} 3 & 0 & 1 \\ 0 & -4 & 0 \end{bmatrix}$

2. $\begin{bmatrix} 0 & 3 \\ -2 & 0 \\ 0 & 5 \end{bmatrix}$

3. $\begin{bmatrix} -2 & 0 \\ 3 & 2 \end{bmatrix}$

4. $\begin{bmatrix} 3 & 1 \\ -1 & 5 \end{bmatrix}$

5. $\begin{bmatrix} 0 & -1 & 2 \\ 4 & 2 & 1 \end{bmatrix}$

6. $\begin{bmatrix} 3 & 1 \\ -1 & 2 \\ 1 & -1 \end{bmatrix}$

7. $\begin{bmatrix} -2 & 5 & 2 \\ 2 & 3 & 2 \end{bmatrix}$

8. $\begin{bmatrix} -1 & 3 & 5 \\ 3 & 1 & -2 \end{bmatrix}$

9. $\begin{bmatrix} 3 & 1 & -2 & 1 \\ 5 & -2 & 3 & -7 \end{bmatrix}$

10 $\begin{bmatrix} 3 & 2 & 4 & 2 \\ 4 & -2 & -3 & 1 \end{bmatrix}$

11. $\begin{bmatrix} 3 & 1 & -2 & 1 \\ 5 & -2 & 3 & -7 \\ 6 & -7 & 11 & 11 \end{bmatrix}$

12. $\begin{bmatrix} 1 & 1 & -2 \\ -1 & 1 & 1 \\ 1 & -2 & 1 \\ 2 & 1 & 1 \end{bmatrix}$

13. $\begin{bmatrix} 1 & i & -2i \\ -i & 1 & 1 \end{bmatrix}$

14. $\begin{bmatrix} -1 & 3 \\ 3 & 1 \\ -5i & 2i \end{bmatrix}$

15. $\begin{bmatrix} -1 & 5-i & 3 \\ 3 & 2+i & 1 \end{bmatrix}$

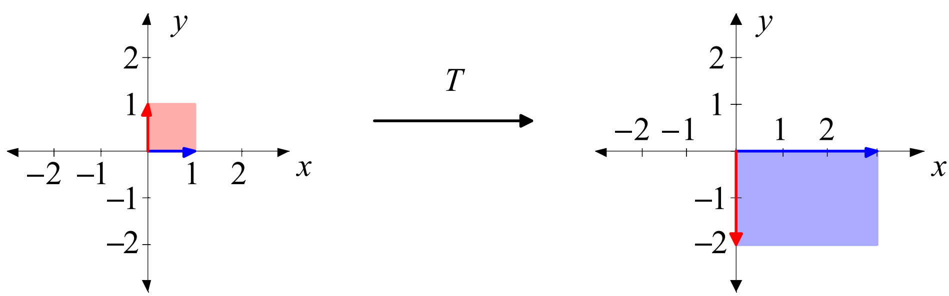
9.8 The Singular Value Decomposition

We are now ready to present a powerful tool in modern applications: the Singular Value Decomposition or SVD. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ can often be a complicated function. However, we do know that the easiest ones to understand are the operators with ***diagonal*** matrices. For example, suppose $T : \mathbb{R}^2 \to \mathbb{R}^2$ is given by:

$$T(\langle x, y \rangle) = \langle 3x, -2y \rangle, \text{ and thus:}$$

$$[T] = \begin{bmatrix} 3 & 0 \\ 0 & -2 \end{bmatrix}.$$

We can immediately draw what happens to our basic box under T:



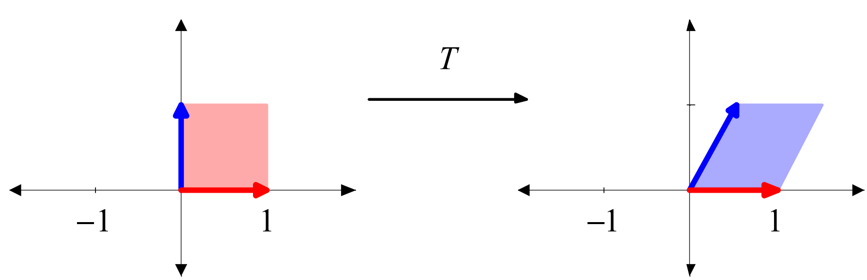
The Image of the Basic Box Under $[T] = Diag(3, -2)$

The image of the basic box is not a square. However, it is still a ***rectangle***. This is the next best thing, even though it was stretched and reflected across the x-axis. In this case, we can think of the x and y axes as ***orthogonal frames*** for our transformation. Now, let's consider the horizontal shear operator given by:

$$T(\langle x, y \rangle) = \langle x + 0.5y, y \rangle, \text{ and thus:}$$

$$[T] = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}$$

We can see its effect on the basic box below:



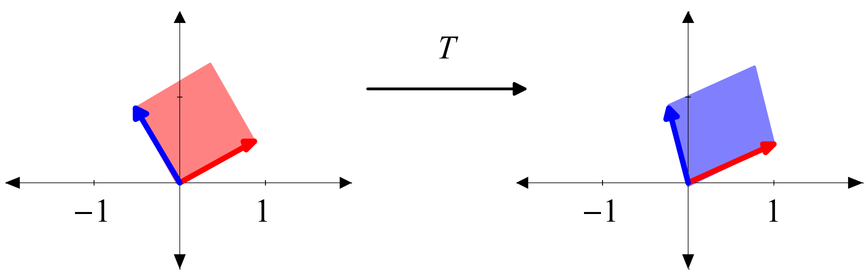
The Image of the Basic Box Under a Horizontal Shear Operator

This time, the image is no longer a rectangle. It is still a ***parallelogram*** because shear operators are linear and invertible. The purpose of the SVD is essentially to find ***new orthogonal frames***, both for the domain and the codomain, that would describe our transformation simply by using scalar multiplication as in the diagonal case above. Intuitively, we can do this by rotating the basic box, little by little, and finding out what its image is under the transformation. Once the image is again a ***rectangle***, we have geometrically found its SVD.

The angle between our image vectors is $tan^{-1}(2) \approx 63^0$. Let us try rotating the basic box by 30^0 and see what happens to the angle between our image vectors. The new orthonormal frame has basis:

$$\left\{ \left\langle \sqrt{3}/2, 1/2 \right\rangle, \left\langle -1/2, \sqrt{3}/2 \right\rangle \right\}.$$

The images of these two vectors are approximately $\langle 1.116, 0.5 \rangle$ and $\langle -0.067, 0.866 \rangle$:

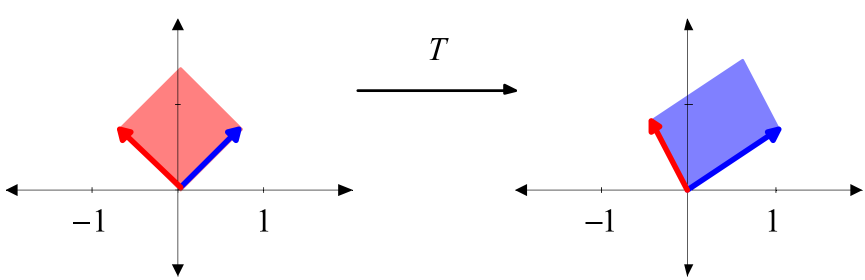


The Image of $\left\{ \left\langle \sqrt{3}/2, 1/2 \right\rangle, \left\langle -1/2, \sqrt{3}/2 \right\rangle \right\}$ Under The Same Shear Operator

Using the Dot Product and the Law of Cosines, we find that the angle between our image vectors is approximately 70.3^0. Let us now rotate the basic box by 45^0 and see what happens. The new orthogonal frame has basis:

$$\left\{ \left\langle 1/\sqrt{2}, 1/\sqrt{2} \right\rangle, \left\langle -1/\sqrt{2}, 1/\sqrt{2} \right\rangle \right\}.$$

The images of these two vectors are approximately $\langle 1.0607, 0.707 \rangle$ and $\langle -0.354, 0.707 \rangle$:



The Image of $\left\{ \left\langle 1/\sqrt{2}, 1/\sqrt{2} \right\rangle, \left\langle -1/\sqrt{2}, 1/\sqrt{2} \right\rangle \right\}$ Under The Same Shear Operator

Similar computations will tell us that the angle between them is now approximately 82.9^0, so we are probably about 7^0 from the correct angle of rotation. It turns out that this estimate is indeed very close to the correct angle of rotation, which we will find more precisely using the SVD.

The Statement of the SVD Theorem

Before we formally state the SVD Theorem, let us motivate it by reviewing some results from recent Sections, especially those found in The Fundamental Theorem of Linear Algebra:

Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation, with $m \times n$ standard matrix $[T] = A$. We can create T^*, the ***adjoint transformation***, $T^* : \mathbb{R}^m \to \mathbb{R}^n$, where $[T^*] = A^\top$, an $n \times m$ matrix. We can form both compositions:

$$T^* \circ T : \mathbb{R}^n \to \mathbb{R}^n, \text{ and } T \circ T^* : \mathbb{R}^m \to \mathbb{R}^m,$$

with standard matrices $[T^* \circ T] = A^\top A$ and $[T \circ T^*] = AA^\top$, respectively. Both $A^\top A$ and $AA^\top$ are ***symmetric*** and ***positive semi-definite***. Furthermore, they share the same positive eigenvalues λ_1, λ_2, …, λ_r, where $r = rank(A^\top A) = rank(AA^\top)$. If $\lambda > 0$ is a shared eigenvalue, then the restrictions:

$$T : Eig(A^\top A, \lambda) \to Eig(AA^\top, \lambda) \text{ and } T^* : Eig(AA^\top, \lambda) \to Eig(A^\top A, \lambda)$$

are both ***isomorphisms*** of eigenspaces, although not necessarily inverses of each other. For the SVD, we will also need to generalize the concept of a diagonal matrix:

Definition: Let D be any $m \times n$ matrix. We say that D is ***diagonal*** if $D_{i,j} = 0$ if $i \neq j$. Thus, D can be written in ***block form*** as:

$$D = \begin{bmatrix} D' \\ \mathbf{0} \end{bmatrix} \text{ if } m > n, \text{ or } D = \begin{bmatrix} D' & \mathbf{0} \end{bmatrix} \text{ if } m < n$$

for some (ordinary) square diagonal matrix D', and some appropriate zero matrix $\mathbf{0}$.
This new definition coincides with the old when $m = n$.

We can now state the Singular Value Decomposition Theorem:

Theorem — The Singular Value Decomposition:
Any $m \times n$ matrix A with rank r can be ***factored*** in the form:

$$A = U\Sigma V^\top, \quad \text{where:}$$

- U is an $m \times m$ ***orthogonal*** matrix whose columns are ***eigenvectors*** of $AA^\top$;
- Σ is an $m \times n$ ***diagonal*** matrix whose diagonal entries consist of the ***singular values*** of A:

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1 \ldots r,$$

  where λ_1 through λ_r are the ***positive eigenvalues*** of $AA^\top$;
- V is an $n \times n$ ***orthogonal*** matrix whose columns are ***eigenvectors*** of $A^\top A$.

Proof: Since $A^\top A$ is ***symmetric*** and ***positive semi-definite***, we can order its eigenvalues in decreasing order:

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r, 0, 0, \ldots, 0,$$

where the number of 0's is $nullity(A^\top A)$, and all the λ_i are ***positive***. By the Spectral Theorem, we can construct an ***orthonormal basis*** $\{\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n\}$ for $\mathbb{R}^n$ consisting of unit eigenvectors of $A^\top A$, where $\vec{v}_i$ is an eigenvector corresponding to λ_i (the last $n - r$ vectors form a basis for the nullspace of $A^\top A$, if need be). However, if $i \leq r$, every $A\vec{v}_i = T(\vec{v}_i)$ is now a member of the λ_i eigenspace of $AA^\top$, where T is the linear transformation represented by A. Since $\lambda_i > 0$, we can define:

$$\vec{u}_i = \frac{1}{\sqrt{\lambda_i}} A\vec{v}_i = \frac{1}{\sigma_i} A\vec{v}_i, \text{ or } A\vec{v}_i = \sigma_i \vec{u}_i, \text{ for } i = 1 \ldots r.$$

The restriction of T on each eigenspace is an isomorphism on the eigenspaces for λ_1 through λ_r, so the set $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_r\}$ of eigenvectors for $AA^\top$ in $\mathbb{R}^m$ is ***linearly independent***. Furthermore, we can prove that this is also an ***orthonormal set:***

$$\begin{aligned} [\vec{u}_i]^\top [\vec{u}_j] &= \frac{1}{\sigma_i} (A[\vec{v}_i])^\top \frac{1}{\sigma_j} A[\vec{v}_j] = \frac{1}{\sigma_i \sigma_j} [\vec{v}_i]^\top (A^\top A[\vec{v}_j]) \\ &= \frac{1}{\sigma_i \sigma_j} [\vec{v}_i]^\top \lambda_j [\vec{v}_j] = \frac{\lambda_j}{\sigma_i \sigma_j} [\vec{v}_i]^\top [\vec{v}_j], \end{aligned}$$

so if $i = j$, we get $\lambda_i / \sigma_i^2 = 1$, and if $i \neq j$, we get 0 since $\vec{v}_i$ is orthogonal to $\vec{v}_j$.

Now, for the final ingredient, let us find an orthonormal basis $\{\vec{u}_{r+1}, \vec{u}_{r+2}, \ldots, \vec{u}_m\}$ for the ***nullspace*** of $AA^\top$. Since eigenvectors from distinct eigenspaces are orthogonal, the combined set $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_m\}$ is an orthonormal set.

We are now ready to assemble our three matrices. Let us start by forming $D' = Diag(\sigma_1, \sigma_2, \ldots, \sigma_n)$, where $\sigma_i = \sqrt{\lambda_i}$, for $i = 1 \ldots n$. Then we form:

$$U = [\vec{u}_1\ \vec{u}_2\ \ldots\ \vec{u}_m],$$

$$V = [\vec{v}_1\ \vec{v}_2\ \ldots\ \vec{v}_n], \text{ and}$$

$$D = \begin{bmatrix} D' \\ \mathbf{0} \end{bmatrix} \text{ if } m \geq n \quad \text{or } D = \begin{bmatrix} D' & \mathbf{0} \end{bmatrix} \text{ if } m \leq n.$$

Note that in either case, all columns of D beyond column r and all rows beyond row r consist of zeroes. Since V is orthogonal, $V^\top = V^{-1}$, so we are done if we can show that $AV = U\Sigma$.

Both products are $m \times n$ matrices. Column i of AV is $A\vec{v}_i$. If $i \leq r$, then $A\vec{v}_i = \sigma_i \vec{u}_i$, which happens to be column i of $U\Sigma$. If $i > r$, then $A\vec{v}_i = \vec{0}_m$, which is likewise column i of $U\Sigma$ (since column i of Σ contains 0). Thus the two products are the same.■

Remarks: The SVD is in the same spirit as some familiar processes:

- We know how to construct the matrix $[T]_{B,B'}$ of a linear transformation T with respect to bases B and B' for the domain and codomain, respectively. This matrix is used via the three steps of encode, multiply and decode. This is essentially what happens in the SVD: we ***encode*** using the eigenvectors of $A^\top A$ (which are also called ***input vectors***) found in the rows of $V^\top$, ***multiply*** this by the singular values in Σ, then ***decode*** using the eigenvectors of $AA^\top$ (also called ***output vectors***) found in the columns of U.
- In Chapter 6, we attempt, when possible, to ***diagonalize*** a square matrix A:

$$A = CDC^{-1}.$$

The SVD is essentially a generalization of the diagonalization process to ***any*** matrix, whether it is diagonalizable or not, and moreover, whether it is ***square*** or not. The big difference, though, is that the entries in D are actual ***eigenvalues*** of A, but the entries of Σ are the ***singular values*** $\sigma_i = \sqrt{\lambda_i}$, where the λ_i are the ***positive eigenvalues*** of $A^\top A$.

- We can also think of the SVD as a ***change of basis:*** A is the standard matrix of some linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. If we let B be the basis for $\mathbb{R}^n$ consisting of the $\vec{v}_i$, and B' the basis for $\mathbb{R}^m$ consisting of the $\vec{u}_i$, then $[B]$ and $[B']$ are both ***orthogonal***, and $[T]_{B,B'}$ is a ***diagonal*** matrix containing the ***singular values*** along the diagonal, so:

$$[T] = [B'][T]_{B,B'}[B]^{-1} = U\Sigma V^{-1} = U\Sigma V^\top.$$

Shear Elegance

The shear transformations represented by:

$$\begin{bmatrix} 1 & k \\ 0 & 1 \end{bmatrix} \text{ and } \begin{bmatrix} 1 & 0 \\ k & 1 \end{bmatrix},$$

where $k \neq 0$, are good examples of operators that are ***not diagonalizable***. However, now that we have the SVD, we can look at them in an orthogonal way. Let us finish our example from the introduction:

Example: Let us find the SVD of our shear transformation T with matrix:

$$[T] = A = \begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}.$$

We will approximate to four decimal places. We form the matrix product:

$$A^\top A = \begin{bmatrix} 1 & 0 \\ 0.5 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1.25 \end{bmatrix}.$$

This matrix has eigenvalues $\lambda_1 \approx 1.6404$ and $\lambda_2 \approx 0.6096$, so the singular values are $\sigma_1 = \sqrt{\lambda_1} \approx 1.2808$ and $\sigma_2 = \sqrt{\lambda_2} \approx 0.7808$. The corresponding unit eigenvectors for $A^\top A$ are:

$$\vec{v}_1 \approx \langle 0.6154, 0.7882\rangle, \text{ and } \vec{v}_2 \approx \langle -0.7882, 0.6154\rangle.$$

The corresponding eigenvectors for $AA^\top$ are:

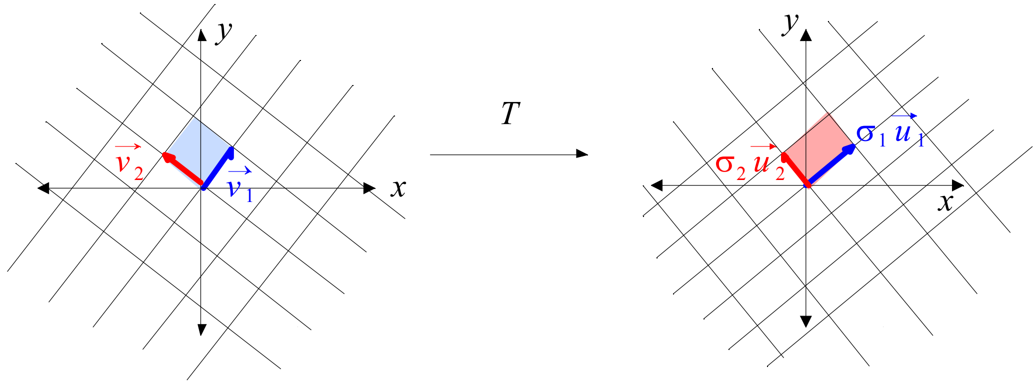
$$\vec{u}_1 \approx \frac{1}{1.2808}\begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 0.6154 \\ 0.7882 \end{bmatrix} = \begin{bmatrix} 0.7882 \\ 0.6154 \end{bmatrix}, \text{ and}$$

$$\vec{u}_2 \approx \frac{1}{0.7808}\begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} -0.7882 \\ 0.6154 \end{bmatrix} = \begin{bmatrix} -0.6154 \\ 0.7882 \end{bmatrix}$$

The final (approximate) SVD is:

$$\begin{bmatrix} 1 & 0.5 \\ 0 & 1 \end{bmatrix} \approx \begin{bmatrix} 0.7882 & -0.6154 \\ 0.6154 & 0.7882 \end{bmatrix}\begin{bmatrix} 1.2808 & 0 \\ 0 & 0.7808 \end{bmatrix}\begin{bmatrix} 0.6154 & 0.7882 \\ -0.7882 & 0.6154 \end{bmatrix}$$

The orthogonal frame $\{\vec{v}_1, \vec{v}_2\}$ is transformed by T onto the orthogonal frame $\{\sigma_1\vec{u}_1, \sigma_2\vec{u}_2\}$, as shown below:



Matching Pairs of Orthogonal Frames for a Horizontal Shear Operator

The $\vec{u}_1$ axis is scaled by $\sigma_1 \approx 1.2808$, and the $\vec{u}_2$ axis is scaled by $\sigma_2 \approx 0.7808$. The angle made by $\vec{v}_1$ with the x-axis is approximately $\tan^{-1}(0.7882/0.6154) \approx 52^0$, so this is the angle by which we need to rotate the basic box to obtain matching orthogonal frames. This explains why the rotation by 45^0 gave us image vectors that were close to orthogonal.□

Examples from the Previous Section

Let us complete the examples that we saw from The Fundamental Theorem of Linear Algebra by finding their SVDs and showing their geometric meaning:

Example: We saw the 2×3 matrix:

$$A = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}, \text{ with } A^\top A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 5 & 3 \\ 0 & 3 & 2 \end{bmatrix}.$$

The characteristic polynomial of $A^\top A$ is:

$$p(\lambda) = \lambda^3 - 9\lambda^2 + 14\lambda = \lambda(\lambda-2)(\lambda-7),$$

so our singular values are $\sigma_1 = \sqrt{7}$ and $\sigma_2 = \sqrt{2}$.

This time, we need a basis of ***unit*** vectors for the eigenspaces of $A^\top A$:

$$Eig(A^\top A, 7) = Span\left(\left\{\langle 1, 5, 3\rangle/\sqrt{35}\right\}\right)$$

$$Eig(A^\top A, 2) = Span\left(\left\{\langle -3, 0, 1\rangle/\sqrt{10}\right\}\right), \text{ and}$$

$$Eig(A^\top A, 0) = Span\left(\left\{\langle 1, -2, 3\rangle/\sqrt{14}\right\}\right).$$

Next, we compute $\vec{u}_1 = \frac{1}{\sigma_1}A\vec{v}_1$ and $\vec{u}_2 = \frac{1}{\sigma_2}A\vec{v}_2$:

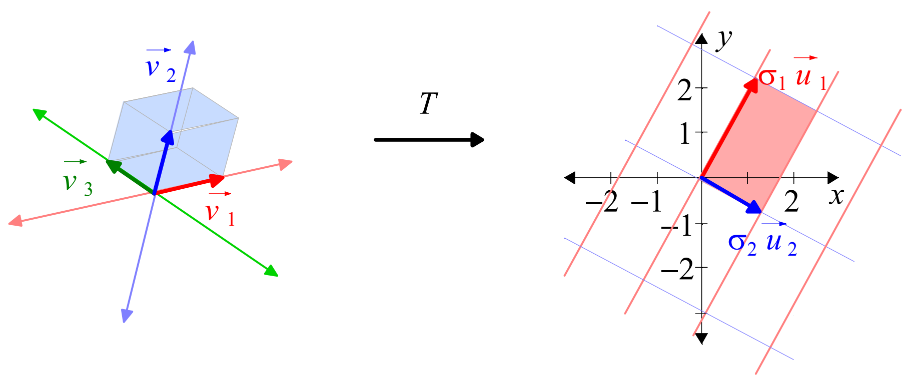
$$\vec{u}_1 = \frac{1}{\sqrt{7}}\begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \cdot \frac{1}{\sqrt{35}}\begin{bmatrix} 1 \\ 5 \\ 3 \end{bmatrix} = \frac{1}{7\sqrt{5}}\begin{bmatrix} 7 \\ 14 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \end{bmatrix}, \text{ and}$$

$$\vec{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} \cdot \frac{1}{\sqrt{10}}\begin{bmatrix} -3 \\ 0 \\ 1 \end{bmatrix} = \frac{1}{2\sqrt{5}}\begin{bmatrix} 4 \\ -2 \end{bmatrix} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ -1 \end{bmatrix}.$$

We can see that these two vectors indeed form an ***orthonormal set***. Our SVD is:

$$A = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{7} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{70}}\begin{bmatrix} \sqrt{2} & 5\sqrt{2} & 3\sqrt{2} \\ -3\sqrt{7} & 0 & \sqrt{7} \\ \sqrt{5} & -2\sqrt{5} & 3\sqrt{5} \end{bmatrix}.$$

We show below the "basic cube" in 3-space formed by $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and the corresponding orthonormal frame $\{\sigma_1\vec{u}_1, \sigma_2\vec{u}_2\}$.



Matching Sets of Orthogonal Frames for a Transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$

The $\vec{u}_1$ axis is scaled by $\sigma_1 = \sqrt{7}$, and the $\vec{u}_2$ axis is scaled by $\sigma_2 = \sqrt{2}$. □

Example: We saw the operator with 2×2 matrix:

$$A = \begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix}, \text{ where } A^{\top}A = \begin{bmatrix} 13 & 8 \\ 8 & 5 \end{bmatrix}.$$

The eigenvalues of $A^{\top}A$ are $\lambda = 9 \pm 4\sqrt{5} \approx 17.9437$ and 0.0557, which we wrote as:

$$\lambda_1 = 8\gamma + 5 = (2\gamma + 1)^2, \text{ and } \lambda_2 = 8\overline{\gamma} + 5 = (-2\overline{\gamma} - 1)^2.$$

where $\gamma = \left(1 + \sqrt{5}\right)/2 \approx 1.618$ and $\overline{\gamma} = \left(1 - \sqrt{5}\right)/2 \approx -0.618$. Thus we get:

$$\sigma_1 = 2\gamma + 1 \approx 4.236 \text{ and } \sigma_2 = -2\overline{\gamma} - 1 \approx 0.236.$$

We need unit vectors as bases for the eigenspaces:

$$Eig(A^\top A, 8\gamma + 5) = Span\left(\left\{\langle \gamma, 1\rangle / \sqrt{\gamma + 2}\right\}\right), \quad \text{and}$$
$$Eig(A^\top A, 8\overline{\gamma} + 5) = Span\left(\left\{\langle \overline{\gamma}, 1\rangle / \sqrt{\overline{\gamma} + 2}\right\}\right),$$

where we used the simplification $\gamma^2 + 1 = \gamma + 1 + 1 = \gamma + 2$, and likewise $\overline{\gamma}^2 + 1 = \overline{\gamma} + 1$. We now multiply the 1st vector by $\frac{1}{\sigma_1}A$ to get:

$$\frac{1}{2\gamma + 1}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \cdot \frac{1}{\sqrt{\gamma + 2}}\begin{bmatrix} \gamma \\ 1 \end{bmatrix}$$
$$= \frac{1}{2\gamma + 1} \cdot \frac{1}{\sqrt{\gamma + 2}}\begin{bmatrix} 2\gamma + 1 \\ 3\gamma + 2 \end{bmatrix} = \frac{1}{\sqrt{\gamma + 2}}\begin{bmatrix} 1 \\ \gamma \end{bmatrix},$$

where we used the fact that $\gamma(2\gamma + 1) = 2\gamma^2 + \gamma = 2\gamma + 2 + \gamma = 3\gamma + 2$.

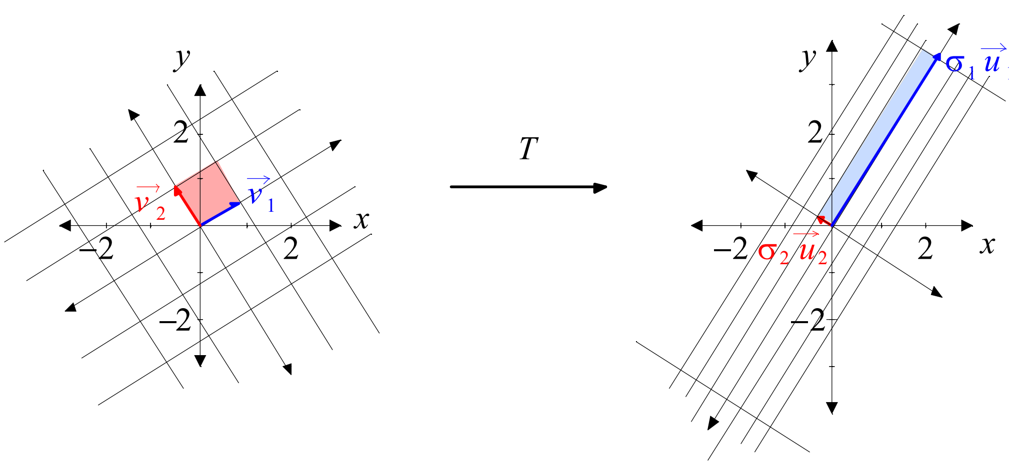
Similarly, for the 2nd vector, using $\frac{1}{\sigma_2}A$, we get:

$$\frac{1}{-2\overline{\gamma} - 1}\begin{bmatrix} 2 & 1 \\ 3 & 2 \end{bmatrix} \cdot \frac{1}{\sqrt{\overline{\gamma} + 2}}\begin{bmatrix} \overline{\gamma} \\ 1 \end{bmatrix} = \frac{-1}{\sqrt{\overline{\gamma} + 2}}\begin{bmatrix} 1 \\ \overline{\gamma} \end{bmatrix}.$$

(This explains the choice of $\langle -1, -\overline{\gamma}\rangle$ as our eigenvector in the previous Section.) Finally, we get the SVD:

$$A = \begin{bmatrix} \frac{1}{\sqrt{\gamma + 2}} & \frac{-1}{\sqrt{\overline{\gamma} + 2}} \\ \frac{\gamma}{\sqrt{\gamma + 2}} & \frac{-\overline{\gamma}}{\sqrt{\overline{\gamma} + 2}} \end{bmatrix}\begin{bmatrix} 2\gamma + 1 & 0 \\ 0 & -2\overline{\gamma} - 1 \end{bmatrix}\begin{bmatrix} \frac{\gamma}{\sqrt{\gamma + 2}} & \frac{1}{\sqrt{\gamma + 2}} \\ \frac{\overline{\gamma}}{\sqrt{\overline{\gamma} + 2}} & \frac{1}{\sqrt{\overline{\gamma} + 2}} \end{bmatrix}.$$

We show the new orthonormal frames below:



Matching Pairs of Orthogonal Frames for an Operator on $\mathbb{R}^2$

The $\vec{u}_1$ axis is scaled by $\sigma_1 \approx 4.236$, and the $\vec{u}_2$ axis is scaled by $\sigma_2 \approx 0.236$. The angle made by $\vec{v}_1$ with the x-axis is $\tan^{-1}(1/\gamma) \approx 31.7^0$, so this is the angle by which we need to rotate the basic box to obtain matching orthogonal frames. □

The Complex Case

The SVD also exists for any $m \times n$ matrix A with complex entries. As in previous generalizations, every ***transpose*** is changed to the ***Hermitian adjoint***, and ***orthogonal matrices*** are replaced by ***unitary*** matrices. The SVD for A becomes:

$$A = U\Sigma V^*, \quad \text{where:}$$

- U is an $m \times m$ ***unitary*** matrix whose columns are ***eigenvectors*** of AA^*;
- Σ is an $m \times n$ ***(real) diagonal*** matrix whose diagonal entries consist of the ***singular values*** of A :

$$\sigma_i = \sqrt{\lambda_i}, \quad i = 1 \ldots r,$$

  where $r = rank(A)$, and λ_1 through λ_r are the ***positive eigenvalues*** of AA^*;
- V is an $n \times n$ ***unitary*** matrix whose columns are ***eigenvectors*** of A^*A.

Example: In the previous Section, we saw:

$$A = \begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix}, \text{ with } A^*A = \begin{bmatrix} 2 & i \\ -i & 2 \end{bmatrix}.$$

The characteristic polynomial of A^*A is $p(\lambda) = (\lambda - 1)(\lambda - 3)$, so our singular values are $\sigma_1 = \sqrt{3}$ and $\sigma_2 = 1$.

We normalize the bases for the eigenspaces of A^*A that we found before, and get:

$$Eig(A^*A, 3) = Span\left(\left\{\langle i, 1\rangle/\sqrt{2}\right\}\right), \text{ and } Eig(A^*A, 1) = Span\left(\left\{\langle -i, 1\rangle/\sqrt{2}\right\}\right).$$

Now we find $\vec{u}_1$ and $\vec{u}_2$:

$$\vec{u}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} i \\ 1 \end{bmatrix} = \begin{bmatrix} 2i/\sqrt{6} \\ -1/\sqrt{6} \\ -1/\sqrt{6} \end{bmatrix}, \text{ and}$$

$$\vec{u}_2 = \frac{1}{1}\begin{bmatrix} 1 & i \\ 0 & -1 \\ i & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{2}}\begin{bmatrix} -i \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}.$$

As opposed to our previous Examples, the reverse product AA^* is ***bigger*** than A^*A, and so we need to find a basis for the nullspace of:

$$AA^* = \begin{bmatrix} 2 & -i & -i \\ i & 1 & 0 \\ i & 0 & 1 \end{bmatrix}, \text{ which has rref: } R = \begin{bmatrix} 1 & 0 & -i \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Thus, a basis for its nullspace is $\langle i, 1, 1\rangle$, which has length $\sqrt{3}$. We normalize this vector and place it in the last column of U, thus obtaining the SVD:

$$A = U\Sigma V^* = \begin{bmatrix} 2i/\sqrt{6} & 0 & i/\sqrt{3} \\ -1/\sqrt{6} & -1/\sqrt{2} & 1/\sqrt{3} \\ -1/\sqrt{6} & 1/\sqrt{2} & 1/\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} -i/\sqrt{2} & 1/\sqrt{2} \\ i/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}. \square$$

9.8 Exercises

For Exercises (1) to (5): For the following matrices A representing shear operators on $\mathbb{R}^2$: (a) Find the SVD of A. You may approximate the entries to 5 decimal places; (b) Draw the orthogonal frames $\{\vec{v}_1, \vec{v}_2\}$ and $\{\sigma_1\vec{u}_1, \sigma_2\vec{u}_2\}$, and show the action of A using these frames.

1. $\begin{bmatrix} 1 & 0.7 \\ 0 & 1 \end{bmatrix}$

2. $\begin{bmatrix} 1 & -1.2 \\ 0 & 1 \end{bmatrix}$

3. $\begin{bmatrix} 1 & 0 \\ 0.3 & 1 \end{bmatrix}$

4. $\begin{bmatrix} 1 & 0 \\ -0.8 & 1 \end{bmatrix}$

5. $\begin{bmatrix} 1 & 2.3 \\ 0 & 1 \end{bmatrix}$

For Exercises (6) to (20): Find the SVD of the matching matrix in Exercises 1 to 15, Section 9.7.

9.9 Applications of The SVD

The Singular Value Decomposition has numerous applications, both within Linear Algebra and in other computational fields. We will look at a sampling of these applications, namely, how to use the SVD to find the constant relating two proportional quantities, to compress data such as digital images, and to solve the Least Squares Problem using a different approach.

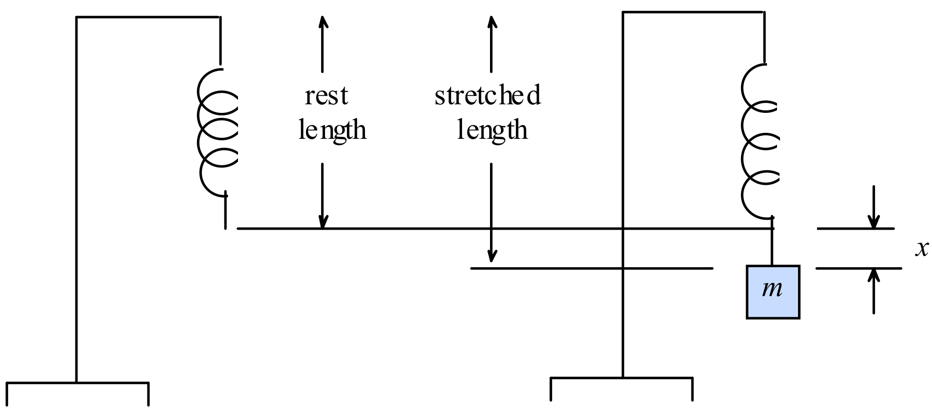
Finding Constants of Proportionality

One of the most basic application of the SVD is to find the constant of proportionality between two quantities, say x and y, if one is in direct proportion to the other. Given a sample of n ordered pairs (x_i, y_i), we can try to solve for the constant of proportionality k in the proportion $y = kx$ by studying the SVD of this data set, and at the same time, determine if it is indeed reasonable to assume that y is proportional to x.

Example: An ideal spring obeys ***Hooke's Law:***

$$F = kx,$$

where F is the magnitude of the force that the spring exerts when it is stretched from its rest length by a distance of x. Suppose we ran the following experiment:



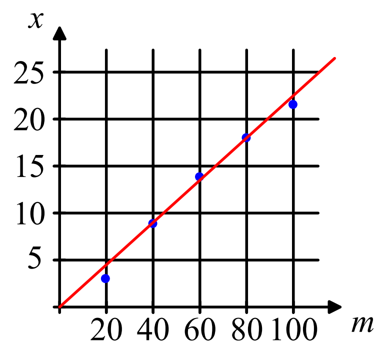
An Experimental Set-Up to Investigate Hooke's Law

1. Hang a spring from a supporting stand and let it dangle freely.
2. Measure its rest length.
3. Hang a small, known mass m, with units in ***grams***, from the bottom of the spring.
4. Measure the stretched length of the spring and subtract the resting length. This difference is x, which we will take to be in ***centimeters***.
5. Repeat steps 3 and 4 using several masses.

Since the weight of the hanging mass is $F = mg$, where g is the acceleration due to gravity, we will simply try to find k in the equation $mg = kx$ (converting m to kilograms first). Now, suppose we obtain the following data:

m	20	40	60	80	100
x	4.7	9.1	13.9	18.2	22.6

The observations seem to indicate that m and x are indeed proportional to each other, as we show on a graph below with an approximate line that is close to the points:



We would like to know the slope of the line that best approximates this data set. Let us create the 2×5 matrix:

$$A = \begin{bmatrix} 20 & 40 & 60 & 80 & 100 \\ 4.7 & 9.1 & 13.9 & 18.2 & 22.6 \end{bmatrix}$$

Using technology, we find that the SVD of A is:

$$A = U\Sigma V^{\mathsf{T}} = \begin{bmatrix} 0.97506 & -0.22196 \\ 0.22196 & 0.97506 \end{bmatrix} \begin{bmatrix} 152.12 & 0 & 0 & 0 & 0 \\ 0 & 0.31908 & 0 & 0 & 0 \end{bmatrix} V^{\mathsf{T}}$$

where V is a 5×5 ***orthogonal matrix*** (fortunately we will have no need for the actual entries of V). Thus, the singular values of A are $\sigma_1 \approx 152.12$ and $\sigma_2 \approx 0.319$. The first singular value is obviously ***dominant***, and it is reasonable to assume that the second singular value (which is practically ***zero***) is only due to small experimental errors (also called ***noise***) or imperfections in the spring. Thus it is reasonable to conclude that m is proportional to x. The dominant singular value, though, does ***not*** give us the spring constant k. The answer lies in the ***eigenvectors*** found in the output matrix U. The first column corresponds to the dominant singular value. This tells us that our data should be clustered close to this eigenvector, $\vec{u}_1 = \langle 0.975, 0.222 \rangle$. This vector defines a line through the origin of slope:

$$\frac{0.222}{0.975} \approx 0.2277$$

Thus:

$$x \approx 0.2277 \cdot m$$

Solving for the spring constant, we get:

$$mg = kx, \text{ or}$$
$$k = \frac{mg}{x} \approx \frac{9.8}{1000 \times 0.2277} \approx 0.04304 \text{ N/cm. or } 4.304 \text{ N/m},$$

where we needed the factor of 1000 because our masses were measured in grams.■

The SVD and Data Compression

One of the most important applications of the SVD is in the compression of data. A good example is a digital photograph. Many digital cameras routinely take pictures of 3000 pixels by 4000 pixels, each pixel being assigned a color from a template of hundreds of possible hues. As a result, a simple picture can be several megabytes in size — large enough to try your patience when e-mailing several of them to your family and friends.

But the same picture can often be easily compressed into a much smaller format, one that is good enough to be seen and admired though not necessarily worth framing. Our megabyte-sized picture can shrink to 300 kilobytes without much visible loss of quality. Many photo editing software have this feature, which you have probably tried yourself. The key idea behind this neat trick is of course the SVD.

To see the principle, let us look at an alternative formula for the SVD. We know that multiplying any matrix U by a diagonal matrix Σ on the right will simply result in each column of U being multiplied by the corresponding diagonal entry in Σ, thus:

$$U\Sigma = \begin{bmatrix} \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & \cdots & \sigma_k\vec{u}_k & \vec{\mathbf{0}}_m & \cdots & \vec{\mathbf{0}}_m \end{bmatrix},$$

since every entry in Σ beyond σ_k is 0. But then, we get:

$$\begin{aligned} A &= (U\Sigma)V^{\top} \\ &= \begin{bmatrix} \sigma_1\vec{u}_1 & \sigma_2\vec{u}_2 & \cdots & \sigma_k\vec{u}_k & \vec{0}_m & \cdots & \vec{0}_m \end{bmatrix} \begin{bmatrix} \vec{v}_1^{\top} \\ \vec{v}_2^{\top} \\ \vdots \\ \vec{v}_n^{\top} \end{bmatrix} \\ &= \sigma_1\vec{u}_1\vec{v}_1^{\top} + \sigma_2\vec{u}_2\vec{v}_2^{\top} + \cdots + \sigma_k\vec{u}_k\vec{v}_k^{\top}. \end{aligned}$$

This final sum is called the ***outer product*** expansion of A. Normally, we view the matrix on the left, $U\Sigma$, in terms of its rows, and the matrix on the right, $V^{\top}$, in terms of its columns. We are doing the exact ***opposite*** of this convention.

Our sum above gives another improvement in our SVD formulation: recall that the $\vec{v}_i$'s are ***eigenvectors*** for $A^{\top}A$. The summation says that we do not need the eigenvectors corresponding to the nullspace $\lambda = 0$. Thus:

$$A = U\Sigma V^{\top} = \sum_{i=1}^{k} \sigma_i\vec{u}_i\vec{v}_i^{\top} = \sigma_1\vec{u}_1\vec{v}_1^{\top} + \sigma_2\vec{u}_2\vec{v}_2^{\top} + \cdots + \sigma_k\vec{u}_k\vec{v}_k^{\top}.$$

Let us look at this equation in terms of the amount of ***essential data*** we need to describe A completely. We need k terms, each term involving a single number (the singular value), an $m \times 1$ vector (the $\vec{u}_i$) and an $n \times 1$ vector (the $\vec{v}_i^{\top}$). Thus each term contains $1 + m + n$ numbers. The total cost of the SVD is thus:

$$(1 + m + n)k \ \text{ numbers.}$$

But recall that $k = rank(A) \leq min(m,n)$, and in some cases, k is a lot ***smaller*** than both m and n. Moreover, suppose that A is a large matrix and we are only interested in an ***approximation*** for A. Since the singular values are arranged in decreasing order, it might be possible that by adding fewer than k terms, we can get a "good" approximation for A.

This is where data compression comes in: Let us say that each pixel could take on a "color" value between 0 and 255, and thus a single byte of memory (8 binary digits) is needed to represent one pixel. Now, we can represent our photograph of 3000 by 4000 pixels to be a 3000×4000 matrix, with each entry costing one byte of memory. This photograph will therefore cost around 12 megabytes of memory. If we keep only 500 terms in the SVD summation formula, we will need only $(1 + 3000 + 4000) \times 500 = 3{,}500{,}500$ bytes, for a savings of about 70%.

Example: Let us represent a "photograph" using a 4×6 matrix whose entries are from 0 through 3 (representing four "colors"). For example, suppose:

$$A = \begin{bmatrix} 3 & 2 & 3 & 2 & 1 & 0 \\ 2 & 3 & 2 & 1 & 0 & 1 \\ 1 & 2 & 2 & 2 & 1 & 0 \\ 1 & 3 & 3 & 2 & 2 & 1 \end{bmatrix}.$$

Notice that the numbers do not change drastically as we move side to side or up and down the matrix. In the same way, colors in pictures do not change much from pixel to pixel, except possible at the edge of an object, such as the wall or roof of a building. Using technology, we find the singular values of A to be approximately:

$$9,\ 1.938,\ 1.67 \text{ and } 0.672$$

Notice that we have a dominant singular value, 9, followed by two smaller singular values, and the last one can definitely be disregarded. Using only the first term $\sigma_1 \vec{u}_1 \vec{v}_1^{\mathsf{T}}$ in our outer product expansion, we would get:

$$9 \begin{bmatrix} 0.555 \\ 0.454 \\ 0.405 \\ 0.568 \end{bmatrix} \begin{bmatrix} 0.394 & 0.554 & 0.565 & 0.39 & 0.233 & 0.113 \end{bmatrix}$$

$$= \begin{bmatrix} 1.968 & 2.767 & 2.82 & 1.948 & 1.164 & 0.564 \\ 1.61 & 2.264 & 2.31 & 1.594 & 0.952 & 0.462 \\ 1.436 & 2.019 & 2.06 & 1.422 & 0.85 & 0.412 \\ 2.014 & 2.832 & 2.89 & 1.994 & 1.19 & 0.578 \end{bmatrix}$$

$$\approx \begin{bmatrix} 2 & 3 & 3 & 2 & 1 & 0 \\ 2 & 2 & 2 & 2 & 1 & 0 \\ 1 & 2 & 2 & 1 & 1 & 0 \\ 2 & 3 & 3 & 2 & 1 & 1 \end{bmatrix},$$

if we round off to the nearest "color." Notice that 16 out of the 24 entries are already correct! Furthermore, all the wrong entries are at most one "color" off. Now, if we use two terms in the outer product expansion, we will get:

$$\begin{bmatrix} 0.555 & -0.545 \\ 0.454 & -0.412 \\ 0.405 & 0.243 \\ 0.568 & 0.69 \end{bmatrix} \begin{bmatrix} 9 & 0 \\ 0 & 1.94 \end{bmatrix} \begin{bmatrix} 0.39 & 0.55 & 0.57 & 0.39 & 0.23 & 0.11 \\ -0.79 & 0.13 & 0.05 & 0.19 & 0.56 & 0.14 \end{bmatrix}$$

$$= \begin{bmatrix} 2.8 & 2.64 & 2.77 & 1.75 & 0.577 & 0.416 \\ 2.24 & 2.17 & 2.27 & 1.44 & 0.507 & 0.349 \\ 1.06 & 2.07 & 2.08 & 1.51 & 1.11 & 0.481 \\ 0.96 & 3 & 2.95 & 2.24 & 1.93 & 0.77 \end{bmatrix}$$

$$\approx \begin{bmatrix} 3 & 3 & 3 & 2 & 1 & 0 \\ 2 & 2 & 2 & 1 & 1 & 0 \\ 1 & 2 & 2 & 2 & 1 & 0 \\ 1 & 3 & 3 & 2 & 2 & 1 \end{bmatrix},$$

which only has 4 incorrect entries. By taking three terms, we get a matrix, which when rounded off to the nearest integer entries, gives us ***precisely*** A. This should not be a surprise because the final singular value is very close to 0. □

The savings and accuracy behind this idea are certainly more impressive when applied on a bigger and more complex photograph using colors coded 0 through 255 instead of just 0 to 3. In this case, the color represented by 173 would not be very different, say, from the color represented by 170 or 175. Thus, minor errors in taking only a few terms in the outer product expansion often do not produce a strikingly different photograph.

The Pseudoinverse and Least Squares Method

In Chapter 6, we discussed the Method of Least Squares that will find an approximate solution to ***any*** system $A\vec{x} = \vec{b}$, especially when this system is inconsistent. The method, however, required us to identify a basis for the columnspace W of A first in order to perform the algorithm. By doing so, we can find the projection $proj_W\left(\vec{b}\right)$ onto the columnspace, and thus the system:

$$A\vec{x} = proj_W\left(\vec{b}\right)$$

is always ***consistent***, and we call a solution $\vec{x}_1$ to this system a ***least squares solution*** or ***best approximation*** to our linear system. The reason for this terminology is that the length $\left\| A\vec{x}_1 - \vec{b} \right\|$ is ***as small as possible***. The SVD gives us an alternative method to find an approximate solution $\vec{x}_1$ to $A\vec{x} = \vec{b}$ so that $A\vec{x}_1 - \vec{b}$ has minimum length.

First, we need to define the pseudoinverse of any matrix:

Definition: Let A be an $m \times n$ matrix with singular value decomposition:

$$A = U\Sigma V^{\top}.$$

Define by Σ^{+} the $n \times m$ diagonal matrix whose entries are $1/\sigma_i$, where $\sigma_i > 0$. In other words, we can obtain Σ^{+} from Σ by taking its ***transpose*** and taking the ***reciprocal*** of its singular values.
Now, define A^{+}, the ***pseudoinverse*** of A, by:

$$A^{+} = V\Sigma^{+}U^{\top}.$$

Example: We saw the SVD of $A = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}$ to be:

$$A = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix} \cdot \begin{bmatrix} \sqrt{7} & 0 & 0 \\ 0 & \sqrt{2} & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{70}}\begin{bmatrix} \sqrt{2} & 5\sqrt{2} & 3\sqrt{2} \\ -3\sqrt{7} & 0 & \sqrt{7} \\ \sqrt{5} & -2\sqrt{5} & 3\sqrt{5} \end{bmatrix}.$$

Thus, its pseudoinverse is:

$$A^{+} = \frac{1}{\sqrt{70}}\begin{bmatrix} \sqrt{2} & -3\sqrt{7} & \sqrt{5} \\ 5\sqrt{2} & 0 & -2\sqrt{5} \\ 3\sqrt{2} & \sqrt{7} & 3\sqrt{5} \end{bmatrix} \cdot \begin{bmatrix} 1/\sqrt{7} & 0 \\ 0 & 1/\sqrt{2} \\ 0 & 0 \end{bmatrix} \cdot \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{4}{7} & \frac{5}{14} \\ \frac{1}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{1}{14} \end{bmatrix}.$$

If we multiply A^{+} and A together, we get:

$$A^{+}A = \begin{bmatrix} -\frac{4}{7} & \frac{5}{14} \\ \frac{1}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{1}{14} \end{bmatrix}\begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{13}{14} & \frac{1}{7} & -\frac{3}{14} \\ \frac{1}{7} & \frac{5}{7} & \frac{3}{7} \\ -\frac{3}{14} & \frac{3}{7} & \frac{5}{14} \end{bmatrix}, \quad \text{and}$$

$$AA^{+} = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 2 & 1 \end{bmatrix}\begin{bmatrix} -\frac{4}{7} & \frac{5}{14} \\ \frac{1}{7} & \frac{2}{7} \\ \frac{2}{7} & \frac{1}{14} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}. \square$$

The second order shows us that the term "pseudoinverse" is justified. However, this happened only because A has rank 2, and thus $\Sigma\Sigma^{+}$ is I_2. Unfortunately:

$$\Sigma^{+}\Sigma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In general, both $\Sigma^{+}\Sigma$ and $\Sigma\Sigma^{+}$ will be diagonal matrices containing k 1's on the main diagonal starting at the upper left corner, where $k = rank(A)$. However, if A is an $n \times n$ ***invertible*** matrix, then $rank(A) = n$, and:

$$\begin{aligned} AA^{+} &= (U\Sigma V^{\top})(V\Sigma^{+}U^{\top}) = U\Sigma(V^{\top}V)\Sigma^{+}U^{\top} \\ &= U\Sigma(\boldsymbol{I_n})\Sigma^{+}U^{\top} = U\Sigma\Sigma^{+}U^{\top} = U\boldsymbol{I_n}U^{\top} = \boldsymbol{I_n}, \quad \text{and so:} \\ A^{+} &= A^{-1}. \end{aligned}$$

For this reason, the pseudoinverse generalizes the concept of an inverse.

Let us now turn to how the pseudoinverse can be used to solve the Least Squares problem. The key idea here is an equivalence was saw in Section 6.6: An $n \times n$ matrix Q is ***orthogonal*** ***if and only if*** for ***any*** $n \times 1$ matrix $\vec{x}$:

$$\|Q\vec{x}\| = \|\vec{x}\|,$$

that is, orthogonal transformations ***preserve lengths***. Since $U^{\top}$ is orthogonal, we get:

$$\begin{aligned} \left\| A\vec{x}_1 - \vec{b} \right\| &= \left\| U\Sigma V^{\top}\vec{x}_1 - \vec{b} \right\| = \left\| U^{\top}\left(U\Sigma V^{\top}\vec{x}_1 - \vec{b} \right) \right\| \\ &= \left\| (U^{\top}U)\Sigma V^{\top}\vec{x}_1 - U^{\top}\vec{b} \right\| = \left\| \Sigma V^{\top}\vec{x}_1 - U^{\top}\vec{b} \right\| \\ &= \left\| \Sigma\vec{y}_1 - U^{\top}\vec{b} \right\|, \end{aligned}$$

where $\vec{y}_1 = V^{\top}\vec{x}_1$. Once again, since $V^{\top}$ is also orthogonal, $\|\vec{y}_1\| = \|\vec{x}_1\|$. Thus, the problem of minimizing $\left\| A\vec{x}_1 - \vec{b} \right\|$ is equivalent to that of minimizing $\left\| \Sigma\vec{y}_1 - U^{\top}\vec{b} \right\|$. The easiest way to do this is to zero out as many components as possible of $\Sigma\vec{y}_1 - U^{\top}\vec{b}$. Unfortunately, every entry in $\Sigma\vec{y}_1$ from the $k+1$ component onward is already 0, so there is not much we can do there. However, if we make every entry of $\Sigma\vec{y}_1$ from the 1st to the kth entry equal to that of $U^{\top}\vec{b}$, then we will get at least k zeroes in $\Sigma\vec{y}_1 - U^{\top}\vec{b}$. In other words, the $\vec{y}_1$ that will minimize $\left\| \Sigma\vec{y}_1 - U^{\top}\vec{b} \right\|$ is:

$$\begin{aligned} \vec{y}_1 &= \Sigma^{+}U^{\top}\vec{b}, \quad \text{thus:} \\ V^{\top}\vec{x}_1 &= \Sigma^{+}U^{\top}\vec{b}, \quad \text{and so:} \\ \vec{x}_1 &= V\Sigma^{+}U^{\top}\vec{b} = A^{+}\vec{b}. \end{aligned}$$

We have thus proven the following:

Theorem: Consider the linear system $A\vec{x} = \vec{b}$. The vector:

$$\vec{x}_1 = A^{+}\vec{b},$$

where A^{+} is the pseudoinverse of A, minimizes $\left\| A\vec{x} - \vec{b} \right\|$ and is therefore a ***least squares solution*** to this system.

As before, all other least squares solution to this system have the form $\vec{x}_1 + \vec{x}_0$, where $\vec{x}_0 \in nullspace(A)$.

Example: Let us re-visit the inconsistent system we saw in Section 7.8:

$$\begin{array}{rcrcrcrcr} 4x_1 & - & 8x_2 & + & 3x_3 & + & 9x_4 & = & 7 \\ 3x_1 & - & 6x_2 & - & 4x_3 & + & 13x_4 & = & 15 \\ -2x_1 & + & 4x_2 & + & 3x_3 & - & 9x_4 & = & -9 \end{array}$$

The coefficient matrix is:

$$A = \begin{bmatrix} 4 & -8 & 3 & 9 \\ 3 & -6 & -4 & 13 \\ -2 & 4 & 3 & -9 \end{bmatrix}.$$

Using technology, we find that the SVD of A is:

$$A = \begin{bmatrix} -0.5514 & 0.8336 & -0.0324 \\ -0.6874 & -0.4319 & 0.5840 \\ 0.4728 & 0.3443 & 0.8111 \end{bmatrix} \times \begin{bmatrix} 21.72 & 0 & 0 & 0 \\ 0 & 6.186 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \times$$

$$\begin{bmatrix} -0.2400 & 0.4800 & 0.1157 & -0.8358 \\ 0.2183 & -0.4366 & 0.8506 & -0.1957 \\ -0.1412 & 0.6862 & 0.5046 & 0.5046 \\ -0.9353 & -0.3287 & 0.0927 & 0.0927 \end{bmatrix}.$$

Its pseudoinverse is thus:

$$\begin{aligned} A^+ &= V\Sigma^+U^\top \\ &= \begin{bmatrix} -0.2400 & 0.2183 & -0.1412 & -0.9353 \\ 0.4800 & -0.4366 & 0.6862 & -0.3287 \\ 0.1157 & 0.8506 & 0.5046 & 0.0927 \\ -0.8358 & -0.1957 & 0.5046 & 0.0927 \end{bmatrix} \times \begin{bmatrix} 0.04604 & 0 & 0 \\ 0 & 0.16166 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \times \\ &\quad \begin{bmatrix} -0.5514 & -0.6874 & 0.4728 \\ 0.8336 & -0.4319 & 0.3443 \\ -0.0324 & 0.5840 & 0.8111 \end{bmatrix} = \begin{bmatrix} 0.03551 & -0.00765 & 0.00693 \\ -0.07102 & 0.01529 & -0.01385 \\ 0.11169 & -0.06305 & 0.04986 \\ -0.00516 & 0.04012 & -0.02909 \end{bmatrix}. \end{aligned}$$

Thus, our least squares solution is:

$$\begin{aligned} \vec{x}_1 &= A^+\vec{b} \\ &= \begin{bmatrix} 0.03551 & -0.00765 & 0.00693 \\ -0.07102 & 0.01529 & -0.01385 \\ 0.11169 & -0.06305 & 0.04986 \\ -0.00516 & 0.04012 & -0.02909 \end{bmatrix} \begin{bmatrix} 7 \\ 15 \\ -9 \end{bmatrix} = \begin{bmatrix} 0.07145 \\ -0.14314 \\ -0.61266 \\ 0.82749 \end{bmatrix}. \end{aligned}$$

We can check that:

$$A\vec{x}_1 = \begin{bmatrix} 4 & -8 & 3 & 9 \\ 3 & -6 & -4 & 13 \\ -2 & 4 & 3 & -9 \end{bmatrix} \begin{bmatrix} 0.07145 \\ -0.14314 \\ -0.61266 \\ 0.82749 \end{bmatrix} = \begin{bmatrix} 7.0404 \\ 14.28 \\ -10.001 \end{bmatrix},$$

which is "close" to $\vec{b} = \langle 7, 15, -9 \rangle$. On the other hand, the solution that we found in Section 6.8 was:

$$\vec{x}_1 = \left\langle \frac{71}{25}, 0, -\frac{36}{25}, 0 \right\rangle = \langle 2.84, 0, -1.44, 0 \rangle.$$

This means that the ***difference*** between our two answers must be a member of the ***nullspace*** of A :

$$\langle 0.07145, -0.14314, -0.61266, 0.82749 \rangle - \langle 2.84, 0, -1.44, 0 \rangle$$
$$= \langle -2.7686, -0.14314, 0.82734, 0.82749 \rangle, \text{ and indeed:}$$

$$\begin{bmatrix} 4 & -8 & 3 & 9 \\ 3 & -6 & -4 & 13 \\ -2 & 4 & 3 & -9 \end{bmatrix} \begin{bmatrix} -2.7686 \\ -0.1431 \\ 0.8273 \\ 0.8275 \end{bmatrix} = \begin{bmatrix} 0.00015 \\ 0.00105 \\ -0.00075 \end{bmatrix} \approx \vec{\mathbf{0}}_3. \square$$

9.9 Exercises

The use of technology is highly recommended for all of the Exercises in this Section.

For Exercises (1) to (3): Assuming that the quantity y is proportional to x, use the SVD of a $2 \times n$ matrix to approximate the constant of proportionality k so that $y = kx$.

1.

x	2.1	3.5	4.2	5.5	6.8
y	9.8	16.3	19.9	26.3	32.4

2.

x	5.2	7.8	9.3	10.7	11.4	14.3
y	37.7	56.7	68.1	78.3	83.4	104.2

3.

x	3.5	8.2	11.3	17.8	22.1	25.7	31.8
y	32.7	77.3	106.7	167.8	207.3	241.3	298.7

For Exercises (4) to (6): Find the SVD of the matrix A, then use only the ***first two terms*** in the outer product expansion to approximate A. Round off your answer to the nearest integer to compare it with the original entries of A.

4. $\begin{bmatrix} 3 & 2 & 1 & 1 & 0 & 1 \\ 2 & 2 & 2 & 0 & 1 & 2 \\ 3 & 3 & 2 & 2 & 2 & 3 \\ 2 & 2 & 1 & 1 & 1 & 2 \end{bmatrix}$

5. $\begin{bmatrix} 2 & 1 & 1 & 2 & 2 & 3 & 2 \\ 1 & 0 & 1 & 1 & 2 & 2 & 3 \\ 2 & 1 & 2 & 2 & 3 & 3 & 2 \\ 3 & 2 & 2 & 3 & 3 & 2 & 2 \\ 3 & 2 & 1 & 2 & 2 & 1 & 1 \end{bmatrix}$

6. $\begin{bmatrix} 5 & 5 & 6 & 6 & 7 & 7 & 7 & 6 & 5 & 5 \\ 5 & 6 & 5 & 5 & 6 & 6 & 7 & 7 & 6 & 5 \\ 4 & 5 & 4 & 4 & 5 & 6 & 6 & 6 & 5 & 4 \\ 3 & 4 & 3 & 4 & 4 & 5 & 5 & 5 & 4 & 3 \\ 2 & 3 & 2 & 3 & 3 & 4 & 4 & 5 & 5 & 4 \\ 1 & 2 & 1 & 2 & 2 & 3 & 3 & 4 & 4 & 3 \end{bmatrix}$

For Exercises (7) to (10): Find the pseudoinverse of the matrix A and use it to find a least-squares solution to the (inconsistent) system $A\vec{x} = \vec{b}$:

7. $A = \begin{bmatrix} 2 & -6 & 3 & 5 \\ 1 & -3 & 0 & 1 \\ -3 & 9 & 2 & -1 \end{bmatrix}; \vec{b} = \begin{bmatrix} -1 \\ 0 \\ 4 \end{bmatrix}$

8. $A = \begin{bmatrix} 3 & 2 & -5 \\ 1 & 4 & 3 \\ -2 & -4 & 0 \\ 7 & 1 & 3 \end{bmatrix}; \vec{b} = \begin{bmatrix} -2 \\ 4 \\ -1 \\ 0 \end{bmatrix}$

9. $A = \begin{bmatrix} -2 & -1 & 1 \\ -1 & -3 & -2 \\ 3 & 4 & 9 \\ 4 & -2 & 3 \\ 6 & 3 & -4 \end{bmatrix}; \vec{b} = \begin{bmatrix} 5 \\ -2 \\ -1 \\ 3 \\ 5 \end{bmatrix}$

10. $A = \begin{bmatrix} 2 & -1 & 3 & 5 \\ -1 & 3 & 0 & 1 \\ 3 & -4 & 2 & -1 \\ 0 & 2 & -3 & 7 \\ 7 & 6 & 4 & 0 \end{bmatrix}; \vec{b} = \begin{bmatrix} 2 \\ -1 \\ -1 \\ 3 \\ -6 \end{bmatrix}$

Appendix A: The Real Number System

Linear Algebra is one attempt to generalize the notion of a real number. We developed the concepts of Euclidean spaces and other vector spaces while assuming that we all know what real numbers are, how to perform operations on them, and what their basic properties are. We took this all for granted. By the time we become educated adults, we instinctively "know" what a real number is.

This is not surprising, considering our education, at least up to this point. We learn to ***count*** using ***natural numbers*** at about the same time we learn the alphabet, colors, and the names of common animals. Soon, we learn to perform simple addition and subtraction, then later basic multiplication or our "times tables." We talk about ***even*** numbers, ***odd*** numbers, and ***prime*** numbers. We learn how to divide two natural numbers, but sometimes they don't come out exactly whole, so we learn about ***rational numbers*** or fractions, mixed numbers and decimals.

We need π to find the circumference and area of a circle, and perhaps by high school, we encounter more unusual numbers, such as $\sqrt{2}$ and Euler's number e. These are ***irrational numbers***, and π and e are special irrational numbers that are called ***transcendental***. And finally (or is it?) we can create ***imaginary*** and ***complex numbers***, built from the imaginary unit $i = \sqrt{-1}$. All these exotic animals beg the question:

Just exactly ***what*** is a ***number***?

In everyday life, numbers are ***concrete*** quantities: we need to know what our credit limit and outstanding balances are before we can buy that HDTV. We do a headcount before setting the table, watch our weight, glance at the speedometer when we spot a cop, frantically compute the average of our exams, and count down the days until we turn 21.

But can a number ***exist*** without anything at all to make them concrete? If so, then what exactly ***is*** a number? Can we make a number even if we have no fingers to count with, no rocks to chisel, no ticking clocks to mark time, nothing at all to make a number physically ***represent*** a quantity? Thanks to Mathematics, of course, the answer is ***yes***.

We can, literally, make not just ***something***, but ***everything*** out of ***nothing***.

The following construction is essentially due to John von Neumann (who is also a major figure in early Computer Science) and is called an ***Axiomatic Development*** of the real number system.

Part I. The Big Bang — How to Create Numbers

It all begins with the ***empty set***. And so, we return to the fundamental Axiom that we encountered in Chapter Zero:

Axiom — The Existence of the Empty Set: The empty set $\phi = \{\ \}$ exists.

This is a reasonable Axiom to accept because we can think of the empty set as a room with nobody in it, and we have all seen a room like that. Once we ***agree*** that the empty set exists, we will say that it ***represents*** the number zero. As usual, we will write this number as the ***symbol*** 0. Thus, technically, the number 0 and the empty set are actually ***the same thing***:

$$0 = \phi.$$

This is the key concept: in order to create numbers without attaching any physical significance to them, we will use ***sets*** to ***represent*** numbers. In the same way that light is both a ***particle*** (a photon) and a ***wave*** (a probability distribution) according to quantum mechanics, sets will have a dual purpose: to represent a ***collection*** of objects, and to represent a ***number***.

So how about the number 1? If the empty set is a good symbol to represent 0 because it has nothing in it, then we would probably want a set with ***one*** object in it to represent the number 1.

What object can we put inside this set? Well, we only know one object — the empty set. Thus we will create:

$$1 = \{\phi\}.$$

Before we move on, let us make something very clear: the symbol ϕ stands for ***the empty set***, and the symbol $\{\phi\}$ stands for ***the set containing the empty set***, and thus it is ***different*** from the empty set.

Thus, we now have two ***distinct*** objects: ϕ and $\{\phi\}$, representing, 0 and 1 respectively.

Great, we now have two distinct objects. This is convenient, because next we want to create the number 2. And so we create:

$$2 = \{\phi, \{\phi\}\}.$$

Hopefully, you can see the pattern emerging:

$$\begin{aligned}
3 &= \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\} \\
4 &= \{\phi, \{\phi\}, \{\phi, \{\phi\}\}, \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\}\} \\
5 &= \{\phi, \{\phi\}, \{\phi, \{\phi\}\}, \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\}, \\
&\quad \{\phi, \{\phi\}, \{\phi, \{\phi\}\}, \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\}\}\} \\
6 &= \ldots
\end{aligned}$$

and so on. More formally, we can ***recursively*** or ***inductively*** create every natural number: if we have created the set representing n, then we can create the set representing $n + 1$ by listing all the sets that we used to create n, followed by the ***entire set*** that we used to represent n.

We can see above, for instance, that the set representing 5 contains the four sets that we used in the set representing 4 (written on the first line), as well as the entire set that we used to represent 4 (written on the second line, since they won't all fit on one line). You are ***not*** seeing double!

Here's a little notational shorthand notation. Notice that we can write our list above as:

$$
\begin{aligned}
0 &= \phi, \\
1 &= \{\phi\} = \{0\}, \\
2 &= \{\phi, \{\phi\}\} = \{0, 1\}, \\
3 &= \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\} = \{0, 1, 2\}, \\
4 &= \{\phi, \{\phi\}, \{\phi, \{\phi\}\}, \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\}\} = \{0, 1, 2, 3\},
\end{aligned}
$$

and thus by induction, we can write:

$$n + 1 = \{0, 1, 2, 3, \dots, n\},$$

Of course, this process is a bit cumbersome, but what is important is the ***idea*** behind this construction, and the realization that we can create ***every*** natural number that we want. Thus, from out of the empty set, we have created the infinite set of ***natural numbers***:

$$\mathbb{N} = \{0, 1, 2, 3, \dots\}$$

Thanks to the evolution of human culture, we have at our disposal a sophisticated method to write or ***notate*** these natural numbers: the ***decimal (or base 10) system***, which uses the ***digits*** 0, 1, . . . 9. By assigning a weight to a place value, the number 5,127 will never be confused with 2,715, and our system allows us to count up to any natural number. There are other useful numerical systems, such as the ***binary (or base 2) system***, that uses only 0 and 1, and its cousin the ***hexadecimal (or base 16) system***, that uses the symbols 0 through 9 but also A through F to represent 10 through 15. Both systems are important in Computer Science.

Operations on $\mathbb{N}$

Now that we have the natural numbers, how will we define ***arithmetic operations***? How will we add, subtract, multiply and divide two natural numbers?

Let's see how we would define addition. This is a ***binary operation***, which means we will take two numbers x and y and produce a single number as the result or ***sum***. We will now need a new symbol for addition, which as usual will be "+". First of all, we want 0 to be our ***additive identity***, so we will agree to the following:

Axiom — The Additive Identity:
For all $x \in \mathbb{N}$ (and all other numbers that we will create in the future):

$$x + 0 = 0 + x = x.$$

How about adding two non-zero numbers? The idea is to ***reverse*** the process of creating the natural numbers. We created 5 from the number 4. We will call 5 the ***successor*** of 4, and we will call 4 the ***predecessor*** of 5. Notice that every natural number has a successor, but every natural number except 0 has a predecessor:

Definition — Successors and Predecessors in $\mathbb{N}$:

For all $x \in \mathbb{N}$, we will write the ***functions***:

$$S(x) \text{ for } x+1, \text{ the } \textbf{\textit{successor}} \text{ of } x, \text{ and}$$
$$P(x) \text{ for } x-1, \text{ the } \textbf{\textit{predecessor}} \text{ of } x, \text{ if } x \neq 0.$$

The number 0 has no predecessor (at least for now).

To define the sum $x + y$, we will ***count down predecessors*** on the second term y as we ***count up successors*** on the first term x, until y has been reduced to 0. Our Additive Identity Axiom gives us the final answer. For example:

$$5+3 = 6+2 = 7+1 = 8+0 = 8.$$

In general, we can thus define addition ***inductively***:

Axiom — Addition in $\mathbb{N}$: ***For all*** $x, y \in \mathbb{N}$:

$$\begin{aligned} x+0 &= x, \text{ and} \\ x+y &= S(x)+P(y) \\ &= (x+1)+(y-1), \text{ where } y \neq 0. \end{aligned}$$

Under this definition of addition, we can prove the following well-known properties:

Theorem — The Commutative and Associative Properties of Addition in $\mathbb{N}$:

For all $x, y, z \in \mathbb{N}$:

$$x+y = y+x, \quad \text{and}$$
$$(x+y)+z = x+(y+z).$$

Now that we know how to add two natural numbers, we can define how to ***multiply*** them. We will use the symbol "•" to represent this operation. Remember that multiplication is nothing more than ***repeated addition***. Thus, for example:

$$5 \bullet 3 = 5+5+5 = 10+5 = 15.$$

Thus, to define the product $x \bullet y$, we repeatedly add x to itself y times. Just like addition, though, we can also define this inductively. However, under this logic, a product involving 0 will have ***no terms*** in the sum (this is called an "empty sum"), which means the product is also 0. Thus we will define:

Axiom — Multiplication in $\mathbb{N}$: For all $x, y \in \mathbb{N}$:

$$\begin{aligned} x \bullet 0 &= 0, \\ x \bullet 1 &= x, \quad \text{and} \\ x \bullet S(y) &= x \bullet (y+1) \\ &= x \bullet y + x. \end{aligned}$$

Incidentally, we will adopt the usual convention in basic algebra that the operation of multiplication takes priority over addition, when both operations appear in an expression. Thus, in the right side of the final line, we multiply x and y first before adding x to this product.

Notice that we built into this definition the property that 1 is the ***multiplicative identity***. Under this definition, we can prove:

Theorem — The Commutative, Associative and Distributive Properties of Multiplication in $\mathbb{N}$:

For all $x,\ y,\ z \in \mathbb{N}$:

$$x \cdot y = y \cdot x,$$
$$(x \cdot y) \cdot z = x \cdot (y \cdot z), \quad \text{and}$$
$$x \cdot (y + z) = x \cdot y + x \cdot z.$$

The Rational Numbers $\mathbb{Q}$

We will next construct the positive rational numbers using the new symbol "/".

Axiom — Definition of Positive Rational Numbers:

We will say that a ***positive rational number*** is a ***symbol***:

$$a/b,$$

where $a,\ b \in \mathbb{N}$ and $a,\ b \neq 0$.

The set of all positive rational numbers is denoted $\mathbb{Q}^+$.

In particular, if $b = 1$, we will agree that the symbol $a/1$ represents the natural number a. Thus, as usual, every non-zero natural number is also a positive rational number.

We will also agree that $0/b = 0$, as long as $b \neq 0$.

However, we know that a fraction can sometimes be ***reduced***, and so we need to know when two symbols ***represent*** the same rational number:

Axiom — Equality in $\mathbb{Q}^+$: Suppose that $a/b,\ c/d \in \mathbb{Q}^+$. We will say that:

$$a/b = c/d \ \textbf{\textit{if and only if}}\ a \cdot d = c \cdot b.$$

As a nice by-product, we can say in particular that $a/b = c/1 = c$ if and only if $a = c \cdot b$. This allows us to decide when the natural number a is ***exactly divisible*** by the natural number b. Thus, as usual, $12/4 = 3$ because $12 = 3 \cdot 4$, and we say that "12 is divisible by $4.$" More generally, we will be able to ***divide*** two natural numbers a and b, where $b \neq 0$:

Axiom — Quotients of Natural Numbers:

Suppose that $a,\ b \in \mathbb{N}$ and $b \neq 0$. We will say that the ***quotient*** of a and b is the ***rational number*** a/b. We call this process the ***division*** of two natural numbers.

Unfortunately, one side effect now is that we need to know how to add and multiply two members of $\mathbb{Q}^+$. Again, we are thankful for what we learned in grade school. Although we were encouraged to use the least common denominator when adding fractions, we are ***not*** required to do so because all we need is a single rational number — ***any*** rational number — that will represent the sum. Thus, we will simply define our two operations as follows:

Axiom — Addition and Multiplication in $\mathbb{Q}^+$:

Suppose that $a/b,\ c/d \in \mathbb{Q}^+$. We will define:

$$a/b + c/d = (a \cdot d + b \cdot c)/(b \cdot d), \text{ and}$$
$$(a/b) \cdot (c/d) = (a \cdot c)/(b \cdot d).$$

For example, $3/4 + 5/6 = (3 \cdot 6 + 4 \cdot 5)/(4 \cdot 6) = (18 + 20)/24 = 38/24 = 19/12$.

The only operation missing now is ***subtraction***. Up to this point, though, we have ***ordered*** only the natural numbers, using the notion of successors and predecessors. We accept in everyday life that 8 is ***bigger*** than 5, and 3/4 is ***smaller*** than 9/10. To avoid any problems with subtracting a bigger number from a smaller number, we will first construct the set of negative rational numbers using, what else, but an Axiom:

Axiom — The Existence of Negative Rational Numbers:

For every $a/b \in \mathbb{Q}^+$, we will create its ***negative*** or ***additive inverse***, denoted by the ***symbol*** $-a/b$, with the property that:

$$a/b + (-a/b) = 0.$$

We will also agree that $-(-a/b) = a/b$.

At this point, we now have the set of all rational numbers, and in so doing, we have also created the set of ***negative integers***, or those rational numbers of the form $-a/1$ (where $a \in \mathbb{N}$ and a is not 0), which we will again simply denote as $-a$. As mentioned in Chapter Zero, the set of ***all integers*** is denoted as:

$$\mathbb{Z} = \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}.$$

The symbol $\mathbb{Z}$ comes from the word ***Zahlen***, which is German for "number." The set of ***all rational numbers*** is denoted by:

$$\mathbb{Q} = \left\{ p/q \,|\, p, q \in \mathbb{Z}, \text{ and } q \neq 0 \right\},$$

where $\mathbb{Q}$ comes from the word ***quotient***. Notice that we have a "nesting" of the sets of numbers we have created so far:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}.$$

All of these numbers came about from the humble empty set (and a few Axioms thrown in here and there). Once again, though, creating new numbers require us to ***expand*** our definition of arithmetic to include them. Let us do so by first defining successors and predecessor on the members of $\mathbb{Z}$:

Axiom — Successors and Predecessors in $\mathbb{Z}$:

The predecessor of 0 is -1. For all $a \in \mathbb{N}$, $a \neq 0$, we have $-a \in \mathbb{Z}$, and we define:

$$S(-a) = -P(a) = -(a-1) \text{ for the } \textbf{\textit{successor}} \text{ of } -a, \text{ and}$$

$$P(-a) = -S(a) = -(a+1) \text{ for the } \textbf{\textit{predecessor}} \text{ of } -a.$$

Note that in the definitions above, $a-1$ and $a+1$ are the predecessor and successor of a in $\mathbb{N}$, respectively.

Thus, all members of $\mathbb{Z}$, whether natural numbers or not, now have a predecessor and a successor. For example, the successor of -3 is $-(3-1) = -2$, and the predecessor of -7 is $-(7+1) = -8$. Since we know arithmetic, we can more naturally say these as $-3+1 = -2$ and $-7-1 = -8$.

We are now in a position to add any two integers:

Axiom — Addition in $\mathbb{Z}$:

Let $a,\ b \in \mathbb{N}$, $a,\ b \neq 0$. We have $-a,\ -b \in \mathbb{Z}$, and we define:

$$(-a) + (-b) = -(a+b),$$

$$(-a) + b = S(-a) + P(b), \text{ and}$$

$$a + (-b) = P(a) + S(-b),$$

where we define the sums inductively in the 2nd and 3rd lines above.

For example, we have as usual:

$$-5+3 = -4+2 = -3+1 = -2+0 = -2.$$

Now that we know how to add two integers, subtraction comes for free:

Axiom — Subtraction in $\mathbb{Z}$:

Let $x,\ y \in \mathbb{Z}$. We define:

$$x - y = x + (-y).$$

Lastly, for integers, we can define how to find any product. Since we have already done so for the members of $\mathbb{Z}$ that are in $\mathbb{N}$, we only have to worry about products involving negative integers. Aside from some special cases, we must also account for signs as we normally do:

Axiom — Multiplication in $\mathbb{Z}$:

Let $a,\ b \in \mathbb{N}$, and thus $-a,\ -b \in \mathbb{Z}$. We define:

$$(-a) \cdot 0 = 0,$$

$$(-a) \cdot 1 = -a,$$

$$(-a) \cdot b = -(a \cdot b),$$

$$a \cdot (-b) = -(a \cdot b), \quad \text{and}$$

$$(-a) \cdot (-b) = a \cdot b.$$

So, as usual, $(-4) \cdot 3 = -12$ and $(-5) \cdot (-7) = 35$.

And now, for our grand finale, we can define arithmetic in $\mathbb{Q}$:

Axiom — Equality, Addition, Subtraction, Multiplication and Division in $\mathbb{Q}$:
Suppose that $p/q,\ r/s \in \mathbb{Q}$ (thus $q,\ s \neq 0$). We will define:

$$p/q = r/s \quad \textbf{\textit{if and only if}} \quad p \cdot s = q \cdot r,$$
$$p/q + r/s = (p \cdot s + q \cdot r)/(q \cdot s),$$
$$p/q - r/s = (p \cdot s - q \cdot r)/(q \cdot s),$$
$$(p/q) \cdot (r/s) = (p \cdot r)/(q \cdot s), \text{ and}$$
$$(p/q)/(r/s) = (p \cdot s)/(q \cdot r),$$

where $r \neq 0$ in the final rule.

It certainly took a lot of work just to create all the rational numbers and the rules on how to perform basic arithmetic on them. Let us review the big steps:

1. Believe in the existence of the empty set $\phi = \{\ \}$.
2. Create the natural numbers $\mathbb{N}$ from the empty set:
$$1 = \{\phi\},\ 2 = \{\phi, \{\phi\}\},\ 3 = \{\phi, \{\phi\}, \{\phi, \{\phi\}\}\} \ldots$$
and define addition and multiplication on this set.
3. Create the positive rational numbers $\mathbb{Q}^+$, and addition and multiplication on this set. This set also allowed us to define division of numbers in $\mathbb{N}$.
4. Create the negative rational numbers, and in so doing, create the full set of rational numbers $\mathbb{Q}$ and the set of integers $\mathbb{Z}$.
5. Define arithmetic in $\mathbb{Z}$, and then in $\mathbb{Q}$.

Of course, this is only one possible way to create the rational numbers. After creating $\mathbb{N}$, we could also have created the negative integers $\{-1, -2, -3, \ldots\}$, thus creating $\mathbb{Z}$. We must then define arithmetic in $\mathbb{Z}$, and from here, create the set $\mathbb{Q}$ using quotients from $\mathbb{Z}$. We would finally define arithmetic in $\mathbb{Q}$. This recipe omits the steps where we created $\mathbb{Q}^+$ and $\mathbb{Q}^-$ separately, so if we are interested in these subsets, we will need to define them.

At this point, the story becomes a lot more complicated. The construction of the ***irrational*** numbers such as $\sqrt{2}$, π and e opens up an enormous can of worms. Several great Mathematicians were instrumental in formalizing this development, among them Karl Weierstrass, Georg Cantor, Richard Dedekind and Giuseppe Peano. One or more of the various constructions are normally discussed in a course in ***Real Analysis*** or ***Logic and Set Theory***, which are usually upper-division level Math subjects.

These complications will require a lot of time and effort, and so we end our discussion of the ***creation*** of real numbers on this note. We will continue our study of the real numbers from an ***axiomatic*** viewpoint.

Part II. The Axioms for the Real Number System

Since we reached a dead-end in Part I, we will now start fresh and ***assume*** that the set of real numbers $\mathbb{R}$ has been ***created*** for us, somehow, along with the binary operations of ***addition*** and ***multiplication*** of real numbers, represented by the symbols + and •. Furthermore, we want $\mathbb{R}$ to enjoy certain nice properties. We will classify these properties into three groups.

The Field Axioms

The first set of Axioms describe 11 properties that the set of real numbers possesses with respect to the operations of addition and multiplication:

Axioms — The Field Axioms for the Set of Real Numbers:
The set of real numbers $\mathbb{R}$ exists, together with an addition operation + and a multiplication operation •. Let $x, y, z \in \mathbb{R}$. Then the following properties are accepted to be true:

1. ***The Closure Property of Addition*** $\quad x + y \in \mathbb{R}$.
2. ***The Closure Property of Multiplication*** $\quad x \cdot y \in \mathbb{R}$.
3. ***The Commutative Property of Addition*** $\quad x + y = y + x$.
4. ***The Commutative Property of Multiplication*** $\quad x \cdot y = y \cdot x$.
5. ***The Associative Property of Addition*** $\quad x + (y + z) = (x + y) + z$.
6. ***The Associative Property of Multiplication*** $\quad x \cdot (y \cdot z) = (x \cdot y) \cdot z$.
7. ***The Distributive Property of Multiplication over Addition*** $\quad x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.
8. ***The Existence of the Additive Identity:***
   There exists $0 \in \mathbb{R}$ such that: $\quad x + 0 = x = 0 + x$.
9. ***The Existence of the Multiplicative Identity:***
   There exists $1 \in \mathbb{R}$, $1 \neq 0$, such that: $\quad x \cdot 1 = x = 1 \cdot x$.
10. ***The Existence of Additive Inverses:***
    There exists $-x \in \mathbb{R}$, such that: $\quad x + (-x) = 0 = (-x) + x$.
11. ***The Existence of Multiplicative Inverses:***
    If $x \neq 0$, ***there exists*** $1/x \in \mathbb{R}$, such that: $\quad x \cdot (1/x) = 1 = (1/x) \cdot x$.

If you look at the list carefully, you might notice something strange. In Part I, the commutative and associative properties for both addition and multiplication were ***Theorems***, that is, we can ***prove*** them to be true, given how we created the natural numbers. However, we were stuck on how to create ***all*** the real numbers, so there is certainly no guarantee that once we have the full set of real numbers, that their addition and multiplication also enjoy these two properties. Thus we will require these properties to be part of the ***Axioms*** that define the real numbers.

The existence of additive and multiplicative identities are what allow us to define the operations of subtraction and division:

$$x - y = x + (-y), \text{ and}$$

$$x/y = x \cdot (1/y), \text{ where } y \neq 0.$$

The Field Axioms can be used to derive many other important properties of the real number system. As such, we call these properties Theorems instead of Axioms: with a little bit of effort, we can prove them to be true, instead of accepting them ***on faith*** as we do with Axioms. In order to do this, though, we would need to ***formally*** know what we can do with equations. Here are the ground rules:

Axioms for Equations:

1. ***The Reflexive Property:*** If $x \in \mathbb{R}$, then $x = x$.
2. ***The Symmetric Property:*** Let $x,\ y \in \mathbb{R}$. If $x = y$, then $y = x$.
3. ***The Transitive Property:*** Let $x,\ y,\ z \in \mathbb{R}$. If $x = y$ and $y = z$, then $x = z$.
4. ***The Substitution Principle:***
   Let $x,\ y \in \mathbb{R}$. If $x = y$, and $F(x)$ is an arithmetic expression concerning x, then:
   $$F(x) = F(y).$$

These properties are commonly accepted when we work to solve an equation in algebra. The last property tells us that if x and y represent the same quantity, then performing the same arithmetic operations on x as we do on y still result in equal quantities. For example:

$$\text{If } x = y, \text{ then } x + 5 = y + 5, \text{ and } 3x - 7 = 3y - 7.$$

Let us now use our formalisms to the prove the following:

Theorem — The Uniqueness of the Additive Identity:

The number 0 is the ***unique*** member of $\mathbb{R}$ with the property that for all $x \in \mathbb{R}$:

$$x + 0 = x = 0 + x.$$

In other words, if $z \in \mathbb{R}$ is any real number such that for all $x \in \mathbb{R}$:

$$x + z = x = z + x$$

as well, then $z = 0$.

Proof: Suppose that $z \in \mathbb{R}$ is a real number such that:

$$x + z = x = z + x,$$

for all $x \in \mathbb{R}$. We must ***show*** that $z = 0$.

One obvious way to do this is to ***solve*** the equation above for z. Our algebra instincts should tell us that we should add $-x$ to all three parts of this compound equation, and get:

$$(-x) + x + z = (-x) + x = z + x + (-x).$$

This equation is true because of the Substitution Principle. We have taken the liberty to apply the Associative and Commutative Properties of Addition in order to simplify our computation. Since $(-x) + x = 0$, we get:

$$0 + z = 0 = z + 0.$$

But The Additive Identity Property tells us that $0 + z = z$, and so we get $z = 0$.

The proof is finished at this point, but we want to show a rather ***clever*** way to get to our conclusion very quickly. We again begin with the equation:

$$x + z = x = z + x,$$

which is valid for ***all*** possible $x \in \mathbb{R}$. But this means that this equation is also valid if we ***replace*** x with 0! Thus we get:

$$0 + z = 0 = z + 0,$$

and once again, since $0 + z = z$, we get $z = 0$ as we should.■

There are many important Theorems the we can derive by just using the Field Axioms, and we mention some of them below:

Theorems:

1. The Uniqueness of Additive Inverses:

For all $x \in \mathbb{R}$, its additive inverse $-x$ is ***unique***.
In other words, if $w \in \mathbb{R}$ is any real number with the property that:

$$x + w = 0 = w + x,$$

then $w = -x$.

In particular, $-0 = 0$, and more generally, $-x = (-1) \cdot x$.

2. The Uniqueness of Multiplicative Inverses:

If $x \in \mathbb{R}$ and $x \neq 0$, then its multiplicative inverse $1/x$ is ***unique***.
In other words, if $y \in \mathbb{R}$ is any real number with the property that:

$$x \cdot y = 1 = y \cdot x,$$

then $y = 1/x$.

3. The Cancellation Law for Addition:

For all $x,\ y,\ c \in \mathbb{R}$:

$$x = y \quad \textbf{\textit{if and only if}} \quad x + c = y + c.$$

4. The Cancellation Law for Multiplication:

For all $x,\ y,\ k \in \mathbb{R}$, if $k \neq 0$, then:

$$x = y \quad \textbf{\textit{if and only if}} \quad k \cdot x = k \cdot y.$$

5. The Multiplicative Property of 0: For all $x \in \mathbb{R}$:

$$0 \cdot x = 0 = x \cdot 0.$$

Consequently, 0 ***cannot*** have a multiplicative inverse.

6. The Zero-Factors Theorem:

For all $x,\ y \in \mathbb{R}$:

$$x \cdot y = 0 \quad \textbf{\textit{if and only if}} \quad x = 0 \ \ \textbf{\textit{or}} \ \ y = 0.$$

The Multiplicative Property of 0 was in fact proven in Chapter Zero using only the 11 Field Axioms. Likewise, the Zero-Products Property was also proven in Chapter Zero using the Multiplicative Property of 0 and a Case-by-Case Analysis.

The Axioms for the Positive Real Numbers and Ordering

The second group of Axioms refer to the existence and properties of positive real numbers:

Axioms for the Positive Real Numbers:
There exists a non-empty subset $\mathbb{R}^+ \subset \mathbb{R}$, consisting of the ***positive real numbers***, such that the following properties are accepted to be true:

P1. ***The Closure Property of Addition . . .*** If $x, y \in \mathbb{R}^+$, then $x + y \in \mathbb{R}^+$,

P2. . . . ***and Multiplication for*** $\mathbb{R}^+$. and $x \cdot y \in \mathbb{R}^+$.

P3. ***Zero is NOT Positive:*** $0 \notin \mathbb{R}^+$.

P4. ***The Dichotomy Property for*** $\mathbb{R}^+$ If $x \neq 0$, then either $x \in \mathbb{R}^+$, or $-x \in \mathbb{R}^+$, but ***not*** both.

Axiom P4 creates another set, $\mathbb{R}^-$, consisting of the ***negative real numbers:***

$$\mathbb{R}^- = \{x \in \mathbb{R} \mid -x \in \mathbb{R}^+\}.$$

Using only the Field Axioms, we can prove that $-0 = 0$, and thus $0 \notin \mathbb{R}^-$ either by Axiom P3. Thus, we can partition $\mathbb{R}$ into three disjoint sets, that is, sets which have no number in common: $\mathbb{R}^+$, $\mathbb{R}^-$, and $\{0\}$. Thus, 0 is neither positive nor negative.

Now, here's a legitimate question: Is 1 a positive number or a negative number? Every fiber of our being would of course answer that 1 is positive. But notice that the four Axioms above do not mention any ***specific*** member of $\mathbb{R}^+$. Thus, we have some proving to do:

Theorem: The set $\mathbb{R}^+$ is ***infinite***, and in particular, $1 \in \mathbb{R}^+$. As a consequence, $\mathbb{R}$ contains a copy of the ***integers*** $\mathbb{Z}$, and the ***rational numbers*** $\mathbb{Q}$.

Proof: Since $1 \neq 0$, either $1 \in \mathbb{R}^+$ or $-1 \in \mathbb{R}^+$, but not both, according to the Dichotomy Property. Now, let us use Proof by Contradiction. Suppose $1 \notin \mathbb{R}^+$. Then by the Dichotomy Property, we must have $-1 \in \mathbb{R}^+$. But then, by the Closure Property, $(-1) \cdot (-1) \in \mathbb{R}^+$, or in other words, $1 \in \mathbb{R}^+$. But now both 1 and -1 are in $\mathbb{R}^+$, which violates the Dichotomy Property. Thus $1 \in \mathbb{R}^+$.

Again, by the Closure Property, $1 + 1 \in \mathbb{R}^+$, that is, $2 \in \mathbb{R}^+$. Continuing inductively, $3 \in \mathbb{R}^+$, $4 \in \mathbb{R}^+$, and so on. Consequently, the set of positive integers (as we call them in everyday life) are in $\mathbb{R}^+$, and so $\mathbb{R}^+$ is an infinite set. By the Dichotomy Property, we can now conclude that the negative integers are in $\mathbb{R}^-$.

Thus the full set of integers $\mathbb{Z}$ is a subset of $\mathbb{R}$. Finally, by dividing any integer by a non-zero integer, we construct the rational numbers $\mathbb{Q}$ as a subset of $\mathbb{R}$. ■

The set $\mathbb{R}^+$ allow us to establish an ***ordering*** of the real numbers:

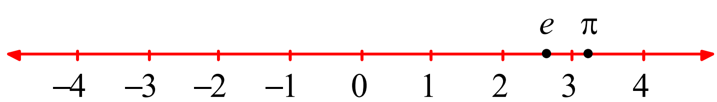
The Order Axioms:
We will say that $x > y$ (in words: x is ***greater than*** y) if $x - y \in \mathbb{R}^+$. Thus, the expression $x > y$ is either ***true*** or ***false***.

In particular, $x > 0$ ***if and only if*** $x \in \mathbb{R}^+$, that is, x is ***positive***.

Similarly, we can define the following expressions:

$$x < y \text{ means } y > x,$$
$$x \le y \text{ means } x < y \text{ or } x = y, \text{ and}$$
$$x \ge y \text{ means } x > y \text{ or } x = y.$$

This leads us to the property that $2 > 1$ because $2 - 1 = 1$ is positive. Similarly, $3 > 2$, $4 > 3$, and so on. Likewise, $-1 > -2$ because $-1 - (-2) = 1$ is positive. Similarly, $-2 > -3$, $-3 > -4$, and so on. Thus, it is perfectly natural for us to visualize the real number system using the ***number line***, as seen in Chapter Zero and in math classrooms near you, where $x > y$ ***if and only if*** x appears to the ***right*** of y on the number line:



The Real Number Line

The Field Axioms allowed us to prove Theorems concerning basic arithmetic properties of real numbers that we take for granted. Similarly, the Order Axiom and Axioms P1 through P4 allow us to prove Theorems concerning $\mathbb{R}^+$, $\mathbb{R}^-$, order properties and inequalities:

Theorems:

1. The Closure Property of $\mathbb{R}^-$ ***Under Addition:***

For all $x, y \in \mathbb{R}^-$: $x + y \in \mathbb{R}^-$.

2. The Reciprocal Property:

For all $x \in \mathbb{R}$, $x \neq 0$:

$$x \in \mathbb{R}^+ \textit{ if and only if } 1/x \in \mathbb{R}^+.$$

Consequently, an analogous statement holds for members of $\mathbb{R}^-$.

3. The Trichotomy Property:

For all $x, y \in \mathbb{R}$, ***exactly one*** of the following three possibilities is true:

$$x = y, \textit{ or } x < y, \textit{ or } y < x.$$

4. The Transitive Property of Inequalities:

For all $x, y, z \in \mathbb{R}$:

$$\textit{if } x < y \textit{ and } y < z, \textit{ then } x < z.$$

5. The Additive Property of Inequalities:

For all $x, y, z, w \in \mathbb{R}$:

$$x < y \textit{ if and only if } x + z < y + z,$$

and more generally:

$$\textit{if } x < y \textit{ and } z < w, \textit{ then } x + z < y + w.$$

6. The Multiplicative Property of Inequalities:

For all $x, y \in \mathbb{R}$ and $z \in \mathbb{R}^+$:

$$\textbf{\textit{if}}\ x < y\ \textbf{\textit{then}}\ x \cdot z < y \cdot z\ \textbf{\textit{and}}\ x \cdot (-z) > y \cdot (-z).$$

7. The Positivity Property of Products:

For all $x, y \in \mathbb{R}$:

$$x \cdot y \in \mathbb{R}^+\ \textbf{\textit{if and only if}}\ x, y \in \mathbb{R}^+\ \textbf{\textit{or}}\ x, y \in \mathbb{R}^-.$$

In particular, for all $x \in \mathbb{R}$, $x \neq 0 : x \cdot x \in \mathbb{R}^+$.

8. The Order Property of Reciprocals:

For all $x, y \in \mathbb{R}$:

$$\textbf{\textit{if}}\ 0 < x\ \textbf{\textit{and}}\ x < y,\ \textbf{\textit{then}}\ 1/y < 1/x.$$

$$\textbf{\textit{if}}\ 0 > y\ \textbf{\textit{and}}\ y > x,\ \textbf{\textit{then}}\ 1/y < 1/x.$$

9. The "Squeeze" Theorem for Real Numbers:

For all $x, y, z \in \mathbb{R}$:

$$\textbf{\textit{if}}\ x \leq y\ \textbf{\textit{and}}\ y \leq z,\ \textbf{\textit{and}}\ x = z,\ \textbf{\textit{then}}\ x = y = z.$$

In particular: ***if*** $x \leq y$ ***and*** $y \leq x$, ***then*** $x = y$.

Notice that the conclusion is ***the same*** in both cases of the Order Property of Reciprocals, although the two ***hypotheses*** are different: $0 < x < y$ in the first, and $x < y < 0$ in the second. However, the ordering of x and y is the same: $x < y$ in both cases, but we require both numbers to be positive or both to be negative. Thus, in the language of Calculus, this property simply says that the function $f(x) = 1/x$ is ***monotonic decreasing*** on the two intervals $(-\infty, 0)$ and $(0, \infty)$.

Let us prove Theorem 2, ***The Reciprocal Property:*** For all $x \in \mathbb{R}$, if $x \neq 0$, then:

$$x \in \mathbb{R}^+\ \textbf{\textit{if and only if}}\ 1/x \in \mathbb{R}^+.$$

Suppose $x \in \mathbb{R}^+$. We know from the Field Axioms that 0 has no reciprocal, and thus either $1/x \in \mathbb{R}^+$ or $-1/x \in \mathbb{R}^+$. Let us use Proof by Contradiction: Suppose $-1/x \in \mathbb{R}^+$. Then:

$$x \cdot (-1/x) \in \mathbb{R}^+,$$

by The Closure of $\mathbb{R}^+$ under Multiplication. But this product is simply -1, which we know is ***not*** a member of $\mathbb{R}^+$. We get a Contradiction, and so $1/x \in \mathbb{R}^+$. Similarly, if $x \in \mathbb{R}^-$, we must show that $1/x \in \mathbb{R}^-$ as well. Suppose $1/x \in \mathbb{R}^+$. Since $-x \in \mathbb{R}^+$, by the Closure Property once again, $(-x)(1/x) = -1 \in \mathbb{R}^+$, getting us the same contradiction (notice that in this portion of the proof, the "–" is in a different location). This tells us that $x \in \mathbb{R}^+$ ***if and only if*** $1/x \in \mathbb{R}^+$. □

Axioms P1 through P4 also tell us that the set of real numbers does ***not*** contain the imaginary unit $i = \sqrt{-1}$. If i were a real number, then by the Dichotomy Property, either i or $-i$ would be positive (since i is obviously not 0). However, if i were positive, then by the closure property for multiplication, $i \cdot i = -1$ would also be positive, which is false, since 1 is positive. Similarly, if $-i$ were positive, so is $(-i) \cdot (-i) = i \cdot i = -1$. Again, this leads to a false statement. Thus, neither i nor $-i$ is positive, which is forbidden by the Dichotomy property.

Now, in Part I, we expended a lot of effort to construct $\mathbb{Q}$, and in the process we created $\mathbb{Q}^+$. We can also easily check that the two sets of Axioms above (all 15 of them) are also satisfied by the set of rational numbers $\mathbb{Q}$, with $\mathbb{R}^+$ replaced by $\mathbb{Q}^+$. Thus, we need another ingredient that will tell us that we are dealing with a number system that does not simply consist of rational numbers. This final, and most crucial ingredient now follows:

The Completeness Axiom

Before we can state The Completeness Axiom, we need to define a few more concepts:

Definitions — Upper and Lower Bounds:

We say that a non-empty ***subset*** $A \subseteq \mathbb{R}$ is ***bounded above*** by x if for all $a \in A$: $a \leq x$. We call x an ***upper bound*** for A.

Similarly, we say that A is ***bounded below*** by y if for all $a \in A$: $a \geq y$. We call y a ***lower bound*** for A.

We say that A is ***bounded*** if it has an upper ***and*** a lower bound.

As an easy example, the set of positive numbers $\mathbb{R}^+$ is bounded below by 0, but is not bounded above. Notice that 0 itself is ***not*** a member of $\mathbb{R}^+$, but it is a lower bound for $\mathbb{R}^+$. On the other hand, the set of rational numbers $\mathbb{Q}$ is neither bounded above nor below. The set of ***prime numbers:***

$$P = \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, \dots\}$$

is certainly bounded below by 2, but it is not clear that there is an upper bound. In fact, this set is ***infinite***, so it is not bounded above.

Now, if a set A is bounded above by x, then $a \leq x$ for all $a \in A$. However, if $a \leq x$, then $a \leq x + 1$ also, so $x + 1$ is again an upper bound for A. In fact, if z is any positive number, then $x + z$ is also an upper bound for A. Thus, a set will have an infinite number of upper bounds, if it has at least one upper bound. Analogously, a set that is bounded below will have an infinite number of lower bounds. It would be nice if ***one*** of these bounds is "special." This is how we will define this special quality:

Definitions — The Least Upper Bound and The Greatest Lower Bound:

Suppose a non-empty subset $A \subset \mathbb{R}$ is bounded above. We say that x is ***the least upper bound*** of A if for any other upper bound x' for A, $x \leq x'$. This means that we ***cannot*** find another upper bound for A that is ***smaller*** than x.

Similarly, if A is bounded below, we say that y is ***the greatest lower bound*** for A if for any other lower bound y' for A, $y \geq y'$.

Think of the least upper bound as the lowest point for the bar in the "limbo dance" (if you have never heard of this dance, you can find videos of it on the Internet). The bar will go lower and lower during the dance, but there comes a point where the dancer can no longer go under the bar without knocking it down if the bar is lowered any further. That point is the least upper bound.

Notice the subtle change in the ***articles*** that we used to define these terms: we say "***an*** upper bound" but we say "***the*** least upper bound." This is because the least upper bound, by its very definition, must be ***unique***, assuming of course that it exists. A set with an upper bound ***cannot*** have two distinct least upper bounds. Similarly, a set with a lower bound cannot have two distinct greatest lower bounds:

Theorem — Uniqueness of The Least Upper Bound:

Suppose a non-empty subset $A \subset \mathbb{R}$ is bounded above. If A has a ***least*** upper bound x, then x is ***unique***.

This means that if $x^{/}$ is another least upper bound for A, then $x = x^{/}$.

Proof: By definition, both x and $x^{/}$ are upper bounds for A. Since x is a least upper bound for A, this means that $x \leq x^{/}$. But similarly, since $x^{/}$ is a least upper bound for A, then $x^{/} \leq x$. Thus $x = x^{/}$ by the special case of the Squeeze Theorem for Real Numbers. □

Now we are ready to state the final Axiom for the set of real numbers:

The Completeness Axiom:

For every non-empty subset $A \subset \mathbb{R}$ that ***has*** an upper bound x, we can find ***the least upper bound*** $\widetilde{x}$ for A.

It turns out that this final Axiom, quite literally, completes ***everything*** that we need in order to ***exactly*** create the set of real numbers. In other words, we are certain that we will get more than just the rational numbers, all the radicals, all the transcendental numbers like π and e, ***but*** we will not get a number system that is ***larger*** than the set of real numbers. Like Goldilocks, we get a system that is ***exactly right***. This is of course not easy to prove, but it is possible to do so.

Example: We will use the Completeness Axiom to ***capture*** $\sqrt{2}$. First, let us construct the set:

$$A = \{x \in \mathbb{R} \,|\, x^2 = x \cdot x \leq 2\},$$

This set is obviously non-empty, since 0 and 1 are members of A. It is also bounded above, say, by the number 2. We can show this by contradiction: suppose $x > 2$. Then: $x^2 > 2^2 = 4$, and so $x \notin A$. Thus A is bounded above by 2. Now, by the ***Completeness Axiom***, A has a least upper bound, and we can call this s. Since $1 \in A$, s must be ***positive***. This s must be $\sqrt{2}$, that is, $s^2 = 2$. We will again show that this is true by ***contradiction***.

Case 1. Suppose $s^2 < 2$. We will create a number w such that $w^2 < 2$ but $s < w$. This will show that s is not an upper bound for A, which would be a contradiction. For w, we can take: $w = (2s+2)/(s+2)$. Since s and $s+2$ are both positive, so is $s(s+2)$ by the closure property of $\mathbb{R}^+$. But then, $s(s+2) = s^2 + 2s < 2 + 2s$, and so by dividing by the positive number $s+2$, we get: $s < (2+2s)/(s+2) = w$. Thus $s < w$. Similarly:

$$4s^2 + 8s + 4 = 2s^2 + 8s + 4 + 2s^2 < 2s^2 + 8s + 4 + 4 = 2(s^2 + 4s + 4),$$

and so $(4s^2 + 8s + 4)/(s^2 + 4s + 4) < 2$. The last step is valid because the denominator is clearly positive. Basic algebra tells us that the left side is w^2. Thus we have checked that $s < w$ and $w^2 < 2$, which shows that s is not an upper bound for A. Contradiction!

Case 2. Suppose $s^2 > 2$. If we let w be the same expression above, then this time, $s^2 > 2$ and $s > w$. This is because all the inequality manipulations are still valid since all the expressions involved are positive. Thus, w is now an upper bound for A that is smaller than s. This contradicts the fact that s is the ***least*** upper bound for A.

Thus, the only possibility left is that $s^2 = 2$. □

Summary: The Full System of Axioms of The Real Number System

There exists a non-empty set of real numbers $\mathbb{R}$ as well as two binary operations: $+$ and $\cdot$. Let x, y, $z \in \mathbb{R}$. Then the following properties, the ***field axioms***, are accepted to be true:

1. ***The Closure Property of Addition*** $\quad x + y \in \mathbb{R}.$
2. ***The Closure Property of Multiplication*** $\quad x \cdot y \in \mathbb{R}.$
3. ***The Commutative Property of Addition*** $\quad x + y = y + x.$
4. ***The Commutative Property of Multiplication*** $\quad x \cdot y = y \cdot x.$
5. ***The Associative Property of Addition*** $\quad x + (y + z) = (x + y) + z.$
6. ***The Associative Property of Multiplication*** $\quad x \cdot (y \cdot z) = (x \cdot y) \cdot z.$
7. ***The Distributive Property of Multiplication over Addition*** $\quad x \cdot (y + z) = (x \cdot y) + (x \cdot z).$
8. ***The Existence of the Additive Identity:*** ***There exists*** $0 \in \mathbb{R}$ such that: $\quad x + 0 = x = 0 + x.$
9. ***The Existence of the Multiplicative Identity:*** ***There exists*** $1 \in \mathbb{R}$, $1 \neq 0$, such that: $\quad x \cdot 1 = x = 1 \cdot x.$
10. ***The Existence of Additive Inverses:*** ***There exists*** $-x \in \mathbb{R}$, such that: $\quad x + (-x) = 0 = (-x) + x.$
11. ***The Existence of Multiplicative Inverses:*** If $x \neq 0$, ***there exists*** $1/x \in \mathbb{R}$, such that: $\quad x \cdot (1/x) = 1 = (1/x) \cdot x.$

Furthermore, there exists a non-empty subset $\mathbb{R}^+ \subset \mathbb{R}$, consisting of the ***positive real numbers***, such that the following properties are also accepted to be true:

12. ***The Closure Property of Addition . . .*** $\quad$ If x, $y \in \mathbb{R}^+$, then $x + y \in \mathbb{R}^+$,
13. . . . ***and Multiplication for*** $\mathbb{R}^+$. $\quad$ and $x \cdot y \in \mathbb{R}^+$.
14. ***Zero is NOT Positive:*** $\quad 0 \notin \mathbb{R}^+.$
15. ***The Dichotomy Property for*** $\mathbb{R}^+$ $\quad$ If $x \neq 0$, then either $x \in \mathbb{R}^+$, or $-x \in \mathbb{R}^+$, but ***not*** both.

Finally, we will accept to be true:

16. ***The Completeness Axiom:*** For every subset $A \subset \mathbb{R}$ that ***has*** an upper bound x, we can find ***the least upper bound*** $\tilde{x}$ for A.

Appendix A Exercises

All of the following Theorems can be proven using only the 16 Axioms of the real number system. Justify each step of the proof (some of them are outlined for you) by stating which Axiom was used to perform each step. Write ***complete sentences*** and not just equations.

1. Prove ***The Uniqueness of Additive Inverses:*** Suppose $x \in \mathbb{R}$. If $w \in \mathbb{R}$ is any real number with the property that:

$$x + w = 0 = w + x,$$

then $w = -x$. Hint: solve for w.

2. Use the previous Exercise to show that $-0 = 0$. Hint: which of the Field Axioms tells us the value of $0 + 0$?

3. Use the Uniqueness of Additive Inverses to prove that $-x = (-1) \cdot x$. Hint: find the value of:

$$x + (-1) \cdot x.$$

4. Use the previous Exercises to prove that $-(-x) = x$.

5. Prove ***The Uniqueness of Multiplicative Inverses:*** Suppose $x \in \mathbb{R}$ and $x \neq 0$. If $y \in \mathbb{R}$ is any real number with the property that:

$$x \cdot y = 1 = y \cdot x,$$

then $y = 1/x$. Hint: solve for y.

6. Prove ***The Cancellation Law for Addition:*** For all $x,\ y,\ c \in \mathbb{R}$:

$$x = y \ \textit{\textbf{if and only if}}\ x + c = y + c.$$

   a. Suppose $x = y$. What Principle tells us that $x + c = y + c$?

   b. Now, suppose $x + c = y + c$. Prove that $x = y$. Hint: what can we do to both sides of the equation?

7. Use the ideas behind the previous exercise to prove ***The Cancellation Law for Multiplication:*** For all $x,\ y,\ k \in \mathbb{R}$, if $k \neq 0$, then:

$$x = y \ \textit{\textbf{if and only if}}\ k \cdot x = k \cdot y.$$

8. Use ***The Multiplicative Property of 0*** (which was proven in the text) to prove that 0 ***cannot*** have a multiplicative inverse. Hint: Use Proof by Contradiction: Suppose 0 ***has*** a multiplicative inverse $1/0$. What happens when we multiply it with 0?

9. Prove ***The Zero-Factors Theorem:*** For all $x,\ y \in \mathbb{R}$:

$$x \cdot y = 0 \ \textit{\textbf{if and only if}}\ x = 0 \ \textit{\textbf{or}}\ y = 0.$$

   a. Let us start with the converse, which is easier. Suppose $x = 0$ ***or*** $y = 0$. Use one of the previous exercises to explain why $x \cdot 0$ and $0 \cdot y$ are both 0.

   b. Now, suppose $x \cdot y = 0$. Perform a Case-by-Case Analysis to show that either $x = 0$ or $y = 0$. Case 1. could be $x = 0$. What should Case 2 be so that all possibilities are covered?

10. The objective of this Exercise is to study the set $\mathbb{R}^-$.

   a. Prove ***The Closure Property of*** $\mathbb{R}^-$ ***Under Addition:*** For all $x,\ y \in \mathbb{R}^-$: $x + y \in \mathbb{R}^-$. Hint: what can you say about $-x$ and $-y$?

b. On the other hand, prove that for all $x,\ y \in \mathbb{R}^-$: $x \cdot y \in \mathbb{R}^+$, and thus $\mathbb{R}^-$ is ***not*** closed under multiplication. The hint for the previous exercise is still applicable.

For Exercises (11) to (20): Use the definition that $x > y$ ***if and only if*** $x - y \in \mathbb{R}^+$ to prove the following properties.

11. Prove ***The Trichotomy Property of Inequalities:*** For all $x,\ y \in \mathbb{R}$, ***exactly one*** of the following three possibilities is true:

$$x = y,\ \ \textbf{\textit{or}}\ \ x < y,\ \ \textbf{\textit{or}}\ \ y < x.$$

Hint: Consider the expression $x - y$ and do a Case-by-Case Analysis.

12. Prove ***The Transitive Property of Inequalities:*** For all $x,\ y,\ z \in \mathbb{R}$:

$$\textbf{\textit{if}}\ \ x < y\ \ \textbf{\textit{and}}\ \ y < z,\ \ \textbf{\textit{then}}\ \ x < z.$$

13. Prove ***The Additive Property of Inequalities:*** For all $x,\ y,\ z \in \mathbb{R}$:

$$x < y\ \ \textbf{\textit{if and only if}}\ \ x + z < y + z.$$

Note that this is an if-and-only-if statement.

14. Prove more generally that: For all $x,\ y,\ z,\ w \in \mathbb{R}$:

$$\textbf{\textit{if}}\ \ x < y\ \ \textbf{\textit{and}}\ z < w,\ \ \textbf{\textit{then}}\ \ x + z < y + w.$$

15. Prove ***The Multiplicative Property of Inequalities:*** For all $x,\ y \in \mathbb{R}$ and $z \in \mathbb{R}^+$:

$$\textbf{\textit{if}}\ \ x < y\ \ \textbf{\textit{then}}\ \ x \cdot z < y \cdot z\ \ \textbf{\textit{and}}\ \ x \cdot (-z) > y \cdot (-z).$$

16. Prove ***The Positivity of Products:*** For all $x,\ y \in \mathbb{R}$:

$$x \cdot y \in \mathbb{R}^+\ \ \textbf{\textit{if and only if}}\ \ x,\ y \in \mathbb{R}^+\ \ \textbf{\textit{or}}\ \ x,\ y \in \mathbb{R}^-.$$

17. Use the previous Exercise to prove that for ***all*** $x \in \mathbb{R}$, $x \neq 0$: $x \cdot x \in \mathbb{R}^+$. Hint: do a Case-by-Case analysis.

18. Prove ***The Order Property of Reciprocals:*** Suppose that: $x,\ y \in \mathbb{R}$:

a. ***If*** $0 < x$ ***and*** $x < y$, show that $1/y < 1/x$.

b. ***If*** $0 > y$ ***and*** $y > x$, show that $1/y < 1/x$.

19. Prove ***The "Squeeze" Theorem for Real Numbers:*** For all $x,\ y,\ z \in \mathbb{R}$:

$$\textbf{\textit{if}}\ x \leq y\ \ \textbf{\textit{and}}\ \ y \leq z,\ \ \textbf{\textit{and}}\ \ x = z,\ \ \textbf{\textit{then}}\ \ x = y = z.$$

Recall that $x \leq y$ ***if and only if*** either $x < y$ ***or*** $x = y$.

20. Use the previous Exercise to prove: For all $x,\ y \in \mathbb{R}$: ***if*** $x \leq y$ ***and*** $y \leq x$, ***then*** $x = y$.

Appendix B: Logical Symbols and Truth Tables

The purpose of this brief Appendix is to introduce and discuss symbols that are often used in basic logic, and how to create and understand truth tables.

Logical Quantifiers and their Symbols

Recall that there are two logical quantifiers:

- The existential quantifier ***there exists***, which is written symbolically as $\exists$, and
- The universal quantifier ***for all***, which is written symbolically as $\forall$.

Examples: We can write the Commutative Property of Addition as:

$$\forall x \in \mathbb{R}, \forall y \in \mathbb{R}: x + y = y + x.$$

This is read as "***For all*** x, a member of $\mathbb{R}$, (and) ***for all*** y a member of $\mathbb{R}$: $x + y = y + x$."

We can write the Axiom for the Additive Identity 0 as:

$$\exists 0 \in \mathbb{R}: \forall x \in \mathbb{R}: x + 0 = x = 0 + x.$$

This is read as "***There exists*** 0, a member of $\mathbb{R}$, such that ***for all*** x that is a member of $\mathbb{R}$: $x + 0 = x = 0 + x$."

On the other hand, we can write the Axiom for the Existence of Negatives as:

$$\forall x \in \mathbb{R}, \exists -x \in \mathbb{R}: x + (-x) = 0 = (-x) + x.$$

This is read as "***For all*** x, a member of $\mathbb{R}$, ***there exists*** $-x$, also a member of $\mathbb{R}$, such that $x + (-x) = 0 = (-x) + x$. $\square$

The Negation, Conjunction and Disjunction

The basic truth tables involve the negation, conjunction and disjunction operations:

- ***not***, which is written symbolically as $\rceil$,
- ***and***, which is written symbolically as $\wedge$, and
- ***or***, which is written symbolically as $\vee$.

Naturally, we abbreviate ***true*** as ***T*** and ***false*** as ***F*** in a truth table. Our first example will be the truth table for the ***not*** operation:

p	$\rceil p$
T	F
F	T

The Truth Table For the ***not*** Operator

The table says that if p is true, then $\rceil p$ is false, and if p is false, then $\rceil p$ is true.
The truth tables for the conjunction $p \wedge q$ and the disjunction $p \vee q$ are as follows:

p	q	$p \wedge q$
T	T	T
T	F	F
F	T	F
F	F	F

The Truth Table For ***and***

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

The Truth Table For ***or***

Notice that there are exactly 4 possible combinations for the values of p and q. As we have seen before, $p \wedge q$ is true only when ***both*** p and q are true, but $p \vee q$ is true when ***either*** p or q is true.

Implications

The truth table for an implication is usually the most baffling to comprehend. Recall that an implication has the form:

$$\textbf{\textit{If}}\ p\ \textbf{\textit{then}}\ q, \quad \text{written symbolically as:}\quad p \Rightarrow q.$$

An implication is true if the conclusion q is true whenever the hypothesis p is true. Therefore, if p is false, then the implication is ***automatically*** true (this is the part that is toughest to swallow). We also say that the implication is ***vacuously satisfied***.

The key to understanding its truth table is therefore the following:

The ***only*** possible way for $p \Rightarrow q$ to be ***false*** is the case when p is ***true*** but q is ***false***.
Thus we have:

p	q	$p \Rightarrow q$
T	T	T
T	F	$\boxed{F}$
F	T	T
F	F	T

The Truth Table For An Implication

We boxed the lone $\boxed{F}$ entry for emphasis. We will use the same technique in the truth tables below that involve other implications.

Contrapositives

We can now tie up a lose end from Chapter Zero. We claimed there that an implication is always logically equivalent to its contrapositive. We said this so that we can prove an implication $p \Rightarrow q$ by proving its contrapositive $\neg q \Rightarrow \neg p$, using the technique appropriately called Proof by Contrapositive. As we can see below, the truth table for the contrapositive is exactly the same for our original implication above, and thus we are justified in saying that they are logically equivalent:

p	q	$\neg q$	$\neg p$	$\neg q \Rightarrow \neg p$
T	T	F	F	T
T	F	T	F	$\boxed{F}$
F	T	F	T	T
F	F	T	T	T

The Truth Table For The Contrapositive

Notice that the lone $\boxed{F}$ appears when $\neg q$ is ***true*** but $\neg p$ is ***false***. This is exactly the same line where p is ***true*** but q is ***false***, and so the table for $p \Rightarrow q$ is the same as the table for $\neg q \Rightarrow \neg p$.

Converses and Inverses

We can see below why the converse $q \Rightarrow p$ and the inverse $\neg p \Rightarrow \neg q$ are not logically equivalent to the original implication $p \Rightarrow q$, but they are logically equivalent to ***each other***:

p	q	q	p	$q \Rightarrow p$
T	T	T	T	T
T	F	F	T	T
F	T	T	F	$\boxed{F}$
F	F	F	F	T

The Truth Table For The Converse

p	q	$\neg p$	$\neg q$	$\neg p \Rightarrow \neg q$
T	T	F	F	T
T	F	F	T	T
F	T	T	F	$\boxed{F}$
F	F	T	T	T

The Truth Table For The Inverse

Notice that the lone $\boxed{F}$ appears on the same line for ***both*** tables, showing they are logically equivalent, but appears in a ***different*** line from the table for $p \Rightarrow q$ and its contrapositive. This shows why, in general, the converse and the inverse of a true implication are not always true.

Equivalences

An equivalence $p \Leftrightarrow q$ is true precisely when ***both*** p and q are ***true***, or when ***both*** are ***false***, and it is ***false*** for the other two combinations. Thus, its truth table is as follows:

p	q	$p \Leftrightarrow q$
T	T	T
T	F	F
F	T	F
F	F	T

The Truth Table For An Equivalence

Example: We can write the Zero-Factors Theorem for Vectors as:

$$\forall k \in \mathbb{R}, \forall \vec{v} \in V\colon k\vec{v} = \vec{0}_V \Leftrightarrow k = 0 \vee \vec{v} = \vec{\mathbf{0}}_V. \square$$

NAND and NOR

The ***nand*** and ***nor*** binary operators appear frequently in Computer Science, and so we show their truth tables below:

p	q	p ***nand*** q
T	T	F
T	F	T
F	T	T
F	F	T

The Truth Table For ***nand***

p	q	p***nor*** q
T	T	F
T	F	F
F	T	F
F	F	T

The Truth Table For ***nor***

As you may have guessed from its truth table, ***nand*** is equivalent to “not and”. We obtained its truth table from that of the ***and*** operator by taking the negation. Similarly, ***nor*** is equivalent to “not or”.

Appendix B: Exercises

For Exercises (1) to (5): Use the symbols for the logical quantifiers to write the following Axioms of the Real Number System symbolically (see Chapter Zero or Appendix A for their full statements):

1. The Closure Property of Addition.
2. The Associative Property of Multiplication.
3. The Distributive Property of Multiplication over Addition.
4. The Existence of the Multiplicative Identity.
5. The Existence of Multiplicative Inverses.
6. Prove ***de Morgan's Laws*** using truth tables: For all logical statements p and q:

   $$\textbf{\textit{not}}\ (p\ \textbf{\textit{and}}\ q) \text{ is logically equivalent to } (\textbf{\textit{not}}\ p)\ \textbf{\textit{or}}\ (\textbf{\textit{not}}\ q)$$

   and likewise:

   $$\textbf{\textit{not}}\ (p\ \textbf{\textit{or}}\ q) \text{ is logically equivalent to } (\textbf{\textit{not}}\ p)\ \textbf{\textit{and}}\ (\textbf{\textit{not}}\ q).$$

7. Notice that the truth table for the disjunction contains a lone F entry in the final column, just like the truth table for the implication. Use this to rewrite $p \Rightarrow q$ in terms of a negation and a disjunction.
8. ***Divisibility:*** We say that $a \in \mathbb{Z}$ divides $b \in \mathbb{Z}$ if $b = ka$ for another integer $k \in \mathbb{Z}$. Rewrite this definition using the universal and existential quantifiers:

   $$\forall a \in \mathbb{Z}, \forall b \in \mathbb{Z}\colon\ a \textbf{ divides } b \iff \exists \underline{\qquad} : \underline{\qquad}.$$

9. ***Prime Numbers:*** We say that $p \in \mathbb{Z}$ is a prime number if $p > 1$ and the only positive integers n that divide p are 1 and p itself. Use the previous Exercise to rewrite this definition using quantifiers.
10. ***Parity:*** Use the existential quantifier to define the concept of an even integer, and likewise to define an odd integer:

    $$\forall a \in \mathbb{Z}\colon\ a \text{ is } \textbf{\textit{even}} \iff \exists n \in \mathbb{Z}\colon\ \underline{\qquad}.$$
    $$\forall a \in \mathbb{Z}\colon\ a \text{ is } \textbf{\textit{odd}} \iff \ldots.$$

11. Use the previous Exercise to rewrite symbolically the following Theorem, which we saw in Chapter Zero, ***without*** using the word ***odd***: For all $a,\ b \in \mathbb{Z}$:

    If the product $a \cdot b$ is odd, then both a and b are odd.

12. Use quantifiers to rewrite ***Goldbach's Conjecture:*** Every ***even*** integer bigger than 2 can be expressed as the ***sum*** of two ***prime numbers***.

Symbols

Chapter Zero

$\emptyset$ or $\{\ \}$	the empty set or null set
$\in$	"an element of" or member of a set
$\mathbb{N}$	the set of natural numbers
$\mathbb{Z}$	the set of integers (from "Zahlen")
$\mathbb{Q}$	the set of rational numbers (from "quotient")
$\mathbb{R}$	the set of real numbers
not p	the negation of a logical statement p
$p \Rightarrow q$	the implication p ***implies*** q
$p \Leftrightarrow q$	the equivalence p ***if and only if*** q
p ***and*** q	the conjunction of p and q
p ***or*** q	the disjunction of p and q
$\forall$	"for all", the universal quantifier
$\exists$	"there exists," the existential quantifier
$X \subseteq Y$	X is a subset of Y
$X \cup Y$	X union Y, or the union of X and Y
$X \cap Y$	X intersection Y, or the intersection of X and Y
$X - Y$	X minus Y, or the difference between X and Y

Chapter 1

$\mathbb{R}^{\boldsymbol{n}}$	Euclidean n-space
$\vec{v} = \langle v_1, v_2, \ldots, v_n \rangle$	an arbitrary vector of $\mathbb{R}^{\boldsymbol{n}}$
$\vec{\mathbf{0}}_{\boldsymbol{n}} = \langle 0, 0, \ldots, 0 \rangle$	the zero vector of $\mathbb{R}^{\boldsymbol{n}}$
$-\vec{v}$	the negative of the vector $\vec{v}$
$\vec{u} + \vec{v}$	the sum of the vectors $\vec{u}$ and $\vec{v}$
$r \cdot \vec{v}$ or $r\vec{v}$	the scalar product of r with $\vec{v}$
$\overrightarrow{PQ}$	the vector from a point P to a point Q
$\lVert \vec{v} \rVert$	the length or norm of a vector $\vec{v}$

$x_1\vec{v}_1 + x_2\vec{v}_2 + \cdots + x_k\vec{v}_k$	a linear combination of vectors
$\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n$	the standard unit vectors in $\mathbb{R}^n$
$\vec{i}, \vec{j}$ and $\vec{k}$	the standard unit vectors in $\mathbb{R}^3$
$Span(S)$	the Span of a set of vectors S
Π	the capital Greek letter "pi," representing a plane
$\vec{u} \circ \vec{v}$	the dot product of the vectors $\vec{u}$ and $\vec{v}$
$d(\vec{u},\vec{v})$	the distance between the vectors $\vec{u}$ and $\vec{v}$
$\left[\vec{v}_1\ \vec{v}_2\ \ldots\ \vec{v}_n \mid \vec{b}\right]$	an augmented matrix
$m \times n$	"m by n," the dimension of a matrix
$R_i \longrightarrow cR_i$	multiply row i by c
$R_i \leftrightarrow R_j$	exchange row i and *row* j
$R_i \longrightarrow R_i + cR_j$	add c times row j to row i
rref	the reduced row echelon form of a matrix
$\boldsymbol{I_n}$	the $n \times n$ identity matrix
$A\vec{x}$	a matrix product with a column vector
$[\vec{v}_1\ \vec{v}_2\ \ldots\ \vec{v}_n]$	a matrix with vectors assembled in columns
$\trianglelefteq$	the subspace symbol
$\vec{b} + W$	a translate (or coset) of a subspace W
$W^{\perp}$	the orthogonal complement of a subspace W
$dim(W)$	the dimension of a space (or subspace) W
$\langle \vec{w} \rangle_S$	the coordinate vector of $\vec{w}$ with respect to the basis S

Chapter 2

$[T]$	the standard matrix of a linear transformation
$Z_{n,m}$	the zero transformation from $\mathbb{R}^n$ into $\mathbb{R}^m$
$\boldsymbol{0_{m \times n}}$	the zero $m \times n$ matrix
$\boldsymbol{I}_{\mathbb{R}^n}$	the identity operator on $\mathbb{R}^n$
$\boldsymbol{S_k}$	the scaling operator $\boldsymbol{S_k}(\vec{v}) = k\vec{v}$
rot_θ	the counterclockwise rotation in $\mathbb{R}^2$ by θ
$proj_x$	the projection operator onto the x-axis (etc.)
$refl_x$	the reflection operator across the x-axis (etc.)
$refl_\Pi$	the reflection operator in $\mathbb{R}^3$ across Π

$T_2 \circ T_1$	the composition of T_2 with T_1
$ker(T)$	the kernel of a linear transformation T
$range(T)$	the range of a linear transformation T
$nullity(T)$	the dimension of $ker(T)$
$rank(T)$	the dimension of $range(T)$
T^{-1}	the inverse of an invertible operator T
A^{-1}	the inverse of an invertible square matrix A
$Diag(d_1, d_2, \ldots, d_n)$	a diagonal matrix with diagonal entries $d_1, d_2, \ldots, d_n$
$A^\top$	the transpose of a matrix A

Chapter 3

$\oplus$	a generalized vector addition
$\odot$	a generalized scalar multiplication
$\vec{\mathbf{0}}_V$	the zero vector of the vector space V
$\mathbb{P}^n$	the space of polynomials of degree at most n
$F(I)$	the space of functions defined on an interval I
$Mat(m, n)$	the space of $m \times n$ matrices
$\mathbb{R}^+$	the space of positive numbers under multiplication and exponentiation
$\|X\|$	the cardinality of a set X
$\aleph_0$	"aleph zero" or "aleph nought," the cardinality of $\mathbb{N}$
$\aleph_1$	"aleph one," the cardinality of $\mathbb{R}$
$C(I)$	the space of continuous functions on I
$C^n(I)$	the space of n-times differentiable functions on I whose derivatives are all continuous
$C^\infty(I)$	the space of infinitely differentiable functions on I
$\boldsymbol{I}_V$	the identity operator on V
E_a	the evaluation transformation: $E_a(f) = f(a)$
$[\vec{v}]_B$	the coordinate matrix of $\vec{v}$ with respect to the basis B
$[T]_{B,B'}$	the matrix of a transformation T relative to B and B'
$[T]_B$	the matrix of an operator T relative to the basis B
$C_{B,B'}$	the change of basis matrix from B to B'

Chapter 4

$V \vee W$	the join of the subspaces V and W
$V \cap W$	the intersection of the subspaces V and W
$V \oplus W$	the direct sum of the subspaces V and W
$T(V)$	the image of a subspace V of the domain of T
$T^{-1}(W)$	the preimage of a subspace W of the codomain of T
$T\vert_U$	the restriction of T to a subspace U of the domain of T
V/W	"V modulo W," the quotient space of V modulo W

Chapter 5

$det(A)$ or $\vert A\vert$	the determinant of the square matrix A
$\sigma = (i_1, i_2, \ldots, i_n)$	"sigma," a permutation of $1 \ldots n$
$sgn(\sigma)$	the sign of the permutation σ
$M_{i,j}$	the i,j-minor of a square matrix A
$C_{i,j}$	the i,j-cofactor of a square matrix A

Chapter 6

λ	"lambda," an eigenvalue of a square matrix A
$p(\lambda)$ or $p_A(\lambda)$	the characteristic polynomial of a square matrix A
$Eig(A, \lambda)$	the eigenspace of a square matrix A associated to λ
$[B]_S$	the matrix with columns $[\vec{v}_i]_S$, where $B = \{\vec{v}_1, \ldots, \vec{v}_n\}$
$det(T)$	the determinant of an operator T
$A \backsim B$	the square matrix A is similar to B
$tr(A)$	the trace of a square matrix A
$Eig(T, \lambda)$	the eigenspace of an operator T associated to λ
$cof(A)$	the cofactor matrix of a square matrix A
$adj(A)$	the adjugate matrix of a square matrix A

Chapter 7

$\langle \vec{u} \vert \vec{v} \rangle$	the inner product of $\vec{u}$ and $\vec{v}$
$\Vert \vec{v} \Vert$	the length of $\vec{v}$: $\Vert \vec{v} \Vert = \sqrt{\langle \vec{v} \vert \vec{v} \rangle}$
$d(\vec{u}, \vec{v})$	the distance between $\vec{u}$ and $\vec{v}$: $d(\vec{u}, \vec{v}) = \Vert \vec{u} - \vec{v} \Vert$
$W^{\perp}$	the orthogonal complement of a subspace W
$proj_W$	the projection operator onto W
$refl_W$	the reflection operator across W

Chapter 8

i	the imaginary unit $\sqrt{-1}$
$\mathbb{C}$	the field of complex numbers $z = a + bi$
$\bar{z}$	the complex conjugate of z: $\bar{z} = a - bi$
$\lVert z \rVert$	the norm or length of z: $\lVert z \rVert = \sqrt{a^2 + b^2}$
$0_{\mathbb{C}}$	the complex zero: $0_{\mathbb{C}} = 0 + 0i$
$1_{\mathbb{C}}$	the complex unit: $1_{\mathbb{C}} = 1 + 0i$
$\mathbf{F}$	an arbitrary field, with zero and unit $0_{\mathbf{F}}$ and $1_{\mathbf{F}}$
$arg(z)$	the argument of z; the angle made by z with the positive real axis
$\mathbb{C}^n$	the complex vector space of all n –tuples of complex numbers
$\mathbb{P}^n(\mathbb{C})$	all polynomials of degree at most n with complex coefficients
$Mat(\mathbb{C}, m, n)$	the space of $m \times n$ matrices whose entries are from $\mathbb{C}$
$\langle \vec{z} \vert \vec{w} \rangle$	the complex inner product of $\vec{z}$ and $\vec{w}$
$Spec(T)$	the spectrum of an operator T; the set of its eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$
A^*	the Hermitian adjoint of A: $A^* = \overline{A^\top}$
T^*	the Hermitian adjoint of T: $[T^*] = [T]^*$
$Herm(n)$	the real vector space of all $n \times n$ Hermitian matrices over $\mathbb{C}$
$SkewHerm(n)$	the real vector space of all $n \times n$ Skew-Hermitian matrices over $\mathbb{C}$
$A \backsim B$	the relation A is unitarily equivalent to B

Chapter 9

Ω	"Ohm," the unit of resistance
γ	"gamma," the golden ratio $\left(1 + \sqrt{5}\right)/2$
$Q(x_1, \ldots, x_n)$	a quadratic form in n variables
$[Q]$	the matrix of the quadratic form Q
Δ	"delta," the discriminant of a binary quadratic form $Q(x,y)$
$A > 0,\ A \geq 0,$ etc.	A is a positive definite (resp. semi-definite) matrix, etc.
$Q > 0,\ Q \geq 0,$ etc.	Q is a positive definite (resp. semi-definite) quadratic form, etc.
$A^{(k)}$	the upper left $k \times k$ submatrix of A
σ_i	a singular value for A: $\sigma_i = \sqrt{\lambda_i}$, where $\lambda_i > 0$
$U \Sigma V^\top$	the singular value decomposition of A
A^+	the pseudoinverse of A: $A^+ = V \Sigma^+ U^\top$

Index

Important Axioms, Definitions, Theorems and Algorithms are in ***bold italics***.