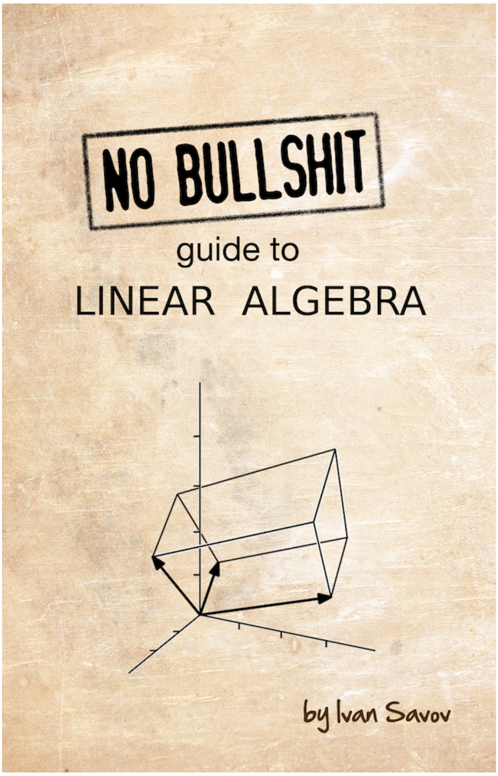
No bullshit guide to linear algebra

PREVIEW and SAMPLE CHAPTER

The full book has 550 pages and includes 128 exercises and 200 solved problems. This preview has been redacted to show only the definitions, chapter intros, and some sample problems. Buy the full book for only $29 at https://gum.co/noBSLA .



April 8, 2017

No bullshit guide to linear algebra
by Ivan Savov

Published by Minireference Co.
Montréal, Québec, Canada
minireference.com | @minireference | fb.me/noBSguide
For inquiries, contact the author at ivan@minireference.com

Second edition
v2.0
git commits preview@880:7a948a0 + preview@642:14731cb

ISBN 978-0-9920010-2-5
10 9 8 7 6 5 4 3 2 1

Contents

Concept maps

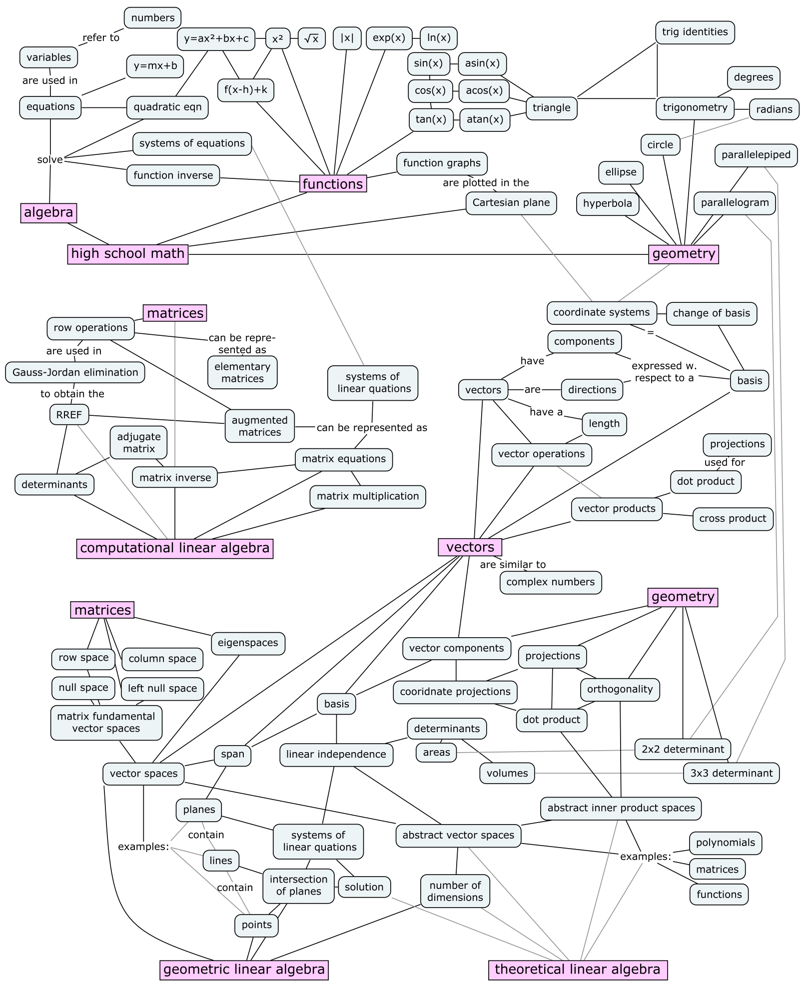


Figure 1: This concept map shows all the topics and concepts covered in this book and illustrates the connections between them. This book is all about linking the concepts together. Since it's a lot of stuff, we'll start slowly by reviewing prerequisite topics of high school math (Chapter 1), and gradually build your knowledge from there.

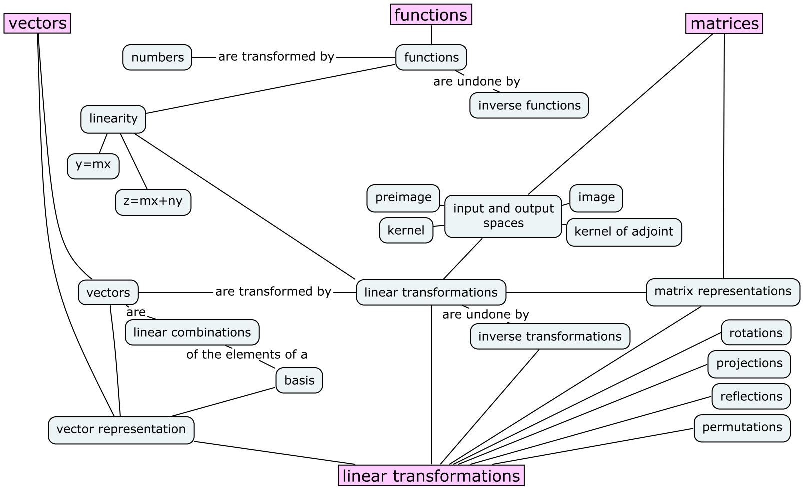


Figure 2: Chapter 5 is about linear transformations and their properties.

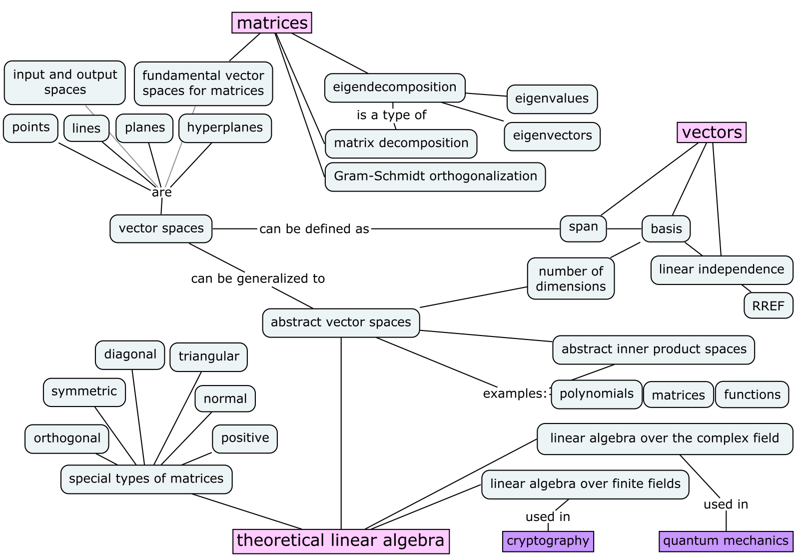


Figure 3: Chapter 6 covers theoretical aspects of linear algebra.

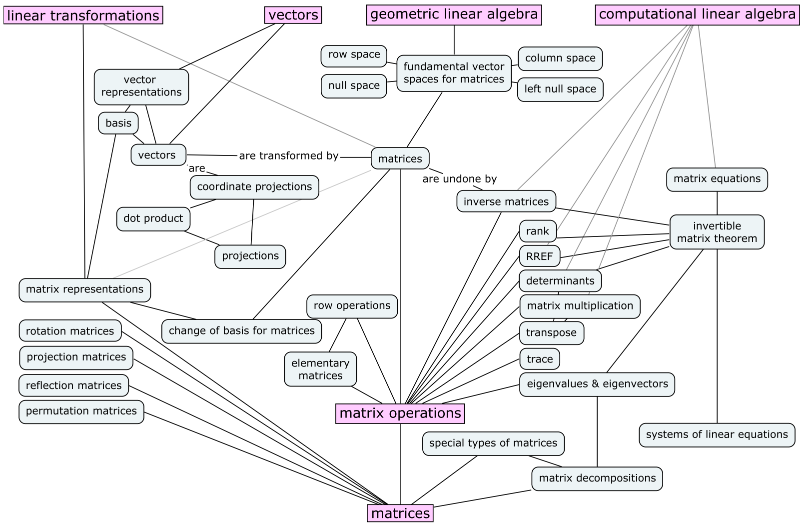


Figure 4: Matrix operations and matrix computations play an important role throughout this book. Matrices are used to represent linear transformations, systems of linear equations, and various geometric computations.

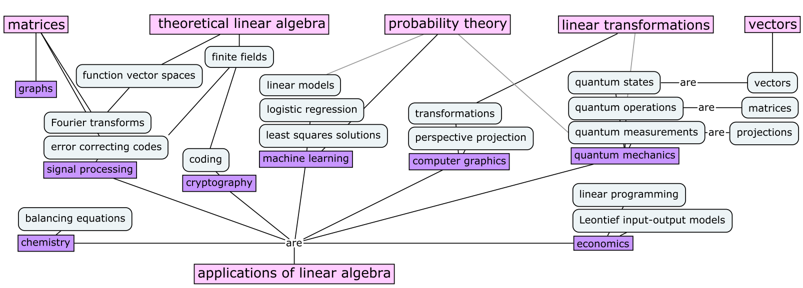


Figure 5: The book concludes with three chapters on linear algebra applications. In Chapter 7 we'll discuss applications to science, economics, business, computing, and signal processing. Chapter 8 on probability theory and Chapter 9 on quantum mechanics serve as examples of advanced subjects that you can access once you learn linear algebra.

Preface

This is a book about linear algebra and its applications. The material is presented at the level of a first-year university course, in an approachable style that cuts to the point. It covers both practical and theoretical aspects of linear algebra, with extra emphasis on explaining the connections between concepts and building a solid understanding of the material.

This book is designed to **give readers access to advanced math modelling tools** regardless of their academic background. Since the book includes all the prerequisites needed to learn linear algebra, it's suitable for readers of any skill level—including those who don't feel comfortable with fundamental math concepts.

Why learn linear algebra?

Linear algebra is one of the most fundamental and all-around useful subjects in mathematics. The practical skills learned by studying linear algebra—such as manipulating vectors and matrices—form an essential foundation for applications in physics, computer science, statistics, machine learning, and many other fields of scientific study. Learning linear algebra can also be a lot of fun. Readers will experience *knowledge buzz* as they learn about the connections between concepts, and it's not uncommon to experience mind-expanding moments while studying this subject.

The powerful concepts and tools of linear algebra form a bridge to more advanced areas of mathematics. For example, learning about *abstract vector spaces* will help students recognize the common "vector space structure" in seemingly unrelated mathematical objects like matrices, polynomials, and functions. Linear algebra techniques apply not only to standard vectors, but to *all* mathematical objects that are vector-like!

What's in this book?

Each section is a self-contained tutorial that covers the definitions, formulas, and explanations associated with a single topic. Check out the concept maps on the preceding pages to see the book's many topics and the connections between them.

The book begins with a review chapter on numbers, algebra, equations, functions, geometry, trigonometry, vectors, and complex numbers (Chapter 1). If you haven't previously studied these concepts, or if you feel your math and vector skills are a little "rusty," read these chapters and work through the exercises and problems provided. If you feel confident in your high school math abilities, jump straight to Chapter 2, where the linear algebra begins.

Chapters 3–6 cover the core topics of linear algebra: vectors, bases, analytical geometry, matrices, linear transformations, matrix representations, vector spaces, inner product spaces, eigenvectors, and matrix decompositions. These chapters contain the material required for every university-level linear algebra course. Each section contains plenty of exercises so you can test your understanding as you read; and each chapter concludes with an extensive list of problems for further practice.

Chapters 7, 8, and 9 discuss various applications of linear algebra. Though this material isn't likely to appear on any linear algebra final exam, these chapters serve to demonstrate the power of linear algebra techniques and their relevance to many areas of science. The mini-course on quantum mechanics (Chapter 9) is unique to this book. Read this chapter to understand the fascinating laws of physics that govern the behaviour of atoms and photons.

Is this book for you?

The quick pace and lively explanations in this book provide interesting reading for students and non-students alike. Whether you're learning linear algebra for a course, reviewing material as a prerequisite for more advanced topics, or generally curious about the subject, this guide will help you find your way in the land of linear algebra. The tutorial format cuts quickly and clearly to the point—because we're all busy people with no time to waste!

Students and educators can use this book as the main textbook for any university-level linear algebra course. It contains everything students need to know to prepare for a linear algebra final exam. Don't be fooled by the book's small size compared to other textbooks: it's all in here. The text is compact because we've distilled the essentials and removed the unnecessary crud.

Publisher

The genesis of the NO BULLSHIT textbook series dates back to my student days, when I was required to purchase expensive course textbooks, which were long and tedious to read. I said to myself, "Something must be done," and started a textbook company to produce textbooks that explain math and physics concepts clearly, concisely, and affordably.

The goal of `Minireference Publishing` is to fix the first-year science textbook problem: mainstream textbooks are too expensive, boring, and limited in how they teach. We're creating a better alternative—one that's redefining readers' expectations about what textbooks should be. Print-on-demand and digital distribution strategies allow us to provide readers with high-quality textbooks at reasonable prices, making advanced math and science knowledge accessible to anyone interested in learning.

The secret behind the effectiveness of the NO BULLSHIT series is the spirit of continuous improvement. All `Minireference` authors are experts with years of teaching experience who co-own the books they write. Our authors maintain a direct connection with their readers by listening and responding to feedback. The combination of skilled authors and small editorial teams equipped with a modern publishing toolchain allows us to quickly respond to the feedback we receive, constantly improving our titles.

About the author

I have been teaching math and physics for more than 15 years as a private tutor. Through teaching, I learned to explain difficult concepts by breaking complicated ideas into smaller chunks. An interesting feedback loop occurs when students learn concepts in small, manageable chunks: they experience *knowledge buzz* whenever concepts "click" into place, and this excitement motivates them to continue learning more. I know this from first-hand experience, both as a teacher and as a student. I completed my undergraduate studies in electrical engineering, then stayed on to earn a M.Sc. in physics, and a Ph.D. in computer science from McGill University. Nowadays I focus on teaching and writing effective lessons that help students and adult learners increase their math power.

Linear algebra played a central role throughout my studies. With this book, I want to share with you some of what I've learned about this expansive subject.

Ivan Savov
Montreal, 2017

Introduction

In recent years we've seen countless advances in science and technology. Modern science and engineering fields have developed advanced models for understanding the real world, predicting the outcomes of experiments, and building useful technology. Although we're still far from obtaining a "theory of everything" that can fully explain reality and predict the future, we do have a significant understanding of the natural world on many levels: physical, chemical, biological, ecological, psychological, and social. And, since mathematical models are leveraged throughout these fields of study, anyone interested in contributing to scientific and technological advances must also understand mathematics.

The linear algebra techniques you'll learn in this book are some of the most powerful mathematical modelling tools that exist. At the core of linear algebra lies a very simple idea: *linearity*. A function f is *linear* if it obeys the equation

$$f(a\mathbf{x}_1 + b\mathbf{x}_2) = af(\mathbf{x}_1) + bf(\mathbf{x}_2),$$

where $\mathbf{x}_1$ and $\mathbf{x}_2$ are any two inputs of the function. We use the term *linear combination* to describe any expression constructed from a set of variables by multiplying each variable by a constant and adding the results. In the above equation, the linear combination $a\mathbf{x}_1 + b\mathbf{x}_2$ of the inputs $\mathbf{x}_1$ and $\mathbf{x}_2$ is transformed into the linear combination $af(\mathbf{x}_1) + bf(\mathbf{x}_2)$ of the outputs of the function $f(\mathbf{x}_1)$ and $f(\mathbf{x}_2)$. **Essentially, linear functions transform linear combinations of inputs into the *same* linear combinations of outputs.** If the input to the linear function f consists of five parts $\mathbf{x}_1$ and three parts $\mathbf{x}_2$, then the output of the function will consist of five parts $f(\mathbf{x}_1)$ and three parts $f(\mathbf{x}_2)$. That's it, that's all! Now you know everything there is to know about linear algebra. The rest of the book is just details.

A significant proportion of the models used by scientists and engineers describe *linear relationships* between quantities. Scientists, engineers, statisticians, business folk, and politicians develop and use linear models to make sense of the systems they study. In fact, linear

models are often used to model even *nonlinear* (more complicated) phenomena.

There are several excellent reasons for using linear models. The first reason is that linear models are very good at *approximating* the real world. Linear models that represent nonlinear phenomena are referred to as *linear approximations.*

The second excellent reason to use linear algebra is that we can describe nonlinear phenomena by combining linear models with nonlinear transformations of the models' inputs or outputs. These techniques are often employed in machine learning: *kernel methods* are arbitrary, nonlinear transformations of the inputs of a linear model, and the *sigmoid activation curve* is used to transform the smoothly-varying output of a linear model into a hard `yes` or `no` decision, an `on` or `off` command, a `0` or `1` value, etc.

Perhaps the main reason linear models are widely used is because they are easy to describe mathematically, and easy to "fit" to real-world systems. We can obtain the parameters of a linear model for a real-world system by analyzing the system's behaviour for relatively few inputs. Let's illustrate this important point with an example.

Example You enter an art gallery. Inside, the screen of a tablet computer is being projected onto a giant wall. Anything you draw on the tablet instantly appears projected onto the wall. However, the tablet's user interface doesn't give any indication about how to hold the tablet "right side up." How can you find the correct orientation of the tablet so your drawing won't appear rotated or upside-down?

This situation is directly analogous to the tasks scientists face every day when trying to model real-world systems. The tablet's screen is a two-dimensional *input space*, and the projection is a two-dimensional *output space*. We're looking for the unknown transformation T that maps the pixels of the tablet screen (the input space) to the projection on the wall (the output space). If the unknown transformation T is a linear transformation, we can learn its parameters very quickly.

Let's describe each pixel in the input space with a pair of coordinates (x, y) and each point on the wall with another pair of coordinates (x', y'). The unknown transformation T describes the mapping of tablet coordinates to wall coordinates:

$$(x, y) \xrightarrow{T} (x', y').$$

To uncover how T transforms (x, y)-coordinates to (x', y')-coordinates, you can use the following three-step procedure. First, put a dot in the lower left corner of the tablet to represent the *origin* $(0, 0)$ of the xy-coordinate system. Observe the location where the dot appears on

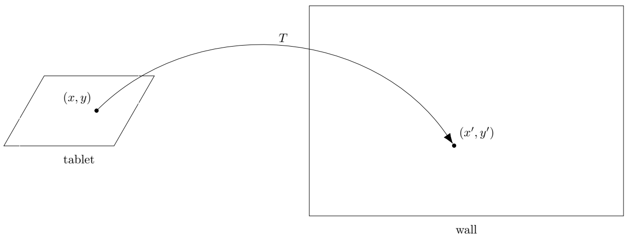


Figure 6: An unknown linear transformation T maps "tablet coordinates" to "wall coordinates." How can we characterize T?

the wall—we'll call this location the origin of the $x'y'$-coordinate system. Next, draw a short, horizontal line on the tablet to represent the x-direction $(1, 0)$, and observe the transformed $T(1, 0)$ that appears on the wall. Last, draw a vertical line in the y-direction $(0, 1)$ on the tablet, and see the transformed $T(0, 1)$ that appears on the wall. By noting how the xy-coordinate system is mapped to the $x'y'$-coordinate system, you can determine the orientation in which you must hold the tablet so your drawing appears upright when projected. **Knowing the outputs of a linear transformation T for all "directions" in its input space is a complete characterization of T.**

In the case of the tablet and the wall, we're looking for an unknown transformation T from a two-dimensional input space to a two-dimensional output space. Since T is a linear transformation, it's possible to completely describe T with only two lines (one line for each dimension). Let's look at the math to see why this is true. Can you predict what will appear on the wall if you draw an angled line on the tablet in the $(2, 3)$-direction? First, locate the point $(2, 3)$ in the input space by moving 2 units in the x-direction and 3 units in the y-direction: $(2, 3) = (2, 0) + (0, 3) = 2(1, 0) + 3(0, 1)$. Then, using the fact that T is a linear transformation, we can predict the output of the transformation when the input is $(2, 3)$:

$$T(2, 3) = T(2(1, 0) + 3(0, 1)) = 2T(1, 0) + 3T(0, 1).$$

The projection of the diagonal line in the $(2, 3)$-direction will have a length equal to 2 times the unit x-direction output $T(1, 0)$ plus 3 times the unit y-direction output $T(0, 1)$. Knowing the outputs of the two lines $T(1, 0)$ and $T(0, 1)$ is sufficient to determine the linear transformation's output for any input (a, b). Any input (a, b) can be expressed as a linear combination: $(a, b) = a(1, 0) + b(0, 1)$. The corresponding output will be $T(a, b) = aT(1, 0) + bT(0, 1)$. Since we know $T(1, 0)$ and $T(0, 1)$, we can calculate $T(a, b)$.

Don't worry if you can't follow *all* the math in this example. It's

the concepts that are essential right now, and we'll have plenty of time to work on the math behind the concepts in the rest of the book.

TL;DR Linearity allows us to analyze multidimensional processes and transformations by studying their effects on a small set of inputs. This is the essential reason linear models are so prominent in science. Probing a linear system with each "input direction" is enough to completely characterize the system. Without this linear structure, characterizing unknown input-output systems is a much harder task. Linear algebra is the study of linear structure, in all its details. The theoretical results and computational procedures you'll learn apply to all things linear and vector-like.

Linear transformations

You can think of linear transformations as "vector functions" and understand their properties as analogous to the properties of the regular functions you're familiar with. The action of a function on a number is similar to the action of a linear transformation on a vector:

$$\begin{aligned}
\text{function } f : \mathbb{R} \to \mathbb{R} \quad &\Leftrightarrow \quad \text{linear transformation } T : \mathbb{R}^n \to \mathbb{R}^m \\
\text{input } x \in \mathbb{R} \quad &\Leftrightarrow \quad \text{input } \vec{x} \in \mathbb{R}^n \\
\text{output } f(x) \in \mathbb{R} \quad &\Leftrightarrow \quad \text{output } T(\vec{x}) \in \mathbb{R}^m \\
\text{inverse function } f^{-1} \quad &\Leftrightarrow \quad \text{inverse transformation } T^{-1} \\
\text{zeros of } f \quad &\Leftrightarrow \quad \text{kernel of } T
\end{aligned}$$

Studying linear algebra will expose you to many topics associated with linear transformations. You'll learn about concepts like vector spaces, projections, and orthogonalization procedures. Indeed, a first linear algebra course introduces many advanced, abstract ideas; yet all the new ideas you'll encounter can be seen as extensions of ideas you're already familiar with. Linear algebra is the vector-upgrade to your high school knowledge of functions.

Prerequisites

To understand linear algebra, you must have some preliminary knowledge of fundamental math concepts like numbers, equations, and functions. For example, you should be able to tell me the meaning of the parameters m and b in the equation $f(x) = mx + b$. If you do not feel confident about your basic math skills, don't worry. Chapter 1 is a prerequisites chapter specially designed to help bring you quickly up to speed on the material of high school math. It also contains a

short summary of vectors concepts usually taught in the first week of Physics 101, and a section on complex numbers (Section 1.16). You should read about complex numbers at some point because we'll use complex numbers in Section 6.7 later in the book.

Executive summary

The book is organized into 10 chapters. Chapters 2–6 are the core of linear algebra. Chapters 7 through 9 contain optional reading about linear algebra applications. The concept maps on pages v, vi, and vii illustrate the connections between the topics we'll cover. I know the maps are teeming with concepts, but don't worry—the book is split into tiny chunks, and we'll navigate the material step by step. It will be like Mario World, but in n dimensions and with a lot of bonus levels.

Chapter 2 is an introduction to the subject of linear algebra. Linear algebra is the math of vectors and matrices, so we'll start by defining the mathematical operations we can perform on vectors and matrices.

In Chapter 3, we'll tackle the computational aspects of linear algebra. By the end of this course, you'll know how to solve systems of equations, transform a matrix into its reduced row echelon form, compute the product of two matrices, and find the determinant and the inverse of a square matrix. Each of these computational tasks can be tedious to carry out by hand and can require lots of steps. There is no way around this; we must do the grunt work before we get to the cool stuff.

In Chapter 4, we'll review the properties and the equations that describe basic geometric objects like points, lines, and planes. We'll learn how to compute projections onto vectors, projections onto planes, and distances between objects. We'll also review the meaning of vector coordinates, which are lengths measured with respect to a basis. We'll learn about linear combinations of vectors, the span of a set of vectors, and formally define what a vector space is. In Section 4.5, we'll learn how to use the *reduced row echelon form* of a matrix, to describe the fundamental spaces associated with the matrix.

Chapter 5 is about linear transformations. Armed with the computational tools from Chapter 3 and the geometric intuition from Chapter 4, we can tackle the core subject of linear algebra: linear transformations. We'll explore in detail the correspondence between linear transformations ($T : \mathbb{R}^n \to \mathbb{R}^m$) and their representation as $m \times n$ matrices. We'll also learn how the coefficients in a matrix representation depend on the choice of basis for the input and output spaces of the transformation. Section 5.4 on the invertible ma-

trix theorem serves as a midway checkpoint for your understanding of linear algebra. This theorem connects several seemingly disparate concepts: reduced row echelon forms, matrix inverses, row spaces, column spaces, and determinants. The invertible matrix theorem links all these concepts and highlights the properties of invertible linear transformations that distinguish them from non-invertible transformations. Invertible transformations are one-to-one correspondences (bijections) between vectors in the input space and vectors in the output space.

Chapter 6 covers more advanced theoretical topics of linear algebra. We'll define the eigenvalues and the eigenvectors of a square matrix. We'll see how the eigenvalues of a matrix tell us important information about the properties of the matrix. We'll learn about some special names given to different types of matrices, based on the properties of their eigenvalues. In Section 6.3 we'll learn about abstract vector spaces. Abstract vectors are mathematical objects that—like vectors—have components and can be scaled, added, and subtracted by manipulating their components. Section 6.7 will discuss linear algebra with complex numbers. Instead of working with vectors with real coefficients, we'll see how to do linear algebra with vectors that have complex coefficients. This section serves as a review of all the material in the book. We'll revisit all the key concepts and find out how they are affected when working with complex numbers.

In Chapter 7, we'll discuss the applications of linear algebra. If you've done your job learning the material in the first seven chapters, you'll get to learn all the cool things you can do with linear algebra. Chapter 8 will introduce the basic concepts of probability theory. Chapter 9 contains an introduction to quantum mechanics.

The sections in the book are self-contained so you can read them in any order. Feel free to skip ahead to the parts that you want to learn first. That being said, the material is ordered to provide an optimal knowing-what-you-need-to-know-before-learning-what-you-want-to-know experience. If you're new to linear algebra, it would be best to read everything in order. If you find yourself stuck on a concept at some point, refer to the concept maps to see if you're missing some prerequisites and flip to the section of the book that will help you fill in your knowledge gap accordingly.

Difficulty level

In terms of difficulty, I must prepare you to get ready for some serious uphill pushes. As your personal "trail guide" up the mountain of linear algebra, it's my obligation to warn you about the difficulties that lie ahead, so that you can mentally prepare for a good challenge.

Linear algebra is a difficult subject because it requires developing your computational skills, your geometric intuition, and your abstract thinking. The computational aspects of linear algebra are not particularly difficult, but they can be boring and repetitive. You'll have to carry out hundreds of steps of basic arithmetic. The geometric problems you'll encounter in Chapter 4 can be tough at first, but they'll get easier once you learn to draw diagrams and develop your geometric reasoning. The theoretical aspects of linear algebra are difficult because they require a new way of thinking, which resembles what doing "real math" is like. You must not only understand and use the material; you must also know how to *prove* mathematical statements using the definitions and properties of math objects.

In summary, much toil awaits you as you learn the concepts of linear algebra, but the effort is totally worth it. All the brain sweat you put into understanding vectors and matrices will lead to mind-expanding insights. You will reap the benefits of your efforts for the rest of your life as your knowledge of linear algebra opens many doors for you.

Chapter 1

Math fundamentals

In this chapter we'll review the fundamental ideas of mathematics which are the prerequisites for learning linear algebra. We'll define the different types of numbers and the concept of a function, which is a transformation that takes numbers as inputs and produces numbers as outputs. Linear algebra is the extension of these ideas to many dimensions: instead of "doing math" with numbers and functions, in linear algebra we'll be "doing math" with vectors and linear transformations.

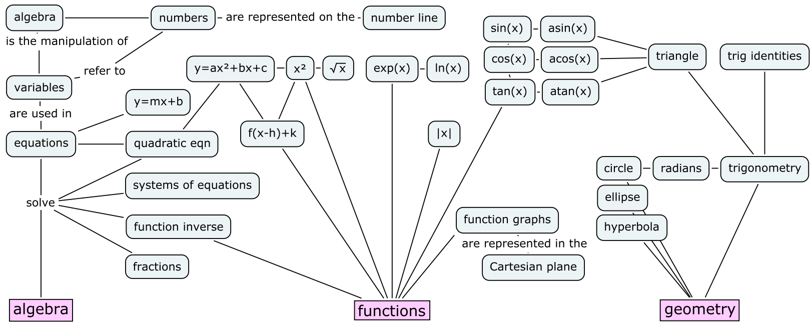


Figure 1.1: A concept map showing the mathematical topics covered in this chapter. We'll learn how to solve equations using algebra, how to model the world using functions, and some important facts about geometry. The material in this chapter is required for your understanding of the more advanced topics in this book.

1.1 Solving equations

Most math skills boil down to being able to manipulate and solve equations. Solving an equation means finding the value of the unknown in the equation.

Check this shit out:

$$x^2 - 4 = 45.$$

To solve the above equation is to answer the question "What is x?" More precisely, we want to find the number that can take the place of x in the equation so that the equality holds. In other words, we're asking,

"Which number times itself minus four gives 45?"

That is quite a mouthful, don't you think? To remedy this verbosity, mathematicians often use specialized symbols to describe math operations. The problem is that these specialized symbols can be very confusing. Sometimes even the simplest math concepts are inaccessible if you don't know what the symbols mean.

What are your feelings about math, dear reader? Are you afraid of it? Do you have anxiety attacks because you think it will be too difficult for you? Chill! Relax, my brothers and sisters. There's nothing to it. Nobody can magically guess the solution to an equation immediately. To find the solution, you must break the problem into simpler steps.

To find x, we can manipulate the original equation, transforming it into a different equation (as true as the first) that looks like this:

$$x \;=\; \text{only numbers}.$$

That's what it means to *solve*. The equation is solved because you can type the numbers on the right-hand side of the equation into a calculator and obtain the numerical value of x that you're seeking.

By the way, before we continue our discussion, let it be noted: the equality symbol ($=$) means that all that is to the left of $=$ is equal to all that is to the right of $=$. To keep this equality statement true, **for every change you apply to the left side of the equation, you must apply the same change to the right side of the equation**.

To find x, we need to correctly manipulate the original equation into its final form, simplifying it in each step. The only requirement is that the manipulations we make transform one true equation into another true equation. Looking at our earlier example, the first simplifying step is to add the number four to both sides of the equation:

$$x^2 - 4 + 4 = 45 + 4,$$

which simplifies to

$$x^2 = 49.$$

Now the expression looks simpler, yes? How did I know to perform this operation? I was trying to "undo" the effects of the operation -4. We undo an operation by applying its *inverse*. In the case where the operation is the subtraction of some amount, the inverse operation is the addition of the same amount. We'll learn more about function inverses in Section 1.4 (page 12).

We're getting closer to our goal, namely to *isolate* x on one side of the equation, leaving only numbers on the other side. The next step is to undo the square x^2 operation. The inverse operation of squaring a number x^2 is to take the square root $\sqrt{\ }$ so this is what we'll do next. We obtain

$$\sqrt{x^2} = \sqrt{49}.$$

Notice how we applied the square root to both sides of the equation? If we don't apply the same operation to both sides, we'll break the equality!

The equation $\sqrt{x^2} = \sqrt{49}$ simplifies to

$$|x| = 7.$$

What's up with the vertical bars around x? The notation $|x|$ stands for the *absolute value* of x, which is the same as x except we ignore the sign. For example $|5| = 5$ and $|-5| = 5$, too. The equation $|x| = 7$ indicates that both $x = 7$ and $x = -7$ satisfy the equation $x^2 = 49$. Seven squared is 49, and so is $(-7)^2 = 49$ because two negatives cancel each other out.

We're done since we isolated x. The final solutions are

$$x = 7 \qquad \text{or} \qquad x = -7.$$

Yes, there are *two* possible answers. You can check that both of the above values of x satisfy the initial equation $x^2 - 4 = 45$.

If you are comfortable with all the notions of high school math and you feel you could have solved the equation $x^2 - 4 = 45$ on your own, then you should consider skipping ahead to Chapter **??**. If on the other hand you are wondering how the squiggle killed the power two, then this chapter is for you! In the following sections we will review all the essential concepts from high school math that you will need to power through the rest of this book. First, let me tell you about the different kinds of numbers.

1.2 Numbers

Definitions

Numbers are the basic objects we use to calculate things. Mathematicians like to classify the different kinds of number-like objects into *sets*:

- The natural numbers: $\mathbb{N} = \{0, 1, 2, 3, 4, 5, 6, 7, \ldots\}$
- The integers: $\mathbb{Z} = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$
- The rational numbers: $\mathbb{Q} = \{\frac{5}{3}, \frac{22}{7}, 1.5, 0.125, -7, \ldots\}$
- The real numbers: $\mathbb{R} = \{-1, 0, 1, \sqrt{2}, e, \pi, \ 4.94\ldots, \ \ldots\}$
- The complex numbers: $\mathbb{C} = \{-1, 0, 1, i, 1 + i, 2 + 3i, \ldots\}$

Operations on numbers

Addition

Multiplication

Division

Exponentiation

Operator precedence

Exercises

Other operations

1.3 Variables

Variable names

There are common naming patterns for variables:

- x: general name for the unknown in equations (also used to denote a function's input, as well as an object's position in physics problems)
- v: velocity in physics problems
- θ, φ: the Greek letters *theta* and *phi* are used to denote angles
- x_i, x_f: denote an object's initial and final positions in physics problems
- X: a random variable in probability theory
- C: costs in business along with P for profit, and R for revenue

Variable substitution

Compact notation

1.4 Functions and their inverses

As we saw in the section on solving equations, the ability to "undo" functions is a key skill for solving equations.

Example Suppose we're solving for x in the equation

$$f(x) = c,$$

where f is some function and c is some constant. Our goal is to isolate x on one side of the equation, but the function f stands in our way.

By using the inverse function (denoted f^{-1}) we "undo" the effects of f. We apply the inverse function f^{-1} to both sides of the equation to obtain

$$f^{-1}(f(x)) = x = f^{-1}(c).$$

By definition, the inverse function f^{-1} performs the opposite action of the function f, so together the two functions cancel each other out. We have $f^{-1}(f(x)) = x$ for any number x.

Provided everything is kosher (the function f^{-1} must be defined for the input c), the manipulation we made above is valid and we have obtained the answer $x = f^{-1}(c)$.

The above example introduces the notation f^{-1} for denoting the function's *inverse*. This notation is borrowed from the notion of inverse numbers: multiplication by the number a^{-1} is the inverse operation of multiplication by the number a: $a^{-1}ax = 1x = x$. In the case of functions, however, the negative-one exponent does not refer to "one over-$f(x)$" as in $\frac{1}{f(x)} = (f(x))^{-1}$; rather, it refers to the function's inverse. In other words, the number $f^{-1}(y)$ is equal to the number x such that $f(x) = y$.

Be careful: sometimes applying the inverse leads to multiple solutions. For example, the function $f(x) = x^2$ maps two input values (x and $-x$) to the same output value $x^2 = f(x) = f(-x)$. The inverse function of $f(x) = x^2$ is $f^{-1}(x) = \sqrt{x}$, but both $x = +\sqrt{c}$ and $x = -\sqrt{c}$ are solutions to the equation $x^2 = c$. In this case, this equation's solutions can be indicated in shorthand notation as $x = \pm\sqrt{c}$.

Formulas

Here is a list of common functions and their inverses:

$$\text{function } f(x) \quad \Leftrightarrow \quad \text{inverse } f^{-1}(x)$$

$$\begin{aligned}
x+2 &\Leftrightarrow x-2\\
2x &\Leftrightarrow \tfrac{1}{2}x\\
-1x &\Leftrightarrow -1x\\
x^2 &\Leftrightarrow \pm\sqrt{x}\\
2^x &\Leftrightarrow \log_2(x)\\
3x+5 &\Leftrightarrow \tfrac{1}{3}(x-5)\\
a^x &\Leftrightarrow \log_a(x)\\
\exp(x)\equiv e^x &\Leftrightarrow \ln(x)\equiv\log_e(x)\\
\sin(x) &\Leftrightarrow \sin^{-1}(x)\equiv\arcsin(x)\\
\cos(x) &\Leftrightarrow \cos^{-1}(x)\equiv\arccos(x)
\end{aligned}$$

The function-inverse relationship is *symmetric*—if you see a function on one side of the above table (pick a side, any side), you'll find its inverse on the opposite side.

Example

Let's say your teacher doesn't like you and right away, on the first day of class, he gives you a serious equation and tells you to find x:

$$\log_5\left(3+\sqrt{6\sqrt{x}-7}\right) = 34+\sin(5.5)-\Psi(1).$$

See what I mean when I say the teacher doesn't like you?

First, note that it doesn't matter what Ψ (the capital Greek letter *psi*) is, since x is on the other side of the equation. You can keep copying $\Psi(1)$ from line to line, until the end, when you throw the ball back to the teacher. "My answer is in terms of *your* variables, dude. *You* go figure out what the hell Ψ is since you brought it up in the first place!" By the way, it's not actually recommended to quote me verbatim should a situation like this arise. The same goes with $\sin(5.5)$. If you don't have a calculator handy, don't worry about it. Keep the expression $\sin(5.5)$ instead of trying to find its numerical value. In general, try to work with variables as much as possible and leave the numerical computations for the last step.

Okay, enough beating about the bush. Let's just find x and get it over with! On the right-hand side of the equation, we have the sum of a bunch of terms with no x in them, so we'll leave them as they are. On the left-hand side, the outermost function is a logarithm base 5. Cool. Looking at the table of inverse functions, we find the exponential function is the inverse of the logarithm: $a^x \Leftrightarrow \log_a(x)$.

To get rid of $\log_5$, we must apply the exponential function base 5 to both sides:

$$5^{\log_5\left(3+\sqrt{6\sqrt{x}-7}\right)} = 5^{34+\sin(5.5)-\Psi(1)},$$

which simplifies to

$$3+\sqrt{6\sqrt{x}-7} = 5^{34+\sin(5.5)-\Psi(1)},$$

since 5^x cancels $\log_5 x$.

From here on, it is going to be as if Bruce Lee walked into a place with lots of bad guys. Addition of 3 is undone by subtracting 3 on both sides:

$$\sqrt{6\sqrt{x}-7} = 5^{34+\sin(5.5)-\Psi(1)} - 3.$$

To undo a square root we take the square:

$$6\sqrt{x}-7 = \left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2.$$

Add 7 to both sides,

$$6\sqrt{x} = \left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2 + 7,$$

divide by 6

$$\sqrt{x} = \frac{1}{6}\left(\left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2 + 7\right),$$

and square again to find the final answer:

$$x = \left[\frac{1}{6}\left(\left(5^{34+\sin(5.5)-\Psi(1)} - 3\right)^2 + 7\right)\right]^2.$$

Did you see what I was doing in each step? Next time a function stands in your way, hit it with its inverse so it knows not to challenge you ever again.

Discussion

The recipe I have outlined above is not universally applicable. Sometimes x isn't alone on one side. Sometimes x appears in several places in the same equation. In these cases, you can't effortlessly work your way, Bruce Lee-style, clearing bad guys and digging toward x—you need other techniques.

The bad news is there's no general formula for solving complicated equations. The good news is the above technique of "digging toward

the x" is sufficient for 80% of what you are going to be doing. You can get another 15% if you learn how to solve the quadratic equation:

$$ax^2 + bx + c = 0.$$

Solving third-degree polynomial equations like $ax^3 + bx^2 + cx + d = 0$ with pen and paper is also possible, but at this point you might as well start using a computer to solve for the unknowns.

There are all kinds of other equations you can learn how to solve: equations with multiple variables, equations with logarithms, equations with exponentials, and equations with trigonometric functions. The principle of "digging" toward the unknown by applying inverse functions is the key for solving all these types of equations, so be sure to practice using it.

Exercises

E1.1 Solve for x in the following equations:

a) $3x = 6$ **b)** $\log_5(x) = 2$ **c)** $\log_{10}(\sqrt{x}) = 1$

E1.2 Find the function inverse and use it to solve the problems.

a) Solve the equation $f(x) = 4$, where $f(x) \equiv \sqrt{x}$.
b) Solve for x in the equation $g(x) = 1$, given $g(x) \equiv e^{-2x}$.

1.5 Basic rules of algebra

Given any four numbers a, b, c, and d, we can apply the following algebraic properties:

1. Associative property: $a + b + c = (a + b) + c = a + (b + c)$ and $abc = (ab)c = a(bc)$
2. Commutative property: $a + b = b + a$ and $ab = ba$
3. Distributive property: $a(b + c) = ab + ac$

Expanding brackets

Factoring

Quadratic factoring

Completing the square

Exercises

1.6 Solving quadratic equations

Claim

The solutions to the equation $ax^2 + bx + c = 0$ are

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \qquad \text{and} \qquad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}.$$

Proof of claim

Alternative proof of claim

Applications

The golden ratio

Explanations

Multiple solutions

Relation to factoring

Exercises

1.7 The Cartesian plane

Named after famous philosopher and mathematician René Descartes, the Cartesian plane is a graphical representation for *pairs* of numbers.

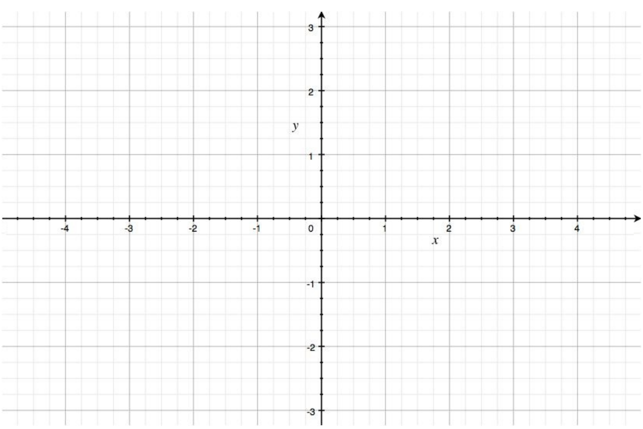


Figure 1.2: The (x, y)-coordinate system, which is also known as the Cartesian plane. Points $P = (P_x, P_y)$, vectors $\vec{v} = (v_x, v_y)$, and graphs of functions $(x, f(x))$ live here.

Vectors and points

Graphs of functions

Discussion

1.8 Functions

We need to have a relationship talk. We need to talk about functions. We use functions to describe the relationships between variables. In particular, functions describe how one variable *depends* on another.

Definitions

A *function* is a mathematical object that takes numbers as inputs and gives numbers as outputs. We use the notation

$$f\colon A \to B$$

to denote a function from the input set A to the output set B. In this book, we mostly study functions that take real numbers as inputs and give real numbers as outputs: $f\colon \mathbb{R} \to \mathbb{R}$.

We'll now define some fancy technical terms used to describe the input and output sets of functions.

- The *domain* of a function is the set of allowed input values.
- The *image* or *range* of the function f is the set of all possible output values of the function.
- The *codomain* of a function describes the type of outputs the function has.

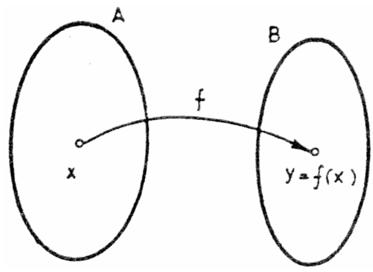
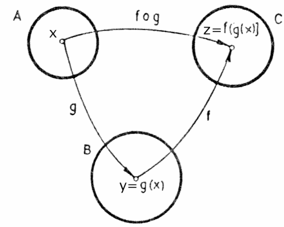


Figure 1.3: An abstract representation of a function f from the set A to the set B. The function f is the arrow which *maps* each input x in A to an output $f(x)$ in B. The output of the function $f(x)$ is also denoted y.

Function composition

We can combine two simple functions by chaining them together to build a more complicated function. This act of applying one function after another is called *function composition.* Consider for example the composition:

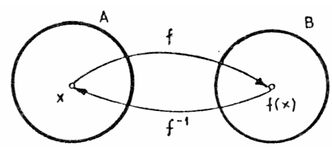


$$f \circ g\,(x) \equiv f(\,g(x)\,) = z.$$

The diagram on the right illustrates what is going on. First, the function $g : A \to B$ acts on some input x to produce an intermediary value $y = g(x)$ in the set B. The intermediary value y is then passed through the function $f : B \to C$ to produce the final output value $z = f(y) = f(g(x))$ in the set C. We can think of the *composite function* $f \circ g$ as a function in its own right. The function $f \circ g : A \to C$ is defined through the formula $f \circ g\,(x) \equiv f(g(x))$.

Inverse function

Recall that a *bijective* function is a one-to-one correspondence between a set of input values and a set of output values. Given a bijective function $f : A \to B$, there exists an inverse function $f^{-1} : B \to A$, which performs the *inverse mapping* of f. If you start from some x, apply f, and then apply f^{-1}, you'll arrive—full circle—back to the original input x:



$$f^{-1}(\,f(x)\,) \equiv f^{-1} \circ f\,(x) = x.$$

This inverse function is represented abstractly as a backward arrow, and it puts the value $f(x)$ back to the x it came from.

Function names

Handles on functions

Table of values

Function graph

Facts and properties

Example

Example 2

Discussion

1.9 Functions reference

Line

Graph

Properties

General equation

Square

Properties

Square root

The square root function is denoted

$$f(x) = \sqrt{x} \equiv x^{\frac{1}{2}}.$$

The square root $\sqrt{x}$ is the inverse function of the square function x^2 for $x \geq 0$. The symbol $\sqrt{c}$ refers to the *positive* solution of $x^2 = c$. Note that $-\sqrt{c}$ is also a solution of $x^2 = c$.

Graph

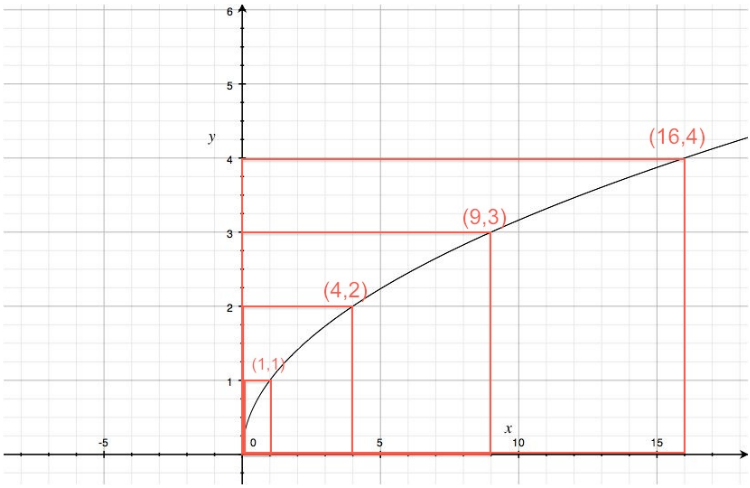


Figure 1.4: The graph of the function $f(x) = \sqrt{x}$. The domain of the function is $x \in [0, \infty)$. You can't take the square root of a negative number.

Properties

- Domain: $x \in [0, \infty)$.
  The function $f(x) = \sqrt{x}$ is only defined for nonnegative inputs $x \geq 0$. There is no real number y such that y^2 is negative, hence the function $f(x) = \sqrt{x}$ is not defined for negative inputs x.
- Image: $f(x) \in [0, \infty)$.
  The outputs of the function $f(x) = \sqrt{x}$ are never negative: $\sqrt{x} \geq 0$, for all $x \in [0, \infty)$.

In addition to *square* root, there is also *cube* root $f(x) = \sqrt[3]{x} \equiv x^{\frac{1}{3}}$, which is the inverse function for the cubic function $f(x) = x^3$. We have $\sqrt[3]{8} = 2$ since $2 \times 2 \times 2 = 8$. More generally, we can define the n^{th}-root function $\sqrt[n]{x}$ as the inverse function of x^n.

Absolute value

Graph

Properties

Polynomial functions

Parameters

Properties

Even and odd functions

Sine

Graph

Properties

Links

Cosine

Graph

Properties

Tangent

Graph

Properties

Exponential

Graph

Properties

Links

Natural logarithm

The natural logarithm function is denoted

$$f(x) = \ln(x) = \log_e(x).$$

The function $\ln(x)$ is the inverse function of the exponential e^x.

Graph

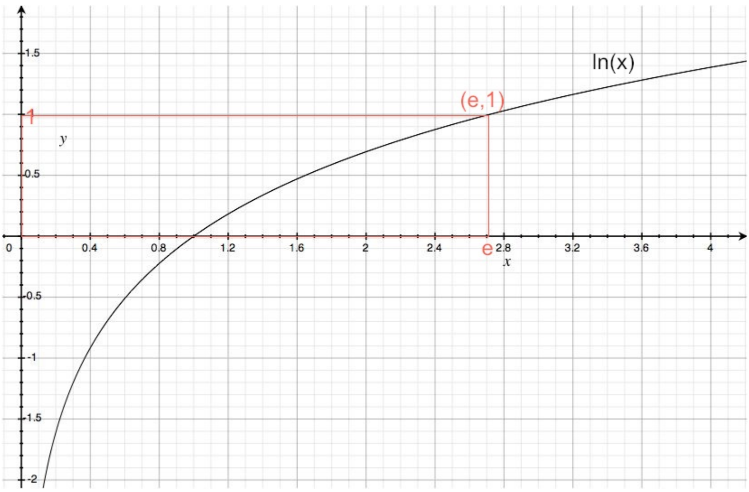


Figure 1.5: The graph of the function $\ln(x)$ passes through the following (x, y) coordinates: $(\frac{1}{e^2}, -2)$, $(\frac{1}{e}, -1)$, $(1, 0)$, $(e, 1)$, $(e^2, 2)$, $(e^3, 3)$, $(148.41\ldots, 5)$, and $(22026.46\ldots, 10)$.

Exercises

1.10 Polynomials

Definitions

- x: the variable
- $f(x)$: the polynomial. We sometimes denote polynomials $P(x)$ to distinguish them from a generic function $f(x)$.
- Degree of $f(x)$: the largest power of x that appears in the polynomial
- Roots of $f(x)$: the values of x for which $f(x) = 0$

Solving polynomial equations

Formulas

First

Second

Higher degrees

Using a computer

When solving real-world problems, you'll often run into much more complicated equations. To find the solutions of anything more complicated than the quadratic equation, I recommend using a computer algebra system like `SymPy`: `http://live.sympy.org`.

To make the computer solve the equation $x^2 - 3x + 2 = 0$ for you, type in the following:

```
>>> solve(x**2 - 3*x + 2, x)          # usage: solve(expr, var)
[1, 2]
```

The function `solve` will find the roots of any equation of the form `expr = 0`. Indeed, we can verify that $x^2 - 3x + 2 = (x - 1)(x - 2)$, so $x = 1$ and $x = 2$ are the two roots.

Substitution trick

Exercises

1.11 Trigonometry

We can put any three lines together to make a triangle. What's more, if one of the triangle's angles is equal to $90°$, we call this triangle a *right-angle triangle*.

In this section we'll discuss right-angle triangles in great detail and get to know their properties. We'll learn some fancy new terms like *hypotenuse*, *opposite*, and *adjacent*, which are used to refer to the different sides of a triangle. We'll also use the functions *sine*, *cosine*, and *tangent* to compute the *ratios of lengths* in right triangles.

Understanding triangles and their associated trigonometric functions is of fundamental importance: you'll need this knowledge for your future understanding of mathematical subjects like vectors and complex numbers, as well as physics subjects like oscillations and waves.

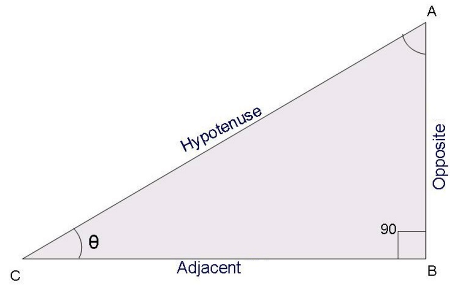


Figure 1.6: A right-angle triangle. The angle θ and the names of the sides of the triangle are indicated.

Concepts

- A, B, C: the three *vertices* of the triangle
- θ: the angle at the vertex C. Angles can be measured in degrees or radians.
- $\text{opp} \equiv \overline{AB}$: the length of the *opposite* side to θ
- $\text{adj} \equiv \overline{BC}$: the length of side *adjacent* to θ
- $\text{hyp} \equiv \overline{AC}$: the *hypotenuse*. This is the triangle's longest side.
- h: the "height" of the triangle (in this case $h = \text{opp} = \overline{AB}$)
- $\sin\theta \equiv \frac{\text{opp}}{\text{hyp}}$: the *sine* of theta is the ratio of the length of the opposite side and the length of the hypotenuse
- $\cos\theta \equiv \frac{\text{adj}}{\text{hyp}}$: the *cosine* of theta is the ratio of the adjacent length and the hypotenuse length
- $\tan\theta \equiv \frac{\sin\theta}{\cos\theta} \equiv \frac{\text{opp}}{\text{adj}}$: the *tangent* is the ratio of the opposite length divided by the adjacent length

Pythagoras' theorem

Sin and cos

The unit circle

Non-unit circles

Calculators

Exercises

Links

1.12 Trigonometric identities

1. Unit hypotenuse

2. sico + sico

3. coco − sisi

Derived formulas

Double angle formulas

Self similarity

Sin is cos, cos is sin

Sum formulas

Product formulas

Discussion

Exercises

1.13 Geometry

Triangles

Sine rule

Cosine rule

Circle

Sphere

Cylinder

Cones and pyramids

Exercises

1.14 Circle

The *circle* is a set of points located a constant distance from a centre point. This geometric shape appears in many situations.

Definitions

- r: the radius of the circle
- A: the area of the circle
- C: the circumference of the circle
- (x, y): a point on the circle
- θ: the angle (measured from the x-axis) of some point on the circle

Formulas

Explicit function

Polar coordinates

Parametric equation

Area

Circumference and arc length

Radians

Exercises

1.15 Vectors

Definitions

The two-dimensional vector $\vec{v} \in \mathbb{R}^2$ is equivalent to a *pair of numbers* $\vec{v} \equiv (v_x, v_y)$. We call v_x the x-component of $\vec{v}$, and v_y is the y-component of $\vec{v}$.

Vector representations

We'll use three equivalent ways to denote vectors:

- $\vec{v} = (v_x, v_y)$: component notation, where the vector is represented as a pair of coordinates with respect to the x-axis and the y-axis.
- $\vec{v} = v_x\hat{\imath} + v_y\hat{\jmath}$: unit vector notation. The vector is expressed in terms of the unit vectors $\hat{\imath} = (1, 0)$ and $\hat{\jmath} = (0, 1)$.
- $\vec{v} = \|\vec{v}\|\angle\theta$: length-and-direction notation, where the vector is expressed in terms of its *length* $\|\vec{v}\|$ and the angle θ that the vector makes with the x-axis.

These three notations describe different aspects of vectors, and we will use them throughout the rest of the book. We'll learn how to convert between them—both algebraically (with pen, paper, and calculator) and intuitively (by drawing arrows).

Vector operations

Consider two vectors, $\vec{u} = (u_x, u_y)$ and $\vec{v} = (v_x, v_y)$, and assume that $\alpha \in \mathbb{R}$ is an arbitrary constant. The following operations are defined for these vectors:

- **Addition:** $\vec{u} + \vec{v} = (u_x + v_x, u_y + v_y)$
- **Subtraction:** $\vec{u} - \vec{v} = (u_x - v_x, u_y - v_y)$
- **Scaling:** $\alpha\vec{u} = (\alpha u_x, \alpha u_y)$
- **Dot product:** $\vec{u} \cdot \vec{v} = u_x v_x + u_y v_y$
- **Length:** $\|\vec{u}\| = \sqrt{\vec{u} \cdot \vec{u}} = \sqrt{u_x^2 + u_y^2}$. We will also sometimes simply use the letter u to denote the length of $\vec{u}$.
- **Cross product:** $\vec{u} \times \vec{v} = (u_y v_z - u_z v_y,\ u_z v_x - u_x v_z,\ u_x v_y - u_y v_x)$. The cross product is only defined for three-dimensional vectors like $\vec{u} = (u_x, u_y, u_z)$ and $\vec{v} = (v_x, v_y, v_z)$.

Pay careful attention to the dot product and the cross product. Although they're called products, these operations behave much differently from taking the product of two numbers. Also note, there is no notion of vector division.

Vector algebra

Addition and subtraction

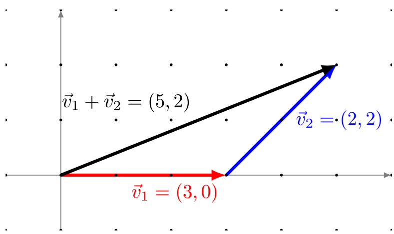
Scaling

Length

Vector as arrows

So far, we described how to perform algebraic operations on vectors in terms of their components. Vector operations can also be interpreted geometrically, as operations on two-dimensional arrows in the Cartesian plane.

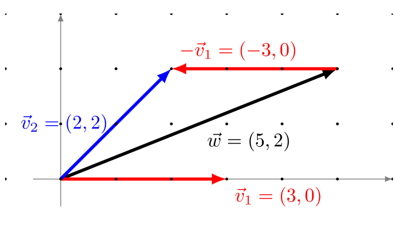
Vector addition The sum of two vectors corresponds to the combined displacement of the two vectors. The diagram on the right illustrates the addition of two vectors, $\vec{v}_1 = (3,0)$ and $\vec{v}_2 = (2,2)$. The sum of the two vectors is the vector $\vec{v}_1 + \vec{v}_2 = (3,0) + (2,2) = (5,2)$.



Vector subtraction Before we describe vector subtraction, note that multiplying a vector by a scaling factor $\alpha = -1$ gives a vector of the same length as the original, but pointing in the opposite direction.

This fact is useful if you want to subtract two vectors using the graphical approach. Subtracting a vector is the same as adding the negative of the vector:

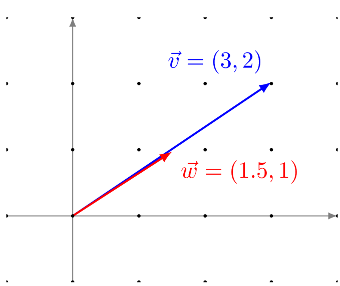
$$\vec{w} - \vec{v}_1 = \vec{w} + (-\vec{v}_1) = \vec{v}_2.$$



The diagram on the right illustrates the graphical procedure for subtracting the vector $\vec{v}_1 = (3,0)$ from the vector $\vec{w} = (5,2)$. Subtraction of $\vec{v}_1 = (3,0)$ is the same as addition of $-\vec{v}_1 = (-3,0)$.

Scaling The scaling operation acts to change the length of a vector. Suppose we want to obtain a vector in the same direction as the vector $\vec{v} = (3,2)$, but half as long. "Half as long" corresponds to a scaling factor of $\alpha = 0.5$. The scaled-down vector is $\vec{w} = 0.5\vec{v} = (1.5,1)$.

Conversely, we can think of the vector $\vec{v}$ as being twice as long as the vector $\vec{w}$.



Length-and-direction representation

Unit vector notation

Examples

Vector dimensions

Coordinate system

Exercises

1.16 Complex numbers

Definitions

Complex numbers have a real part and an imaginary part:

- i: the unit imaginary number $i \equiv \sqrt{-1}$ or $i^2 = -1$
- bi: an imaginary number that is equal to b times i
- $\mathbb{R}$: the set of real numbers
- $\mathbb{C}$: the set of complex numbers $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$
- $z = a + bi$: a complex number
  - ▷ $\text{Re}\{z\} = a$: the real part of z
  - ▷ $\text{Im}\{z\} = b$: the imaginary part of z
- $\bar{z}$: the *complex conjugate* of z. If $z = a + bi$, then $\bar{z} = a - bi$.

The polar representation of complex numbers:

- $z = |z|\angle\varphi_z = |z|\cos\varphi_z + i|z|\sin\varphi_z$
- $|z| = \sqrt{\bar{z}z} = \sqrt{a^2 + b^2}$: the *magnitude* of $z = a + bi$
- $\varphi_z = \tan^{-1}(b/a)$: the *phase* or *argument* of $z = a + bi$
- $\text{Re}\{z\} = |z|\cos\varphi_z$
- $\text{Im}\{z\} = |z|\sin\varphi_z$

Formulas

Addition and subtraction

Polar representation

Multiplication

Division

Cardano's example

Fundamental theorem of algebra

Euler's formula

De Moivre's formula

Links

1.17 Solving systems of linear equations

Concepts

- x, y: the two unknowns in the equations
- $eq1, eq2$: a system of two equations that must be solved *simultaneously*. These equations will look like

$$a_1 x + b_1 y = c_1,$$
$$a_2 x + b_2 y = c_2,$$

  where as, bs, and cs are given constants.

Principles

Solution techniques

Solving by substitution

Solving by subtraction

Solving by equating

Discussion

Exercises

1.18 Set notation

Definitions

- *set*: a collection of mathematical objects
- S, T: the usual variable names for sets
- $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$: some important sets of numbers: the naturals, the integers, the rationals, and the real numbers, respectively.
- $\{$ definition $\}$: the curly brackets surround the definition of a set, and the expression inside the curly brackets describes what the set contains.

Set operations:

- $S \cup T$: the *union* of two sets. The union of S and T corresponds to the elements in either S or T.
- $S \cap T$: the *intersection* of the two sets. The intersection of S and T corresponds to the elements that are in both S and T.
- $S \setminus T$: *set difference* or *set minus*. The set difference $S \setminus T$ corresponds to the elements of S that are not in T.

Set relations:

- $\subset$: is a strict subset of
- $\subseteq$: is a subset of or equal to

Special mathematical shorthand symbols and their corresponding meanings:

- $\forall$: for all
- $\exists$: there exists
- $\nexists$: there doesn't exist
- $|$: such that
- $\in$: element of
- $\notin$: not an element of

Sets

Example 1: The nonnegative real numbers

Example 2: Odd and even integers

Number sets

Set relations and set operations

Example 3: Set operations

Example 4: Word problem

New vocabulary

Simple example

Less simple example: Square root of 2 is irrational

Sets related to functions

Discussion

Exercises

1.19 Math problems

We've now reached the first section of problems in this book. The purpose of these problems is to give you a way to comprehensively practice your math fundamentals. In the real world, you'll rarely have to solve equations by hand; however, knowing how to solve math equations and manipulate math expressions will be very useful in later chapters of this book. At times, honing your math chops might seem like tough mental work, but at the end of each problem, you'll gain a stronger foothold on all the subjects you've been learning about. You'll also experience a small *achievement buzz* after each problem you vanquish.

I have a special message to readers who are learning math just for fun: you can either try the problems in this section or skip them. Since you have no upcoming exam on this material, you could skip ahead to Chapter **??** without any immediate consequences. However (and it's a big however), those readers who don't take a crack at these problems will be missing a significant opportunity.

Sit down to do them later today, or another time when you're properly caffeinated. If you take the initiative to make time for math, you'll find yourself developing lasting comprehension and true math fluency. Without the practice of solving problems, however, you're extremely likely to forget most of what you've learned within a month, simple as that. You'll still remember the big ideas, but the details will be fuzzy and faded. Don't break the pace now: with math, it's very much *use it or lose it*!

By solving some of the problems in this section, you'll remember a lot more stuff. Make sure you step away from the pixels while you're solving problems. You don't need fancy technology to do math; grab a pen and some paper from the printer and you'll be fine. Do yourself a favour: put your phone in airplane-mode, close the lid of your laptop, and move away from desktop computers. Give yourself some time to think. Yes, I know you can look up the answer to any question in five seconds on the Internet, and you can use `live.sympy.org` to solve any math problem, but that is like outsourcing the thinking. Descartes, Leibniz, and Riemann did most of their work with pen and paper and they did well. Spend some time with math the way the masters did.

P1.1 Solve for x in the equation $x^2 - 9 = 7$.

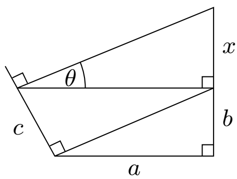
P1.2 Solve for x in the equation $\cos^{-1}\left(\frac{x}{A}\right) - \phi = \omega t$.

P1.3 Solve for x in the equation $\frac{1}{x} = \frac{1}{a} + \frac{1}{b}$.

P1.4 Use a calculator to find the values of the following expressions:

a) $\sqrt[4]{3^3}$ **b)** 2^{10} **c)** $7^{\frac{1}{4}} - 10$ **d)** $\frac{1}{2}\ln(e^{22})$

P1.5 Find x. Express your answer in terms of a, b, c and θ.

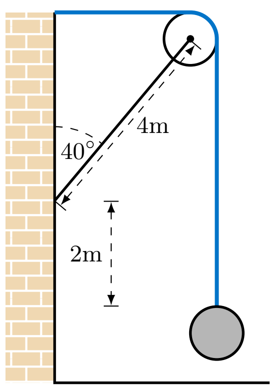


Hint: Use Pythagoras' theorem twice; then use the function tan.

P1.6 Satoshi likes warm saké. He places 1 litre of water in a sauce pan with diameter 17 cm. How much will the height of the water level rise when Satoshi immerses a saké bottle with diameter 7.5 cm?

Hint: You'll need the volume conversion ratio 1 litre $= 1000\,\text{cm}^3$.

P1.7 In preparation for the shooting of a music video, you're asked to suspend a wrecking ball hanging from a circular pulley. The pulley has a radius of 50 cm. The other lengths are indicated in the figure. What is the total length of the rope required?



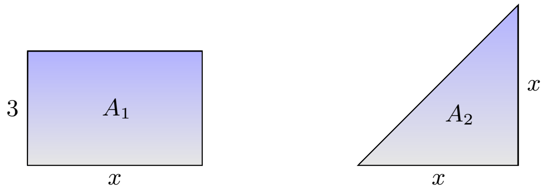
Hint: The total length of rope consists of two straight parts and the curved section that wraps around the pulley.

P1.8 Let B be the set of people who are bankers and C be the set of crooks. Rewrite the math statement $\exists b \in B \mid b \notin C$ in plain English.

P1.9 Let M denote the set of people who run Monsanto, and H denote the people who ought to burn in hell for all eternity. Write the math statement $\forall p \in M, p \in H$ in plain English.

P1.10 When starting a business, one sometimes needs to find investors. Define M to be the set of investors with money, and C to be the set of investors with connections. Describe the following sets in words: **a)** $M \setminus C$, **b)** $C \setminus M$, and the most desirable set **c)** $M \cap C$.

P1.11 Write the formulas for the functions $A_1(x)$ and $A_2(x)$ that describe the areas of the following geometric shapes.



Chapter 2

Intro to linear algebra

The first chapter reviewed core ideas of mathematics. Now that we're done with the prerequisites, we can begin the main discussion of linear algebra: the study of vectors and matrices.

2.1 Definitions

Vectors and matrices are the objects of study in linear algebra, and in this chapter we'll define them and learn the basic operations we can perform on them.

We denote the set of n-dimensional vectors as $\mathbb{R}^n$. A vector $\vec{v} \in \mathbb{R}^n$ is an n-tuple of real numbers.[1] For example, a three-dimensional vector is a triple of the form

$$\vec{v} = (v_1, v_2, v_3) \quad \in \quad (\mathbb{R}, \mathbb{R}, \mathbb{R}) \equiv \mathbb{R}^3.$$

To specify the vector $\vec{v}$, we must specify the values for its three components: v_1, v_2, and v_3.

A matrix $A \in \mathbb{R}^{m \times n}$ is a rectangular array of real numbers with m rows and n columns. For example, a 3×2 matrix looks like this:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \in \begin{bmatrix} \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} \end{bmatrix} \equiv \mathbb{R}^{3 \times 2}.$$

To specify the matrix A, we need to specify the values of its six components, a_{11}, a_{12}, ... , a_{32}.

In the remainder of this chapter we'll learn about the mathematical operations we can perform on vectors and matrices. Many problems in science, business, and technology can be described in terms of

[1] The notation "$s \in S$" is read "s is an element of S" or "s in S."

vectors and matrices, so it's important you understand how to work with these math objects.

Context

To illustrate what's new about vectors and matrices, let's begin by reviewing the properties of something more familiar: the set of real numbers $\mathbb{R}$. The basic operations for real numbers are:

- Addition (denoted $+$)
- Subtraction, the inverse of addition (denoted $-$)
- Multiplication (denoted implicitly)
- Division, the inverse of multiplication (denoted by fractions)

You're familiar with these operations and know how to use them to evaluate math expressions and solve equations.

You should also be familiar with *functions* that take real numbers as inputs and give real numbers as outputs, denoted $f : \mathbb{R} \to \mathbb{R}$. Recall that, by definition, the *inverse function* f^{-1} *undoes* the effect of f. If you are given $f(x)$ and want to find x, you can use the inverse function as follows: $f^{-1}(f(x)) = x$. For example, the function $f(x) = \ln(x)$ has the inverse $f^{-1}(x) = e^x$, and the inverse of $g(x) = \sqrt{x}$ is $g^{-1}(x) = x^2$.

Having reviewed the basic operations for real numbers $\mathbb{R}$, let's now introduce the basic operations for vectors $\mathbb{R}^n$ and matrices $\mathbb{R}^{m\times n}$.

Vector operations

The operations we can perform on vectors are:

- Addition (denoted $+$)
- Subtraction, the inverse of addition (denoted $-$)
- Scaling (denoted implicitly)
- Dot product (denoted $\cdot$)
- Cross product (denoted $\times$)

We'll discuss each of these vector operations in Section 2.2. Although you should already be familiar with vectors and vector operations from Section **??**, it's worth revisiting these concepts in greater depth, because vectors are the foundation of linear algebra.

Matrix operations

The mathematical operations defined for matrices A and B are:

- Addition (denoted $A + B$)

- Subtraction, the inverse of addition (denoted $A - B$)
- Scaling by a constant α (denoted αA)
- Matrix product (denoted AB)
- Matrix-vector product (denoted $A\vec{v}$)
- Matrix inverse (denoted A^{-1})
- Trace (denoted $\mathrm{Tr}(A)$)
- Determinant (denoted $\det(A)$ or $|A|$)

We'll define each of these operations in Section 2.4, and we'll learn about the various computational, geometric, and theoretical considerations associated with these matrix operations throughout the remainder of the book.

Let's now examine one important matrix operation in closer detail: the matrix-vector product $A\vec{x}$.

Matrix-vector product

Consider the matrix $A \in \mathbb{R}^{m\times n}$ and the vector $\vec{v} \in \mathbb{R}^n$. The matrix-vector product $A\vec{x}$ produces a linear combination of the columns of the matrix A with coefficients $\vec{x}$. For example, the product of a 3×2 matrix A and a 2×1 vector $\vec{x}$ results in a 3×1 vector, which we'll denote $\vec{y}$:

$$\vec{y} = A\vec{x},$$

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \equiv \underbrace{\begin{bmatrix} x_1a_{11} + x_2a_{12} \\ x_1a_{21} + x_2a_{22} \\ x_1a_{31} + x_2a_{32} \end{bmatrix}}_{\text{row picture}} \equiv \underbrace{x_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + x_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix}}_{\text{column picture}}.$$

The key thing to observe in the above formula is the dual interpretation of the matrix-vector product $A\vec{x}$ in the "row picture" and in the "column picture." In the row picture, we obtain the vector $\vec{y}$ by computing the dot product of the vector $\vec{x}$ with each of the rows of the matrix A. In the column picture, we interpret the vector $\vec{y}$ as x_1 times the first column of A plus x_2 times the second column of A. In other words, $\vec{y}$ is a linear combination of the columns of A. For example, if you want to obtain the linear combination consisting of three times the first column of A and four times the second column of A, you can multiply A by the vector $\vec{x} = \left[\begin{smallmatrix} 3 \\ 4 \end{smallmatrix}\right]$.

Linear combinations as matrix products

Consider some set of vectors $\{\vec{e}_1, \vec{e}_2\}$, and a third vector $\vec{y}$ that is a *linear combination* of the vectors $\vec{e}_1$ and $\vec{e}_2$:

$$\vec{y} \;=\; \alpha\vec{e}_1 \;+\; \beta\vec{e}_2.$$

The numbers $\alpha, \beta \in \mathbb{R}$ are the coefficients in this linear combination.

The matrix-vector product is defined expressly for the purpose of studying linear combinations. We can describe the linear combination $\vec{y} = \alpha\vec{e}_1 + \beta\vec{e}_2$ as the following matrix-vector product:

$$\vec{y} \;=\; \begin{bmatrix} | & | \\ \vec{e}_1 & \vec{e}_2 \\ | & | \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}.$$

The matrix E has $\vec{e}_1$ and $\vec{e}_2$ as columns. The dimensions of the matrix E will be $n \times 2$, where n is the dimension of the vectors $\vec{e}_1$, $\vec{e}_2$, and $\vec{y}$.

Linear transformations

Dear readers, we've reached the key notion in the study of linear algebra. This is the crux. The essential fibre. The main idea. I know you're ready to handle it because you're familiar with functions of a real variable $f : \mathbb{R} \to \mathbb{R}$, and you just learned the definition of the matrix-vector product (in which the variables were chosen to subliminally remind you of the standard conventions for the function input x and the function output $y = f(x)$). Without further ado, I present to you the concept of a *linear transformation*.

The matrix-vector product corresponds to the abstract notion of a *linear transformation*, which is one of the key notions in the study of linear algebra. Multiplication by a matrix $A \in \mathbb{R}^{m\times n}$ can be thought of as computing a linear transformation T_A that takes n-vectors as inputs and produces m-vectors as outputs:

$$T_A : \mathbb{R}^n \to \mathbb{R}^m.$$

Instead of writing $\vec{y} = T_A(\vec{x})$ to denote the linear transformation T_A applied to the vector $\vec{x}$, we can write $\vec{y} = A\vec{x}$. Since the matrix A has m rows, the result of the matrix-vector product is an m-vector. Applying the linear transformation T_A to the vector $\vec{x}$ corresponds to the product of the matrix A and the column vector $\vec{x}$. We say T_A is *represented by* the matrix A.

Inverse When a matrix A is square and invertible, there exists an inverse matrix A^{-1} which *undoes* the effect of A to restore the original input vector:

$$A^{-1}(A\vec{x}) = A^{-1}A\vec{x} = \vec{x}.$$

Using the matrix inverse A^{-1} to undo the effects of the matrix A is analogous to using the inverse function f^{-1} to undo the effects of the function f.

Example 1 Consider the linear transformation that multiplies the first components of input vectors by 3 and multiplies the second components by 5, as described by the matrix

$$A = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}, \qquad A\vec{x} = \begin{bmatrix} 3 & 0 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3x_1 \\ 5x_2 \end{bmatrix}.$$

The inverse of the matrix A is

$$A^{-1} = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{5} \end{bmatrix}, \qquad A^{-1}(A\vec{x}) = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{5} \end{bmatrix}\begin{bmatrix} 3x_1 \\ 5x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{x}.$$

The inverse matrix multiplies the first component by $\frac{1}{3}$ and the second component by $\frac{1}{5}$, which effectively undoes what A did.

Example 2 Things get a little more complicated when matrices *mix* the different components of the input vector, as in this example:

$$B = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}, \quad \text{which acts as} \quad B\vec{x} = \begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 + 2x_2 \\ 3x_2 \end{bmatrix}.$$

Make sure you understand how to compute $B\vec{x}$ using both the *row picture* and the *column picture* of the matrix-vector product.

The inverse of the matrix B is the matrix

$$B^{-1} = \begin{bmatrix} 1 & \frac{-2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}.$$

Multiplication by the matrix B^{-1} is the "undo action" for multiplication by B:

$$B^{-1}(B\vec{x}) = \begin{bmatrix} 1 & \frac{-2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 & \frac{-2}{3} \\ 0 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} x_1 + 2x_2 \\ 3x_2 \end{bmatrix} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \vec{x}.$$

By definition, the inverse A^{-1} *undoes* the effects of the matrix A. The cumulative effect of applying A^{-1} after A is the *identity matrix* $\mathbb{1}$, which has 1s on the diagonal and 0s everywhere else:

$$A^{-1}A\vec{x} = \mathbb{1}\vec{x} = \vec{x} \qquad \Rightarrow \qquad A^{-1}A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \mathbb{1}.$$

Note that $\mathbb{1}\vec{x} = \vec{x}$ for any vector $\vec{x}$.

We'll discuss matrix inverses and how to compute them in more detail later (Section 3.5). For now, it's important you know they exist.

An overview of linear algebra

In the remainder of the book, we'll learn all about the properties of vectors and matrices. Matrix-vector products play an important role in linear algebra because of their relation to *linear transformations*.

Functions are transformations from an input space (the domain) to an output space (the image). A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a function that takes n-vectors as inputs and produces m-vectors as outputs. If the function T is linear, the output $\vec{y} = T(\vec{x})$ of T applied to $\vec{x}$ can be computed as the matrix-vector product $A_T\vec{x}$, for some matrix $A_T \in \mathbb{R}^{m\times n}$. We say T is *represented by* the matrix A_T. Equivalently, every matrix $A \in \mathbb{R}^{m\times n}$ corresponds to some linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$. Given the equivalence between matrices and linear transformations, we can reinterpret the statement "linear algebra is about vectors and matrices" by saying "linear algebra is about vectors and linear transformations."

You can adapt your existing knowledge about functions to the world of linear transformations. The action of a function on a number is similar to the action of a linear transformation on a vector. The table below summarizes several useful correspondences between functions and linear transformations.

function $f : \mathbb{R} \to \mathbb{R}$	$\Leftrightarrow$	linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$ represented by the matrix $A \in \mathbb{R}^{m\times n}$
input $x \in \mathbb{R}$	$\Leftrightarrow$	input $\vec{x} \in \mathbb{R}^n$
output $f(x) \in \mathbb{R}$	$\Leftrightarrow$	output $T_A(\vec{x}) \equiv A\vec{x} \in \mathbb{R}^m$
$g \circ f\ (x) = g(f(x))$	$\Leftrightarrow$	$T_B(T_A(\vec{x})) \equiv BA\vec{x}$
function inverse f^{-1}	$\Leftrightarrow$	matrix inverse A^{-1}
roots of f	$\Leftrightarrow$	kernel of $T_A \equiv$ null space of $A \equiv \mathcal{N}(A)$
image of f	$\Leftrightarrow$	image of $T_A \equiv$ column space of $A \equiv \mathcal{C}(A)$

Table 2.1: Correspondences between functions and linear transformations.

This table of correspondences serves as a roadmap for the rest of the material in this book. You'll notice the table introduces several new linear algebra concepts like *kernel*, *null space*, and *column space*, but not too many. You can totally do this!

Remember to always connect the new concepts of linear algebra to concepts you're already familiar with. For example, the roots of a function $f(x)$ are the set of inputs for which the function's output is zero. Similarly, the *kernel* of a linear transformation T is the set of inputs that T sends to the zero vector. The roots of a function $f : \mathbb{R} \to \mathbb{R}$ and the kernel of a linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$

are essentially the same concept; we're just upgrading functions to vector inputs.

In Chapter 1, I explained why functions are useful tools for modelling the real world. Well, linear algebra is the "vector upgrade" to your real-world modelling skills. With linear algebra you'll be able to model complex relationships between multivariable inputs and multivariable outputs. To build modelling skills, you must first develop your geometric intuition about lines, planes, vectors, bases, linear transformations, vector spaces, vector subspaces, etc. It's a lot of work, but the effort you invest will pay dividends.

Links

[Linear algebra lecture series by Prof. Strang from MIT]
`http://bit.ly/1ayRcrj` (row and column picture example)

[A system of equations in the row picture and column picture]
`https://www.youtube.com/watch?v=uNKDw46_Ev4`

Exercises

E2.1 Find the inverse matrix A^{-1} for the matrix $A = \left[\begin{smallmatrix} 7 & 0 \\ 0 & 2 \end{smallmatrix}\right]$. Verify that $A^{-1}(A\vec{v}) = \vec{v}$ for any vector $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$.

E2.2 Given the matrices $A = \left[\begin{smallmatrix} 1 & 3 \\ 4 & 5 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} -1 & 0 \\ 3 & 3 \end{smallmatrix}\right]$, and the vectors $\vec{v} = \left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ and $\vec{w} = \left[\begin{smallmatrix} -3 \\ -4 \end{smallmatrix}\right]$, compute the following expressions.

a) $A\vec{v}$ **b)** $B\vec{v}$ **c)** $A(B\vec{v})$ **d)** $B(A\vec{v})$ **e)** $A\vec{w}$ **f)** $B\vec{w}$

E2.3 Find the coefficients v_1 and v_2 of the vector $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$ so that $E\vec{v} = 3\vec{e}_2 - 2\vec{e}_1$, where E is the following matrix:

$$E = \begin{bmatrix} | & | \\ \vec{e}_1 & \vec{e}_2 \\ | & | \end{bmatrix}.$$

What next?

We won't bring geometry, vector spaces, algorithms, and the applications of linear algebra into the mix all at once. Instead, let's start with the basics. Since linear algebra is about vectors and matrices, let's define vectors and matrices precisely, and describe the math operations we can perform on them.

2.2 Vector operations

Formulas

Consider the vectors $\vec{u} = (u_1, u_2, u_3)$ and $\vec{v} = (v_1, v_2, v_3)$, and an arbitrary constant $\alpha \in \mathbb{R}$. Vector algebra can be summarized as the following operations:

- Addition: $\vec{u} + \vec{v} \equiv (u_1 + v_1, u_2 + v_2, u_3 + v_3)$
- Subtraction: $\vec{u} - \vec{v} \equiv (u_1 - v_1, u_2 - v_2, u_3 - v_3)$
- Scaling: $\alpha\vec{u} \equiv (\alpha u_1, \alpha u_2, \alpha u_3)$
- Dot product: $\vec{u} \cdot \vec{v} \equiv u_1 v_1 + u_2 v_2 + u_3 v_3$
- Cross product: $\vec{u} \times \vec{v} \equiv (u_2 v_3 - u_3 v_2,\ u_3 v_1 - u_1 v_3,\ u_1 v_2 - u_2 v_1)$
- Length: $||\vec{u}|| \equiv \sqrt{u_1^2 + u_2^2 + u_3^2}$

In the next few pages we'll see what these operations can do for us.

Notation

Addition and subtraction

Scaling by a constant

2.3 Vector products

Dot product

The *dot product* takes two vectors as inputs and produces a single, real number as an output:

$$\cdot : \mathbb{R}^3 \times \mathbb{R}^3 \quad \to \quad \mathbb{R}.$$

The dot product between two vectors can be computed using either the algebraic formula,

$$\vec{v} \cdot \vec{w} \ \equiv \ v_x w_x + v_y w_y + v_z w_z,$$

or the geometric formula,

$$\vec{v} \cdot \vec{w} \ \equiv \ \|\vec{v}\|\|\vec{w}\| \cos(\varphi),$$

where φ is the angle between the two vectors. Note the value of the dot product depends on the vectors' lengths and cosine of the angle between them.

Cross product

The *cross product* takes two vectors as inputs and produces another vector as the output:

$$\times : \mathbb{R}^3 \times \mathbb{R}^3 \quad \to \quad \mathbb{R}^3.$$

The cross product of two vectors is perpendicular to both vectors:

$$\vec{v} \times \vec{w} = \{ \text{ a vector perpendicular to both } \vec{v} \text{ and } \vec{w} \ \} \quad \in \mathbb{R}^3.$$

If you take the cross product of one vector pointing in the x-direction with another vector pointing in the y-direction, the result will be a vector in the z-direction: $\hat{\imath} \times \hat{\jmath} = \hat{k}$. The name *cross product* comes from the symbol used to denote it. It is also sometimes called the *vector product*, since the output of this operation is a vector.

The cross products of individual basis elements are defined as

$$\hat{\imath} \times \hat{\jmath} = \hat{k}, \qquad \hat{\jmath} \times \hat{k} = \hat{\imath}, \qquad \hat{k} \times \hat{\imath} = \hat{\jmath}.$$

The right-hand rule

Length of a vector

Unit vectors

Projection

Discussion

Links

Exercises

2.4 Matrix operations

Addition and subtraction

Multiplication by a constant

Matrix-vector multiplication

Matrix-matrix multiplication

Transpose

Vectors as matrices

Outer product

Matrix inverses

Trace

Determinant

Discussion

Exercises

2.5 Linearity

Introduction

Example of composition of linear functions

Definition

Linear functions map any linear combination of inputs to the same linear combination of outputs. A function f is *linear* if it satisfies the equation

$$f(\alpha x_1 + \beta x_2) = \alpha f(x_1) + \beta f(x_2),$$

for any two inputs x_1 and x_2, and for all constants α and β.

Lines are not linear functions!

Multivariable functions

Linear expressions

Linear equations

Applications

Geometric interpretation of linear equations

First-order approximations

Discussion

Exercises

2.6 Overview of linear algebra

In linear algebra, you'll learn new computational techniques and develop new ways of thinking about math. With these new tools, you'll be able to use linear algebra techniques for many applications. Let's look at what lies ahead in this book.

Computational linear algebra

The first steps toward understanding linear algebra will seem a little tedious. In Chapter 3 you'll develop basic skills for manipulating vectors and matrices. Matrices and vectors have many components and performing operations on them involves many arithmetic steps—there is no way to circumvent this complexity. Make sure you understand the basic algebra rules (how to add, subtract, and multiply vectors and matrices) because they are a prerequisite for learning more advanced material. You should be able to perform all the matrix algebra operations with pen and paper for small and medium-sized matrices.

The good news is, with the exception of your homework assignments and final exam, you won't have to carry out matrix algebra by hand. It is much more convenient to use a computer for large matrix calculations. The more you develop your matrix algebra skills, the deeper you'll be able to delve into the advanced topics.

Geometric linear algebra

So far, we've described vectors and matrices as arrays of numbers. This is fine for the purpose of doing *algebra* on vectors and matrices, but this description is not sufficient for understanding their geometric

properties. The components of a vector $\vec{v} \in \mathbb{R}^n$ can be thought of as distances measured along a coordinate system with n axes. The vector $\vec{v}$ can therefore be said to "point" in a particular direction with respect to the coordinate system. The fun part of linear algebra starts when you learn about the geometric interpretation of the algebraic operations on vectors and matrices.

Consider some unit vector that specifies a direction of interest $\hat{r}$. Suppose we're given some other vector $\vec{v}$, and we're asked to find *how much of $\vec{v}$ is in the $\hat{r}$ direction.* The answer is computed using the dot product: $v_r = \vec{v} \cdot \hat{r} = \|\vec{v}\| \cos\theta$, where θ is the angle between $\vec{v}$ and $\hat{r}$. The technical term for the quantity v_r is "the length of the projection of $\vec{v}$ in the $\hat{r}$ direction." By "projection," I mean we ignore all parts of $\vec{v}$ that are not in the $\hat{r}$ direction. Projections are used in mechanics to calculate the x- and y-components of forces in force diagrams. In Section 4.2 we'll learn how to calculate all kinds of projections using the dot product.

To further consider the geometric aspects of vector operations, imagine the following situation. Suppose I gave you two vectors $\vec{u}$ and $\vec{v}$, and asked you to find a third vector $\vec{w}$ that is perpendicular to both $\vec{u}$ and $\vec{v}$. A priori this sounds like a complicated question to answer, but in fact the required vector $\vec{w}$ can easily be obtained by computing the cross product $\vec{w} = \vec{u} \times \vec{v}$.

In Section 4.1 we'll learn how to describe lines and planes in terms of points, direction vectors, and normal vectors. Consider the following geometric problem: given the equations of two planes in $\mathbb{R}^3$, find the equation of the line where the two planes intersect. There is an algebraic procedure called *Gauss–Jordan elimination* we can use to find the solution.

The determinant of a matrix has a geometric interpretation (Section 3.4). The determinant tells us something about the relative orientation of the vectors that make up the rows of the matrix. If the determinant of a matrix is zero, it means the rows are not *linearly independent*, in other words, at least one of the rows can be written in terms of the other rows. Linear independence, as we'll see shortly, is an important property for vectors to have. The determinant is a convenient way to test whether vectors are linearly independent.

As you learn about geometric linear algebra, practice *visualizing* each new concept you learn about. Always keep a picture in your head of what is going on. The relationships between two-dimensional vectors can be represented in vector diagrams. Three-dimensional vectors can be visualized by pointing pens and pencils in different directions. Most of the intuition you build about vectors in two and three dimensions are applicable to vectors with more dimensions.

Theoretical linear algebra

Linear algebra will teach you how to reason about vectors and matrices in an abstract way. By thinking abstractly, you'll be able to extend your geometric intuition of two and three-dimensional problems to problems in higher dimensions. Much *knowledge buzz* awaits as you learn about new mathematical ideas and develop new ways of thinking.

You're no doubt familiar with the normal coordinate system made of two orthogonal axes: the x-axis and the y-axis. A vector $\vec{v} \in \mathbb{R}^2$ is specified in terms of its coordinates (v_x, v_y) with respect to these axes. When we say $\vec{v} = (v_x, v_y)$, what we really mean is $\vec{v} = v_x\hat{\imath} + v_y\hat{\jmath}$, where $\hat{\imath}$ and $\hat{\jmath}$ are unit vectors that point along the x- and y-axes. As it turns out, we can use many other kinds of coordinate systems to represent vectors. A *basis* for $\mathbb{R}^2$ is any set of two vectors $\{\hat{e}_1, \hat{e}_2\}$ that allows us to express all vectors $\vec{v} \in \mathbb{R}^2$ as linear combinations of the basis vectors: $\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2$. The same vector $\vec{v}$ corresponds to two different coordinate pairs, depending on which basis is used for the description: $\vec{v} = (v_x, v_y)$ in the basis $\{\hat{\imath}, \hat{\jmath}\}$ and $\vec{v} = (v_1, v_2)$ in the basis $\{\hat{e}_1, \hat{e}_2\}$. We'll learn about bases and their properties in great detail in the coming chapters. The choice of basis plays a fundamental role in all aspects of linear algebra. Bases relate the real-world to its mathematical representation in terms of vector and matrix components.

In the text above, I explained that computing the product between a matrix and a vector $A\vec{x} = \vec{y}$ can be thought of as a linear transformation, with input $\vec{x}$ and output $\vec{y}$. Any linear transformation (Section 5.1) can be represented (Section 5.2) as a multiplication by a matrix A. Conversely, every $m \times n$ matrix $A \in \mathbb{R}^{m\times n}$ can be thought of as performing a linear transformation $T_A \colon \mathbb{R}^n \to \mathbb{R}^m$. The equivalence between matrices and linear transformations allows us to identify certain matrix properties with properties of linear transformations. For example, the *column space* $\mathcal{C}(A)$ of the matrix A (the set of vectors that can be written as a combination of the columns of A) corresponds to the image space of the linear transformation T_A (the set of possible outputs of T_A).

The eigenvalues and eigenvectors of matrices (Section 6.1) allow us to describe the actions of matrices in a natural way. The set of eigenvectors of a matrix are special input vectors for which the action of the matrix is described as a *scaling*. When a matrix acts on one of its eigenvectors, the output is a vector in the same direction as the input vector scaled by a constant. The scaling constant is the *eigenvalue* (own value) associated with this eigenvector. By specifying all the eigenvectors and eigenvalues of a matrix, it is possible to obtain a complete description of what the matrix does. Thinking of matrices

in terms of their eigenvalues and eigenvectors is a powerful technique for describing their properties and has many applications.

Linear algebra is useful because linear algebra techniques can be applied to all kinds of “vector-like” objects. The abstract concept of a vector space (Section 6.3) captures precisely what it means for some class of mathematical objects to be “vector-like.” For example, the set of polynomials of degree at most two, denoted $P_2(x)$, consists of all functions of the form $f(x) = a_0 + a_1x + a_2x^2$. Polynomials are vector-like because it’s possible to describe each polynomial in terms of its coefficients (a_0, a_1, a_2). Furthermore, the sum of two polynomials and the multiplication of a polynomial by a constant both correspond to vector-like calculations of coefficients. Once you realize polynomials are vector-like, you’ll be able to use linear algebra concepts like *linear independence*, *dimension*, and *basis* when working with polynomials.

Useful linear algebra

One of the most useful skills you’ll learn in linear algebra is the ability to solve systems of linear equations. Many real-world problems are expressed as linear equations in multiple unknown quantities. You can solve for n unknowns simultaneously if you have a set of n linear equations that relate the unknowns. To solve this system of equations, eliminate the variables one by one using basic techniques such as substitution and subtraction (see Section 1.17); however, the procedure will be slow and tedious for many unknowns. If the system of equations is linear, it can be expressed as an *augmented matrix* built from the coefficients in the equations. You can then use the Gauss–Jordan elimination algorithm to solve for the n unknowns (Section 3.1). The key benefit of the augmented matrix approach is that it allows you to focus on the coefficients without worrying about the variable names. This saves time when you must solve for many unknowns. Another approach for solving n linear equations in n unknowns is to express the system of equations as a matrix equation (Section 3.2) and then solve the matrix equation by computing the matrix inverse (Section 3.5).

In Section 6.6 you’ll learn how to *decompose* a matrix into a product of simpler matrices. Matrix decompositions are often performed for computational reasons: certain problems are easier to solve on a computer when the matrix is expressed in terms of its simpler constituents. Other decompositions, like the decomposition of a matrix into its eigenvalues and eigenvectors, give you valuable information about the properties of the matrix. Google’s original PageRank algorithm for ranking webpages by “importance” can be explained as the search for an eigenvector of a matrix. The matrix in question contains information about all hyperlinks that exist between webpages. The

eigenvector we're looking for corresponds to a vector that describes the relative importance of each page. So when I tell you eigenvectors are *valuable information*, I'm not kidding: a little 350-billion-dollar company called Google started as an eigenvector idea.

The techniques of linear algebra find applications in many areas of science and technology. We'll discuss applications such as *modelling* multidimensional real-world problems, finding *approximate solutions* to equations (curve fitting), solving constrained optimization problems using *linear programming*, and many other in Chapter 7. As a special bonus for readers interested in physics, a short introduction to quantum mechanics can be found in Chapter 9; if you have a good grasp of linear algebra, you can understand matrix quantum mechanics at no additional mental cost.

Our journey into the land of linear algebra will continue in the next chapter with the study of computational aspects of linear algebra. We'll learn how to solve large systems of linear equations, practice computing matrix products, discuss matrix determinants, and compute matrix inverses.

2.7 Introductory problems

We've been having fun learning about vector and matrix operations, and we've also touched upon linear transformations. I've summarized what linear algebra is about; now it's time for you to put in the effort and check whether you understand the definitions of the operations.

Don't cheat yourself by thinking my summaries are enough; you can't magically understand everything about linear algebra merely by reading about it. Learning doesn't work that way! The only way to truly "get" math—especially advanced math—is to solve problems using the new concepts you've learned. Indeed, the only math I remember from my university days is math that I practiced by solving lots of problems. There's no better way to test whether you understand than testing yourself. Of course, it's your choice whether you'll dedicate the next hour of your life to working through the problems in this section. All I'll say is that you'll have something to show for your efforts; and it's totally worth it.

P2.1 Which of the following functions are linear?

a) $q(x) = x^2$ **b)** $f(x) = g(h(x))$, where $g(x) = \sqrt{3}x$ and $h(x) = -4x$
c) $i(x) = \frac{1}{mx}$ **d)** $j(x) = \frac{x-a}{x-b}$

P2.2 Find the sum of the vectors $(1, 0, 1)$ and the vector $(0, 2, 2)$.

P2.3 Given unit vectors $\hat{\imath} = (1,0,0)$, $\hat{\jmath} = (0,1,0)$, and $\hat{k} = (0,0,1)$, find the following cross products: **a)** $\hat{\imath} \times \hat{\imath}$, **b)** $\hat{\imath} \times \hat{\jmath}$, **c)** $(-\hat{\imath}) \times \hat{k} + \hat{\jmath} \times \hat{\imath}$, **d)** $\hat{k} \times \hat{\jmath} + \hat{\imath} \times \hat{\imath} + \hat{\jmath} \times \hat{k} + \hat{\jmath} \times \hat{\imath}$.

P2.4 Given $\vec{v} = (2,-1,3)$ and $\vec{w} = (1,0,1)$, compute the following vector products: **a)** $\vec{v} \cdot \vec{w}$, **b)** $\vec{v} \times \vec{w}$, **c)** $\vec{v} \times \vec{v}$, and **d)** $\vec{w} \times \vec{w}$.

P2.5 Compute the product $M\vec{v}$ where $M = \left[\begin{smallmatrix} \alpha & \beta \\ \gamma & \delta \end{smallmatrix}\right]$ and $\vec{v} = \left[\begin{smallmatrix} z_1 \\ z_2 \end{smallmatrix}\right]$.

P2.6 Consider the following matrices of different dimensions:

$$A = \begin{bmatrix} 2 & 3 \\ 5 & 6 \end{bmatrix}, \; B = \begin{bmatrix} 7 & 8 \\ 1 & 4 \end{bmatrix}, \; C = \begin{bmatrix} -1 & 4 \\ 2 & 1 \\ -1 & 2 \end{bmatrix}, \; U = \begin{bmatrix} 2 & 1 \end{bmatrix}, \text{ and } V = \begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Compute the following matrix expressions:

a) $A - B$ **b)** $5C^{\mathsf{T}}$ **c)** AB **d)** AC **e)** CA **f)** UA **g)** AV
h) UV **i)** VU **j)** $\mathrm{Tr}(A)$ **k)** $\mathrm{Tr}(B)$ **l)** $\det(A)$ **m)** $\det(B)$

Chapter 3

Computational linear algebra

This chapter covers the computational aspects of performing matrix calculations. Understanding matrix computations is important because the rest of the chapters in this book depend on them. Suppose we're given a huge matrix $A \in \mathbb{R}^{n \times n}$ with $n = 1000$. Behind the innocent-looking mathematical notation of the matrix inverse A^{-1}, the matrix product AA, and the matrix determinant $\det(A)$, are hidden monster computations involving all the $1000 \times 1000 = 1$ million entries of the matrix A. Millions of arithmetic operations must be performed... so I hope you have at least a thousand pencils ready!

Okay, calm down. I won't *actually* make you calculate millions of arithmetic operations. In fact, to learn linear algebra, it is sufficient to know how to carry out calculations with 3×3 and 4×4 matrices. Yet, even for such moderately sized matrices, computing products, inverses, and determinants by hand are serious computational tasks. If you're ever required to take a linear algebra final exam, you'll need to make sure you can do these calculations quickly. And even if no exam looms in your imminent future, it's still important you practice matrix operations by hand to get a feel for them.

This chapter will introduce you to four important computational tasks involving matrices.

Gauss–Jordan elimination Suppose we're trying to solve two equations in two unknowns x and y:

$$
\begin{aligned}
ax + by &= c,\\
dx + ey &= f.
\end{aligned}
$$

If we add α-times the first equation to the second equation, we obtain an equivalent system of equations:

$$\begin{aligned} ax \; + \qquad\quad by &= \qquad c \\ (d+\alpha a)x + (e+\alpha b)y &= f + \alpha c. \end{aligned}$$

This is called a *row operation*: we added α-times the first row to the second row. Row operations change the coefficients of the system of equations, but leave the solution unchanged. Gauss–Jordan elimination is a systematic procedure for solving systems of linear equations using row operations.

Matrix product The product AB between matrices $A \in \mathbb{R}^{m\times\ell}$ and $B \in \mathbb{R}^{\ell\times n}$ is the matrix $C \in \mathbb{R}^{m\times n}$ whose coefficients c_{ij} are defined by the formula $c_{ij} = \sum_{k=1}^{\ell} a_{ik}b_{kj}$ for all $i \in [1,\ldots,m]$ and $j \in [1,\ldots,n]$. In Section 3.3, we'll unpack this formula and learn about its intuitive interpretation: that computing $C = AB$ is computing all the dot products between the rows of A and the columns of B.

Determinant The determinant of a matrix A, denoted $\det(A)$, is an operation that gives us useful information about the linear independence of the rows of the matrix. The determinant is connected to many notions of linear algebra: linear independence, geometry of vectors, solving systems of equations, and matrix invertibility. We'll discuss these aspects of determinants in Section 3.4.

Matrix inverse In Section 3.5, we'll build upon our knowledge of Gauss–Jordan elimination, matrix products, and determinants to derive three different procedures for computing the matrix inverse A^{-1}.

3.1 Reduced row echelon form

Solving equations

Augmented matrix

Row operations

We can manipulate the rows of an augmented matrix without changing its solutions. We're allowed to perform the following three types of row operations:

- Add a multiple of one row to another row
- Swap the position of two rows
- Multiply a row by a constant

Definitions

- The *solution* to a system of linear equations in the variables $x_1, x_2, \ldots, x_n$ is the set of values $\{(x_1, x_2, \ldots, x_n)\}$ that satisfy *all* the equations.
- The *pivot* for row j of a matrix is the left-most nonzero entry in the row j. Any *pivot* can be converted into a *leading one* by an appropriate scaling of that row.
- *Gaussian elimination* is the process of bringing a matrix into *row echelon form.*
- A matrix is said to be in *row echelon form* (REF) if all entries below the leading ones are zero. This form can be obtained by adding or subtracting the row with the leading one from the rows below it.
- *Gaussian-Jordan elimination* is the process of bringing a matrix into *reduced row echelon form.*
- A matrix is said to be in *reduced row echelon form* (RREF) if all the entries below *and above* the pivots are zero. Starting from the REF, we obtain the RREF by subtracting the row containing the pivots from the rows above that row.
- $\text{rank}(A)$: the *rank* of the matrix A is the number of pivots in the RREF of A.

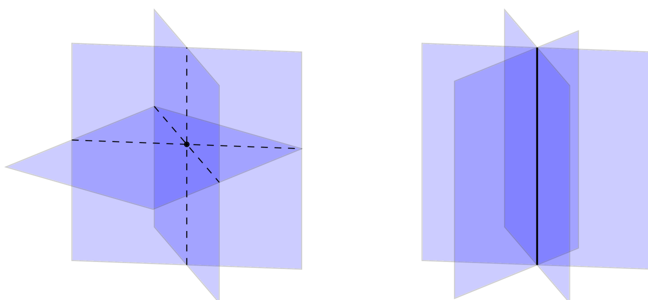
Gauss–Jordan elimination algorithm

Number of solutions

Geometric interpretation

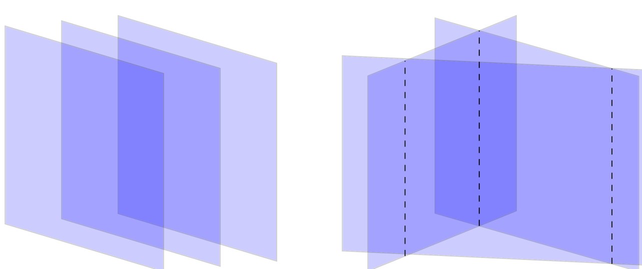
Lines in two dimensions

Planes in three dimensions



(a) One solution (b) Infinite solutions

Figure 3.1: Three planes can intersect at a unique point, as in figure (a); or along a line, as in figure (b). In the first case, there is a unique point (x_o, y_o, z_o) common to all three planes. In the second case, all points on the line $\{p_o + t\vec{v},\ \forall t \in \mathbb{R}\}$ are shared by the planes.



(a) No solution (b) No solution

Figure 3.2: These illustrations depict systems of three equations in three unknowns that have no solution. No common points of intersection exist.

Computer power

The computer algebra system at `http://live.sympy.org` can be used to compute the reduced row echelon form of any matrix.

Here is an example of how to create a sympy `Matrix` object:

```
>>> from sympy.matrices import Matrix
>>> A = Matrix([[1, 2,  5],      # use SHIFT+ENTER for newline
                [3, 9, 21]])
```

In `Python`, we define lists using the square brackets `[` and `]`. A matrix is defined as a list of lists.

To compute the reduced row echelon form of `A`, call its `rref()` method:

```
>>> A.rref()
( [1, 0, 1] # RREF of A          # locations of pivots
  [0, 1, 2],                     [0, 1]                )
```

The `rref()` method returns a tuple containing the RREF of `A` and an array that tells us the 0-based indices of the columns that contain leading ones. Usually, we'll want to find the RREF of `A` and ignore the pivots; to obtain the RREF without the pivots, select the first (index zero) element in the result of `A.rref()`:

```
>>> Arref = A.rref()[0]
>>> Arref
[1, 0, 1]
[0, 1, 2]
```

The `rref()` method is the fastest way to obtain the reduced row echelon form of a `SymPy` matrix. The computer will apply the Gauss–Jordan elimination procedure and show you the answer. If you want to see the intermediary steps of the elimination procedure, you can also manually apply row operations to the matrix.

Example Let's compute the reduced row echelon form of the same augmented matrix by using row operations in `SymPy`:

```
>>> A = Matrix([[1, 2,  5],
                [3, 9, 21]])

>>> A[1,:] = A[1,:] - 3*A[0,:]
>>> A
[1, 2, 5]
[0, 3, 6]
```

We use the notation `A[i,:]` to refer to entire rows of the matrix. The number `i` specifies the 0-based row index: the first row of `A` is `A[0,:]` and the second row is `A[1,:]`. The code example above implements the row operation $R_2 \leftarrow R_2 - 3R_1$.

To obtain the reduced row echelon form of the matrix `A`, we carry out two more row operations, $R_2 \leftarrow \frac{1}{3}R_2$ and $R_1 \leftarrow R_1 - 2R_2$, using the following commands:

```
>>> A[1,:] = S(1)/3*A[1,:]
>>> A[0,:] = A[0,:] - 2*A[1,:]
>>> A
[1, 0, 1]            # the same result as A.rref()[0]
[0, 1, 2]
```

Note we represent the fraction $\frac{1}{3}$ as `S(1)/3` in order to obtain the exact rational expression `Rational(1,3)`. If we were to input $\frac{1}{3}$ as `1/3`, `SymPy` would interpret this either as integer or floating point division, which is not what we want. The single-letter helper function `S` is an alias for the function `sympify`, which ensures a `SymPy` object is produced. Another way to input the exact fraction $\frac{1}{3}$ is `S('1/3')`.

If you need to swap two rows of a matrix, you can use the standard Python tuple assignment syntax. To swap the position of the first and second rows, use

```
>>> A[0,:], A[1,:] = A[1,:], A[0,:]
>>> A
[0, 1, 2]
[1, 0, 1]
```

Using row operations to compute the reduced row echelon form of a matrix allows you to see the intermediary steps of a calculation; which is useful, for instance, when checking the correctness of your homework problems.

There are other applications of matrix methods that use row operations (see Section 7.6), so it's good idea to know how to use `SymPy` for this purpose.

Discussion

Exercises

3.2 Matrix equations

Introduction

Matrix times vector

Matrix times matrix

Matrix times matrix variation

Exercises

3.3 Matrix multiplication

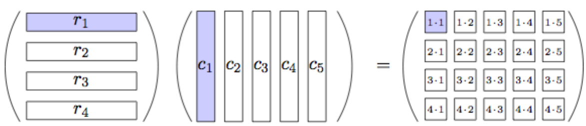


Figure 3.3: Matrix multiplication is performed rows-times-columns. The first-row, first-column entry of the product is the dot product of r_1 and c_1.

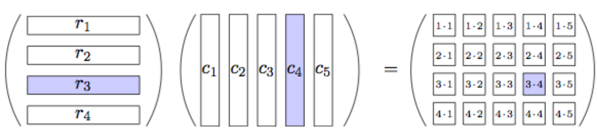


Figure 3.4: The third-row, fourth-column entry of the product is computed by taking the dot product of r_3 and c_4.

Matrix multiplication rules

- Matrix multiplication is associative:

$$(AB)C = A(BC) = ABC.$$

- The touching dimensions of the matrices must be the same. For the triple product ABC to exist, the rows of A must have the same dimension as the columns of B, and the rows of B must have the same dimension as the columns of C.

- Given two matrices $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{n\times k}$, the product AB is an $m \times k$ matrix.
- Matrix multiplication is not a commutative operation.

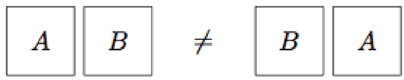


Figure 3.5: The order of multiplication matters for matrices: the product AB does not equal the product BA.

Example

Applications

Composition of linear transformations

Row operations as matrix products

Exercises

3.4 Determinants

Formulas

The determinant of a 2×2 matrix is

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) \equiv \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

The formulas for the determinants of larger matrices are defined recursively. For example, the determinant of a 3×3 matrix is defined in terms of 2×2 determinants:

$$\begin{aligned}
&\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \\
&\quad = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\
&\quad = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{13}(a_{21}a_{32} - a_{22}a_{31}) \\
&\quad = a_{11}a_{22}a_{33} \qquad - a_{12}a_{21}a_{33} \qquad + a_{13}a_{21}a_{32} \\
&\qquad\quad - a_{11}a_{23}a_{32} \qquad + a_{12}a_{23}a_{31} \qquad - a_{13}a_{22}a_{31}.
\end{aligned}$$

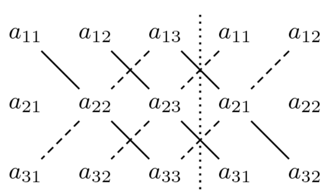


Figure 3.6: Computing the determinant using the extended array trick. The solid lines indicate the positive terms while the dashed lines indicate the negative terms in the determinant calculation.

Geometric interpretation

Area of a parallelogram

Volume of a parallelepiped

Sign and absolute value of the determinant

Properties

Let A and B be two square matrices of the same dimension. The determinant operation has the following properties:

- $\det(AB) = \det(A)\det(B) = \det(B)\det(A) = \det(BA)$

- If $\det(A) \neq 0$, the matrix is invertible and $\det(A^{-1}) = \frac{1}{\det(A)}$
- $\det\left(A^{\mathsf{T}}\right) = \det(A)$
- $\det(\alpha A) = \alpha^n \det(A)$, for an $n \times n$ matrix A
- $\det(A) = \prod_{i=1}^{n} \lambda_i$, where $\{\lambda_i\} = \operatorname{eig}(A)$ are the eigenvalues of A

The effects of row operations on determinants

Add a multiple of one row to another row

Adding a multiple of one row of a matrix to another row does not change the determinant of the matrix.

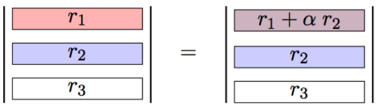


Figure 3.7: Row operations of the form $\mathcal{R}_\alpha : R_i \leftarrow R_i + \alpha R_j$ do not change the value of the matrix determinant.

Swapping rows

Multiply a row by a constant

Zero-vs-nonzero determinant property

Applications

Cross product as a determinant

Cramer's rule

Linear independence test

Eigenvalues

Exercises

Links

3.5 Matrix inverse

Existence of an inverse

Adjugate matrix approach

Using row operations

Example

Using elementary matrices

Using a computer algebra system

You can use a computer algebra system to specify matrices and compute their inverses. Let's illustrate how to find the matrix inverse using the computer algebra system at `http://live.sympy.org`.

```
>>> from sympy.matrices import Matrix
>>> A = Matrix( [ [1,2],[3,9] ] )   # define a Matrix object
>>> A.inv()                          # call the inv method on A
[ 3, -2/3]
[-1,  1/3]
```

Discussion

Invertibility

Exercises

3.6 Computational problems

We've reached the problem section where you're supposed to practice all the computational techniques of linear algebra. This is not going to be the most exciting three hours of your life, but you'll get through it. You need to know how to solve computational problems by hand and apply the Gauss–Jordan elimination procedure; and you need to know how to multiply matrices, calculate determinants, and find matrix inverses. These computational techniques enable all the advanced procedures we'll develop later in the book. If you skip these practice problems, you'll have trouble later when it comes to mastering more advanced topics that rely on these basic matrix operations as building blocks. Do this important work now, and you'll be on your way to becoming fluent in linear algebra computations... plus, the rest of the book will be much more pleasant.

P3.1 Mitchell is on a new diet. His target is to eat exactly 25 grams of fat and 32 grams of protein for lunch today. There are two types of food in the fridge, x and y. One serving of food x contains one gram of fat and two grams of protein, while a serving of food y contains five grams of fat and one gram of protein. To figure out how many servings of each type of food he should eat, Mitchell writes the following system of equations:

$$\begin{array}{rcrcl} x & + & 5y & = & 25 \\ 2x & + & y & = & 32 \end{array} \qquad \Rightarrow \qquad \left[\begin{array}{cc|c} 1 & 5 & 25 \\ 2 & 1 & 32 \end{array}\right].$$

Help Mitchell find how many servings of x and y he should eat.

Hint: Find the reduced row echelon form of the augmented matrix.

P3.2 Alice, Bob, and Charlotte are solving this system of equations:

$$\begin{aligned} 3x \quad &+3y &= 6 \\ 2x \quad &+\tfrac{3}{2}y &= 5 \end{aligned} \qquad \Rightarrow \qquad \left[\begin{array}{cc|c} 3 & 3 & 6 \\ 2 & \frac{3}{2} & 5 \end{array}\right].$$

Alice follows the standard procedure to obtain a leading one by performing the row operation $R_1 \leftarrow \frac{1}{3}R_1$. Bob starts with a different row operation, applying $R_1 \leftarrow R_1 - R_2$ to obtain a leading one. Charlotte takes a third approach by swapping the first and second rows: $R_1 \leftrightarrow R_2$. Their respective versions of the augmented matrix are shown below.

a) $\left[\begin{array}{cc|c} 1 & 1 & 2 \\ 2 & \frac{3}{2} & 5 \end{array}\right]$ **b)** $\left[\begin{array}{cc|c} 1 & \frac{3}{2} & 1 \\ 2 & \frac{3}{2} & 5 \end{array}\right]$ **c)** $\left[\begin{array}{cc|c} 2 & \frac{3}{2} & 5 \\ 3 & 3 & 6 \end{array}\right]$

Help Alice, Bob, and Charlotte finish solving the system of equations by writing the remaining row operations each of them must perform to bring their version of the augmented matrix into reduced row echelon form.

P3.3 Find the solutions to the systems of equations that correspond to the following augmented matrices.

a) $\left[\begin{array}{cc|c} -1 & -2 & -2 \\ 3 & 3 & 0 \end{array}\right]$ **b)** $\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ -2 & 3 & 3 & -1 \\ -1 & 0 & 1 & 2 \end{array}\right]$ **c)** $\left[\begin{array}{ccc|c} 2 & -2 & 3 & 2 \\ 1 & -2 & -1 & 0 \\ -2 & 2 & 2 & 1 \end{array}\right]$

P3.4 Solve for C in the matrix equation $ABCD = AD$.

P3.5 Compute the product of three matrices:

$$\begin{bmatrix} 2 & 10 & -5 & 0 \\ 0 & 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 0 & 2 \\ 5 & 1 \\ -3 & -4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}.$$

P3.6 Given an unknown variable $\alpha \in \mathbb{R}$ and the matrices

$$A = \begin{bmatrix} \cos(\alpha) & 1 \\ -1 & -\sin(\alpha) \end{bmatrix}; \quad B = \begin{bmatrix} \sin(\alpha) & 0 \\ 0 & -\sin(\alpha) \end{bmatrix}; \quad C = \begin{bmatrix} 1 & -\cos(\alpha) \\ \sin(\alpha) & 1 \end{bmatrix},$$

compute the value of **a)** $A^2 + B^2$, **b)** $A^2 + C$, and **c)** $A^2 + C - B^2$. Give your answer in terms of α and use the double-angle formulas as needed.

P3.7 Find the determinants of the following matrices.

a) $\begin{bmatrix} 2 & 1 \\ 3 & 0 \end{bmatrix}$ **b)** $\begin{bmatrix} 0 & 5 & 3 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}$ **c)** $\begin{bmatrix} 1 & 2 & 0 \\ 3 & 1 & 1 \\ 4 & -2 & 0 \end{bmatrix}$

P3.8 Find the area of a parallelogram that has vectors $\vec{v} = (3, -5)$ and $\vec{w} = (1, -1)$ as its sides.

Hint: Use the formula from Section 3.4 (page 60).

P3.9 Find the inverses of the following matrices:

a) $\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ **b)** $\begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix}$ **c)** $\begin{bmatrix} 2 & 3 \\ 2 & 4 \end{bmatrix}$

Chapter 4

Geometric aspects of linear algebra

In this section, we'll study geometric objects like lines, planes, and vector spaces. We'll use what we learned about vectors and matrices in the previous chapters to perform geometric calculations such as projections and distance measurements.

Developing your intuition about the geometric problems of linear algebra is very important: of all the things you learn in this course, your geometric intuition will stay with you the longest. Years from now, you may not recall the details of the Gauss–Jordan elimination procedure, but you'll still remember that the solution to three linear equations in three variables corresponds to the intersection of three planes in $\mathbb{R}^3$.

4.1 Lines and planes

Points, *lines*, and *planes* are the basic building blocks of geometry. In this section, we'll explore these geometric objects, the equations that describe them, and their visual representations.

Concepts

- $p = (p_x, p_y, p_z)$: a *point* in $\mathbb{R}^3$
- $\vec{v} = (v_x, v_y, v_z)$: a *vector* in $\mathbb{R}^3$
- $\hat{v} = \frac{\vec{v}}{\|\vec{v}\|}$: the *unit vector* in the same direction as the vector $\vec{v}$
- An infinite line ℓ is a one-dimensional space defined in one of several possible ways:
  - ▷ $\ell : \{p_{\text{o}} + t\,\vec{v}, t \in \mathbb{R}\}$: a *parametric equation* of a line with direction vector $\vec{v}$ passing through the point p_{o}
  - ▷ $\ell : \left\{\frac{x-p_{\text{o}x}}{v_x} = \frac{y-p_{\text{o}y}}{v_y} = \frac{z-p_{\text{o}z}}{v_z}\right\}$: a *symmetric equation*

- An infinite plane P is a two-dimensional space defined in one of several possible ways:
  - ▷ $P : \{Ax + By + Cz = D\}$: a *general equation*
  - ▷ $P : \{p_\text{o} + s\,\vec{v} + t\,\vec{w},\ s, t \in \mathbb{R}\}$: a *parametric equation*
  - ▷ $P : \{\vec{n} \cdot [(x, y, z) - p_\text{o}] = 0\}$: a *geometric equation* of the plane that contains point p_o and has normal vector $\hat{n}$
- $d(a, b)$: the shortest *distance* between geometric objects a and b

Points

Lines

Lines as intersections of planes

Planes

Distance formulas

Distance between points

Distance between a line and the origin

Distance between a plane and the origin

Discussion

Exercises

4.2 Projections

Concepts

- $S \subseteq \mathbb{R}^n$: S is a *vector subspace* of $\mathbb{R}^n$. In this chapter, we assume $S \subseteq \mathbb{R}^3$. The subspaces of $\mathbb{R}^3$ are lines ℓ and planes P that pass through the origin.
- $S^\perp$: the orthogonal space to S, $S^\perp \equiv \{\vec{w} \in \mathbb{R}^n \mid \vec{w} \cdot S = 0\}$. The symbol $^\perp$ stands for *perpendicular to*.
- Π_S: the *projection* onto the subspace S.
- $\Pi_{S^\perp}$: the projection onto the orthogonal space $S^\perp$.

Definitions

The projection operation onto the subspace S is a linear transformation that takes as inputs vectors in $\mathbb{R}^3$, and produces outputs in the subspace S:

$$\Pi_S : \mathbb{R}^3 \;\to\; S.$$

The transformation Π_S, pronounced "projection onto S," cuts off all parts of the input that do not lie within the subspace S. We can understand Π_S by analyzing its action for different inputs:

- If $\vec{v} \in S$, then $\Pi_S(\vec{v}) = \vec{v}$.
- If $\vec{w} \in S^\perp$, then $\Pi_S(\vec{w}) = \vec{0}$.
- Linearity and the above two conditions imply that, for any vector $\vec{u} = \alpha\vec{v} + \beta\vec{w}$ with $\vec{v} \in S$ and $\vec{w} \in S^\perp$, we have

$$\Pi_S(\vec{u}) = \Pi_S(\alpha\vec{v} + \beta\vec{w}) = \alpha\vec{v}.$$

The *orthogonal subspace* to S is the set of vectors that are perpendicular to all vectors in S:

$$S^\perp \equiv \{\, \vec{w} \in \mathbb{R}^3 \mid \vec{w} \cdot \vec{s} = 0,\ \forall \vec{s} \in S \,\}.$$

Projection onto a line

Projection onto a plane

Distances formulas revisited

Projections matrices

Discussion

Exercises

4.3 Coordinate projections

Concepts

We can define three different types of bases for an n-dimensional vector space V:

- A generic basis $B_f = \{\vec{f}_1, \vec{f}_2, \ldots, \vec{f}_n\}$ consists of any set of linearly independent vectors in V.
- An orthogonal basis $B_e = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ consists of n mutually orthogonal vectors in V obeying $\vec{e}_i \cdot \vec{e}_j = 0$, $\forall i \neq j$.
- An orthonormal basis $B_{\hat{e}} = \{\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n\}$ is an orthogonal basis of unit vectors: $\hat{e}_i \cdot \hat{e}_j = 0$ and $\|\hat{e}_i\| = 1$, $\forall i \in \{1, 2, \ldots, n\}$.

A vector $\vec{v}$ is expressed as coordinates v_i with respect to any basis B:

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_n\vec{e}_n = (v_1, v_2, \ldots, v_n)_B.$$

We can use two different bases, B and B', to express the same vector:

- $\vec{v}$: a vector
- $[\vec{v}]_B = (v_1, v_2, \ldots, v_n)_B$: the vector $\vec{v}$ expressed in the basis B
- $[\vec{v}]_{B'} = (v'_1, v'_2, \ldots, v'_n)_{B'}$: the same vector $\vec{v}$ expressed in a different basis B'
- ${}_{B'}[\mathbb{1}]_B$: the change-of-basis matrix that converts from B coordinates to B' coordinates: $[\vec{v}]_{B'} = {}_{B'}[\mathbb{1}]_B \, [\vec{v}]_B$

Components with respect to a basis

Definition of a basis

A basis $B = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ for the vector space V has the following two properties:

- **Spanning property**: Any vector $\vec{v} \in V$ can be expressed as a linear combination of the basis vectors:
$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_n\vec{e}_n.$$
This property guarantees that the vectors in the basis B are *sufficient* to represent any vector in V.
- **Linear independence property**: The vectors that form the basis $B = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ are linearly independent. The linear independence of the vectors in the basis guarantees that none of the vectors $\vec{e}_i$ are redundant.

If a set of vectors $B = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ satisfies both properties, we say B is a basis for V. In other words, B can serve as a coordinate system for V.

Coordinates with respect to an orthonormal basis

Coordinates with respect to an orthogonal basis

Coordinates with respect to a generic basis

Change of basis

Change of basis to the standard basis

Links

Exercises

4.4 Vector spaces

Definitions

- V: a *vector space*

- $\vec{v}$: a *vector*. We use the notation $\vec{v} \in V$ to indicate the vector $\vec{v}$ is part of the vector space V.
- W: a *vector subspace*. We use the notation $W \subseteq V$ to indicate the vector space W is a subspace of the vector space V.
- *span*: the span of a set of vectors is the set of vectors that can be constructed as linear combinations of these vectors:

$$\text{span}(\vec{v}_1, \ldots, \vec{v}_n) \equiv \{\vec{v} \in V \mid \vec{v} = \alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n,\ \alpha_i \in \mathbb{R}\}.$$

For every matrix $M \in \mathbb{R}^{m\times n}$, we define the following *fundamental vector spaces* associated with the matrix M:

- $\mathcal{R}(M) \subseteq \mathbb{R}^n$: the *row space* of the matrix M consists of all possible linear combinations of the rows of the matrix M.
- $\mathcal{C}(M) \subseteq \mathbb{R}^m$: the *column space* of the matrix M consists of all possible linear combinations of the columns of the matrix M.
- $\mathcal{N}(M) \subseteq \mathbb{R}^n$: the *null space* of M is the set of vectors that go to the zero vector when multiplying M from the right: $\mathcal{N}(M) \equiv \{\vec{v} \in \mathbb{R}^n \mid M\vec{v} = \vec{0}\}$.
- $\mathcal{N}(M^{\mathsf{T}})$: the *left null space* of M is the set of vectors that go to the zero vector when multiplying M from the left: $\mathcal{N}(M^{\mathsf{T}}) \equiv \{\vec{w} \in \mathbb{R}^m \mid \vec{w}^{\mathsf{T}}M = \vec{0}^{\mathsf{T}}\}$.

The dimensions of the column space and the row space of a matrix are equal. We call this dimension the *rank* of the matrix: $\text{rank}(M) \equiv \dim(\mathcal{C}(M)) = \dim(\mathcal{R}(M))$.

Vector space

Span

Vector subspaces

Subspaces specified by constraints

Subspaces specified as a span

Subsets vs. subspaces

Matrix fundamental spaces

Matrices and systems of linear equations

Matrices and linear transformations

Matrix-vector and vector-matrix products

Left and right input spaces

Matrix rank

Summary

Linear independence

Basis

The rank–nullity theorem

Rank–nullity theorem. *For any matrix $M \in \mathbb{R}^{m\times n}$, the following statement holds:*

$$\operatorname{rank}(M) + \operatorname{nullity}(M) = n,$$

where the rank of M is $\operatorname{rank}(M) \equiv \dim(\mathcal{R}(M)) = \dim(\mathcal{C}(M))$ *and its nullity is defined as* $\operatorname{nullity}(M) \equiv \dim(\mathcal{N}(M))$.

Distilling bases

Links

Exercises

4.5 Vector space techniques

Finding a basis

Definitions

- $B = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$. A *basis* for an n-dimensional vector space S is a set of n linearly independent vectors that span S. Any vector $\vec{v} \in S$ can be written as a linear combination of the basis vectors:
$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_n\vec{e}_n.$$
A basis for an n-dimensional vector space contains exactly n vectors.
- $\dim(S)$: the dimension of the vector space S is equal to the number of vectors in a basis for S.

Bases for the fundamental spaces of matrices

Basis for the row space

Basis for the column space

Basis for the null space

Examples

Discussion

Dimensions

Importance of bases

Exercises

4.6 Geometric problems

So far, we've defined all the important linear algebra concepts like vectors and matrices, and we've learned some useful computational techniques like the Gauss–Jordan elimination procedure. It's now time to apply what you've learned to solve geometric problems.

Points, lines, and planes can be difficult to understand and conceptualize. But now that you're armed with the tools of vectors,

projections, and geometric intuition, you can solve all kinds of complicated geometric analysis problems—such as those waiting for you at the end of this paragraph. Remember to always sketch a diagram before you begin to write equations. Diagrams are great for visualizing and determining the steps you'll need to solve each problem.

P4.1 Find the intersections of the these pairs of lines: **a)** ℓ_1: $2x + y = 4$ and ℓ_2: $3x - 2y = -1$, **b)** ℓ_1: $y + x = 2$ and ℓ_2: $2x + 2y = 4$, **c)** ℓ_1: $y + x = 2$ and ℓ_2: $y - x = 0$.

P4.2 Find the lines of intersection between these pairs of planes: **a)** P_1: $3x - 2y - z = 2$ and P_2: $x + 2y + z = 0$, **b)** P_3: $2x + y - z = 0$ and P_4: $x + 2y + z = 3$.

P4.3 Given two vectors $\vec{u} = (2, 1, -1)$ and $\vec{v} = (1, 1, 1)$, find the projection of $\vec{v}$ onto $\vec{u}$, and the projection of $\vec{u}$ onto $\vec{v}$.

P4.4 Find a projection of $\vec{v} = (3, 4, 1)$ onto the plane $P : 2x - y + 4z = 4$.

P4.5 Find the coordinates of the vector $\vec{v} = 9\hat{\imath} + 5\hat{\jmath} + 4\hat{k} = (9, 5, 4)_{B_s}$ with respect to the basis $W = \{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$, which consists of the vectors $\vec{w}_1 \equiv (0, 0, 2)_{B_s}$, $\vec{w}_2 \equiv (0, 5, 0)_{B_s}$, and $\vec{w}_3 \equiv (3, 0, 0)_{B_s}$.

P4.6 Consider the vector space of three-dimensional vectors $V \equiv \mathbb{R}^3$. Which of the following sets are subspaces of V?

a) $W_1 \equiv \{(v_x, v_y, v_z) \in V \mid v_x + v_y = 0\}$
b) $W_2 \equiv \{(v_x, v_y, v_z) \in V \mid v_y v_z = 0\}$
c) $W_3 \equiv \{(v_x, v_y, v_z) \in V \mid v_x = v_y = v_z\}$
d) $W_4 \equiv \{(v_x, v_y, v_z) \in V \mid v_x \geq 0\}$
e) $W_5 \equiv \{(v_x, v_y, v_z) \in V \mid v_x + v_z = 3\}$

Hint: To form a subspace, a set must be closed under addition, closed under scalar multiplication, and contain the zero element.

P4.7 Suppose the set of vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ spans the vector space V. Prove that $\{\vec{v}_1, \vec{v}_2 - \vec{v}_1, \vec{v}_3 - \vec{v}_2, \vec{v}_4 - \vec{v}_3\}$ also spans V.

P4.8 Suppose the set of vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is linearly independent. Prove that $\{\vec{v}_1, \vec{v}_2 - \vec{v}_1, \vec{v}_3 - \vec{v}_2, \vec{v}_4 - \vec{v}_3\}$ is also a linearly independent set.

P4.9 Let $\vec{u}$ and $\vec{v}$ be distinct vectors from the vector space V, and assume that $\{\vec{u}, \vec{v}\}$ is a basis for V. Show that $\{\vec{u} + \vec{v}, a\vec{u}\}$ and $\{a\vec{u}, b\vec{v}\}$ are also bases for V, for any choice of nonzero constants a and b.

P4.10 Find a vector that is perpendicular to the vectors $\vec{v}_1 = (1, 2, 3, 0)$, $\vec{v}_2 = (0, 1, 1, 1)$, and $\vec{v}_3 = (1, 0, 2, 1)$.

Chapter 5

Linear transformations

Linear transformations are a central idea of linear algebra—they form the cornerstone that connects all the seemingly unrelated concepts we've studied so far. We previously introduced linear transformations, informally describing them as "vector functions." In this chapter, we'll formally define linear transformations, describe their properties, and discuss their applications.

In Section 5.2, we'll learn how matrices can be used to *represent* linear transformations. We'll show the matrix representations of important types of linear transformations like projections, reflections, and rotations. Section 5.3 discusses the relation between bases and matrix representations. We'll learn how the bases chosen for the input and output spaces determine the coefficients of matrix representations. A single linear transformation can correspond to many different matrix representations, depending on the choice of bases for the input and output spaces.

Section 5.4 discusses and characterizes the class of *invertible linear transformations*. This section serves to connect several topics we covered previously: linear transformations, matrix representations, and the fundamental vector spaces of matrices.

5.1 Linear transformations

Linear transformations take vectors as inputs and produce vectors as outputs. A transformation T that takes n-dimensional vectors as inputs and produces m-dimensional vectors as outputs is denoted $T\colon \mathbb{R}^n \to \mathbb{R}^m$.

The class of linear transformations includes most of the useful transformations of analytical geometry: stretchings, projections, reflections, rotations, and combinations of these. Since linear transfor-

mations describe and model many real-world phenomena in physics, chemistry, biology, and computer science, learning the theory behind them is worthwhile.

Concepts

Linear transformations are mappings between *vector inputs* and *vector outputs*. The following concepts describe the input and output spaces:

- V: the input vector space of T
- W: the output vector space of T
- $\dim(U)$: the dimension of the vector space U
- $T : V \to W$: a linear transformation that takes vectors $\vec{v} \in V$ as inputs and produces vectors $\vec{w} \in W$ as outputs. The notation $T(\vec{v}) = \vec{w}$ describes T acting on $\vec{v}$ to produce the output $\vec{w}$.

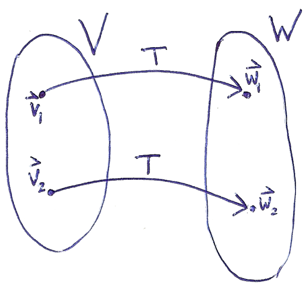


Figure 5.1: An illustration of the linear transformation $T : V \to W$.

- $\text{Im}(T)$: the *image space* of the linear transformation T is the set of vectors that T can output for some input $\vec{v} \in V$. The mathematical definition of the image space is

  $$\text{Im}(T) \equiv \{\vec{w} \in W \mid \vec{w} = T(\vec{v}), \text{ for some } \vec{v} \in V\} \quad \subseteq \quad W.$$

  The image space is the vector equivalent of the *image set* of a single-variable function $\text{Im}(f) \equiv \{y \in \mathbb{R} \,|\, y = f(x), \forall x \in \mathbb{R}\}$.

- $\text{Ker}(T)$: the *kernel* of the linear transformation T; the set of vectors mapped to the zero vector by T. The mathematical definition of the kernel is

  $$\text{Ker}(T) \equiv \{\vec{v} \in V \mid T(\vec{v}) = \vec{0}\} \quad \subseteq \quad V.$$

  The kernel of a linear transformation is the vector equivalent of the roots of a function: $\{x \in \mathbb{R} \,|\, f(x) = 0\}$.

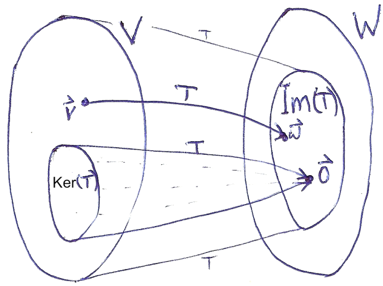


Figure 5.2: Two key properties of a linear transformation $T : V \to W$; its kernel $\text{Ker}(T) \subseteq V$, and its image space $\text{Im}(T) \subseteq W$.

Matrix representations

Given bases for the input and output spaces of a linear transformation T, the transformation's action on vectors can be represented as a matrix-vector product:

- $B_V = \{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$: a basis for the input vector space V
- $B_W = \{\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_m\}$: a basis for the output vector space W
- $M_T \in \mathbb{R}^{m \times n}$: a matrix representation of the linear transformation T:
  $$\vec{w} = T(\vec{v}) \qquad \Leftrightarrow \qquad \vec{w} = M_T\vec{v}.$$
  To be precise, we denote the matrix representation as ${}_{B_W}[M_T]_{B_V}$ to show it depends on the input and output bases.
- $\mathcal{C}(M_T)$: the *column space* of the matrix M_T
- $\mathcal{R}(M_T)$: the *row space* of the matrix M_T
- $\mathcal{N}(M_T)$: the *null space* the matrix M_T

Properties of linear transformations

Linearity

Linear transformations as black boxes

Input and output spaces

Linear transformations as matrix multiplications

Finding the matrix

Input and output spaces

Composition of linear transformations

Importance of the choice of bases

Invertible transformations

Affine transformations

Discussion

The most general linear transformation

Links

Exercises

5.2 Finding matrix representations

Concepts

The previous section covered linear transformations and their matrix representations:

- $T : \mathbb{R}^n \to \mathbb{R}^m$: a linear transformation that takes inputs in $\mathbb{R}^n$ and produces outputs in $\mathbb{R}^m$
- $M_T \in \mathbb{R}^{m\times n}$: the matrix representation of T

The action of the linear transformation T is equivalent to multiplication by the matrix M_T:

$$\vec{w} = T(\vec{v}) \qquad \Leftrightarrow \qquad \vec{w} = M_T\vec{v}.$$

Theory

Projections

X projection

Y projection

Projection onto a vector

Projection onto a plane

Projections as outer products

Projections are idempotent

Subspaces

Reflections

X reflection

Y reflection

Diagonal reflection

Reflections through lines and planes

Rotations

We'll now find the matrix representations for *rotation* transformations. The counterclockwise rotation by the angle θ is denoted R_θ. Figure 5.3 illustrates the action of the rotation R_θ: the point A is rotated around the origin to become the point B.

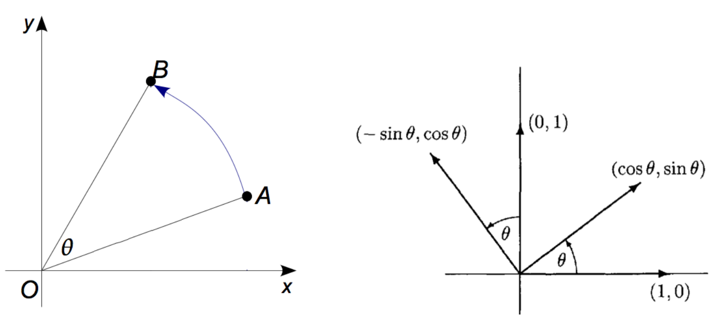


Figure 5.3: The linear transformation R_θ rotates every point in the plane by the angle θ in the counterclockwise direction. Note the effect of R_θ on the basis vectors $(1, 0)$ and $(0, 1)$.

To find the matrix representation of R_θ, probe it with the standard basis as usual:

$$M_{R_\theta} = \left[R_\theta\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix} \right) \; R_\theta\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix} \right) \right].$$

To compute the values in the first column, observe that R_θ rotates the vector $(1,0)^{\mathsf{T}} = 1\angle 0$ to the vector $1\angle\theta = (\cos\theta, \sin\theta)^{\mathsf{T}}$. The second input $\hat{e}_2 = (0,1)^{\mathsf{T}} = 1\angle\frac{\pi}{2}$ is rotated to $1\angle(\frac{\pi}{2}+\theta) = (-\sin\theta, \cos\theta)^{\mathsf{T}}$. Therefore, the matrix for R_θ is

$$M_{R_\theta} = \begin{bmatrix} | & | \\ 1\angle\theta & 1\angle(\frac{\pi}{2}+\theta) \\ | & | \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

Finding the matrix representation of a linear transformation is like a colouring-book activity for mathematicians. Filling in the columns is just like colouring inside the lines—nothing too complicated.

Inverses

Nonstandard-basis probing

Eigenspaces

Links

Exercises

5.3 Change of basis for matrices

Concepts

You should already be familiar with the concepts of vector spaces, bases, vector coefficients with respect to different bases, and the change-of-basis transformation:

- V: an n-dimensional vector space
- $\vec{v}$: a vector in V
- $B = \{\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n\}$: an orthonormal basis for V
- $[\vec{v}]_B = (v_1, v_2, \ldots, v_n)_B$: the vector $\vec{v}$ expressed in the basis B
- $B' = \{\hat{e}'_1, \hat{e}'_2, \ldots, \hat{e}'_n\}$: another orthonormal basis for V
- $[\vec{v}]_{B'} = (v'_1, v'_2, \ldots, v'_n)_{B'}$: the vector $\vec{v}$ expressed in the basis B'
- ${}_{B'}[\mathbb{1}]_B$: the change-of-basis matrix that converts from B coordinates to B' coordinates, $[\vec{v}]_{B'} = {}_{B'}[\mathbb{1}]_B \, [\vec{v}]_B$
- ${}_B[\mathbb{1}]_{B'}$: the inverse change-of-basis matrix $[\vec{v}]_B = {}_B[\mathbb{1}]_{B'} \, [\vec{v}]_{B'}$ (note that ${}_B[\mathbb{1}]_{B'} = \left({}_{B'}[\mathbb{1}]_B\right)^{-1}$)

Matrix components

Change of basis for matrices

Similarity transformation

Exercises

5.4 Invertible matrix theorem

Invertible matrix theorem. *For an $n \times n$ matrix A, the following statements are equivalent:*

(**1**) *A is invertible*

(**2**) *The equation $A\vec{x} = \vec{b}$ has exactly one solution for each $\vec{b} \in \mathbb{R}^n$*

(**3**) *The null space of A contains only the zero vector $\mathcal{N}(A) = \{\vec{0}\}$*

(**4**) *The equation $A\vec{x} = \vec{0}$ has only the trivial solution $\vec{x} = \vec{0}$*

(**5**) *The columns of A form a basis for $\mathbb{R}^n$:*
- *The columns of A are linearly independent*
- *The columns of A span $\mathbb{R}^n$; $\mathcal{C}(A) = \mathbb{R}^n$*

(**6**) *The rank of the matrix A is n*

(**7**) *The RREF of A is the $n \times n$ identity matrix $\mathbb{1}_n$*

(**8**) *The transpose matrix A^{T} is invertible*

(**9**) *The rows of A form a basis for $\mathbb{R}^n$:*
- *The rows of A are linearly independent*
- *The rows of A span $\mathbb{R}^n$; $\mathcal{R}(A) = \mathbb{R}^n$*

(**10**) *The determinant of A is nonzero* $\det(A) \neq 0$

Proof of the invertible matrix theorem

Proofs by contradiction

Review of definitions

Proof of the invertible matrix theorem. □

Nice work! By proving the chain of implications (**1**) ⇒ (**2**) ⇒ ··· ⇒ (**7**) ⇒ (**1**), we've shown that the first seven statements are equivalent. If one of these statements is true, then all others are true—just follow

the arrows of implication. Alternatively, if one statement is false, all statements are false, as we see by following the arrows of implication in the backward direction.

We also "attached" statements (**8**), (**9**), and (**10**) to the main loop of implications using "if and only if" statements. Thus, we've shown the equivalence of all 10 statements, which completes the proof.

Invertible linear transformations

Kernel and null space

Linear transformations as functions

Links

Exercises

Discussion

5.5 Linear transformations problems

Understanding linear transformations is extremely important for your overall understanding of linear algebra. This is why it's crucial for you to solve all the problems in this section. By working on these problems, you'll discover whether you really understand all the new material covered in this chapter. Remember in the book's introduction, when I mentioned that linear algebra is all about vectors and linear transformations? Well, if you can solve all the problems in this section, you're 80% of the way to understanding all of linear algebra.

P5.1 Determine whether each of the following transformations are linear.

a) $T_1(x, y) = (y, x + y)$ **b)** $T_2(x, y) = (x + 3, y - 3)$
c) $T_3(x, y) = (|x|, |y|)$ **d)** $T_4(x, y, z) = (3x - 2y + z, 2x + y - 4z)$
e) $T_5(x) = (x, 2x, 3x)$ **f)** $T_6(x, y, z, w) = (5x, 4y, 3z, 2w, 1)$

If the transformation is linear, find its matrix representation. If the transformation is nonlinear, find an example of a calculation where the linear property fails.

P5.2 Find image space $\text{Im}(T)$ for the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(x, y) = (x, x - y, 2y)$.

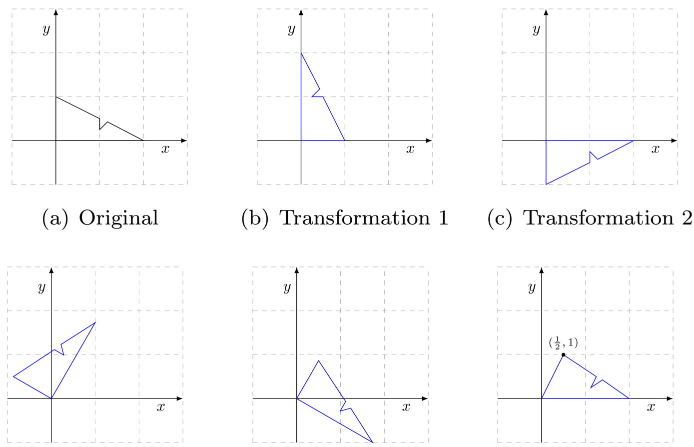
P5.3 Consider the transformation defined by $T(\vec{v}) \equiv \vec{a} \times \vec{v}$, where $\vec{a} = (a_x, a_y, a_z)$. Find the matrix representation of T. What is the kernel of T?

P5.4 Given the linear transformation $T(x, y, z) = (x, x + y, x + y + z)$, the standard basis for $\mathbb{R}^3$ $B = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$, and the alternative basis $B' = \{(1, 0, 0), (0, 1, 1), (0, 1, -1)\}$, find the following matrix representations of T:

a) ${}_B[M_T]_B$: the representation of T with respect to the standard basis
b) ${}_{B'}[M_T]_{B'}$: the representation of T with respect to the basis B'
c) ${}_B[M_T]_{B'}$: the mixed representation of T with respect to input vectors expressed in the basis B' and output vectors in the standard basis B

P5.5 Find the matrix representations of each of the transformations shown in Figure 5.4. The input to each transformation is the triangle with vertices $(0,0)$, $(2,0)$, $(0,1)$ that is shown in Figure 5.4 image (a).

Hint: Your answers should be 2×2 matrices. Recall that $\sin(\frac{\pi}{6}) = \frac{1}{2}$.



(a) Original (b) Transformation 1 (c) Transformation 2

(d) Transformation 3 (e) Transformation 4 (f) Transformation 5

Figure 5.4: The effects of various linear transformations on a triangle.

P5.6 Prove that statement (**2**) implies statement (**1**) in the invertible matrix theorem.

P5.7 Suppose $T : V \to W$ is an injective linear transformation, and $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a linearly independent set in V. Prove that $\{T(\vec{v}_1), \ldots, T(\vec{v}_n)\}$ is a linearly independent set in W.

Chapter 6

Theoretical linear algebra

Let's take a trip down memory lane: 150 pages ago, we embarked on a mind-altering journey through the land of linear algebra. We encountered vector and matrix operations. We studied systems of linear equations, solving them with row operations. We covered miles of linear transformations and their matrix representations. With the skills you've acquired to reach this point, you're ready to delve into the abstract, theoretical aspects of linear algebra—that is, since you know all the useful stuff, you can officially move on to the cool stuff. The lessons in this chapter are less concerned with calculations and more about mind expansion.

In math, we often use abstraction to find the commonalities between different mathematical objects. These parallels give us a deeper understanding of the mathematical structures we compare. This chapter extends what we know about the vector space $\mathbb{R}^n$ to the realm of abstract vector spaces of vector-like mathematical objects (Section 6.3). We'll discuss linear independence, find bases, and count dimensions for these abstract vector spaces. We'll define abstract inner product operations and use them to generalize the concept of orthogonality for abstract vectors (Section 6.4). We'll explore the Gram–Schmidt orthogonalization procedure for distilling orthonormal bases from non-orthonormal bases (Section 6.5). Finally, we'll introduce vectors and matrices with complex coefficients (Section 6.7). This section also reviews everything we've learned in this book, so be sure to read it even if complex numbers are not required for your course. Along the way, we'll develop a taxonomy for the different types of matrices according to their properties and applications (Section 6.2). We'll also investigate matrix decompositions—techniques for splitting matrices into products of simpler matrices (Section 6.6). The chapter begins by discussing the most important decomposition technique of them all: the *eigendecomposition*, which is a way to uncover the "natural basis" for any matrix.

6.1 Eigenvalues and eigenvectors

Definitions

- A: an $n \times n$ square matrix. The entries of A are denoted as a_{ij}.
- $\text{eig}(A) \equiv (\lambda_1, \lambda_2, \ldots, \lambda_n)$: the list of *eigenvalues* of A. Eigenvalues are usually denoted by the Greek letter *lambda*. Note that some eigenvalues could be repeated in the list.
- $p(\lambda) = \det(A - \lambda \mathbb{1})$: the *characteristic polynomial* of A. The eigenvalues of A are the roots of the characteristic polynomial.
- $\{\vec{e}_{\lambda_1}, \vec{e}_{\lambda_2}, \ldots, \vec{e}_{\lambda_n}\}$: the set of *eigenvectors* of A. Each eigenvector is associated with a corresponding eigenvalue.
- $\Lambda \equiv \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$: the diagonalized version of A. The matrix Λ contains the eigenvalues of A on the diagonal:

$$\Lambda = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & 0 \\ 0 & 0 & \lambda_n \end{bmatrix}.$$

  The matrix Λ is the matrix A expressed in its eigenbasis.
- Q: a matrix whose columns are eigenvectors of A:

$$Q \equiv \begin{bmatrix} | & & | \\ \vec{e}_{\lambda_1} & \cdots & \vec{e}_{\lambda_n} \\ | & & | \end{bmatrix} = {}_{B_s}[\mathbb{1}]_{B_\lambda}.$$

  The matrix Q corresponds to the *change-of-basis* matrix from the eigenbasis $B_\lambda = \{\vec{e}_{\lambda_1}, \vec{e}_{\lambda_2}, \vec{e}_{\lambda_3}, \ldots\}$ to the standard basis $B_s = \{\hat{\imath}, \hat{\jmath}, \hat{k}, \ldots\}$.
- $A = Q\Lambda Q^{-1}$: the *eigendecomposition* of the matrix A
- $\Lambda = Q^{-1}AQ$: the *diagonalization* of the matrix A

Eigenvalues

Eigenvectors

Eigendecomposition

Explanations

Applications

6.2 Special types of matrices

Diagonal matrices

Symmetric matrices

Upper triangular matrices

Identity matrix

Orthogonal matrices

Rotation matrices

Reflections

Permutation matrices

Positive matrices

Projection matrices

Normal matrices

Discussion

Exercises

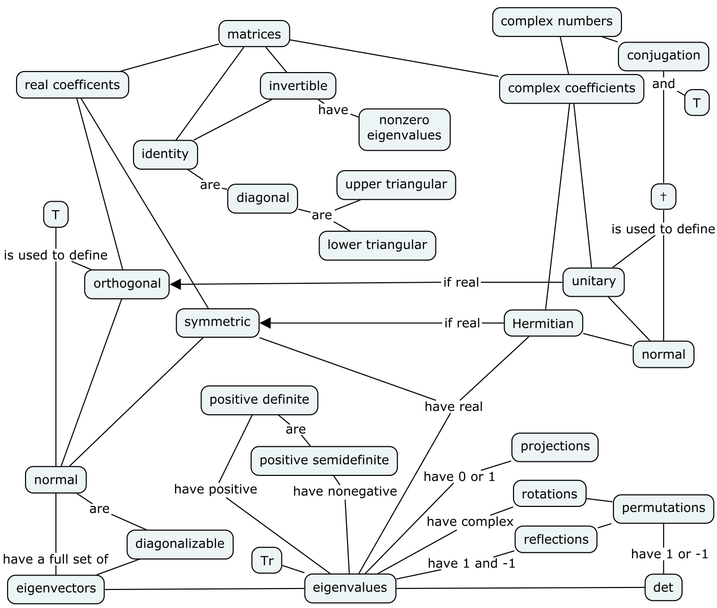


Figure 6.1: This concept map illustrates the connections and relations between special types of matrices. We can understand matrices through the constraint imposed on their eigenvalues or their determinants. This diagram shows only a subset of the many connections between the different types of matrices. We'll discuss matrices with complex coefficients in Section 6.7.

6.3 Abstract vector spaces

Definitions

An abstract vector space $(V, F, +, \cdot)$ consists of four things:

- A set of vector-like objects $V = \{\mathbf{u}, \mathbf{v}, \ldots\}$
- A field F of scalar numbers, usually $F = \mathbb{R}$
- An addition operation "+" for elements of V that dictates how to add vectors: $\mathbf{u} + \mathbf{v}$
- A scalar multiplication operation "$\cdot$" for scaling a vector by an element of the field. Scalar multiplication is usually denoted implicitly $\alpha\mathbf{u}$ (without the dot).

A vector space satisfies the following eight axioms, for all scalars $\alpha, \beta \in F$ and all vectors $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$:

1. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ (associativity of addition)
2. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ (commutativity of addition)
3. There exists a zero vector $\mathbf{0} \in V$, such that $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u} \in V$.
4. For every $\mathbf{u} \in V$, there exists an inverse element $-\mathbf{u}$ such that $\mathbf{u} + (-\mathbf{u}) = \mathbf{u} - \mathbf{u} = \mathbf{0}$.
5. $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ (distributivity I)
6. $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$ (distributivity II)
7. $\alpha(\beta\mathbf{u}) = (\alpha\beta)\mathbf{u}$ (associativity of scalar multiplication)
8. There exists a unit scalar 1 such that $1\mathbf{u} = \mathbf{u}$.

Theory

Examples

Matrices

Symmetric 2x2 matrices

Polynomials of degree n

Functions

Discussion

Links

Exercises

6.4 Abstract inner product spaces

Definitions

We'll work with vectors from an abstract vector space $(V, \mathbb{R}, +, \cdot)$ where:

- V is the set of vectors in the vector space.
- $\mathbb{R}$ is the *field* of real numbers. The coefficients of the abstract vectors are taken from this field.
- $+$ is the addition operation defined for elements of V.
- $\cdot$ is the scalar multiplication operation between an element of the field $\alpha \in \mathbb{R}$ and a vector $\mathbf{u} \in V$.

We define a new operation called *abstract inner product* for that space:

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}.$$

The abstract inner product takes as inputs two vectors $\mathbf{u}, \mathbf{v} \in V$ and produces real numbers as outputs: $\langle \mathbf{u}, \mathbf{v} \rangle \in \mathbb{R}$.

We define the following related quantities in terms of the inner product operation:

- $\|\mathbf{u}\| \equiv \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$: the *norm* or *length* of an abstract vector
- $d(\mathbf{u}, \mathbf{v}) \equiv \|\mathbf{u} - \mathbf{v}\|$: the *distance* between two vectors

Orthogonality

Norm

Distance

Examples

Matrix inner product

Hilbert–Schmidt norm

Function inner product

Generalized dot product

Valid and invalid inner product spaces

Discussion

Exercises

6.5 Gram–Schmidt orthogonalization

Definitions

- V: an n-dimensional vector space
- $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$: a generic basis for the space V
- $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$: an *orthogonal basis* for V. Each vector $\mathbf{e}_i$ is orthogonal to all other vectors: $\mathbf{e}_i \cdot \mathbf{e}_j = 0$, for $i \neq j$.
- $\{\hat{\mathbf{e}}_1, \hat{\mathbf{e}}_2, \ldots, \hat{\mathbf{e}}_n\}$: an *orthonormal* basis for V. An orthonormal basis is an orthogonal basis of unit vectors.

We assume the vector space V is equipped with an inner product operation:

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}.$$

The following operations are defined in terms of the inner product:

- The *length* of a vector $\|\mathbf{v}\| = \langle \mathbf{v}, \mathbf{v} \rangle$
- The *projection* operation. The projection of the vector $\mathbf{u}$ onto the subspace spanned by the vector $\mathbf{e}$ is denoted $\Pi_{\mathbf{e}}(\mathbf{u})$ and is computed using

$$\Pi_{\mathbf{e}}(\mathbf{u}) = \frac{\langle \mathbf{u}, \mathbf{e} \rangle}{\|\mathbf{e}\|^2}\mathbf{e}.$$

- The *projection complement* of the projection $\Pi_{\mathbf{e}}(\mathbf{u})$ is the vector $\mathbf{w}$ that we must add to $\Pi_{\mathbf{e}}(\mathbf{u})$ to recover the original vector $\mathbf{u}$:

$$\mathbf{u} = \Pi_{\mathbf{e}}(\mathbf{u}) + \mathbf{w} \qquad \Rightarrow \qquad \mathbf{w} = \mathbf{u} - \Pi_{\mathbf{e}}(\mathbf{u}).$$

  The vector $\mathbf{w}$ is orthogonal to the vector $\mathbf{e}$, $\langle \mathbf{w}, \mathbf{e} \rangle = 0$.

Orthonormal bases are nice

Orthogonalization

Gram–Schmidt orthogonalization procedure

Discussion

Exercises

6.6 Matrix decompositions

Eigendecomposition

Singular value decomposition

LU decomposition

Cholesky decomposition

QR decomposition

Discussion

Links

[Cool retro video showing the steps of the SVD procedure]
`http://www.youtube.com/watch?v=R9UoFyqJca8`

Exercises

6.7 Linear algebra with complex numbers

Definitions

Recall the basic notions of complex numbers introduced in Section 1.16:

- i: the unit imaginary number; $i \equiv \sqrt{-1}$ or $i^2 = -1$
- $z = a + bi$: a complex number z whose real part is a and whose imaginary part is b
- $\mathbb{C}$: the set of complex numbers $\mathbb{C} = \{a + bi \mid a, b \in \mathbb{R}\}$
- $\mathrm{Re}\{z\} = a$: the *real* part of $z = a + bi$
- $\mathrm{Im}\{z\} = b$: the *imaginary* part of $z = a + bi$
- $\bar{z}$: the *complex conjugate* of z. If $z = a + bi$ then $\bar{z} = a - bi$
- $|z| = \sqrt{\bar{z}z} = \sqrt{a^2 + b^2}$: the *magnitude* or *length* of $z = a + bi$
- $\arg(z) = \tan^{-1}(\frac{b}{a})$: the *phase* or *argument* of $z = a + bi$

Complex vectors

Complex matrices

Hermitian transpose

Complex inner product

Linear algebra over the complex field

Complex eigenvalues

Special types of matrices

Unitary matrices

Hermitian matrices

Normal matrices

Inner product for complex vectors

Complex inner product spaces

Explanations

Linear algebra over other fields

Discussion

The adjoint operator

Matrix quantum mechanics

Exercises

6.8 Theory problems

It's now time to test your understanding of the theoretical concepts we discussed in this chapter. The eigenvector equation $A\vec{e}_\lambda = \lambda\vec{e}_\lambda$ is one of the deepest ideas in linear algebra. I've prepared several problems so you can challenge yourself and test your understanding of eigenvalues and eigenvectors. The problems will test your theoretical understanding as well as your stamina, because computing eigenvectors requires many steps of arithmetic and takes a long time. The first eigenvector problem you'll solve might take you up to an hour. Don't be alarmed by this—that's totally normal. After solving a few eigenvector problems, your problem-solving time will drop to 30 minutes; and quickly after that you'll able to solve eigenvalue problems easily in 15 minutes.

It's up to you how fluent you want to become. Certainly if you have a linear algebra exam coming up, it would be good to solve all the problems and maybe even solve problems in other books, too. If you're just reading about linear algebra for fun, you probably don't need to suffer through the steps of finding eigenvalues using only pen and paper. Solve the problems using `SymPy` instead—you can't say no to that!

In this chapter we also learned about abstract vector spaces, another important theoretical idea in linear algebra. All the techniques you've learned about vectors can be applied to polynomials, matrices, functions, and other vector-like objects. That's all nice in theory, but we're going to move beyond passive appreciation and get into the nitty gritty by solving problems that involve bases, linear independence, dimensions, and orthogonality in abstract vector spaces. It might seem like crazy stuff, but if you trust the idea of equivalent representations and the abstract notion of a linear transformation, you'll see it's all good and that you can work with abstract vectors.

Finally, the problems that involve linear algebra over the complex field will serve as the final review of what you've learned in this book. This is the final boss. You'll be asked to review and combine your computational, geometric, and theoretical linear algebra skills, applying them to vectors and matrices with complex coefficients. Are you ready for this?

I'm not going to lie to you and say the problems are easy, but this is the final push, so hang in there and you'll be done with all the linear algebra theory in just a few hours. After finishing the problems in this chapter, the rest of the book winds down with three chapters of cool applications, which are much lighter reading. So grab a pen, pull out some paper and kick some problem ass!

P6.1 Show that the vectors $\vec{e}_1 = (1, \frac{1}{\varphi})^{\mathsf{T}}$ and $\vec{e}_2 = (1, -\varphi)^{\mathsf{T}}$ are eigenvectors of the matrix $A = \left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$. What are the eigenvalues associated with these eigenvectors?

Hint: Compute $A\vec{e}_1$ and $A\vec{e}_2$ to see what happens. Use the fact that φ satisfies the equation $\varphi^2 - \varphi - 1 = 0$ to simplify expressions.

P6.2 An unknown matrix $A \in \mathbb{R}^{3\times 3}$ has eigenvalues $\lambda_1 = 2$, $\lambda_2 = -3$, and $\lambda_3 = 5$. Calculate the value of the following expressions:

a) $\det(2A)$ **b)** $\det(A^2)$ **c)** $\det(A^{-1})$ **d)** $\mathrm{Tr}(A + 15A^{-1} + A^{\mathsf{T}})$

P6.3 Given A and B are two positive semidefinite matrices, show that the sum $A + B$ is also a positive semidefinite matrix.

P6.4 Let V be the set of two-dimensional vectors of real numbers, with addition defined as $(a_1, a_2) + (b_1, b_2) = (a_1 + b_1, a_2 b_2)$ and scalar multiplication defined as $c \cdot (a_1, a_2) = (ca_1, a_2)$. Is $(V, \mathbb{R}, +, \cdot)$ a vector space? Justify your answer.

Hint: Check whether scaling by zero obeys the vector space axioms.

P6.5 Determine whether the following subsets of $\mathbb{R}^3$ are subspaces:

a) $\{(x, y, z) \in \mathbb{R} \mid x + y + z = 3\}$
b) $\{(x, y, z) \in \mathbb{R} \mid x + y + z = 0\}$
c) $\{(x, y, z) \in \mathbb{R} \mid x = 2y = 3z\}$

P6.6 Consider the linear transformation $T : P_2(x) \to P_2(x)$ defined as $T(a_0 + a_1 x + a_2 x) = a_2 x^2$. Find the matrix representation of T with respect to the basis $\{1, x, x^2\}$, and compute the eigenvalues of T.

P6.7 Find the matrix representation of the derivative operator $Dp(x) \equiv \frac{d}{dx}p(x)$ for the vector space of polynomials of degree three $p(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3$, represented as coefficient $(a_0, a_1, a_2, a_3)^{\mathsf{T}}$.

P6.8 Show that the following functions, called Laguerre polynomials, are orthogonal with respect to the inner product $\langle f(x), g(x)\rangle = \int_0^\infty f(x)g(x)e^{-x}\,dx$.

$$L_0(x) = 1, \qquad L_1(x) = 1 - x, \qquad L_2(x) = \tfrac{1}{2}(x^2 - 4x + 2).$$

Hint: Use the formula $\int_0^\infty f(x)g'(x)\,dx = f(x)g(x)\big|_0^\infty - \int_0^\infty f'(x)g(x)\,dx$.

P6.9 Perform Gram–Schmidt orthogonalization on vectors $\vec{v}_1 = (1, 1)$ and $\vec{v}_2 = (0, 1)$ to obtain an orthonormal basis.

P6.10 Find the eigendecomposition of the matrix $A = \begin{bmatrix} 2 & 0 & -5 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix}$.

P6.11 Given complex matrices $A = \left[\begin{smallmatrix} 2+i & -1+2i \\ 3+2i & -2i \end{smallmatrix}\right]$, $B = \left[\begin{smallmatrix} 2-i & 3-2i \\ 5+i & -5+5i \end{smallmatrix}\right]$, and $C = \left[\begin{smallmatrix} 1+2i & i \\ 3-i & 8 \\ 4+2i & 1-i \end{smallmatrix}\right]$, find $A + B$, CB, and $(2 + i)B$.

Chapter 7

Applications

In this chapter, we'll learn about applications of linear algebra. We'll cover a wide range of topics from different areas of science, business, and technology to give you an idea of the spectrum of possible calculations based on vector and matrix algebra. Don't worry if you're not able to follow all the details in each section—we're taking a broad approach here, covering many different topics in the hope that some will interest you. Note that most of the material covered in this chapter is not likely to show up on your linear algebra final, so no pressure—this is just for fun.

Before we start, I want to say a few words about scientific ethics. Linear algebra is a powerful tool for solving problems and modelling the real world. But with great power comes great responsibility. I hope you'll make an effort to think about the ethical implications when you use linear algebra to solve problems. Certain applications of linear algebra, like building weapons, interfering with crops, and building mathematically-complicated financial scams are clearly evil, so you should avoid them. Other areas where linear algebra can be applied are not so clear-cut: perhaps you're building a satellite localization service to find missing people in emergency situations, but the same technology could be used by governments to spy on and persecute your fellow citizens. Do you want to be the person responsible for bringing about an Orwellian state? All I ask of you is to run a quick "System check" before you set to work on a project: ask yourself "Am I working for the System?" Don't just say "It's my job" and proceed without caution. If you find what you're doing for your employer is unethical, then maybe you should find a different job. There are a lot of jobs out there for people who know math, and if the bad guys can't hire qualified people like you, their power will decrease—and that's a good thing.

Our System check is complete. On to the applications!

7.1 Balancing chemical equations

Exercises

7.2 Input–output models in economics

Links

[History of the Leontief input-output model in economics]
`https://en.wikipedia.org/wiki/Input-output_model`

Exercises

7.3 Electric circuits

Background

Using linear algebra to solve circuits

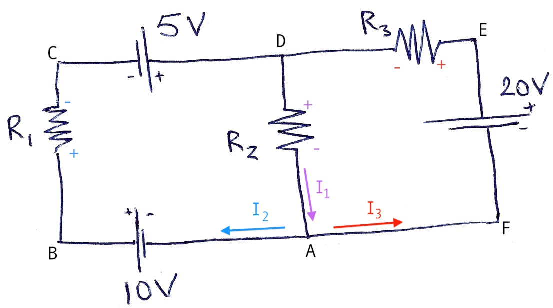


Figure 7.1: The circuit with branch currents labelled. Each resistor is assigned a *polarity* relative to the current flowing through it.

$$\begin{aligned} +10 - R_1 I_2 + 5 - R_2 I_1 &= 0, \\ +20 - R_3 I_3 - R_2 I_1 &= 0, \\ I_1 &= I_2 + I_3. \end{aligned}$$

Do you see where this is going?

$$\begin{aligned} R_2 I_1 \quad + R_1 I_2 \qquad\qquad &= 15, \\ R_2 I_1 \qquad\qquad + R_3 I_3 &= 20, \\ I_1 \quad - I_2 \quad - I_3 &= 0. \end{aligned}$$

$$\left[\begin{array}{ccc|c} R_2 & R_1 & 0 & 15 \\ R_2 & 0 & R_3 & 20 \\ 1 & -1 & -1 & 0 \end{array}\right].$$

Other network flows

Exercises

7.4 Graphs

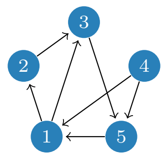


Figure 7.2: A simple graph with five vertices and seven edges.

The graph in Figure 7.2 is represented mathematically as $G = (V, E)$, where $V = \{1, 2, 3, 4, 5\}$ is the set of vertices, and $E = \{(1, 2), (1, 3), (2, 3), (3, 5), (4, 1), (4, 5), (5, 1)\}$ is the set of edges. Note the edge from vertex i to vertex j is represented as the pair (i, j).

Adjacency matrix

The *adjacency matrix* representation of the graph in Figure 7.2 is a 5×5 matrix A that contains information about the edges in the graph. Specifically, $A_{ij} = 1$ if the edge (i, j) exists, otherwise $A_{ij} = 0$ if the edge doesn't exist:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Applications

Discussion

Links

Exercises

7.5 Fibonacci sequence

Links

Exercises

7.6 Linear programming

A linear program with n variables and m constraints is expressed as a maximization problem,

$$\max_{x_1,x_2,\ldots,x_n} g(x_1,x_2,\ldots,x_n) \;=\; c_1x_1+c_2x_2+\cdots+c_nx_n,$$

subject to m linear constraints,

$$\begin{aligned}
a_{11}x_1+a_{12}x_2+\cdots+a_{1n}x_n &\leq\; b_1,\\
a_{21}x_1+a_{22}x_2+\cdots+a_{2n}x_n &\leq\; b_2,\\
&\;\;\vdots\\
a_{m1}x_1+a_{m2}x_2+\cdots+a_{mn}x_n &\leq\; b_m.
\end{aligned}$$

Since the details of the simplex algorithm might not be of interest to all readers of the book, I split the topic of linear programming into a separate tutorial, which you can read online at the link below.

[Linear programming tutorial]
`https://minireference.github.io/linear_programming/tutorial.pdf`

7.7 Least squares approximate solutions

Statement of the problem

Linear model

Linear algebra formulation

When faced with an unfamiliar problem (such as finding the approximate solution to a system of equations $A\vec{x}=\vec{b}$ using a quadratic error model) don't become alarmed. Stand your ground and try relating

the problem to something you're more familiar with. Thinking about problems in terms of linear algebra can often unlock your geometric intuition and show you a path toward the solution.

Finding the least squares approximate solution

Pseudoinverse

Geometric interpretation

Affine models

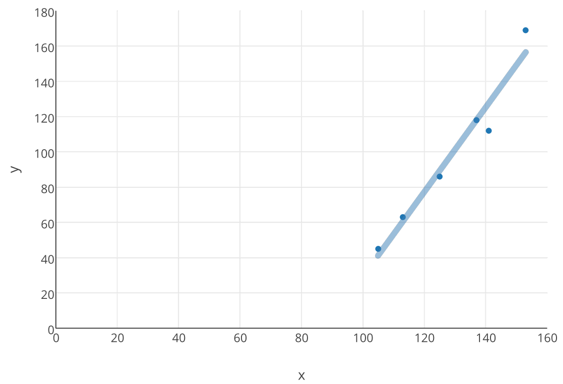


Figure 7.3: The affine model ($y_{\vec{m}'}(x) = m_0 + m_1 x$) that best fits the data points is the line $y(x) = -210.4 + 2.397x$. Allowing for one extra parameter, m_0 (the y-intercept), results in a model that fits the data much better, as compared to the fit in Figure **??**.

Example 2

Quadratic models

Links

[More about the Moore–Penrose pseudoinverse]
`https://en.wikipedia.org/wiki/Moore-Penrose_pseudoinverse`

Exercises

7.8 Computer graphics

Affine transformations

Homogeneous coordinates

Affine transformations in homogeneous coordinates

Graphics transformations in 2D

Linear transformations

Orthogonal projections

Translation

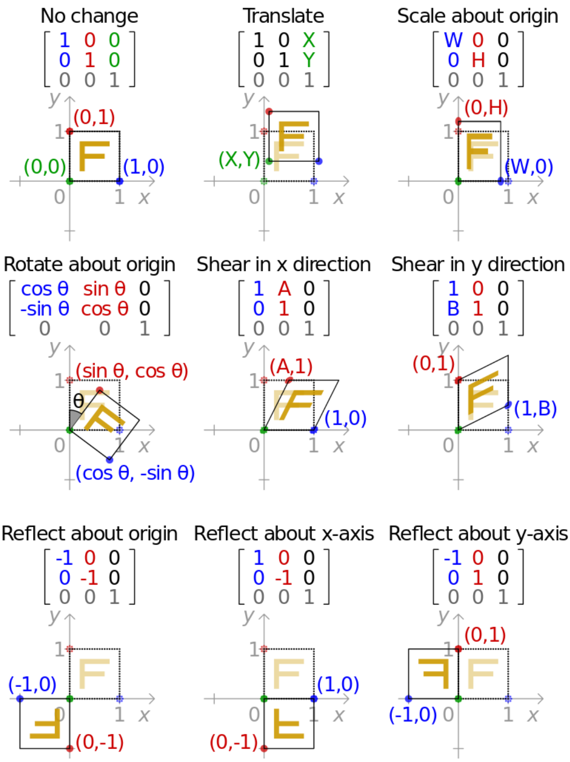


Figure 7.4: Illustration of the different transformations on a sample shape. Source: `wikipedia File:2D_affine_transformation_matrix.svg`

Perspective projections

General perspective transformation

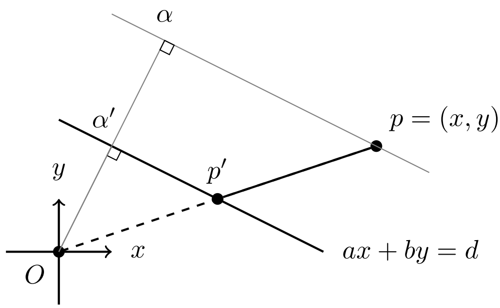


Figure 7.5: The point $p' = (x', y')$ is the projection of the point $p = (x, y)$ onto the line with equation $ax + by = d$. We define points α' and α in the direction of line's normal vector $\vec{n} = (a, b)$. The distances from the origin to these points are ℓ' and ℓ respectively. We have $\ell'/\ell = x'/x = y'/y$.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} X' \\ Y' \\ W' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{a}{d} & \frac{b}{d} & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

Graphics transformations in 3D

3D graphics programming

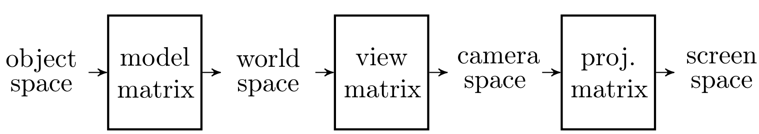


Figure 7.6: A graphics processing pipeline for drawing 3D objects on the screen. A 3D model is composed of polygons expressed with respect to a coordinate system centred on the object. The model matrix positions the object in the scene, the view matrix positions the camera in the scene, and finally the projection matrix computes what should appear on the screen.

We can understand the graphics processing pipeline as a sequence of matrix transformations: the model matrix M, the view matrix V,

and the projection matrix Π_s.

$$\begin{bmatrix} x' \\ y' \end{bmatrix}_s = \Pi_s V M \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}_m \qquad \Rightarrow \qquad (x,y,z,1)_m M^{\mathsf{T}} V^{\mathsf{T}} \Pi_s^{\mathsf{T}} = (x',y')_s.$$

Discussion

Links

7.9 Cryptography

Context

The secure communication scenarios we'll discuss in this section involve three parties:

- Alice is the message sender
- Bob is the message receiver
- Eve is the eavesdropper

Alice wants to send a private message to Bob, but Eve has the ability to see all communication between Alice and Bob.

Definitions

A cryptographic protocol consists of an encryption function, a decryption function, and a procedure for generating the secret key $\vec{k}$. For simplicity we assume messages and keys are all binary strings:

- $\vec{m} \in \{0,1\}^n$: the *message* or *plaintext* is a bitstring of length n
- $\vec{k} \in \{0,1\}^n$: the *key* is a shared secret between Alice and Bob
- $\vec{c} \in \{0,1\}^n$: the *ciphertext* is the encrypted message
- $\text{Enc}(\vec{m},\vec{k})$: the encryption function that takes as input a message $\vec{m} \in \{0,1\}^n$ and the key $\vec{k} \in \{0,1\}^n$ and produces a ciphertext $\vec{c} \in \{0,1\}^n$ as output
- $\text{Dec}(\vec{c},\vec{k})$: the decryption function that takes as input a ciphertext $\vec{c} \in \{0,1\}^n$ and the key $\vec{k} \in \{0,1\}^n$ and produces the decrypted message $\vec{m} \in \{0,1\}^n$ as output

We consider the protocol to be secure if Eve cannot gain any information about the messages $\vec{m}_1, \vec{m}_2, \ldots$ from the ciphertexts $\vec{c}_1, \vec{c}_2, \ldots$ she intercepts.

Binary

The XOR operation

One-time pad cryptosystem

One-time pad encryption

One-time pad decryption

Discussion

One-time pad security

Definition: Indistinguishability under chosen-plaintext attack (IND-CPA) A cryptosystem is considered secure in terms of indistinguishability if no eavesdropper can distinguish the ciphertexts of two messages $\vec{m}_a$ and $\vec{m}_b$ chosen by the eavesdropper, with a probability greater than guessing randomly.

Sketch of security proof

Public key cryptography

Definitions

Encryption

Digital signatures

Example: ssh keys for remote logins

Discussion

Links

Exercises

7.10 Error-correcting codes

Definitions

An *error-correcting code* is a prescription for encoding *binary* information. Recall that bits are elements of the binary field, $\mathbb{F}_2 = \{0, 1\}$. A bitstring of length n is an n-dimensional vector of bits $\vec{v} \in \{0, 1\}^n$. For example, 0010 is a bitstring of length 4.

We use several parameters to characterize error-correcting codes:

- k: the size, or length, of the messages for the code.

- $\vec{x}_i \in \{0,1\}^k$: a *message*. Any bitstring of length k is a valid message.
- n: the size of the codewords in the code.
- $\vec{c}_i \in \{0,1\}^n$: the *codeword* that corresponds to message $\vec{x}_i$.
- A *code* consists of 2^k codewords $\{\vec{c}_1, \vec{c}_2, \ldots\}$, one for each of the possible messages $\{\vec{x}_1, \vec{x}_2, \ldots\}$.
- $d(\vec{c}_i, \vec{c}_j)$: the *Hamming distance* between codewords $\vec{c}_i$ and $\vec{c}_j$.
- An (n,k,d) code is a procedure for encoding messages into codewords; $\text{Enc} : \{0,1\}^k \to \{0,1\}^n$, which guarantees the *minimum distance* between any two codewords is at least d.

The *Hamming distance* between two bitstrings $\vec{x}, \vec{y} \in \{0,1\}^n$ counts the number of bits where the two bitstrings differ:

$$d(\vec{x}, \vec{y}) \equiv \sum_{i=1}^{n} \delta(x_i, y_i), \quad \text{where } \delta(x_i, y_i) = \begin{cases} 0 & \text{if } x_i = y_i, \\ 1 & \text{if } x_i \neq y_i. \end{cases}$$

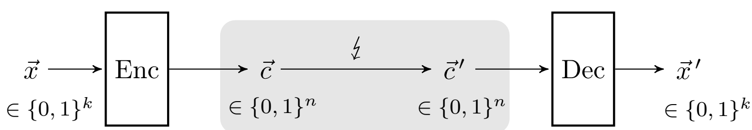


Figure 7.7: An error-correcting scheme using the encoding function Enc and the decoding function Dec to protect against the effect of noise (denoted $\lightning$). Each message $\vec{x}$ is encoded into a codeword $\vec{c}$. The codeword $\vec{c}$ is transmitted through a *noisy channel* that can corrupt the codeword by transforming it into another bitstring $\vec{c}\,'$. The decoding function Dec looks for a valid codeword $\vec{c}$ that is close in Hamming distance to $\vec{c}\,'$. If the protocol is successful, the decoded message will match the transmitted message $\vec{x}\,' = \vec{x}$, despite the noise ($\lightning$).

Linear codes

Coding theory

Observation 1

Repetition code

The Hamming code

Encoding

Decoding with error correction

Discussion

Error-detecting codes

Links

Exercises

7.11 Fourier analysis

Way back in the 17^{th} century, Isaac Newton carried out a famous experiment using light beams and glass prisms. He showed that when a beam of white light passes through a prism, it splits into a rainbow of colours: the rainbow is red at one end, followed by orange, yellow, green, blue, and finally violet at the other end. This experiment showed that white light is made of *components* with different colours. Using the language of linear algebra, we can say that white light is a "linear combination" of different colours.

Today we know that different colours of light correspond to electromagnetic waves with different frequencies: red light has a frequency around 450 THz, while violet light has a frequency around 730 THz. We can therefore say that white light is made of components with different frequencies. The notion of describing complex phenomena in terms of components with different frequencies is the main idea behind *Fourier analysis*.

Fourier analysis is used to describe sounds, vibrations, electric signals, radio signals, light signals, and many other phenomena. The Fourier transform allows us to represent all these "signals" in terms of components with different frequencies. Indeed, the Fourier transform can be understood as a change-of-basis operation that converts a signal from a time basis to a frequency basis:

$$[\mathbf{v}]_t \qquad \Leftrightarrow \qquad [\mathbf{v}]_f .$$

For example, if $\mathbf{v}$ represents a musical vibration, then $[\mathbf{v}]_t$ corresponds to the vibration as a function of time, while $[\mathbf{v}]_f$ corresponds to the frequency content of the vibration. Depending on the properties of the signal in the time domain and the choice of basis for the frequency domain, different Fourier transformations are possible.

Table 7.1 shows a summary of these three Fourier-type transformations. The table indicates the class of functions for which the transform applies, the Fourier basis for the transform, and the frequency-domain representation used.

Fourier transformations

Name	Time domain	Fourier basis	Frequency domain
FS	$f(t) \in \{\mathbb{R} \to \mathbb{R}\}$ s.t. $f(t) = f(t+T)$	$1, \{\cos\left(\frac{2\pi n}{T}\right)\}_{n\in\mathbb{N}_+}$, $\{\sin\left(\frac{2\pi n}{T}\right)\}_{n\in\mathbb{N}_+}$	$(a_0, a_1, b_1, \ldots)$
FT	$f(t) \in \{\mathbb{R} \to \mathbb{R}\}$ s.t. $\int_{-\infty}^{\infty} \lvert f(t)\rvert^2 \, dt < \infty$	$\{e^{i\omega t}\}_{\omega\in\mathbb{R}}$	$f(\omega) \in \{\mathbb{R} \to \mathbb{C}\}$
DFT	$f[t] \in \{[N] \to \mathbb{R}\}$	$\{e^{\frac{iwt}{N}}\}_{w\in[N]}$	$f[w] \in \{[N] \to \mathbb{C}\}$

Table 7.1: This table shows the time domains, Fourier bases, and frequency domains for the three Fourier transformations that we'll discuss. The *Fourier series* (**FS**) converts periodic continuous time signals into Fourier coefficients. The *Fourier transform* (**FT**) converts finite-power continuous signal into continuous functions of frequency. The *discrete Fourier transform* (**DFT**) is the discretized version of the Fourier transform.

Example 1: Describing the vibrations of a string

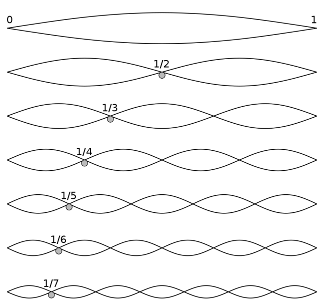


Figure 7.8: Standing waves on a string with length $L = 1$. The longest vibration is called the *fundamental*. Other vibrations are called *overtones*.

Depending how you pluck the string, the shape of the vibrating

string $f(x)$ will be some linear combination of the vibrations $\mathbf{e}_n(x)$:

$$\begin{aligned} f(x) &= a_1 \sin(\tfrac{\pi}{L}x) + a_2 \sin(\tfrac{2\pi}{L}x) + a_3 \sin(\tfrac{3\pi}{L}x) + \cdots \\ &= a_1\, \mathbf{e}_1(x) \quad + a_2\, \mathbf{e}_2(x) \quad + a_3\, \mathbf{e}_3(x) \quad + \cdots \end{aligned}$$

The main idea

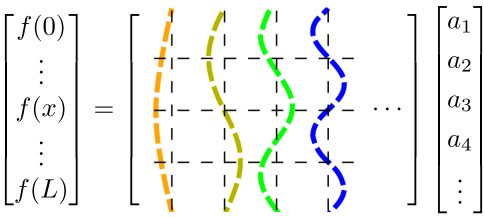


Figure 7.9: Any string vibration $f(x)$ can be represented as coefficients $(a_1, a_2, a_3, a_4, \ldots)$ with respect to the basis of functions $\mathbf{e}_n(x) \equiv \sin(\frac{n\pi}{L}x)$.

Change-of-basis review

Analysis and synthesis

Fourier series

Fourier transform

Discrete Fourier transform

Digital signal processing

Discussion

Links

[Excellent video tutorial about digital audio processing]
`http://xiph.org/video/vid2.shtml`

Exercises

Discussion

More linear algebra applications

7.12 Applications problems

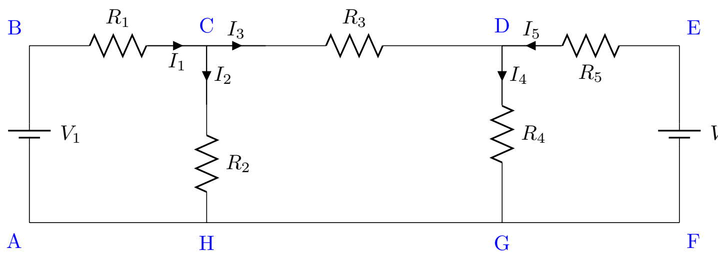
It would be easy to think of all the applications of linear algebra presented in this chapter as a TV program, designed to entertain

rather than teach. Certainly you can continue to the next chapter without solving any problems, but do you really want to do that to yourself?

Presented next are a number of practice problems that will test your understanding of the new concepts and give you a great opportunity to practice your linear algebra skills. The linear algebra techniques we learned in previous chapters are key building blocks for applications. So don't sit on your laurels thinking, "Yay, I'm in Chapter 7 and I know linear algebra now, I'm so good." Prove it.

P7.1 Consider the following chemical equation that describes how your body burns fat molecules: $C_{55}H_{104}O_6 + O_2 \ \rightarrow \ CO_2 + H_2O$. Balance this chemical equation.

P7.2 Check out this circuit containing two batteries and five resistors:



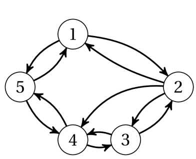
a) Label the polarity of each resistor in the circuit.
b) Write three KVL equations and two KCL equations.
c) Rewrite the equations in the form $R\vec{I} = \vec{V}$, where R is a 5×5 matrix, $\vec{I} = (I_1, I_2, I_3, I_4, I_5)^{\mathsf{T}}$, and $\vec{V}$ is a vector of constants.
d) Find the value of the currents I_1 and I_5 given $V_1 = 15[\text{V}]$, $V_3 = 10[\text{V}]$, $R_1 = 1[\Omega]$, $R_2 = 1[\Omega]$, $R_3 = 4[\Omega]$, $R_4 = 2[\Omega]$, $R_5 = 2[\Omega]$.

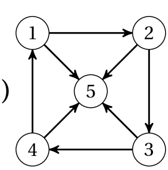
Hint: The direction of the voltage drop across a resistor depends on the direction of the current flowing through it.

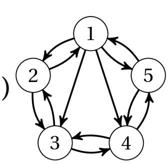
P7.3 Given the (x, y) pairs $(0, 0.9)$, $(1, 1.6)$, $(2, 2.1)$, and $(3, 2.4)$, find the best-fitting affine model $y = b + mx$ for this data.

Hint: Find the Moore–Penrose pseudoinverse.

P7.4 Find the adjacency matrix representation of the following graphs:

a) 

b) 

c) 

P7.5 Draw the graphs that correspond to these adjacency matrices:

a) $A = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$ **b)** $A = \begin{bmatrix} 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$

P7.6 Alice and Bob share the key $\vec{k} = 10010111\ 01010011\ 10011110$. Bob receives the ciphertext $\vec{c} = 11100100\ 00100110\ 11101110$ sent by Alice. What is the message sent by Alice? Find an ASCII lookup table on the web and use it to convert the binary message into characters.

P7.7 Find the Fourier transform of the function $f(t) = e^{-|at|}$.

Hint: Split $e^{-i\omega t}$ into sine and cosine terms and use symmetry to convert the two-sided integral $\int_{-\infty}^{\infty} \cdots dt$ into a one-sided integral $\int_{0}^{\infty} \cdots dt$.

Chapter 8

Probability theory

In this chapter, we'll use linear algebra concepts to explore the world of probability theory. Think of this as bonus material because the topics we'll discuss are not normally part of a linear algebra course. Given the general usefulness of probabilistic reasoning and the fact that you have already covered all the prerequisites, it would be a shame *not* to learn a bit about probability theory and its applications.

The chapter is structured as follows. In Section 8.1, we'll discuss probability distributions, which are mathematical models for describing random events. Section 8.2 introduces the concept of a *Markov chain*, which can be used to characterize the random transitions between different states of a system. Of the myriad of topics in probability theory, we've chosen to discuss probability distributions and Markov chains because they correspond one-to-one with vectors and matrices. This means you should feel right at home. In Section 8.3, we'll describe Google's PageRank algorithm for ranking webpages, which is an interesting application of Markov chains.

8.1 Probability distributions

Many phenomena in the world are inherently unpredictable. When you throw a six-sided die, one of the outcomes $\{1, 2, 3, 4, 5, 6\}$ will result, but you don't know which one. Similarly, when you toss a coin, you know the outcome will be either `heads` or `tails` but you can't predict which outcome will result. Probabilities are used to describe events where uncertainty plays a role. We can assign probabilities to the different outcomes of a dice roll, the outcomes of a coin toss, and also to many real-world systems. For example, we can build a probabilistic model of hard drive failures using past observations. We can then calculate the probability that your family photo albums will

survive the next 10 or 20 years. Backups my friends, backups.

Probabilistic models can help us better understand random events. The fundamental concept in probability theory is that of a *probability distribution*, which describes the likelihood of different outcomes of a random event. For example, the probability distribution for the roll of a fair die is $p_X = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})^{\mathsf{T}}$, and the probability distribution for a coin toss is $p_Y = (\frac{1}{2}, \frac{1}{2})^{\mathsf{T}}$. Each entry of a probability distribution corresponds to the *probability mass* of a given outcome. This terminology borrows from the concept of mass distribution used in physics. The entries of a probability distribution satisfy the following conditions: each entry is a nonnegative number, and the sum of the entries is one. These two conditions are known as the *Kolmogorov axioms* of probability.

Strictly speaking, understanding linear algebra is not required for understanding probability theory. However, vector notation is very effective for describing probability distributions. Your existing knowledge of vectors and the rules for matrix multiplication will allow you to quickly understand many concepts in probability theory. Probabilistic reasoning is highly useful, so it's totally worth taking the time to learn about it.

Random variables

Probability distributions

The probability distribution of a discrete random variable $X \in \mathcal{X}$ is a vector of $|\mathcal{X}|$ nonnegative numbers that sum to one. Using mathematically precise notation, we write the definition of p_X as follows:

$$p_X \in \mathbb{R}^{|\mathcal{X}|} \quad \text{such that} \quad p_X(x) \geq 0, \forall x \in \mathcal{X} \;\; \text{and} \;\; \sum_{x \in \mathcal{X}} p_X(x) = 1.$$

A probability distribution is a vector in $\mathbb{R}^{|\mathcal{X}|}$ that satisfies two special requirements: its entries must be nonnegative and the sum of the entries must be one.

Events

Expectations

Expected value and variance of random variables

The *expected value* of the random variable X is computed using the formula

$$\mu_X \equiv \mathbb{E}_X[X] \equiv \sum_x x\, p_X(x).$$

The expected value is a single number that tells us what value of X we can expect to obtain on average from the random variable X. The expected value is also called the *average* or the *mean* of the random variable X.

The *variance* of the random variable X is defined as follows:

$$\sigma_X^2 \equiv \mathbb{E}_X\big[(X - \mu_X)^2\big] = \sum_x (x - \mu_X)^2 \, p_X(x).$$

The variance formula computes the expectation of the squared distance of the random variable X from its expected value. The variance σ_X^2, also denoted $\text{var}(X)$, gives us an indication of how clustered or spread the values of X are. A small variance indicates the outcomes of X are tightly clustered near the expected value μ_X, while a large variance indicates the outcomes of X are widely spread.

Conditional probability distributions

Interpretations of probability theory

Discussion

Links

Exercises

8.2 Markov chains

Example

Stationary distribution

Discussion

Links

[Awesome visual representation of states and transitions]
`http://setosa.io/blog/2014/07/26/markov-chains/index.html`

Exercises

8.3 Google's PageRank algorithm

The random surfer model

The PageRank Markov chain

Example: micro-web

We'll now study the micro-web illustrated in Figure 8.1. This is a *vastly* simplified version of the link structure between webpages on the web. Rather than include billions of webpages, the micro-web contains only eight webpages $\{1, 2, 3, 4, 5, 6, 7, 8\}$. Instead of trillions of links between webpages, the micro-web contains only fourteen links $\{(1, 2)$, $(1, 5)$, $(2, 3)$, $(2, 5)$, $(3, 1)$, $(3, 5)$, $(4, 5)$, $(5, 6)$, $(5, 7)$, $(6, 3)$, $(6, 7)$, $(7, 5)$, $(7, 6)$, $(7, 8)\}$. Simple as it may be, this example is sufficient to illustrate the main idea of the PageRank algorithm. Scaling the solution from the case $n = 8$ to the case $n = 1\,000\,000\,000$ is left as an exercise for the reader.

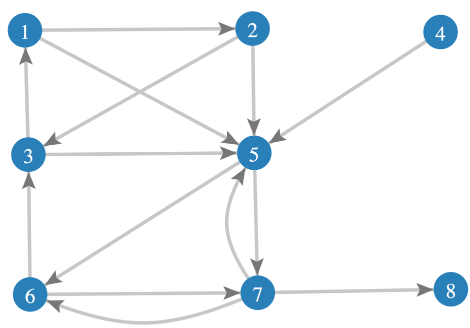


Figure 8.1: A graph showing the links between the pages on the micro-web. Page 5 seems to be an important page because many pages link to it. Since Page 5 links to pages 6 and 7, these pages will probably get a lot of eyeballs, too. Page 4 is the least important, since no links lead to it. Page 8 is an example of the unlikely case of a webpage with no outbound links.

Page ID	PageRank
Page 5	0.22678
Page 7	0.20793
Page 6	0.18642
Page 3	0.13229
Page 8	0.08437
Page 1	0.08152
Page 2	0.05868
Page 4	0.02199

Table 8.2: The PageRanks of the pages from the graph in Figure 8.1.

According to their PageRank score, the top two pages in the micro-web are Page 5 with PageRank 0.22678 and Page 7 with PageRank 0.20793. Page 6 is not far behind with PageRank 0.18642. Looking at Figure 8.1, we can confirm this ranking makes sense, since Page 5 has the most links pointing to it, and since Page 5 links to Page 6 and Page 7. As expected, Page 4 ranks as the least important page on the micro-web since no pages link to it.

Discussion

Links

[The original PageRank paper]
`http://ilpubs.stanford.edu/422/1/1999-66.pdf`

Exercises

8.4 Probability problems

To better understand random variables and probability distributions, you need to practice using these concepts to solve real-world problems. It just so happens there are some practice problems on this very topic in this section—how convenient is that? Don't skip them!

Solving practice problems will help you understand probability theory and Markov chains. If you haven't played with `SymPy` yet, now is a great chance to get to know this powerful computer algebra system because Markov chain calculations are difficult to do by hand.

P8.1 Given a random variable X with three possible outcomes $\{1, 2, 3\}$ and probability distribution $p_X = (p_1, p_2, p_3)$, prove that $p_1 \leq 1$.

Hint: Use the Kolmogorov's axioms and build a proof by contradiction.

P8.2 The probability of `heads` for a fair coin is $p = \frac{1}{2}$. The probability of getting `heads` n times in a row is given by the expression p^n. What is the probability of getting `heads` four times in a row?

P8.3 You have a biased coin that lands on `heads` with probability p, and consequently lands on `tails` with probability $(1 - p)$. Suppose you want to flip the coin until you get `heads`. Define the random variable N as the number of tosses required until the first `heads` outcome. What is the probability mass function $P_N(n)$ for success on the n^{th} toss? Confirm that the formula is a valid probability distribution by showing $\sum_{n=1}^{\infty} P_N(n) = 1$.

Hint: Find the probabilities for cases $n = 1, 2, 3, \ldots$ and look for a pattern.

P8.4 A mathematician walks over to a roulette table in a casino. The roulette wheel has 101 numbers: 50 are black, 50 are red, and the number zero is green. If the mathematician bets \$1 on black and the roulette ball stops on a black number, the payout is \$2, otherwise the bet is lost. Calculate the expected payout for playing this game, and determine whether it's worth playing.

P8.5 Consider the following variation of the six-sided die game. You pay \$1 to play one round of the game and the payout for the game is as follows. If you roll a ⚀, a ⚁, or a ⚂, you win nothing. If you roll a ⚃ or a ⚄, you win \$1. If you roll a ⚅, you win \$5. Should you play this game?

P8.6 Consider the weather in a city which has "good" and "bad" years. Suppose the weather conditions over the years form a Markov chain where a good year is equally likely to be followed by a good or a bad year, while a bad year is three times as likely to be followed by a bad year as by a good year. Given that last year, call it Year 0, was a good weather year, find the probability distribution that describes the weather in Year 1, Year 2, and Year ∞.

P8.7 Consider the network of webpages shown in Figure 8.2. Find the Markov chain transition matrices M_1 and M_2 for Randy's two browsing strategies; then combine the strategies using $\alpha = 0.1$ to obtain the PageRank matrix M. Compute the PageRank vector. Which pages are the most important?

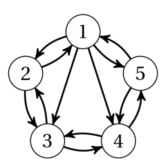


Figure 8.2: Graph showing five webpages and the links between them.

Hint: Use `SymPy` for the calculation. Consult the micro-web calculation example (`bit.ly/microWebPR`) for the `SymPy` commands you'll need.

Chapter 9

Quantum mechanics

By the end of the 19^{th} century, physicists thought they had figured out most of what there is to know about the laws of nature. Newton's laws of mechanics described the motion of objects in space, and Maxwell's equations described electricity and magnetism. Wave phenomena—including the propagation of sound, light, and waves on the surface of liquids—were also well understood. Only a few small inconsistencies between theory and experiments with atoms and radiation remained unsolved.

> "[...] it seems probable that most of the grand underlying principles have now been firmly established and that further advances are to be sought chiefly in the rigorous application of these principles to all the phenomena which come under our notice."
>
> —Albert A. Michelson in 1894

Physicists like Michelson were worried about the future of physics research. It was as if they were wondering, "What are we going to do now that we've figured everything out?" Little did they know about the quantum storm that was about to hit physics, and with it, the complete rewrite of our understanding of nature at the smallest scale.

Understanding the structure of atoms—the smallest constituents of matter known at the time—was no trivial task. Describing the absorption of electromagnetic radiation by metals also turned out to be quite complicated. In both cases, the physical theories of the time predicted that the energy of physical systems could take on any value; yet experimental observations showed *discrete* energy levels. Imagine you throw a (very-very tiny) ball, and the laws of physics force you to choose an initial velocity for the ball from a list of "allowed" values: $0\,\text{m/s}$, $1\,\text{m/s}$, $2\,\text{m/s}$, $3\,\text{m/s}$, and so forth. That would be weird, no? Weird indeed, and this is the situation physicists were facing in the

beginning of the 20^{th} century: their theories described the energy levels of atoms as real numbers $E \in \mathbb{R}$, but experiments showed that only a discrete set of energy levels exist. For example, the energy levels that the electrons of the hydrogen atom can take on are:

$$\begin{aligned} E \in \{\, & 21.8\times 10^{-19}\,\text{J},\ 5.44\times 10^{-19}\,\text{J},\ 2.42\times 10^{-19}\,\text{J}, \\ & 13.6\times 10^{-20}\,\text{J},\ 8.71\times 10^{-20}\,\text{J},\ 6.05\times 10^{-20}\,\text{J},\ \ldots\, \}. \end{aligned}$$

Other experimental observations suggested that electromagnetic radiation is not a continuous wave, but comes in discrete "wave packets," which we call *photons* today. The theory of *quantum mechanics* was born out of a need to explain these observations. The term *quantum*, from the Latin *quantus* for quantity, was coined to describe the discrete nature of the phenomena that physicists were trying to explain.

During the first half of the 20^{th} century, in experiment after experiment, quantum principles were used to correctly predict many previously-unexplained observations. During the second half of the 20^{th} century, biologists, chemists, engineers, and physicists applied quantum principles to all areas of science. This process of "upgrading" classical models to quantum models led to a better understanding of the laws of nature, and the discovery of useful things like transistors and lasers.

The fundamental principles of quantum mechanics can be explained in the space on the back of an envelope. Understanding quantum mechanics is a matter of combining a little knowledge of linear algebra (vectors, inner products, projections) with some probability theory (Chapter 8). In this chapter, we'll take a little excursion to the land of physics to learn about the ideas of great scientists like Bohr, Planck, Dirac, Heisenberg, and Pauli. Your linear algebra skills will allow you to learn about some fascinating 20^{th}-century discoveries. This chapter is totally optional reading, reserved for readers who *insist* on learning about the quantum world. If you're not interested in quantum mechanics, it's okay to skip this chapter, but I recommend you check out Section 9.3 on *Dirac notation* for vectors and matrices. Learning Dirac notation serves as an excellent review of the core concepts of linear algebra.

9.1 Introduction

The principles of quantum mechanics have far-reaching implications for many areas of science: physics, chemistry, biology, engineering, philosophy, and many other fields of study. Each field of study has its own view on quantum mechanics, and has developed a specialized language for describing quantum concepts. We'll formally introduce

the postulates of quantum mechanics in Section 9.5, but before we get there, let's look at some of the disciplines where quantum principles are used.

Physics Physicists use the laws of quantum mechanics as a toolbox to understand and predict the outcomes of atomic-scale physics experiments. By "upgrading" classical physics models to reflect the ideas of quantum mechanics, physicists (and chemists) obtain more accurate models that lead to better predictions.

For example, in a *classical* physics model, the motion of a particle is described by its position $x(t)$ and velocity $v(t)$ as functions of time:

$$\text{classical state} = (x(t), v(t)), \ \text{ for all times } t.$$

At any given time t, the particle is at position $x(t)$ and moving with velocity $v(t)$. Using Newton's laws of motion and calculus, we can predict the position and the velocity of a particle at all times.

In a quantum description of the motion of a particle in one dimension, the state of a particle is represented by a *wave function* $|\psi(x,t)\rangle$, which is a complex-valued function of position x and time t:

$$\text{quantum state} = |\psi(x,t)\rangle, \ \text{ for all times } t.$$

At any given time t, the state of the particle corresponds to a complex-valued function of a real variable $|\psi(x)\rangle \in \{\mathbb{R} \to \mathbb{C}\}$. The wave function $|\psi(x)\rangle$ is also called the *probability-amplitude* function. The probability of finding the particle at position x_a is proportional to the value of the squared norm of the wave function:

$$\Pr(\{\text{particle position} = x_a\}) \quad \propto \quad \big||\psi(x_a)\rangle\big|^2.$$

Instead of having a definite position $x(t)$ as in the classical model, the position of the particle in a quantum model is described by a probability distribution calculated from its wave function $|\psi(x)\rangle$. Instead of having a definite momentum $p(t)$, the momentum of a quantum particle is another function calculated based on its wave function $|\psi(x)\rangle$.

Classical models provide accurate predictions for physics problems involving macroscopic objects, but fail to predict the physics of atomic-scale phenomena. Much of 20^{th}-century physics research efforts were dedicated to the study of quantum concepts like ground states, measurements, spin angular momentum, polarization, uncertainty, entanglement, and non-locality.

Computer science Computer scientists understand quantum mechanics using principles of information. Quantum principles impose a

fundamental change to the "data types" used to represent information. Classical information is represented as *bits*, elements of the finite field of size two $\mathbb{Z}_2$:

$$\text{bit:}\ \ x = 0 \ \ \text{or} \ \ x = 1.$$

In the quantum world, the fundamental unit of information is the *qubit*, which is a two-dimensional, unit vector in a complex inner product space:

$$\text{qubit:}\ \ |x\rangle = \alpha|0\rangle + \beta|1\rangle.$$

This change to the underlying information model requires reconsidering fundamental information processing tasks like computation, data compression, encryption, and communication.

Philosophy Philosophers have also updated their conceptions of the world to incorporate the laws of quantum mechanics. Observations of physics experiments forced them to reconsider the fundamental question, "What are things made of?" Another interesting question philosophers have considered is whether the quantum state $|\psi\rangle$ of a physical system really exists, or if $|\psi\rangle$ is a representation of our knowledge about the system.

A third central philosophy concept that quantum mechanics calls into question is *determinism*—the clockwork-model of the universe, where each effect has a cause we can trace, like the connections between gears in a mechanical clock. The laws of physics tell us that the next state of the universe is determined by the current state of the universe, and the state changes according to the equations of physics. However, representing the universe as a quantum state has implications for our understanding of how the universe "ticks." Clockwork (deterministic) models of the universe are not wrong—they just require a quantum upgrade.

Many scientists are also interested in the philosophical aspects of quantum mechanics. Physicists call these types of questions *foundations* or *interpretations*. Since different philosophical interpretations of quantum phenomena cannot be tested experimentally, these questions are considered outside the scope of physics research. Nevertheless, these questions are so deep and fascinating that physicists continue to pursue them, and contribute interesting philosophical work.

[Philosophical issues in quantum theory]
`http://plato.stanford.edu/entries/qt-issues/`

Physical models of the world

Quantum model peculiarities

Chapter overview

In the next section, we'll describe a tabletop experiment involving lasers and polarization lenses, with an outcome that's difficult to explain using classical physics. The remainder of the chapter will introduce the tools needed to explain the outcome of this experiment in terms of quantum physics. We'll start by introducing a special notation for vectors that is used to describe quantum phenomena (Section 9.3).

In Section 9.5, we'll formally define the "rules" of quantum mechanics, also known as the *postulates* of quantum mechanics. We'll learn the "rules of the game" using the simplest possible quantum systems (qubits), and define how quantum systems are prepared, how we manipulate them using *quantum operations*, and how we extract information from them using *quantum measurements*. This part of the chapter is based on the notes from the introductory lectures of a graduate-level quantum information course, so don't think you'll be getting some watered-down, hand-wavy version of quantum mechanics. You'll learn the real stuff, because I know you can handle it.

In Section 9.6 we'll apply the quantum formalism to the polarizing lenses experiment, showing that a quantum model leads to the correct qualitative and quantitative prediction for the observed outcome. We'll close the chapter with short explanations of different applications of quantum mechanics with pointers for further exploration about each topic.

Throughout the chapter, we'll focus on *matrix* quantum mechanics and use computer science language to describe quantum phenomena. A computer science approach allows us to discuss the fundamental aspects of quantum theory without introducing all the physics required to understands atoms. Finally, I just might throw in a sample calculation using the wave function of the hydrogen atom, to give you an idea of what that's like.

9.2 Polarizing lenses experiment

Background

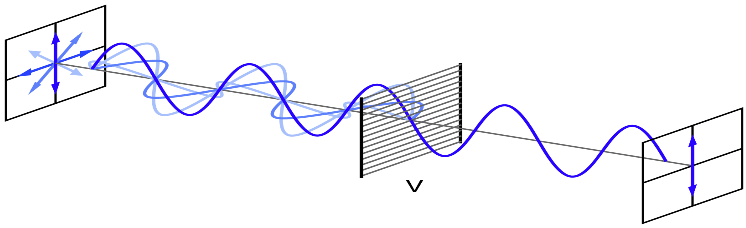


Figure 9.1: Incoming photons interact with the horizontal conductive bands of a polarizing filter. The horizontal bands of the filter reflect the horizontal component of the photons's electric field. Vertically-polarized photons pass through the filter because the conductive bands are perpendicular to their electric field. Thus, a vertically polarizing filter denoted V allows only vertically polarized light to pass through.

Consider the illustration in Figure 9.2. The effect of a vertically polarizing lens on a beam of light is to only allow vertically polarized light to pass through.

$$\text{unpolarized light} \rightarrow \boxed{V} \rightarrow \text{vertically polarized light}$$

Figure 9.2: A vertically polarizing lens (V) allows only vertically polarized light particles to pass through.

Classical physics paradigm

In the setup shown in Figure 9.2, each photon that passes through the lens must have `tag="V"`, because we know by definition that a V-polarizing lens only allows vertically polarized photons to pass through. Readers familiar with SQL syntax will recognize the action of the vertically polarizing lens as the following query:

```
SELECT photon FROM photons WHERE tag="V";
```

In other words, out of all the incoming photons, only the vertically polarized photons pass through the lens.

Polarizing lenses experiment

The initial setup for the experiment consists of an H-polarizing lens followed by a V-polarizing lens, as shown in Figure 9.3.

$$\text{light} \longrightarrow \boxed{H} \longrightarrow P = 1 \longrightarrow \boxed{V} \longrightarrow \; P = 0$$

Figure 9.3: The initial setup for the polarizing lenses experiment consists of an H-polarizing lens followed by a V-polarizing lens. Only photons with `tag="H"` can pass through the first lens, so no photons with `tag="V"` pass through the first lens. No photons can pass through both lenses since the V-polarizing lens accepts only photons with `tag="V"`.

$$\text{light} \rightarrow \boxed{H} \rightarrow P = 1 \rightarrow \boxed{D} \rightarrow \; P = 0.5 \; \rightarrow \boxed{V} \rightarrow \; P = 0.25$$

Figure 9.4: Adding an additional polarizing filter in the middle of the circuit causes light to appear at the end of the optical circuit.

Adding a third lens

Classical analysis

9.3 Dirac notation for vectors

The standard basis

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \qquad \ldots, \quad |d-1\rangle \equiv \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Vectors

In Dirac notation, a vector in $\mathbb{C}^2$ is denoted as a *ket*:

$$|v\rangle = \alpha|0\rangle + \beta|1\rangle \qquad \Leftrightarrow \qquad \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

where $\alpha \in \mathbb{C}$ and $\beta \in \mathbb{C}$ are the *coefficients* of $|v\rangle$ and $\{|0\rangle, |1\rangle\}$ is the standard basis for $\mathbb{C}^2$:

$$|0\rangle \equiv \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad |1\rangle \equiv \begin{bmatrix} 0 \\ 1 \end{bmatrix}.$$

Why do we call the angle-bracket thing a "ket," you ask? Let me tell you about the *bra* part, and then it will start to make sense.

The Hermitian transpose of the ket-vector $|v\rangle = \alpha|0\rangle + \beta|1\rangle$ is the *bra*-vector $\langle v|$:

$$\langle v| = \overline{\alpha}\langle 0| + \overline{\beta}\langle 1| \qquad \Leftrightarrow \qquad [\overline{\alpha}, \overline{\beta}] = \overline{\alpha}[1, 0] + \overline{\beta}[0, 1].$$

Vector coefficients

Change of basis

Outer products

Matrices

Summary

Exercises

9.4 Quantum information processing

Digital signal processing

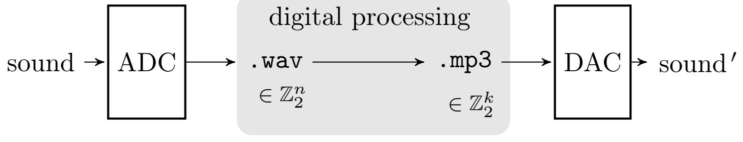


Figure 9.5: A digital information processing pipeline for sound recording and playback. Sound vibrations are captured by a microphone and converted to digital form using an analog-to-digital converter (ADC). Next the digital `wav` file is converted to the more compact `mp3` format using digital processing. In the last step, sound is converted back into analog sound vibrations by a digital-to-analog converter (DAC).

Quantum information processing

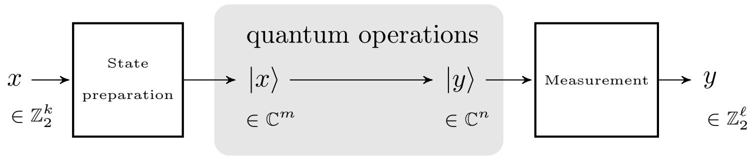


Figure 9.6: A quantum information processing pipeline. A classical bitstring x of length k is used as instructions for preparing an m-dimensional quantum state $|x\rangle$. Next, quantum operations are performed on the state $|x\rangle$ to convert it to the output state $|y\rangle$. Finally, the state $|y\rangle$ is *measured* to obtain the classical bitstring y as output.

9.5 Postulates of quantum mechanics

The *postulates* of quantum mechanics dictate the rules for working within the "quantum world." The four postulates define:

- What quantum states are
- Which quantum operations can be performed on quantum states
- How to extract information from quantum systems by measuring them
- How to represent composite quantum systems

These postulates specify the structure that all quantum theories must have. Together, the four postulates are known as the *quantum formalism*, and describe the math structure common to all fields that use quantum mechanics: physics, chemistry, engineering, and quantum information. Note the postulates are not provable or derivable from a more basic theory: scientists simply take the postulates as facts and make sure their theories embody these principles.

Quantum states

Postulate 1. To every isolated quantum system is associated a complex inner product space (Hilbert space) called the *state space*. A state is described by a unit vector in state space.

The qubit

Quantum state preparation

Quantum operations

Postulate 2. Time evolution of an isolated quantum system is unitary. If the state at time t is $|\psi\rangle$ and at time t' is $|\psi'\rangle$, then there exists a unitary operator U such that $|\psi'\rangle = U|\psi\rangle$.

Example 1: phase gate

Example 2: NOT gate

Example 3: Hadamard gate

Links

[Wikipedia article on quantum gates]
`https://en.wikipedia.org/wiki/Quantum_gate`

Exercises

Quantum measurements

Postulate 3. A quantum measurement is modelled by a collection of projection operators $\{\Pi_i\}$ that act on the state space of the system being measured and satisfy $\sum_i \Pi_i = \mathbb{1}$. The index i labels the different measurement outcomes.

The probability of outcome i when performing measurement $\{\Pi_i\}$ on a quantum system in the state $|\psi\rangle$ is given by the squared length of the state after applying the i^{th} projection operator:

$$\Pr(\{\text{outcome } i \text{ given state } |\psi\rangle\}) \;\equiv\; \left\| \Pi_i|\psi\rangle \right\|^2 \qquad \text{(Born's rule).}$$

When outcome i occurs, the post-measurement state of the system is

$$|\psi_i'\rangle \equiv \frac{\Pi_i|\psi\rangle}{\|\Pi_i|\psi\rangle\|}.$$

Born's rule

Post-measurement state

Composite quantum systems

Postulate 4. The state space of a composite quantum system is equal to the tensor product of the state spaces of the individual systems. If systems $1, 2, \ldots, n$ exist in states $|\varphi_1\rangle, |\varphi_2\rangle, \cdots, |\varphi_n\rangle$, then the state of the composite system is $|\varphi_1\rangle \otimes |\varphi_2\rangle \otimes \cdots \otimes |\varphi_n\rangle$.

Tensor product space

Tensor product of two vectors

State spaces and dimension counting

Exercises

Quantum entanglement

Summary

We can summarize the new concepts of quantum mechanics we learned in this chapter, and relate them to the standard concepts of linear algebra:

$$\begin{aligned} \text{quantum state} &\Leftrightarrow \text{vector } |v\rangle \in \mathbb{C}^d \\ \text{evolution} &\Leftrightarrow \text{unitary operations} \\ \text{measurement} &\Leftrightarrow \text{projections} \\ \text{composite system} &\Leftrightarrow \text{tensor product} \end{aligned}$$

Exercises

E9.1 Find the matrix representation of the projection matrices $\Pi_+ \equiv |+\rangle\langle+|$ and $\Pi_- \equiv |-\rangle\langle-|$. Show that $\Pi_+ + \Pi_- = \mathbb{1}$.

E9.2 Compute the probability of outcome "$-$" for the measurement $\{\Pi_+, \Pi_-\} = \{|+\rangle\langle+|, |-\rangle\langle-|\}$ performed on the quantum state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

E9.3 Given the state $|\theta\rangle = (\frac{1}{\sqrt{2}}, \frac{e^{i\theta}}{\sqrt{2}})^{\mathsf{T}}$, find a quantum state $|\theta^\perp\rangle$ that is orthogonal to $|\theta\rangle$. Find the projection operators Π_θ and $\Pi_{\theta^\perp}$ that correspond to the measurements in the basis $\{|\theta\rangle, |\theta^\perp\rangle\}$. Verify that $\Pi_\theta + \Pi_{\theta^\perp} = \mathbb{1}$. Compute the probability of outcome θ when performing the measurement $\{\Pi_\theta, \Pi_{\theta^\perp}\}$ on the state $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

Hint: The state $|\theta^\perp\rangle$ satisfies $\langle\theta^\perp|\theta\rangle = 0$.

E9.4 Given the two qubits $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and $|\phi\rangle = \gamma|0\rangle + \delta|1\rangle$, compute the tensor product state $|\psi\rangle \otimes |\phi\rangle$.

Links

9.6 Polarizing lenses experiment revisited

Discussion

[The Stern–Gerlach experiment]
`https://en.wikipedia.org/wiki/Stern-Gerlach_experiment`
`https://youtube.com/watch?v=rg4Fnag4V-E`

9.7 Quantum physics is not that weird

Quantum superposition

Interference

Measurement of a system affects the system's state

Wave functions

9.8 Quantum mechanics applications

Particle physics

Solid state physics

Superconductors

Quantum optics

Quantum cryptography

Quantum computing

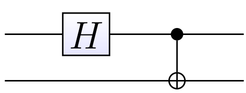
The idea of quantum computing has existed since the early days of quantum physics. Richard Feynman originally proposed the idea of a *quantum simulator* in 1982, which is a quantum apparatus that can simulate the quantum behaviour of another physical system. Imagine a device that can simulate the behaviour of physical systems that would otherwise be too difficult and expensive to build. The quantum simulator would be much better at simulating quantum phenomena than any simulation of quantum physics on a classical computer.

Another possible application of a quantum simulator could be to encode classical mathematical optimization problems as constraints in a quantum system, then let the quantum evolution of the system "search" for good solutions. Using a quantum simulator in this way,

it might be possible to find solutions to optimization problems much faster than any classical optimization algorithm could.

Once computer scientists started thinking about quantum computing, they weren't satisfied with studying optimization problems alone, and they set out to qualify and quantify all the computational tasks that are possible with qubits. A quantum computer stores and manipulates information that is encoded as quantum states. It's possible to perform certain computational tasks on a quantum computer much faster than on any classical computer. We'll discuss *Grover's search algorithm* and *Shor's factoring algorithm* below, but first let's introduce the basic notions of quantum computing.

Quantum circuits Computer scientists like to think of quantum computing tasks as series of "quantum gates," in analogy with the logic gates used to construct classical computers. Figure 9.7 shows an example of a quantum circuit that takes two qubits as inputs and produces two qubits as outputs.



$$\Leftrightarrow \quad \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & -\frac{1}{\sqrt{2}} \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

Figure 9.7: A quantum circuit that applies the Hadamard gate to the first qubit, then applies the controlled-NOT gate from the first qubit to the second qubit.

This circuit in Figure 9.7 is the combination of two quantum gates. The first operation is to apply the Hadamard gate H on the first qubit, leaving the second qubit untouched. This operation is equivalent to multiplying the input state by the matrix $H\otimes\mathbb{1}$. The second operation is called the *controlled-NOT* (or controlled-X) gate, which applies the X operator (also known as the NOT gate) to the second qubit whenever the first qubit is $|1\rangle$, and does nothing otherwise:

$$\text{CNOT}(|0\rangle\otimes|\varphi\rangle) = |0\rangle\otimes|\varphi\rangle, \qquad \text{CNOT}(|1\rangle\otimes|\varphi\rangle) = |1\rangle\otimes X|\varphi\rangle.$$

The circuit illustrated in Figure 9.7 can be used to create entangled quantum states. If we input the quantum state $|00\rangle \equiv |0\rangle\otimes|0\rangle$ into the circuit, we obtain the maximally entangled state $|\Phi_+\rangle \equiv \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ as output, as depicted in Figure 9.8.

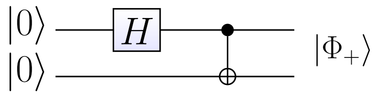


Figure 9.8: Inputting $|0\rangle \otimes |0\rangle$ into the circuit produces an EPR state $|\Phi_+\rangle \equiv \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$ on the two output wires of the circuit.

Quantum circuits can also represent quantum measurements. Figure 9.9 shows how a quantum measurement in the standard basis is represented.

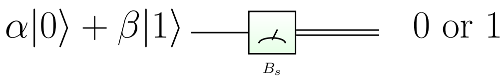


Figure 9.9: Measurement in the standard basis $B_s = \{|0\rangle, |1\rangle\}$. The projectors of this measurement are $\Pi_0 = |0\rangle\langle 0|$ and $\Pi_1 = |1\rangle\langle 1|$.

We use double lines to represent the flow of classical information in the circuit.

Quantum registers Consider a quantum computer with a single register $|R\rangle$ that consists of three qubits. The quantum state of this quantum register is a vector in $\mathbb{C}^2 \otimes \mathbb{C}^2 \otimes \mathbb{C}^2$:

$$|R\rangle = \quad (\alpha_1|0\rangle + \beta_1|1\rangle) \;\otimes\; (\alpha_2|0\rangle + \beta_2|1\rangle) \;\otimes\; (\alpha_3|0\rangle + \beta_3|1\rangle)\,,$$

where the tensor product $\otimes$ is used to combine the quantum states of the individual qubits. We'll call this the "physical representation" of the register and use 0-based indexing for the qubits. Borrowing language from classical computing, we'll call the rightmost qubit the *least significant* qubit, and the leftmost qubit the *most significant* qubit.

The tensor product of three vectors with dimension two is a vector with dimension eight. The quantum register $|R\rangle$ is thus a vector in an eight-dimensional vector space. The quantum state of a three-qubit register can be written as:

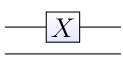
$$|R\rangle = a_0|0\rangle + a_1|1\rangle + a_2|2\rangle + a_3|3\rangle + a_4|4\rangle + a_5|5\rangle + a_6|6\rangle + a_7|7\rangle,$$

where a_i are complex coefficients. We'll call this eight-dimensional vector space the "logical representation" of the quantum register. Part of the excitement about quantum computing is the huge size of the "logical space" where quantum computations take place. The logical space of a 10-qubit quantum register has dimension $2^{10} = 1024$.

That's 1024 complex coefficients we're talking about. That's a big state space for just a 10-qubit quantum register. Compare this with a 10-bit classical register, which can store one of $2^{10} = 1024$ discrete values.

We won't discuss quantum computing further here, but I still want to show you some examples of single-qubit quantum operations and their effect on the tensor product space, so you'll have an idea of the craziness that is possible.

Quantum gates Let's say you've managed to construct a quantum register; what can you do with it? Recall the single-qubit quantum operations Z, X, and H we described earlier. We can apply any of these operations on individual qubits in the quantum register. For example, applying the $X = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ gate to the first (most significant) qubit of the quantum register corresponds to the following quantum operation:



$$\Leftrightarrow \qquad X \otimes \mathbb{1} \otimes \mathbb{1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The operator $X \otimes \mathbb{1} \otimes \mathbb{1}$ "toggles" the first qubit in the register while leaving all other qubits unchanged.

Okay, so what?

Quantum computers give us access to a very large state space. The fundamental promise of quantum computing is that a small set of simple quantum operations (quantum gates) can be used to perform interesting computational tasks. Sure it's difficult to interact with and manipulate quantum systems, but the space is so damn big that it's worth checking out what kind of computing you can do in there. It turns out there are already several useful things you can do using a quantum computer. The two flagship applications for quantum computing are Grover's search algorithm and Shor's factoring algorithm.

Grover's search algorithm Suppose you're given an unsorted list of n items and you want to find a particular item in that list. This is called an *unstructured search problem.* This is a hard problem to solve for a classical computer since the algorithm must parse through the

entire list, which takes roughly n steps. In contrast, the unstructured problem can be solved in roughly $\sqrt{n}$ steps on a quantum computer using *Grover's algorithm*.

The quantum speed for the unstructured search problem sure is nice, but it's really nothing to get excited about. The real money-maker for the field of quantum computing has been Shor's factoring algorithm for factoring products of prime numbers.

Shor's factoring algorithm The security of the RSA cryptosystem we discussed in Section 7.9 is based on the assumption that factoring products of large prime numbers is computationally intractable. Given the product de of two unknown prime numbers d and e, it is computationally difficult to find the factors e and d. No classical algorithm is known that can factor large numbers; even the letter agencies will have a hard time finding the factors of de when d and e are chosen to be sufficiently large prime numbers. Thus, if an algorithm that could quickly factor large numbers existed, attackers would be able to break many of the current security systems. *Shor's factoring algorithm* fits the bill, theoretically speaking.

Shor's algorithm reduces the factoring problem to the problem of *period finding*, which can be solved efficiently using the quantum Fourier transform. Shor's algorithm can factor large numbers efficiently (in polynomial time). This means RSA encryption would be easily hackable by running Shor's algorithm on a sufficiently large, and sufficiently reliable quantum computer. The letter agencies are excited about this development since they'd love to be able to hack all present-day cryptography. Can you imagine not being able to log in securely to any website because Eve is listening in, hacking your crypto using her quantum computer?

Currently, Shor's algorithm is only a *theoretical* concern. Despite considerable effort, no quantum computers exist today that can manipulate quantum registers with thousands of qubits.

Discussion

Quantum teleportation Figure 9.10 illustrates a surprising aspect of quantum information: we can "teleport" a quantum state $|\psi\rangle$ from one lab to another. The quantum state $|\psi\rangle$ starts in the first qubit of the register, which is held by Alice, and ends in the third qubit, which is in Bob's lab, but there is no quantum communication channel between the two labs. This is why the term "quantum teleportation" was coined to describe this communication task, since the state $|\psi\rangle$ seems to materialize in Bob's lab like the teleportation machines used in Star Trek.

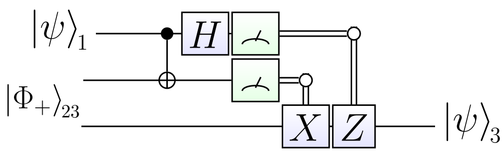


Figure 9.10: The first two qubits are in Alice's lab. The state of the first qubit $|\psi\rangle_1$ is transferred into the third qubit $|\psi\rangle_3$, which Bob controls. We say ψ is "teleported" from Alice's lab to Bob's lab because the quantum state ends up in Bob's lab, but there is no quantum communication channel connecting the labs. The state teleportation happens thanks to the pre-shared entanglement and the two bits of classical information.

Links

[Shor's algorithm for factoring products of prime integers]
`https://en.wikipedia.org/wiki/Shor's_algorithm`

[Emerging insights on limitations of quantum computing]
`https://www.siam.org/pdf/news/100.pdf`

Quantum error-correcting codes

Quantum information theory

Conclusion

9.9 Quantum mechanics problems

Let's recap what just happened here. Did we really cover all the topics of an introductory quantum mechanics course? Yes, we did! Thanks to your solid knowledge of linear algebra, learning the postulates of quantum mechanics took only a few dozen pages. Sure we went quickly and skipped the more physics-y topics, but we covered all the core ideas of quantum theory.

But surely it's impossible to learn quantum mechanics in such a short time? Well, you tell me. You're here. The problems are here. Prove to me you've really learned quantum mechanics by tackling the practice problems presented in this section like a boss. It's the end of the book, so don't be saving your energy. Solve these problems and then you're done.

P9.1 You work in a quantum computing startup and your boss asks you to implement the quantum gate $Q = \frac{1}{\sqrt{2}}\left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right]$. Can you do it?

Hint: Recall the requirements for quantum gates.

P9.2 The Y gate is defined as $Y = \left[\begin{smallmatrix} 0 & -i \\ i & 0 \end{smallmatrix}\right]$. Compute the effect of the operator YY on the elements of the standard basis $\{|0\rangle, |1\rangle\}$.

P9.3 Compute the products of the quantum gates HXH and HZH.

P9.4 Show that the functions $\psi_1(x) = 2x - 1$ and $\psi_2(x) = 6x^2 - 6x + 1$ are orthogonal with respect to the inner product $\langle f, g\rangle = \int_0^1 \overline{f(x)}g(x)\,dx$.

P9.5 Consider a model of a particle in a one-dimensional box of width one. The state of the particle is described by the wave function $\psi(x)$, where $x \in [0, 1]$. Find the probability of observing x in the first quarter of the box (x between 0 and $\frac{1}{4}$) for the following wave functions: **a)** $\psi_a(x) = \sqrt{3}(2x - 1)$, **b)** $\psi_b(x) = \sqrt{5}(6x^2 - 6x + 1)$, **c)** a constant wave function ψ_c.

Hint: The probability of finding the particle somewhere in the interval $[a, b]$ is computed using the integral $\Pr(\{a \leq x \leq b\}|\psi) = \int_a^b |\psi(x)|^2\,dx$.

End matter

Conclusion

By tackling the linear algebra concepts in this book, you've proven you can handle computational complexity, develop geometric intuition, and understand abstract math ideas. These are precisely the types of skills you'll need in order to understand more advanced math concepts, build scientific models, and develop useful applications. Congratulations on taking this important step toward your mathematical development. Throughout this book, we learned about vectors, linear transformations, matrices, abstract vector spaces, and many other math concepts that are useful for building math models.

Mathematical models serve as a highly useful common core for all sciences, and the techniques of linear algebra are some of the most versatile modelling tools that exist. Every time you use an equation to characterize a real-world phenomenon, you're using your math modelling skills. Whether you're applying some well-known scientific model to describe a phenomenon or developing a new model specifically tailored to a particular application, the deeper your math knowledge, the better the math models you'll be able to leverage. Let's review and catalogue some of the math modelling tools we've learned about, and see how linear algebra fits into a wider context.

To learn math modelling, you must first understand basic math concepts such as numbers, equations, and functions $f : \mathbb{R} \to \mathbb{R}$. Once you know about functions, you can use different formulas $f(x)$ to represent, model, and predict the values of real-world quantities. Working with functions is the first modelling superpower conferred on people who become knowledgeable in math. For example, understanding the properties of the function $f(x) = Ae^{-x/B}$ in the abstract enables you to describe the expected number of atoms remaining in a radioactive reaction $N(t) = N_o\, e^{-\gamma t}$, predict the voltage of a discharging capacitor over time $v(t) = V_o\, e^{-\frac{t}{RC}}$, and understand the exponential probability distribution $p_X(x) = \lambda e^{-\lambda x}$.

To further develop your math modelling skills, the next step is to

generalize the concepts of inputs x, outputs y, and functions f to other input-output relationships. In linear algebra, we studied functions of the form $T : \mathbb{R}^n \to \mathbb{R}^m$ that obey the linear property:

$$T(\alpha\vec{x}_1 + \beta\vec{x}_2) = \alpha T(\vec{x}_1) + \beta T(\vec{x}_2).$$

This linear structure enables us to study the properties of many functions, solve equations involving linear transformations, and build useful models for many applications (some of which we discussed in Chapter 7). The mathematical structure of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ can be represented as multiplication by a matrix $M_T \in \mathbb{R}^{m\times n}$. The notion of matrix *representations* ($T \Leftrightarrow M_T$) was central throughout this book. Even if you forget the computational procedures we learned, the idea of representations should stick with you, and you should be able to recognize representations in many contexts. That's a big deal, because most advanced math topics involve studying the parallels between different abstract notions. Understanding linear transformations and their concrete representations as matrices is an important step in your math development.

The computational skills you learned in Chapter 3 are also useful; though you probably won't be solving any problems by hand using row operations from this point forward, since computers outclass humans on matrix arithmetic tasks. Good riddance. Until now, you did all the work and used `SymPy` to check your answers. From now on, you can let `SymPy` do all the calculations and your job will be to chill.

If you didn't skip the sections on abstract vector spaces, you know about the parallels between the vector space $\mathbb{R}^4$ and the abstract vector spaces of third-degree polynomials $a_0 + a_1x + a_2x^2 + a_3x^3$ and 2×2 matrices $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$. This is another step up the ladder of abstraction, as it deepens your understanding of all math objects with vector-like structure.

It was my great pleasure to be your guide through the subject of linear algebra. I hope you walk away from this book with a solid understanding of how the concepts of linear algebra fit together. In the book's introduction, I likened linear algebra to playing with LEGOs. Indeed, if you feel comfortable manipulating vectors and matrices, performing change-of-basis operations, and using the matrix decomposition techniques to see inside matrices, you'll be able to "play" with all kinds of complex systems and problems. For example, consider the linear transformation T that you want to apply to an input vector $\vec{v}$. Suppose the linear transformation T is most easily described in the basis B', but the vector $\vec{v}$ is expressed with respect to the basis B. "No problem," you can say, and proceed to build the following chain of matrices that compute the output vector $\vec{w}$:

$$[\vec{w}]_B = {}_B[\mathbb{1}]_{B'}\ {}_{B'}[A_T]_{B'}\ {}_{B'}[\mathbb{1}]_B[\vec{v}]_B.$$

Do you see how matrices and vectors fit together neatly like LEGOs?

I can't tell you what the next step on your journey will be. With your new linear algebra modelling skills, a thousand doors have opened for you; now you must explore and choose. Will you learn how to code and start a software company? Maybe you'll use your analytical skills to go to Wall Street and destroy the System from the inside. Or perhaps you'll apply your modelling skills to revolutionize energy generation, thus making human progress sustainable. Regardless of your choice of career, I hope you'll stay on good terms with math and continue learning whenever you have the chance. Good luck with your studies!

Social stuff

Be sure to contact me if you have any feedback about this book. It helps to hear which parts of the book readers like, hate, or don't understand. I consider all feedback in updating and improving future editions of this book. This is how the book got good in the first place—lots of useful feedback from readers. You can reach me by email at `ivan@minireference.com`.

Another appreciated thing you can do to help us is to write a review of the book on Amazon.com, Goodreads, Google Books, or otherwise spread the word about the NO BULLSHIT textbook series. Talk to your friends and let them in on the math buzz.

If you want to know what `Minireference Co.` has been up to, check out our blog at `minireference.com/blog/`. The blog is a mix of 30% publishing technology talk, 50% startup business talk, and 20% announcements. Checking the blog is the easiest way to follow the progress of our revolution in the textbook industry. For real; we're totally serious about making education accessible through affordable textbooks. You can also connect via Twitter `@minireference` and Facebook `fb.me/noBSguide`.

Acknowledgements

This book would not have been possible without the support of my parents, and my teachers: Claude Crépeau, Patrick Hayden, Paul Kenton, Benoit Larose, and Ioannis Psaromiligkos. It's by listening to them that I gained the confidence to learn and later teach advanced math topics. Patrick Hayden deserves a particular mention in this regard since his ability to explain concepts clearly inspired me to tackle complicated problems using a steady, methodical approach. Patrick also showed me that it's possible to trick students into learning even

the most advanced topics by presenting the material in a logical and structured manner. Many thanks also to David Avis, Arlo Breault, Michael Hilke, Igor Khavkine, Felix Kwok, Juan Pablo Di Lelle, Ivo Panayotov, and Mark M. Wilde for their support with the book.

The errorlessness and consistency of the text would not have been possible without the help of my editor Sandy Gordon, who did a great job at polishing the text until it flowed. Truly no bullshit is allowed into the book when Sandy's on watch. Many thanks to Polina Anis'kina who helped me to create the problem sets for the book.

General linear algebra links

Below are some useful links to resources where you can learn more about linear algebra. We covered a lot of ground, but linear algebra is endless. Don't sit on your laurels and think you're the boss now that you've completed this book and its problem sets. You have the tools, but you need to practice using them. Try reading about the same topics from some other sources. See if you can do the problem sets in another linear algebra textbook. Try to use linear algebra in the coming year and further solidify your understanding of the material.

[Video lectures of Gilbert Strang's linear algebra class at MIT]
`http://ocw.mit.edu/courses/mathematics/18-06-linear-algebra-spring-2010`

[The essence of linear algebra video playlist by 3Blue1Brown]
`http://bit.ly/essence_of_LA`

[A free online textbook with amazing interactive visualizations]
`http://immersivemath.com/ila/index.html`

[Lecture notes by Terrence Tao]
`http://www.math.ucla.edu/~tao/resource/general/115a.3.02f/`

[Wikipedia overview on matrices]
`https://en.wikipedia.org/wiki/Matrix_(mathematics)`

[Linear algebra wikibook (with solved problems)]
`https://en.wikibooks.org/wiki/Linear_Algebra`

[Proofs involving linear algebra]
`http://proofwiki.org/wiki/Category:Linear_Algebra`

[Linear algebra from first principles using diagrams only]
`https://graphicallinearalgebra.net/`

Appendix A

Answers and solutions

Chapter 1 solutions

Answers to exercises

E1.1 a) $x = 2$; **b)** $x = 25$; **c)** $x = 100$. **E1.2 a)** $f^{-1}(x) = x^2$, $x = 16$. **b)** $g^{-1}(x) = -\frac{1}{2}\ln(x)$, $x = 0$.

Solutions to selected exercises

Answers to problems

P1.1 $x = \pm 4$. **P1.2** $x = A\cos(\omega t + \phi)$. **P1.3** $x = \frac{ab}{a+b}$. **P1.4 a)** 2.2795. **b)** 1024. **c)** -8.373. **d)** 11. **P1.5** $x = \tan\theta\sqrt{a^2+b^2+c^2}$. **P1.6** 1.06 cm. **P1.7** $\ell_{\text{rope}} = 8.42\,\text{m}$. **P1.11** $A_1(x) = 3x$ and $A_2(x) = \frac{1}{2}x^2$.

Solutions to selected problems

P1.6 The volume of the water stays constant and is equal to $1000\,\text{cm}^3$. Initially the height of the water h_1 can be obtained from the formula for the volume of a cylinder $1000\,\text{cm}^3 = h_1\pi(8.5\,\text{cm})^2$, so $h_1 = 4.41\,\text{cm}$. After the bottle is inserted, the water has the shape of a cylinder with a cylindrical part missing. The volume of water is $1000\,\text{cm}^3 = h_2\left(\pi(8.5\,\text{cm})^2 - \pi(3.75\,\text{cm})^2\right)$. We find $h_2 = 5.47\,\text{cm}$. The change in height is $h_2 - h_1 = 5.47 - 4.41 = 1.06\,\text{cm}$.

P1.7 The length of the horizontal part of the rope is $\ell_h = 4\sin 40$. The circular portion of the rope that hugs the pulley has length $\frac{1}{4}$ of the circumference of a circle with radius $r = 50\,\text{cm} = 0.5\,\text{m}$. Using the formula $C = 2\pi r$, we find $\ell_c = \frac{1}{4}(2\pi(0.5)) = \frac{\pi}{4}$. The vertical part of the rope has length $\ell_v = 4\cos 40 + 2$. The total length of rope is $\ell_h + \ell_c + \ell_v = 8.42\,\text{m}$.

P1.8 There exists at least one banker who is not a crook. Another way of saying the same thing is "not all bankers are crooks"—just *most* of them.

P1.9 Everyone steering the ship at Monsanto ought to burn in hell, forever.

P1.10 a) Investors with money but without connections. **b)** Investors with connections but no money. **c)** Investors with both money and connections.

Chapter 2 solutions

Answers to exercises

E2.1 $A^{-1} = \begin{bmatrix} \frac{1}{7} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$. **E2.2 a)** $A\vec{v} = \left[\begin{smallmatrix} 7 \\ 14 \end{smallmatrix}\right]$; **b)** $B\vec{v} = \left[\begin{smallmatrix} -1 \\ 9 \end{smallmatrix}\right]$; **c)** $A(B\vec{v}) = \left[\begin{smallmatrix} 26 \\ 41 \end{smallmatrix}\right]$; **d)** $B(A\vec{v}) = \left[\begin{smallmatrix} -7 \\ 63 \end{smallmatrix}\right]$; **e)** $A\vec{w} = \begin{bmatrix} -15 \\ -32 \end{bmatrix}$; **f)** $B\vec{w} = \left[\begin{smallmatrix} 3 \\ -21 \end{smallmatrix}\right]$. **E2.3** $v_1 = -2$, $v_2 = 3$.

Solutions to selected exercises

E2.1 To find A^{-1} we must consider the action of $A = \left[\begin{smallmatrix} 7 & 0 \\ 0 & 2 \end{smallmatrix}\right]$ on an arbitrary vector $\vec{v} = \left[\begin{smallmatrix} v_1 \\ v_2 \end{smallmatrix}\right]$, and perform the inverse action. Since A multiplies the first component by 7, A^{-1} must divide the first component by 7. Since A multiplies the second component by 2, A^{-1} must divide the second component by 2. Thus $A^{-1} = \begin{bmatrix} \frac{1}{7} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}$.

Answers to problems

P2.1 a) $q(x)$ is nonlinear; **b)** $f(x)$, $g(x)$, and $h(x)$ are all linear; **c)** $i(x)$ is nonlinear; **d)** $j(x)$ is nonlinear. **P2.2** $(1, 2, 3)$. **P2.3 a)** 0; **b)** $(0, 0, 1)$; **c)** $(0, 1, -1)$; **d)** $(0, 0, -1)$. **P2.4 a)** 5; **b)** $(-1, 1, 1)$; **c)** $(0, 0, 0)$; **d)** $(0, 0, 0)$. **P2.5** $M\vec{v} = \begin{bmatrix} \alpha z_1 + \beta z_2 \\ \gamma z_1 + \delta z_2 \end{bmatrix}$. **P2.6 a)** $\left[\begin{smallmatrix} -5 & -5 \\ 4 & 2 \end{smallmatrix}\right]$; **b)** $\left[\begin{smallmatrix} -5 & 10 & -5 \\ 20 & 5 & 10 \end{smallmatrix}\right]$; **c)** $\left[\begin{smallmatrix} 17 & 28 \\ 41 & 64 \end{smallmatrix}\right]$; **d)** Doesn't exist; **e)** $\left[\begin{smallmatrix} 18 & 21 \\ 9 & 12 \\ 8 & 9 \end{smallmatrix}\right]$; **f)** $\left[\begin{smallmatrix} 9 & 12 \end{smallmatrix}\right]$; **g)** $\begin{bmatrix} -4 \\ -7 \end{bmatrix}$; **h)** 0; **i)** $\left[\begin{smallmatrix} 2 & 1 \\ -4 & -2 \end{smallmatrix}\right]$; **j)** 8; **k)** 11; **l)** -3; **m)** 20.

Solutions to selected problems

P2.1 A function is linear in x if it contains x raised only to the first power. Basically, $f(x) = mx$ (for some constant m) is the only possible linear function of one variable.

Chapter 3 solutions

Answers to exercises

Solutions to selected exercises

Answers to problems

P3.1 $x = 15$ and $y = 2$. **P3.2 a)** $R_2 \leftarrow R_2 - 2R_1$, $R_2 \leftarrow -2R_2$, $R_1 \leftarrow R_1 - R_2$; **b)** $R_2 \leftarrow R_2 - 2R_1$, $R_2 \leftarrow \frac{-2}{3}R_2$, $R_1 \leftarrow R_1 - \frac{3}{2}R_2$; **c)** $R_1 \leftarrow \frac{1}{2}R_1$, $R_2 \leftarrow R_2 - 3R_1$, $R_2 \leftarrow \frac{4}{3}R_2$, $R_1 \leftarrow R_1 - \frac{3}{4}R_2$. **P3.3 a)** $(-2, 2)$; **b)** $(-4, -1, -2)$; **c)** $(\frac{-2}{5}, \frac{-1}{2}, \frac{3}{5})$. **P3.4** $C = B^{-1}$. **P3.5** $\begin{bmatrix} -2 & -2 \\ -15 & -15 \end{bmatrix}$. **P3.6 a)** $\begin{bmatrix} 0 & \cos(\alpha) - \sin(\alpha) \\ \sin(\alpha) - \cos(\alpha) & -\cos(2\alpha) \end{bmatrix}$; **b)** $\begin{bmatrix} \cos^2(\alpha) & \sin(\alpha) \\ -\cos(\alpha) & \sin^2(\alpha) \end{bmatrix}$; **c)** $\begin{bmatrix} \cos(2\alpha) & \sin(\alpha) \\ -\cos(\alpha) & 0 \end{bmatrix}$. **P3.7 a)** -3; **b)** 0; **c)** 10. **P3.8** Area $= 2$. **P3.9 a)** The inverse doesn't exist; **b)** $\begin{bmatrix} 5 & -2 \\ -2 & 1 \end{bmatrix}$; **c)** $\begin{bmatrix} 2 & -\frac{3}{2} \\ -1 & 1 \end{bmatrix}$.

Solutions to selected problems

P3.4 First simplify the equation by multiplying with A^{-1} from the left, and with D^{-1} from the right, to obtain $BC = \mathbb{1}$. Now we can isolate C by multiplying with B^{-1} from the left. We obtain $B^{-1} = C$.

Chapter 4 solutions

Answers to exercises

Solutions to selected exercises

Answers to problems

P4.1 **a)** $(1, 2)$. **b)** Lines overlap so every point on the lines is an intersection point. **c)** $(1, 1)$. **P4.2** **a)** $\{(1, -\frac{1}{2}, 0) + s(0, -\frac{1}{2}, 1), \forall s \in \mathbb{R}\}$; **b)** $\{(-1, 2, 0) + t(1, -1, 1), \forall t \in \mathbb{R}\}$. **P4.3** $\Pi_{\vec{u}}(\vec{v}) = \frac{\vec{v}\cdot\vec{u}}{\|\vec{u}\|^2}\vec{u} = \frac{(1,1,1)\cdot(2,1,-1)}{2^2+1^2+(-1)^2}(2, 1, -1) = \frac{1}{3}(2, 1, -1)$; $\Pi_{\vec{v}}(\vec{u}) = \frac{\vec{v}\cdot\vec{u}}{\|\vec{v}\|^2}\vec{v} = \frac{(1,1,1)\cdot(2,1,-1)}{1^2+1^2+1^2}(1, 1, 1) = \frac{2}{3}(1, 1, 1)$. **P4.4** $\Pi_P(\vec{v}) = \frac{1}{7}(17, 30, -1)$. **P4.5** $\vec{v} = (2, 1, 3)_W$. **P4.6** **a)** Yes; **b)** No; **c)** Yes; **d)** No; **e)** No. **P4.10** $\vec{v}_4 = (5, 2, -3, 1)$.

Solutions to selected problems

P4.1 To find the intersection point, solve the two equations simultaneously.

P4.4 Let's first find $\Pi_{P^\perp}(\vec{v}) = \frac{\vec{v}\cdot\vec{n}}{\|\vec{n}\|^2}\vec{n} = \frac{(3,4,1)\cdot(2,-1,4)}{2^2+(-1)^2+4^2}(2, -1, 4) = \frac{2}{7}(2, -1, 4)$. Then $\Pi_P(\vec{v}) = \vec{v} - \Pi_{P^\perp}(\vec{v}) = (3, 4, 1) - \frac{2}{7}(2, -1, 4) = \frac{1}{7}(17, 30, -1)$. We can verify $\Pi_P(\vec{v}) + \Pi_{P^\perp}(\vec{v}) = \frac{1}{7}(17, 30, -1) + \frac{2}{7}(2, -1, 4) = (3, 4, 1) = \vec{v}$. This shows that the projection we found is correct.

P4.6 In each case, we must either recognize the set as being a subspace or explain which of the three subset conditions fails. **a)** W_1 corresponds to the plane $x + y = 0$ which is a subspace of $\mathbb{R}^3$. **b)** Consider the vectors $(0, 1, 0) \in W_2$ and $(0, 0, 1) \in W_2$. The sum $(0, 1, 0) + (0, 0, 1) = (0, 1, 1) \notin W_2$, so W_2 is not closed under addition. **c)** W_3 is equivalent to the line $\{(0, 0, 0) + t(1, 1, 1), \forall t \in \mathbb{R}\}$, which is a subspace of $\mathbb{R}^3$. **d)** Consider the vector $(1, 0, 0) \in W_4$. Multiplying this vector by -3 results in $(-3, 0, 0) \notin W_4$, so the set W_4 is not closed under scalar multiplication. **e)** The set W_5 does not contain the zero element $(0, 0, 0)$.

P4.7 Define $W = \text{span}(\vec{v}_1, \vec{v}_2 - \vec{v}_1, \vec{v}_3 - \vec{v}_2, \vec{v}_4 - \vec{v}_3)$. Every vector $\vec{w} \in W$ can be written in the following two equivalent ways:

$$\begin{aligned} \vec{w} &= \alpha_1\vec{v}_1 + \alpha_2(\vec{v}_2 - \vec{v}_1) + \alpha_3(\vec{v}_3 - \vec{v}_2) + \alpha_4(\vec{v}_4 - \vec{v}_3) \\ &= (\alpha_1 + \alpha_2)\vec{v}_1 + \alpha_2\vec{v}_2 + \alpha_3(\vec{v}_3 - \vec{v}_2) + \alpha_4(\vec{v}_4 - \vec{v}_3). \end{aligned}$$

This implies $W = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3 - \vec{v}_2, \vec{v}_4 - \vec{v}_3)$. Using this approach we can show $W = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4 - \vec{v}_3)$, and then $W = \text{span}(\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4) = V$.

P4.8 We are told the set $S = \{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is linearly independent, which implies $\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}$ has only trivial solution $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 0, 0, 0)$. Define the set $S' = \{\vec{v}_1, \vec{v}_2 - \vec{v}_1, \vec{v}_3 - \vec{v}_2, \vec{v}_4 - \vec{v}_3\}$. To find whether S' is a linearly independent set, consider the following equation:

$$\beta_1\vec{v}_1 + \beta_2(\vec{v}_2 - \vec{v}_1) + \beta_3(\vec{v}_3 - \vec{v}_2) + \beta_4(\vec{v}_4 - \vec{v}_3) = 0.$$

Rearranging the terms, we obtain the equivalent expression

$$\underbrace{(\beta_1 - \beta_2 - \beta_3 - \beta_4)}_{\alpha_1}\vec{v}_1 + \underbrace{(\beta_2 - \beta_3)}_{\alpha_2}\vec{v}_2 + \underbrace{(\beta_3 - \beta_4)}_{\alpha_3}\vec{v}_3 + \underbrace{\beta_4}_{\alpha_4}\vec{v}_4 = 0.$$

We recognize this form of equation from the definition of linear independence. Since we're told $\{\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{v}_4\}$ is a linearly independent set, we know the only solution to the above equation is $(\alpha_1, \alpha_2, \alpha_3, \alpha_4) = (0, 0, 0, 0)$, or

$$0 = \beta_1 - \beta_2 - \beta_3 - \beta_4, \quad 0 = \beta_2 - \beta_3 - \beta_4, \quad 0 = \beta_3 - \beta_4, \quad 0 = \beta_4.$$

Solving for the unknowns in the order β_4, β_3, β_2, β_1, we find the only solution is $(\beta_1, \beta_2, \beta_3, \beta_4) = (0, 0, 0, 0)$; therefore, $\{\vec{v}_1, \vec{v}_2 - \vec{v}_1, \vec{v}_3 - \vec{v}_2, \vec{v}_4 - \vec{v}_3\}$ is a linearly independent set.

P4.9 Since the set $\{\vec{u}, \vec{v}\}$ is a basis for V, the dimension of V is two. The set $\{\vec{u} + \vec{v}, a\vec{u}\}$ contains two vectors so it is sufficient to check whether the vectors in the set are linearly independent. Consider the following equation:

$$\beta_1(\vec{u} + \vec{v}) + \beta_2 a\vec{u} = 0 = \underbrace{(\beta_1 + \beta_2 a)}_{\alpha_1}\vec{u} + \underbrace{\beta_1}_{\alpha_2}\vec{v}.$$

Given $\{\vec{u}, \vec{v}\}$ is a linearly independent set, the only solution to $\alpha_1\vec{u} + \alpha_2\vec{v} = 0$ is $(\alpha_1, \alpha_2) = (0, 0)$; therefore $\beta_1 + \beta_2 a = \beta_1 = 0$ and $\beta_1 = \beta_2 = 0$. This implies $\{\vec{u} + \vec{v}, a\vec{u}\}$ is a basis for V. The proof for $\{a\vec{u}, b\vec{v}\}$ is analogous.

P4.10 Construct the matrix $V = \begin{bmatrix} -\vec{v}_1- \\ -\vec{v}_2- \\ -\vec{v}_3- \end{bmatrix}$ and find a vector in its null space.

Chapter 5 solutions

Answers to exercises

Solutions to selected exercises

Answers to problems

P5.1 **a)** Linear, $M_{T_1} = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right]$; **b)** Nonlinear, $T_2(2,2) \neq 2T_2(1,1)$; **c)** Nonlinear, $T_3(-1,-1) + T_3(1,1) \neq T_2(0,0)$; **d)** Linear, $M_{T_4} = \left[\begin{smallmatrix} 3 & -2 & 1 \\ 2 & 1 & -4 \end{smallmatrix}\right]$; **e)** Linear, $M_{T_5} = \left[\begin{smallmatrix} \frac{1}{2} \\ 3 \end{smallmatrix}\right]$; **f)** Nonlinear, $T_6(2,0,0,0) \neq 2T_6(1,0,0,0)$. **P5.2** $\text{Im}(T) = \text{span}((1,1,0),(0,-1,2))$. **P5.3** $M_T = \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix}$; $\text{Ker}(T) = \text{span}((a_x, a_y, a_z)^{\mathsf{T}})$.

P5.4 **a)** ${}_B[M_T]_B = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{smallmatrix}\right]$; **b)** ${}_{B'}[M_T]_{B'} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & \frac{3}{2} & \frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$; **c)** ${}_B[M_T]_{B'} = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 0 \end{smallmatrix}\right]$.

P5.5 **a)** $\mathbb{1}$; **b)** $M_1 = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$; **c)** $M_2 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right]$; **d)** $M_3 = \begin{bmatrix} \cos(\frac{\pi}{3}) & -\sin(\frac{\pi}{3}) \\ \sin(\frac{\pi}{3}) & \cos(\frac{\pi}{3}) \end{bmatrix}$; **e)** $M_4 = \begin{bmatrix} \cos(-\frac{\pi}{6}) & -\sin(-\frac{\pi}{6}) \\ \sin(-\frac{\pi}{6}) & \cos(-\frac{\pi}{6}) \end{bmatrix}$; **f)** $M_5 = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$.

Solutions to selected problems

P5.2 Applying T to the input vector $(1,0)$ produces $(1, 1-0, 2\cdot 0) = (1,1,0)$, and the input vector $(0,1)$ produces the output $(0, 0-1, 2\cdot 1) = (0,-1,2)$. Thus, $\text{Im}(T) = \text{span}((1,1,0),(0,-1,2)) \subseteq \mathbb{R}^3$.

P5.3 Use the standard probing-with-the-standard-basis approach to obtain the matrix representation. See `youtu.be/BaM7OCEm3G0` for an interesting discussion about the cross product viewed as a linear transformation.

P5.4 The change-of-basis transformation from B' to the standard basis is ${}_{B}[\mathbb{1}]_{B'} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & -1 \end{bmatrix}$. We find ${}_{B}[M_T]_{B} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}$ directly by observing the formula for T. Next we compute ${}_{B'}[M_T]_{B'}$ using the "sandwich" formula ${}_{B'}[\mathbb{1}]_{B}\, {}_{B}[M_T]_{B}\, {}_{B}[\mathbb{1}]_{B'}$. Finally, we compute ${}_{B}[M_T]_{B'}$ by changing only the right basis of the transformation: ${}_{B}[M_T]_{B'} = {}_{B}[M_T]_{B}\, {}_{B}[\mathbb{1}]_{B'}$.

P5.5 First, you must visually recognize the type of transformation that is acting in each case; then, use the appropriate matrix formula. Transformations 1 and 2 are reflections. Transformations 3 and 4 are rotations. Transformation 5 is a *shear* that maps $T_5(\hat{\imath}) = \hat{\imath}$ and $T_5(\hat{\jmath}) = (\frac{1}{2}, 1)^{\mathsf{T}}$. You can obtain the matrix representation from these two observations.

P5.6 Let $B_s = \{\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_n\}$ be the standard basis for $\mathbb{R}^n$. Since $A\vec{x} = \vec{b}$ has a solution $\vec{x}$ for every possible $\vec{b}$, it is possible to find the solutions $\vec{x}_i$ in n equations of the form $A\vec{x}_i = \hat{e}_i$, for $i \in \{1, 2, \ldots, n\}$. Now construct the matrix B that contains the solutions $\vec{x}_i$ as columns: $B = [\vec{x}_1, \ldots, \vec{x}_n]$. Observe that $AB = A[\vec{x}_1, \ldots, \vec{x}_n]$ $= [A\vec{x}_1, \ldots, A\vec{x}_n] = [\hat{e}_1, \ldots, \hat{e}_n] = \mathbb{1}_n$. The equation $BA = \mathbb{1}_n$ implies A is invertible.

P5.7 If T is injective, then its kernel contains only the zero vector $\text{Ker}(T) = \{\vec{0}\}$. Since we're interested in determining the linear independence of the set $\{T(\vec{v}_1), \ldots, T(\vec{v}_n)\}$, we must consider the solutions to the equation

$$\alpha_1 T(\vec{v}_1) + \cdots + \alpha_n T(\vec{v}_n) = \vec{0}.$$

Knowing T is linear, we can rewrite this as $T(\alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n) = \vec{0}$. Since $\text{Ker}(T) = \{\vec{0}\}$, the equation we started with is equivalent to the equation $\alpha_1\vec{v}_1 + \cdots + \alpha_n\vec{v}_n = \vec{0}$. But we're told $\{\vec{v}_1, \ldots, \vec{v}_n\}$ is a linearly independent set, so the only solution to this equation is the trivial solution $\alpha_i = 0$. Therefore $\{T(\vec{v}_1), \ldots, T(\vec{v}_n)\}$ is a linearly independent set.

Chapter 6 solutions

Answers to exercises

Solutions to selected exercises

Answers to problems

P6.1 $\lambda_1 = \varphi$ and $\lambda_2 = -\frac{1}{\varphi}$. **P6.2** **a)** -240; **b)** 900; **c)** $\frac{-1}{30}$; **d)** $\frac{27}{2}$. **P6.4** No. **P6.5** **a)** No; **b)** Yes; **c)** Yes. **P6.6** $M_T = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$. Eigenvalues: $\lambda = 1$ with multiplicity one, and $\lambda = 0$ with multiplicity two. **P6.7** $D = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}$. **P6.8** $\langle L_0(x), L_1(x)\rangle = 0$, $\langle L_0(x), L_2(x)\rangle = 0$, and $\langle L_1(x), L_2(x)\rangle = 0$. **P6.9** $\hat{e}_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, $\hat{e}_2 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. **P6.10** $A = Q\Lambda Q^{-1} = \begin{bmatrix} 2 & 0 & -5 \\ 0 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} -3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$. **P6.11** $A+B = \begin{bmatrix} 4 & 2 \\ 8+3i & -5+3i \end{bmatrix}$; $CB = \begin{bmatrix} 3+8i & 2-i \\ 45+3i & -33+31i \\ 16-4i & 16+8i \end{bmatrix}$; $(2+i)B = \begin{bmatrix} 5 & 8-i \\ 9+7i & -15+5i \end{bmatrix}$.

Solutions to selected problems

P6.1 The vector $\vec{e}_1$ is an eigenvector of A because $A\vec{e}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ \frac{1}{\varphi} \end{bmatrix} = \begin{bmatrix} 1+\frac{1}{\varphi} \\ 1 \end{bmatrix}$. Now observe the following interesting fact: $\frac{1}{\varphi}(1 + \frac{1}{\varphi}) = \frac{1}{\varphi} + \frac{1}{\varphi^2} = \frac{\varphi+1}{\varphi^2} = 1$.

This means we can write $A\vec{e}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ \frac{1}{\varphi} \end{bmatrix} = \varphi\begin{bmatrix} 1 \\ \frac{1}{\varphi} \end{bmatrix}$, which shows that $\vec{e}_1$ is an eigenvector of A and it corresponds to eigenvalue $\lambda_1 = \varphi$. Similar reasoning shows $A\vec{e}_2 = -\frac{1}{\varphi}\vec{e}_2$ so $\vec{e}_2$ is an eigenvector of A that corresponds to eigenvalue $\lambda_2 = -\frac{1}{\varphi}$.

P6.2 Use the facts that $\text{Tr}(A) = \lambda_1 + \lambda_2 + \lambda_3$ and $\det(A) = \lambda_1\lambda_2\lambda_3$, and the properties of trace and determinant (page 45) to compute the expressions.

P6.3 Show this directly: $\vec{u}^{\mathsf{T}}(A + B)\vec{u} = \vec{u}^{\mathsf{T}}A\vec{u} + \vec{u}^{\mathsf{T}}B\vec{u} \geq 0$, for all $\vec{u}$.

P6.4 A vector space would obey $0 \cdot (a_1, a_2) = (0, 0)$ (the zero vector), but we have $0 \cdot (a_1, a_2) = (0, a_2) \neq (0, 0)$, so $(V, \mathbb{R}, +, \cdot)$ is not a vector space.

P6.5 A subspace of $\mathbb{R}^3$ must be closed under addition and scalar multiplication, and must contain the zero element.

P6.6 Given an arbitrary input $\vec{v} = v_0 + v_1x + v_2x^2$, the effect of T is to select the quadratic term, so it corresponds to a projection matrix onto the third dimension. The eigenvalue $\lambda = 1$ corresponds to the subspace spanned by $(0, 0, 1)$. The eigenvalue $\lambda = 0$ corresponds to the two-dimensional subspace spanned by $(1, 0, 0)$ and $(0, 1, 0)$.

P6.9 We're given the vectors $\vec{v}_1 = (1, 1)$ and $\vec{v}_2 = (0, 1)$ and want to perform the Gram–Schmidt procedure. We pick $\vec{e}_1 = \vec{v}_1 = (1, 1)$, and after normalization we have $\hat{e}_1 = (\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. Next we compute $\vec{e}_2 = \vec{v}_2 - \Pi_{\hat{e}_1}(\vec{v}_2) = \vec{v}_2 - (\hat{e}_1 \cdot \vec{v}_2)\hat{e}_1 = (\frac{-1}{2}, \frac{1}{2})$. Normalizing $\vec{e}_2$ we obtain $\hat{e}_2 = (-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$.

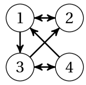
P6.10 First find the eigenvalues of the matrix. Then find an eigenvector for each eigenvalue and construct a matrix Q composed of the three eigenvectors. Compute Q^{-1} and write the eigendecomposition as $A = Q\Lambda Q^{-1}$.

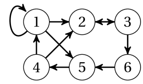
Chapter 7 solutions

Answers to exercises

Solutions to selected exercises

Answers to problems

P7.1 $C_{55}H_{104}O_6 + 78O_2 \;\rightarrow\; 55CO_2 + 52H_2O$. **P7.2** $I_1 = 5$[A] and $I_5 = 20$[A]. **P7.3** $y = 1 + \frac{1}{2}x$. **P7.4** **a)** $A = \begin{bmatrix} 0&1&0&0&1 \\ 1&0&1&1&0 \\ 0&1&0&1&0 \\ 0&0&1&0&1 \\ 1&0&0&1&0 \end{bmatrix}$; **b)** $A = \begin{bmatrix} 0&1&0&0&1 \\ 0&0&1&0&1 \\ 0&0&0&1&1 \\ 1&0&0&0&1 \\ 0&0&0&0&0 \end{bmatrix}$; **c)** $A = \begin{bmatrix} 0&1&1&1&1 \\ 1&0&1&0&0 \\ 0&1&0&1&0 \\ 0&0&1&0&1 \\ 1&0&0&1&0 \end{bmatrix}$. **P7.5** **a)** 

; **b)** 

. **P7.6** Alice sent $\vec{m} = 01110011\ 01110101\ 01110000$. After ASCII decoding each byte of this binary string, we find it corresponds to the message "sup." **P7.7** $f(\omega) = \frac{1}{\sqrt{2\pi}}\frac{2|a|}{|a|^2+\omega^2}$.

Solutions to selected problems

P7.1 The reaction we're studying is the burning of triglycerides. Read more about that here: `http://bmj.com/content/349/bmj.g7257`.

P7.2 The KVL equation for the clockwise loop starting at junction B is $V_1 - R_1I_1 - R_2I_2 = 0$. The equation for the loop starting at C is $-R_3I_3 - R_4I_4 + R_2I_2 = 0$. The equation starting at D is $+R_5I_5 - V_2 + R_4I_4 = 0$. The KCL equation

for junction C is $I_1 = I_2 + I_3$ and that for D is $I_3 + I_5 = I_4$. After combining the equations, we obtain the matrix equation $R\vec{I} = \vec{V}$, where $\vec{V} = \begin{bmatrix} 15 \\ 0 \\ 10 \\ 0 \\ 0 \end{bmatrix}$ and $R = \begin{bmatrix} R_1 & R_2 & 0 & 0 & 0 \\ 0 & -R_2 & R_3 & R_4 & 0 \\ 0 & 0 & 0 & -R_4 & R_5 \\ 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}$. We get $R = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 4 & 2 & 0 \\ 0 & 0 & 0 & -2 & 2 \\ 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 \end{bmatrix}$ if we substitute the values provided. The solution is $\vec{I} = (5, 10, -5, 15, 20)^{\mathsf{T}}$.

P7.3 The dataset consists of a matrix of inputs x, $A = (0, 1, 2, 3)^{\mathsf{T}}$, and a vector of outputs y, $\vec{b} = (0.9, 1.6, 2.1, 2.4)^{\mathsf{T}}$. Since we're interested in fitting an affine model $y = b + mx$, we must augment the matrix A with a column of ones to obtain A', and then compute $A'^{\mathsf{T}}A'$ and its inverse:

$$A' = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad A'^{\mathsf{T}}A' = \begin{bmatrix} 4 & 6 \\ 6 & 14 \end{bmatrix} \quad \Rightarrow \quad (A'^{\mathsf{T}}A')^{-1} = \begin{bmatrix} \frac{7}{10} & -\frac{3}{10} \\ -\frac{3}{10} & \frac{1}{5} \end{bmatrix}.$$

We can now compute the Moore–Penrose pseudoinverse A'^{+} and obtain the approximate solution as follows:

$$A'^{+} = (A'^{\mathsf{T}}A')^{-1}A'^{\mathsf{T}} = \begin{bmatrix} \frac{7}{10} & \frac{2}{5} & \frac{1}{10} & -\frac{1}{5} \\ -\frac{3}{10} & -\frac{1}{10} & \frac{1}{10} & \frac{3}{10} \end{bmatrix} \quad \Rightarrow \quad A'^{+}\vec{b} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}.$$

Thus the best-fitting affine model for the dataset is $y = 1 + \frac{1}{2}x$.

P7.4 The i^{th} row of the adjacency matrix contains the information of the outgoing edges for vertex i in the graph. If the edge (i, j) exists, then $A_{ij} = 1$, otherwise $A_{ij} = 0$.

P7.5 The number of rows (columns) of the adjacency matrix tells us how many vertices the graph contains. The i^{th} row of the adjacency matrix contains the information of the outgoing edges for vertex i in the graph. If you see $A_{ij} = 1$ then you must draw the edge (i, j) in the graph.

P7.7 We start from the definition $f(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\omega t} e^{-|at|}\, dt$. Next we rewrite $e^{-i\omega t}$ as $(\cos(\omega t) - i\sin(\omega t))$ and observe that sine is an odd function so this term vanishes. The integral becomes $f(\omega) = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} \cos(\omega t) e^{-|at|}\, dt$, which we can tackle using integration by parts.

Chapter 8 solutions

Answers to exercises

Solutions to selected exercises

Answers to problems

P8.2 $\Pr(\{\texttt{heads}, \texttt{heads}, \texttt{heads}, \texttt{heads}\}) = \frac{1}{16}$. **P8.4** $\mathbb{E}_X[W] = \$\frac{100}{101} < \$1$. Not worth it. **P8.5** $\mathbb{E}_X[f(X)] = \frac{\$7}{6} > \$1$ so it's worth playing this game. **P8.6** $p_{X_1} = (\frac{1}{2}, \frac{1}{2})^{\mathsf{T}}$, $p_{X_2} = (\frac{3}{8}, \frac{5}{8})^{\mathsf{T}}$, $p_{X_\infty} = (\frac{1}{3}, \frac{2}{3})^{\mathsf{T}}$. **P8.7** $p_{X_\infty} = (0.1721, 0.169, 0.245, 0.245, 0.169)^{\mathsf{T}}$.

Solutions to selected problems

P8.1 Assume, for a contradiction, that $p_1 > 1$. We know from Kolmogorov's axioms that $p_i \geq 0$ for all i. Observe that $\sum_i p_i = p_1 + p_2 + p_3 \geq p_1 > 1$, which means the vector (p_1, p_2, p_3) is not a valid probability distribution. Therefore, it must be that $p_1 > 1$ is false, and $p_1 \leq 1$ is true.

P8.2 Substitute $p = \frac{1}{2}$ and $n = 4$ into the expression p^n.

P8.3 The biased coin flip is modelled by a random variable Y, and different coin flips correspond to random variables Y_1, Y_2, Y_3, ...which are independent copies of Y. The probability of getting `heads` on the first flip is $P_N(1) = \Pr(\{Y_1 = \texttt{heads}\}) = p$. The probability of getting `heads` on the second flip corresponds to the event $\{Y_1 = \texttt{tails}\}$ AND $\{Y_2 = \texttt{heads}\}$. We assumed the coin flips are independent, so $P_N(2) = (1-p)p$. Similarly $P_N(3) = (1-p)^2 p$. The general formula is $P_N(n) = (1-p)^{n-1}p$.

P8.4 We'll model each spin of the roulette wheel as a random variable X with sample space {red, black, green} and probability distribution $p_X = (\frac{50}{101}, \frac{50}{101}, \frac{1}{101})$. When placing a bet on black, the payout is $W = \$2$ if the outcome is black, and zero for other outcomes. The expected value of the payout is $\mathbb{E}_X[W] = 0 \cdot \frac{50}{101} + \$2 \cdot \frac{50}{101} + 0 \cdot \frac{1}{101} = \$\frac{100}{101}$. Since $\frac{100}{101} < 1$, the house has an advantage, so the mathematician shouldn't play.

P8.5 The payout function for this game is defined as follows:

$$f(⚀) = f(⚁) = f(⚂) = \$0, \quad f(⚃) = f(⚄) = \$1, \quad f(⚅) = \$5.$$

The die is described by the distribution $p_X = (\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6})^{\mathsf{T}}$. The expected payout is $\mathbb{E}_X[f(X)] = \sum_x f(x)\, p_X(x) = \frac{\$1+\$1+\$5}{6} = \frac{\$7}{6} = \$1.1\overline{6}$. The expected payout is greater than the cost to play, so you'll win on average.

P8.6 Define X_i to be the probability distribution of the weather in Year i. The transition matrix for the weather Markov chain is $M = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{3}{4} \end{bmatrix}$. We obtain the weather in Year 1 using $p_{X_1} = M(1,0)^{\mathsf{T}} = (\frac{1}{2}, \frac{1}{2})^{\mathsf{T}}$. The weather in Year 2 is $p_{X_2} = M^2(1,0)^{\mathsf{T}} = (\frac{3}{8}, \frac{5}{8})^{\mathsf{T}}$. The long term (stationary) distribution is $p_{X_\infty} = M^\infty(1,0)^{\mathsf{T}} = (\frac{1}{3}, \frac{2}{3})^{\mathsf{T}}$.

P8.7 Construct M_1 by counting the outbound links for each webpage, then mix in 0.9 of it with 0.1 of $\frac{1}{5}\mathbb{J}$ to obtain the Markov chain matrix M:

$$M_1 = \begin{bmatrix} 0 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & 0 \end{bmatrix}, \quad M_2 = \frac{1}{5}\mathbb{J}_5, \quad M = \begin{bmatrix} \frac{1}{50} & \frac{47}{100} & \frac{1}{50} & \frac{1}{50} & \frac{47}{100} \\ \frac{49}{200} & \frac{1}{50} & \frac{47}{100} & \frac{1}{50} & \frac{1}{50} \\ \frac{49}{200} & \frac{47}{100} & \frac{1}{50} & \frac{47}{100} & \frac{1}{50} \\ \frac{49}{200} & \frac{1}{50} & \frac{47}{100} & \frac{1}{50} & \frac{47}{100} \\ \frac{49}{200} & \frac{1}{50} & \frac{1}{50} & \frac{47}{100} & \frac{1}{50} \end{bmatrix}.$$

Solving $(M - \mathbb{1})\,\vec{e} = \vec{0}$, we find $p_{X_\infty} = (0.1721, 0.169, 0.245, 0.245, 0.169)^{\mathsf{T}}$.

Chapter 9 solutions

Answers to exercises

E9.1 $M_{\Pi_+} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ and $M_{\Pi_-} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$. **E9.2** $\Pr(\{-\}|\psi) = \frac{|\alpha-\beta|^2}{2}$. **E9.3** $|\theta^\perp\rangle = (\frac{1}{\sqrt{2}}, \frac{e^{-i(\pi-\theta)}}{\sqrt{2}})^{\mathsf{T}}$. $M_{\Pi_\theta} = \begin{bmatrix} \frac{1}{2} & \frac{e^{-i\theta}}{2} \\ \frac{e^{i\theta}}{2} & \frac{1}{2} \end{bmatrix}$ and $M_{\Pi_{\theta^\perp}} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2}e^{i(\pi-\theta)} \\ \frac{1}{2}e^{-i(\pi-\theta)} & \frac{1}{2} \end{bmatrix}$. $\Pr(\{-\}|\psi) = \sqrt{\left|\frac{\alpha}{2} + \frac{\beta}{2}e^{-i\theta}\right|^2 + \left|\frac{\alpha}{2}e^{i\theta} + \frac{\beta}{2}\right|^2}$. **E9.4** $|\psi\rangle \otimes |\phi\rangle = \alpha\gamma|00\rangle + \alpha\delta|01\rangle + \beta\gamma|10\rangle + \beta\delta|11\rangle$.

Solutions to selected exercises

Answers to problems

P9.1 No, since Q is not unitary. **P9.2** $YY|0\rangle = |0\rangle$ and $YY|1\rangle = |1\rangle$. **P9.3** $HXH = Z$ and $HZH = X$. **P9.4** $\langle\psi_1, \psi_2\rangle = 0$. **P9.5** $\Pr(\{0 \leq x \leq \frac{1}{4}\}|\psi_a) = \frac{7}{16}$; $\Pr(\{0 \leq x \leq \frac{1}{4}\}|\psi_b) = \frac{79}{256}$; $\Pr(\{0 \leq x \leq \frac{1}{4}\}|\psi_c) = \frac{1}{4}$.

Solutions to selected problems

P9.1 Quantum gates correspond to unitary operators. Since $Q^\dagger Q \neq \mathbb{1}$, Q is not unitary and it cannot be implemented by any physical device. The boss is *not* always right!

P9.2 Let's first see what happens to $|0\rangle$ when we apply the operator YY. The result of the first Y applied to $|0\rangle$ is $Y|0\rangle = i|1\rangle$. Then, applying the second Y operator, we get $YY|0\rangle = Y(i|1\rangle) = iY|1\rangle = i(-i)|0\rangle = |0\rangle$. So $YY|0\rangle = |0\rangle$. A similar calculation shows that $YY|1\rangle = |1\rangle$.

P9.4 The inner product $\langle\psi_1, \psi_2\rangle = \int_0^1 \overline{\psi_1(x)}\psi_2(x)\,dx$ corresponds to the integral $\int_0^1 (2x-1)(6x^2-6x+1)\,dx = \int_0^1 (12x^3 - 18x^2 + 8x - 1)\,dx$. Using the formula $\int x^n\,dx = \frac{x^{n+1}}{n+1}$, we find $\langle\psi_1, \psi_2\rangle = \left[\frac{12}{4}x^4 - \frac{18}{3}x^3 + \frac{8}{2}x^2 - x\right]_0^1 = 0$. The functions $\psi_1(x)$ and $\psi_2(x)$ are called the *Shifted Legendre polynomials*.

P9.5 You can perform the required integrals by hand or use the `SymPy` command `integrate((sqrt(3)*(2*x-1))**2,(x,0,1/4))` for part **a)**, and the command `integrate((sqrt(5)*(6*x**2-6*x+1))**2,(x,0,1/4))` for part **b)**. For part **c)**, the constant wave function is $\psi_c(x) = 1$, and the interval $[0, \frac{1}{4}]$ contains exactly $\frac{1}{4}$ of its probability mass.

Appendix B

Notation

This appendix contains a summary of the notation used in this book.

Math notation

Expression	Read as	Used to denote
a, b, x, y		variables
$=$	is equal to	expressions that have the same value
$\equiv$	is defined as	new variable definitions
$a+b$	a plus b	the combined lengths of a and b
$a-b$	a minus b	the difference in lengths between a and b
$a \times b \equiv ab$	a times b	the area of a rectangle
$a^2 \equiv aa$	a squared	the area of a square of side length a
$a^3 \equiv aaa$	a cubed	the volume of a cube of side length a
a^n	a exponent n	a multiplied by itself n times
$\sqrt{a} \equiv a^{\frac{1}{2}}$	square root of a	the side length of a square of area a
$\sqrt[3]{a} \equiv a^{\frac{1}{3}}$	cube root of a	the side length of a cube with volume a
$a/b \equiv \frac{a}{b}$	a divided by b	a parts of a whole split into b parts
$a^{-1} \equiv \frac{1}{a}$	one over a	division by a
$f(x)$	f of x	the function f applied to input x
f^{-1}	f inverse	the inverse function of $f(x)$
$f \circ g$	f compose g	function composition; $f \circ g(x) \equiv f(g(x))$
e^x	e to the x	the exponential function base e
$\ln(x)$	natural log of x	the logarithm base e
a^x	a to the x	the exponential function base a
$\log_a(x)$	log base a of x	the logarithm base a
θ, ϕ	*theta*, *phi*	angles
$\sin, \cos, \tan$	sin, cos, tan	trigonometric ratios
$\%$	percent	proportions of a total; $a\% \equiv \frac{a}{100}$

Set notation

You don't need a lot of fancy notation to do math, but it really helps if you know a little bit of set notation.

Symbol	Read as	Denotes
$\{\ \dots\ \}$	the set ...	definition of a set
$\vert$	such that	describe or restrict the elements of a set
$\mathbb{N}$	the naturals	the set $\mathbb{N} \equiv \{0, 1, 2, \dots\}$. Also $\mathbb{N}_+ \equiv \mathbb{N} \backslash \{0\}$.
$\mathbb{Z}$	the integers	the set $\mathbb{Z} \equiv \{\dots, -2, -1, 0, 1, 2, 3, \dots\}$
$\mathbb{Q}$	the rationals	the set of fractions of integers
$\mathbb{R}$	the reals	the set of real numbers
$\mathbb{C}$		the set of complex numbers
$\mathbb{F}_q$	finite field	the set $\{0, 1, 2, 3, \dots, q-1\}$
$\subset$	subset	one set strictly contained in another
$\subseteq$	subset or equal	containment or equality
$\cup$	union	the combined elements from two sets
$\cap$	intersection	the elements two sets have in common
$S \setminus T$	S set minus T	the elements of S that are not in T
$a \in S$	a in S	a is an element of set S
$a \notin S$	a not in S	a is not an element of set S
$\forall x$	for all x	a statement that holds for all x
$\exists x$	there exists x	an existence statement
$\nexists x$	there doesn't exist x	a non-existence statement
$S \times T$	Cartesian product	all pairs (s, t) where $s \in S$ and $t \in T$

An example of a condensed math statement that uses set notation is "$\nexists m, n \in \mathbb{Z}$ such that $\frac{m}{n} = \sqrt{2}$," which reads "there don't exist integers m and n whose fraction equals $\sqrt{2}$." Since we identify the set of fractions of integers with the rationals, this statement is equivalent to the shorter "$\sqrt{2} \notin \mathbb{Q}$," which reads "$\sqrt{2}$ is irrational."

Vectors notation

Expression	Denotes
$\mathbb{R}^n$	the set of n-dimensional real vectors
$\vec{v}$	a vector
(v_x, v_y)	vector in component notation
$v_x\hat{\imath} + v_y\hat{\jmath}$	vector in unit vector notation
$\\|\vec{v}\\|\angle\theta$	vector in length-and-direction notation
$\\|\vec{v}\\|$	length of the vector $\vec{v}$
θ	angle the vector $\vec{v}$ makes with the x-axis
$\hat{v} \equiv \frac{\vec{v}}{\\|\vec{v}\\|}$	unit vector in the same direction as $\vec{v}$
$\vec{u} \cdot \vec{v}$	dot product of the vectors $\vec{u}$ and $\vec{v}$
$\vec{u} \times \vec{v}$	cross product of the vectors $\vec{u}$ and $\vec{v}$

Complex numbers notation

Expression	Denotes
$\mathbb{C}$	the set of complex numbers $\mathbb{C} \equiv \{a+bi \mid a,b \in \mathbb{R}\}$
i	the unit imaginary number $i \equiv \sqrt{-1}$ or $i^2 = -1$
$\mathrm{Re}\{z\} = a$	real part of $z = a + bi$
$\mathrm{Im}\{z\} = b$	imaginary part of $z = a + bi$
$\|z\|\angle\varphi_z$	polar representation of $z = \|z\|\cos\varphi_z + i\|z\|\sin\varphi_z$
$\|z\| = \sqrt{a^2+b^2}$	magnitude of $z = a + bi$
$\varphi_z = \tan^{-1}(b/a)$	phase or argument of $z = a + bi$
$\bar{z} = a - bi$	complex conjugate of $z = a + bi$
$\mathbb{C}^n$	the set of n-dimensional complex vectors

Vector space notation

Expression	Denotes
U, V, W	vector spaces
$W \subseteq V$	vector space W subspace of vector space V
$\{\vec{v} \in V \mid \langle\texttt{cond}\rangle\}$	subspace of vectors in V satisfying condition $\langle\texttt{cond}\rangle$
$\mathrm{span}(\vec{v}_1, \ldots, \vec{v}_n)$	span of vectors $\vec{v}_1, \ldots, \vec{v}_n$
$\dim(U)$	dimension of vector space U
$\mathcal{R}(M)$	row space of M
$\mathcal{N}(M)$	null space of M
$\mathcal{C}(M)$	column space of M
$\mathcal{N}(M^{\mathsf{T}})$	left null space of M
$\mathrm{rank}(M)$	rank of M; $\mathrm{rank}(M) \equiv \dim(\mathcal{R}(M)) = \dim(\mathcal{C}(M))$
$\mathrm{nullity}(M)$	nullity of M; $\mathrm{nullity}(M) \equiv \dim(\mathcal{N}(M))$
B_s	the standard basis
$\{\vec{e}_1, \ldots, \vec{e}_n\}$	an orthogonal basis
$\{\hat{e}_1, \ldots, \hat{e}_n\}$	an orthonormal basis
${}_{B'}[\mathbb{1}]_B$	the change-of-basis matrix from basis B to basis B'
Π_S	projection onto subspace S
$\Pi_{S^\perp}$	projection onto the orthogonal complement of S

Abstract vector spaces notation

Expression	Denotes
$(V, F, +, \cdot)$	abstract vector space of vectors from the set V, whose coefficients are from the field F, addition operation "+" and scalar-multiplication operation "$\cdot$"
$\mathbf{u}, \mathbf{v}, \mathbf{w}$	abstract vectors
$\langle \mathbf{u}, \mathbf{v} \rangle$	inner product of vectors $\mathbf{u}$ and $\mathbf{v}$
$\\|\mathbf{u}\\|$	norm of $\mathbf{u}$
$d(\mathbf{u}, \mathbf{v})$	distance between $\mathbf{u}$ and $\mathbf{v}$

Notation for matrices and matrix operations

Expression	Denotes
$\mathbb{R}^{m\times n}$	the set of $m \times n$ matrices with real coefficients
A	a matrix
a_{ij}	entry in the i^{th} row and j^{th} column of A
$\|A\|$	determinant of A, also denoted $\det(A)$
A^{-1}	matrix inverse
A^{T}	matrix transpose
$\mathbb{1}$	identity matrix; $\mathbb{1}A = A\mathbb{1} = A$ and $\mathbb{1}\vec{v} = \vec{v}$
AB	matrix-matrix product
$A\vec{v}$	matrix-vector product
$\vec{w}^{\mathsf{T}}A$	vector-matrix product
$\vec{u}^{\mathsf{T}}\vec{v}$	vector-vector inner product; $\vec{u}^{\mathsf{T}}\vec{v} \equiv \vec{u} \cdot \vec{v}$
$\vec{u}\vec{v}^{\mathsf{T}}$	vector-vector outer product
$\text{ref}(A)$	row echelon form of A
$\text{rref}(A)$	reduced row echelon form of A
$\text{rank}(A)$	rank of $A \equiv$ number of pivots in $\text{rref}(A)$
$A \sim A'$	matrix A' obtained from matrix A by row operations
$\mathcal{R}_1, \mathcal{R}_2, \ldots$	row operations, of which there are three types:
	$\rightarrow$ $R_i \leftarrow R_i + kR_j$: add k-times row j to row i
	$\rightarrow$ $R_i \leftrightarrow R_j$: swap rows i and j
	$\rightarrow$ $R_i \leftarrow mR_i$: multiply row i by constant m
$E_{\mathcal{R}}$	elementary matrix for row operation $\mathcal{R}$; $\mathcal{R}(M) \equiv E_{\mathcal{R}}M$
$[\,A \mid \vec{b}\,]$	augmented matrix containing matrix A and vector $\vec{b}$
$[\,A \mid B\,]$	augmented matrix array containing matrices A and B
M_{ij}	minor associated with entry a_{ij}. See page 61.
$\text{adj}(A)$	adjugate matrix of A. See page **??**.
$(A^{\mathsf{T}}A)^{-1}A^{\mathsf{T}}$	Moore–Penrose pseudoinverse of A. See page 95.
$\mathbb{C}^{m\times n}$	the set of $m \times n$ matrices with complex coefficients
$A^{\dagger}$	Hermitian transpose; $A^{\dagger} \equiv (\overline{A})^{\mathsf{T}}$

Notation for linear transformations

Expression	Denotes
$T : \mathbb{R}^n \to \mathbb{R}^m$	linear transformation T
$M_T \in \mathbb{R}^{m \times n}$	matrix representation of T
$\text{Dom}(T) \equiv \mathbb{R}^n$	domain of T
$\text{CoDom}(T) \equiv \mathbb{R}^m$	codomain of T
$\text{Im}(T) \equiv \mathcal{C}(M_T)$	the image space of T
$\text{Ker}(T) \equiv \mathcal{N}(M_T)$	the kernel of T
$S \circ T(\vec{x})$	composition of linear transformations; $S \circ T(\vec{x}) \equiv S(T(\vec{x})) \equiv M_S M_T \vec{x}$
$M \in \mathbb{R}^{m \times n}$	an $m \times n$ matrix
$T_M : \mathbb{R}^n \to \mathbb{R}^m$	the linear transformation defined as $T_M(\vec{v}) \equiv M\vec{v}$
$T_{M^\mathsf{T}} : \mathbb{R}^m \to \mathbb{R}^n$	the adjoint linear transformation $T_{M^\mathsf{T}}(\vec{a}) \equiv \vec{a}^\mathsf{T} M$

Matrix decompositions

Expression	Denotes
$A \in \mathbb{R}^{n \times n}$	a matrix (assume diagonalizable)
$p_A(\lambda) \equiv \lvert A - \lambda \mathbb{1} \rvert$	characteristic polynomial of A
$\lambda_1, \ldots, \lambda_n$	eigenvalues of A = roots of $p_A(\lambda) \equiv \prod_{i=1}^n (\lambda - \lambda_i)$
$\Lambda \in \mathbb{R}^{n \times n}$	diagonal matrix of eigenvalues of A
$\vec{e}_{\lambda_1}, \ldots, \vec{e}_{\lambda_n}$	eigenvectors of A
$Q \in \mathbb{R}^{n \times n}$	matrix whose columns are eigenvectors of A
$A = Q \Lambda Q^{-1}$	eigendecomposition of A
$A = O \Lambda O^\mathsf{T}$	eigendecomposition of a normal matrix
$B \in \mathbb{R}^{m \times n}$	a generic matrix
$\sigma_1, \sigma_2, \ldots$	singular values of B
$\Sigma \in \mathbb{R}^{m \times n}$	matrix of singular values of B
$\vec{u}_1, \ldots, \vec{u}_m$	left singular vectors of B
$U \in \mathbb{R}^{m \times m}$	matrix of left singular vectors of B
$\vec{v}_1, \ldots, \vec{v}_n$	right singular vectors of B
$V \in \mathbb{R}^{n \times n}$	matrix of right singular vectors of B
$B = U \Sigma V^\mathsf{T}$	singular value decomposition of B