# No bullshit guide to linear algebra

Ivan Savov

March 1, 2014

Minireference Publishing. Pre-release v0.31
ISBN 978-0-9920010-1-8
The author can be reached at `ivan.savov@gmail.com`.
hg changeset: 31:9f656a9b42b1.

# Contents


**Introduction** **1**

**1 Math fundamentals** **1**

**2 Vectors** **3**

**3 Intro to linear algebra** **5**
3.1 Introduction . . . . . . . . . . 5
3.2 Review of vector operations . . . . . . . . . . 12
3.3 Matrix operations . . . . . . . . . . 15
3.4 Linearity . . . . . . . . . . 19

**4 Computational linear algebra** **23**
4.1 Reduced row echelon form . . . . . . . . . . 24
4.2 Matrix equations . . . . . . . . . . 31
4.3 Matrix multiplication . . . . . . . . . . 34
4.4 Determinants . . . . . . . . . . 37
4.5 Matrix inverse . . . . . . . . . . 44

**5 Geometrical aspects of linear algebra** **51**
5.1 Lines and planes . . . . . . . . . . 51
5.2 Projections . . . . . . . . . . 56
5.3 Coordinate projections . . . . . . . . . . 60
5.4 Vector spaces . . . . . . . . . . 63
5.5 Vector space techniques . . . . . . . . . . 67

**6 Linear transformations** **75**
6.1 Linear transformations . . . . . . . . . . 75
6.2 Finding matrix representations . . . . . . . . . . 83
6.3 Change of basis for matrices . . . . . . . . . . 91
6.4 Invertible matrix theorem . . . . . . . . . . 94

**7 Theoretical linear algebra** **97**
7.1 Eigenvalues and eigenvectors . . . . . . . . . . 97
7.2 Special types of matrices . . . . . . . . . . 107
7.3 Abstract vector spaces . . . . . . . . . . 111
7.4 Abstract inner product spaces . . . . . . . . . . 114
7.5 Gram–Schmidt orthogonalization . . . . . . . . . . 119
7.6 Matrix decompositions . . . . . . . . . . 121
7.7 Linear algebra with complex numbers . . . . . . . . . . 125

**8 Applications** **137**
8.1 Solving systems of equations . . . . . . . . . . 137
8.2 Linear programming . . . . . . . . . . 139
8.3 Least squares approximate solution . . . . . . . . . . 140
8.4 Error correcting codes . . . . . . . . . . 141

**End matter** **143**
Links . . . 143

**A Notation** **145**
Math notation . . . 145
Set notation . . . 145
Vectors notation . . . 146
Matrix notation . . . 146

**B Intro to quantum physics** **147**
B.1 Introduction to quantum information . . . 148
B.2 Dirac notation . . . 149
B.3 Change of basis . . . 150
B.4 Superoperators = operators on matrices . . . 151
B.5 Quantum states . . . 151
B.6 Composite quantum systems . . . 155
B.7 Lasers demonstration . . . 156
B.8 Quantum information theory . . . 157
B.9 Quantum computing . . . 157

# Concept map

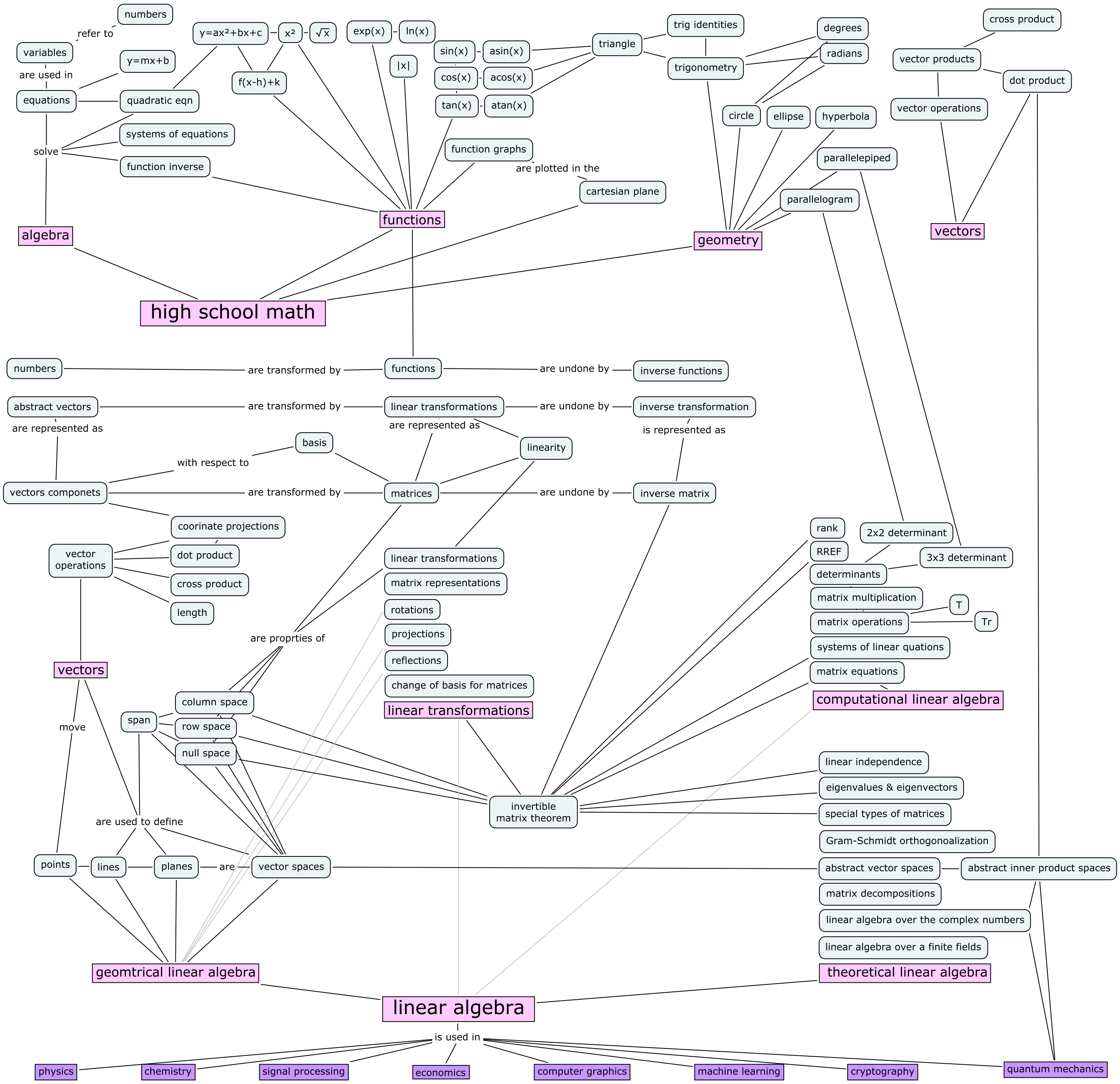



**Figure 1:** Each concept in this diagram corresponds to one section in the book.

# Preface

This book is about linear algebra and its applications. The coverage is at a level appropriate for a first-year university course, but more advanced concepts are also presented when understanding the bigger picture shows interesting parallel structure between concepts. This is not a popular science book based on simplified analogies. This is a proper textbook that covers the real stuff and all the complex stuff too. Each section is a self contained-tutorial on one topic, written in a clean, approachable style.

The teaching style puts particular emphasis on explaining the connections between concepts. It is much easier to learn new concepts when you see how they "attach" to things you already know. The best part is that, you'll feel a lot of knowledge buzz if you choose to follow the path of understanding, instead of trying to memorize formulas and blindly follow procedures.

The tools of linear algebra open the gateway to the study of more advanced areas of mathematics. Many problems in science, business, and technology can be described using techniques from linear algebra so it is important that you learn about this subject.

## Why this book?

The genesis of the No bullshit guide series of textbooks dates back to my student days when I was required to purchase expensive textbooks for my courses, which were long and tedious to read. I said to myself, "something must be done," so I started writing books on math and physics that explain concepts clearly, concisely, and affordably:

After finishing my studies, I started the `Minireference Co.` to help other authors find their audience of interested readers. We're taking over the textbook publishing industry.

> "First we take Manhattan then we take Berlin."
> —Leonard Cohen

Print-on-demand and digital distribution technologies, enable a new shorter, author-centred publishing value chain:

$$\text{author} \quad \to \quad \text{print shop} \quad \to \quad \text{reader}.$$

Removing all the middlemen from the value chain allows us to offer good prices for the readers and reasonable margins for authors.

## What's in this book?

Each section in this book is a *self-contained tutorial* that covers the definitions, formulas, and explanations associated with a single topic. Consult the concept map in the front of the book to see the connections between the concepts and feel free to read the chapters in any order you find logical.

To learn linear algebra, you need to know your high school math and it helps if you've seen vectors before. In order to make the book accessible for all readers, the book begins with a review chapter on numbers, algebra, equations, functions, and trigonometry (Chapter 1) and a review chapter on vectors and vector operations (Chapter 2). If you feel a little rusty on those concepts, be sure to read through these chapters.

Readers who feel confident in their high school math abilities, can jump straight to Chapter 3 where the fun begins.

## Is this book for you?

This book is intended for first-year university students taking a linear algebra class, You can learn everything you need to know to pass a final exam. Don't be fooled by the small size of the book: everything is here. The book is small because I've removed all the unnecessary parts. Speaking of unnecessary things, your fear of midterms and finals is totally unnecessary. How hard could it be? If you understand the material from this book, the teacher's got nothing on you!

The quick pace of the explanations makes for interesting reading even for non-students. Whether you're learning linear algebra to help your children, reviewing linear algebra as a prerequisite for some more advanced course, or generally curious about the subject, this guide will help you find your way in the land of linear algebra. The short tutorial format intended for students with short attention spans is also well suited for adults learners who don't have time to waste.

Those with an analytical mind will have a good time reading this book because linear algebra is one of the deepest subjects in mathematics. The study of linear algebra comes with a number of mind-expanding experiences. Additionally, understanding the basic language of vectors and matrices will allow you to learn about many other subjects.

## About the author

I completed my undergraduate studies at McGill University in electrical engineering, then did a M.Sc. in physics, and recently completed a Ph.D. in computer science. Linear algebra played a central role throughout my researcher career.

I have been teaching math and physics for more than 12 years as a private tutor. As a tutor I learned to explain concepts that are difficult to understand. Over time, I figured how to brake down complicated ideas into smaller chunks. An interesting feedback loop occurs when students learn concepts in small chunks: the *knowledge buzz* they experience when concepts start to connect in their heads motivates them to keep on reading.

With this book, I want to share with you some of the things I've learned about linear algebra. I hope to convince you that reading a math textbook can be fun. Let me know if I've succeeded.

Ivan Savov
Montreal, 2014

# Introduction

A key role in the day-to-day occupations of scientists and engineers is to build mathematical models of the real world. A significant proportion of these models describe linear relationships between quantities. A function $f$ is *linear* if it obeys the equation

$$f(ax_1 + bx_2) = af(x_1) + bf(x_2).$$

Functions that do not obey this property are called *nonlinear*. Most real processes and phenomena of science are described by nonlinear equations. Why are scientists, engineers, statisticians, business folk, and politicians so concentrated on developing and using linear models if the real world is nonlinear?

There are several good reasons for using linear models to model nonlinear phenomena. The first reason is that linear models are very good at *approximating* the real world. Linear models for nonlinear phenomena are also referred to as *linear approximations*. The multivariable equivalent of the *tangent line* approximation is a multivariable linear function. A second reason is that we can "outsource nonlinearity" by combining a linear model with nonlinear transformations of the inputs and/or outputs.

Perhaps the main reason why linear models are so 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 by analyzing the behaviour of the system for very few inputs. Let's illustrate with an example.

**Example** At an art event, you enter a room with a multimedia setup. The contents of a drawing canvas on a tablet computer are projected on a giant screen. Anything drawn on the tablet will instantly appear on the screen. The user interface on the tablet screen doesn't give any indication about how to hold the tablet "right side up." What is the fastest way to find the correct orientation of the tablet so your drawing will not appear rotated or upside-down?

This situation is directly analogous to the task scientists face every day when trying to model real-world systems. The canvas on the tablet describes a two-dimensional *input space*, the wall projection is a two-dimensional *output space*. We're looking for the unknown transformation $T$ that maps the pixels of the input space to coloured dots on the wall.

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 by a pair of coordinates $(x, y)$ and each point on the wall by another pair of coordinates $(x', y')$. The unknown transformation transforms pixels to wall coordinates $T(x, y) = (x', y')$. To understand how $T$ transforms $(x, y)$-coordinates to $(x', y')$-coordinates, proceed as follows. First put a dot in the lower left corner of the tablet to represent the *origin* $(0, 0)$ of the $xy$-coordinate system. Observe where the dot appears on the screen; we'll call this the origin of the

$x'y'$-coordinate system. Next, make a short horizontal swipe on the screen to represent the $x$-direction $(1,0)$ and observe the transformed $T(1,0)$ that appears on the wall screen. The third and final step is to make a vertical swipe in the $y$-direction $(0,1)$ and see the transformed $T(0,1)$ that appears on the wall screen. **By knowing how the origin, the $x$-direction, and the $y$-direction get mapped by the transformation $T$, you know $T$ completely.**

By seeing how the $xy$-coordinate system gets mapped to the wall screen you will be able to figure out what orientation you must hold the tablet so your drawing appears upright. There is a deeper, mathematical sense in which your knowledge of $T$ is *complete*. Rotations and reflections are linear transformations, and it is precisely the linearity property that allows us to completely understand the linear transformation $T$ with only two swipes.

Can you predict what will appear on the wall if you make an angled swipe in the $(2,3)$-direction? Observe the point $(2,3)$ in the input space can be obtained by moving 2 units in the $x$-direction and 3 units in the $y$-direction: $(2,3) = (2,0) + (0,3) = 2(1,0) + 3(0,1)$. 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 wall projection of the diagonal swipe in the $(2,3)$-direction will appear at a length equal to 2 times the $x$-direction output $T(1,0)$ plus 3 times the $y$-direction output $T(0,1)$. Knowledge of the transformed directions $T(1,0)$ and $T(0,1)$ is sufficient to figure out the output of the transformation for any input $(a,b)$ since the input can be written as a linear combination $(a,b) = a(1,0) + b(0,1)$.

Linearity allows us to study complicated, multidimensional processes and transformations by studying their effects on a very small set of inputs. This is the essential reason why linear models are so prominent in science. If the system we are styling is linear, the probing it with each "input direction" is enough to characterize it completely. Without this structure, characterizing an unknown system would be a much harder task.

## Why learn linear algebra?

Linear algebra is one of the coolest undergraduate math subjects. Students don't necessarily know it when they take the class, but the practical skills of manipulating vectors will come in handy for physics, computer graphics, statistics, machine learning, information theory, quantum mechanics, and many other areas of science. Linear algebra is very important if you want to do stuff in science.

The second reason why you should learn linear algebra is because the mind-expanding conceptual ideas of linear algebra form a bridge toward more advanced mathematical topics. You can think of certain properties of matrices, polynomials, and other functions as abstract vectors in a vector space. The techniques of linear algebra can be used not only on regular vectors, but on *all* mathematical objects that are vector-like!

The abstract modelling skills that students pick up in linear algebra are similar to the abstract modelling powers students learn in mechanics, which is another often misunderstood undergraduate course. Learning mechanics is important not because it teaches students to simulate the flight of a ball in the air, but because it teaches students how to start from a set of first principles (the laws of physics) and use mathematics to model the real world, and predict the future for *all* situations[1].

## Prerequisites

Understanding linear algebra requires that students have preliminary knowledge of fundamental math concepts like the real numbers and functions of a real variable. 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 specially designed to help you quickly come up to speed on the material from high school math.

It's not a requirement, but it helps if you've preciously used vectors in physics class. In case you haven't taken a mechanics course and seen velocities and forces represented as vectors then you should read Chapter 2 because it provides a short summary of the vectors concepts usually taught in the first week of Physics 101. The last section in the vectors chapter (Section **??**) is about complex numbers. You should definitely read that section at some point for your general culture, but we won't need complex numbers until much later in the book (Section 7.7).

## Executive summary

Chapter 3 is an introduction to the subject of linear algebra. Linear algebra is the math of vectors and matrices, so the first thing we'll do is define the mathematical operations we can do on vectors and matrices.

In Chapter 4, we'll tackle the computational aspects of linear algebra. By the end of this course, you're supposed to know how to: solve systems of equations, how to transform a matrix into its reduced row echelon form, how to compute the product of two matrices, and how to find the determinant and the inverse of a square matrix. Each of these computational tasks can be tedious to carry out by hand and requires lots of steps. There is no way around this; we must do the grunt work before we get to the advanced stuff.

In Chapter 5, we'll review the properties and equations of basic gemetrical obejcts like points, lines, and planes. We'll also 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 out with respect to a basis. We'll learn about linear combinations of vectors, the span of vectors, and formally define what a vector space is. In Section 5.5 we'll learn how to use the reduced row echelon form of a matrix, to describe the fundamental spaces associated with the matrix.

Chapter 6 is about linear transformations. Armed with the computational tools from Chapter 4 and the geometrical intuition from Chapter 5, we can tackle the core subject of linear algebra: linear transformations. We'll explore in detail the correspondence between linear transformations (vectors functions $f : \mathbb{R}^n \to \mathbb{R}^m$) and their representation as $m \times n$ matrices. We'll see how the coefficients in a matrix representation depend on the choice of basis for the input and output spaces of the transformation

Section 6.4 on the invertible matrix theorem serves as a midway checkpoint for your understanding of linear algebra. This theorem connects several seemingly disparate concepts: RREF, matrix inverse, row space, column space, determinants, and connects them as a single

[1] The models of Newtonian mechanics apply for medium sized objects, moving at slow speeds, and are far away from black holes.

result that singles out the properties of invertible linear transformations and distinguishes them from non-linear transformations. Invertible transformations are one-to-one correspondences between the vectors in the input space and the vectors in the output space.

Chapter 7 covers more advanced topics of linear algebra. We'll define the eigenvalues and the eigenvectors of a square matrix. We'll see that the characteristics of the eigenvalues of a matrix tell us a lot of information about the properties of the matrix. We'll learn about some standard special names given to different types of matrices, based on the properties of the eigenvalues of the matrix. In Section 7.3 we'll learn about abstract vector spaces. Abstract vectors are mathematical object that—like vectors—have components and can be scaled and added and subtracted componentwise.

Section 7.7 will discuss linear algebra with the complex numbers. Instead of working with vectors with real coefficients, we can do linear algebra with vectors that have complex coefficients. This section serves as a review of all the material in the book, because it revisits all the key concepts discussed in order to check how they are affected by the change to complex numbers.

Finally in Chapter 8 we'll discuss the applications of linear algebra. If you've done your job learning the material in the first severn chapters, now you get to learn all the cool things you can do with linear algebra.

**Mar 1$^{\text{st}}$ 2013:** Note to my readers: the applications chapter of the book is not finished yet. The appendix on quantum mechanics is also a work in progress. I will complete them shortly, but the material on applications and quantum probably won't be on your exam...

## Difficulty level

In terms of *difficulty* of the content, I would say that you should get ready for some serious uphills. As your personal "mountain guide" to the "mountain" of linear algebra, it is my obligation to warn you about the difficulties that lie ahead so that you will be mentally prepared.

The computational aspects will be difficult in a boring and repetitive kind of way as you have to go through thousands of steps where you multiply things together and add up the results. The theoretical aspects will be difficult in a very different kind of way: you'll learn about various properties of vectors and matrices and have to develop an intuition about using these properties to *prove* things. In this respect, linear algebra is the undergraduate course that most resembles what doing "real math" is like: you start from axioms and basic facts about mathematical objects and derive new mathematical results using proofs based on the axioms and facts. It's fancy stuff, but very powerful.

In summary, a lot of work and toil awaits you as you learn about the concepts from linear algebra, but the effort is definitely worth it. All the effort you put into understanding vectors and matrices will lead to mind-expanding insights. You will reap the benefits of your effort for the rest of your life: understanding linear algebra will open many doors for you.

# Chapter 1

# Math fundamentals

In this chapter we'll review the fundamental ideas of mathematics, including numbers, equations, and functions. We need this review to make sure you are comfortable with the fundamental concepts of math. Linear algebra is the extension of these ideas to many dimensions. In high school math you "do math" with numbers and functions, in linear algebra we'll "do math" with vectors and linear transformations.

You can think of linear transformations as "vector functions" and describe their properties in analogy with the regular functions you are familiar with. The action of a function on a number is similar to the action of a matrix on a vector:

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

To understand linear algebra, you need to understand the basics first.

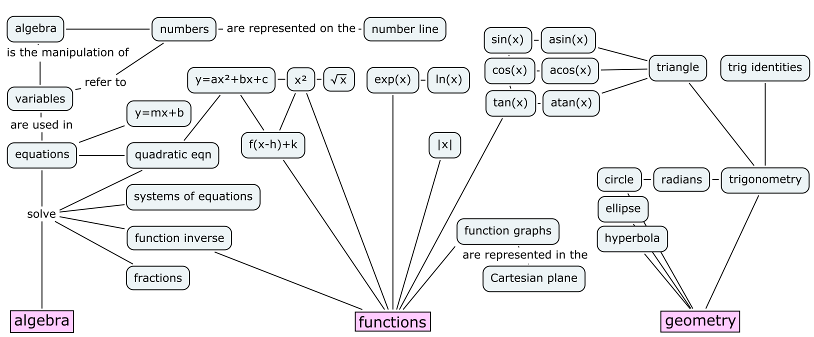


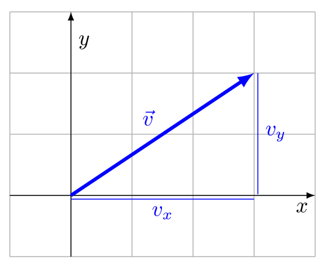

**Figure 1.1:** A concept map showing the mathematical topics that we will cover in this chapter. We'll learn about how to solve equations using algebra, how to model the world using functions, and how to think geometrically. The material in this chapter is required for your understanding of the more advanced topics in this book.

# Chapter 2

# Vectors

In this chapter we will learn how to manipulate multi-dimensional objects called vectors. Vectors are the precise way to describe directions in space. We need vectors in order to describe physical quantities like the velocity of an object, its acceleration, and the net force acting on the object.

Vectors are built from ordinary numbers, which form the *components* of the vector. You can think of a vector as a list of numbers, and *vector algebra* as operations performed on the numbers in the list. Vectors can also be manipulated as geometrical objects, represented by arrows in space. The arrow that corresponds to the vector $\vec{v} = (v_x, v_y)$ starts at the origin $(0, 0)$ and ends at the point $(v_x, v_y)$. The word vector comes from the Latin *vehere*, which means *to carry*. Indeed, the vector $\vec{v}$ takes the point $(0, 0)$ and carries it to the point $(v_x, v_y)$.




This chapter will introduce you to vectors, vector algebra, and vector operations, which are very useful for solving physics problems. What you'll learn here applies more broadly to problems in computer graphics, probability theory, machine learning, and other fields of science and mathematics. It's all about vectors these days, so you better get to know them.

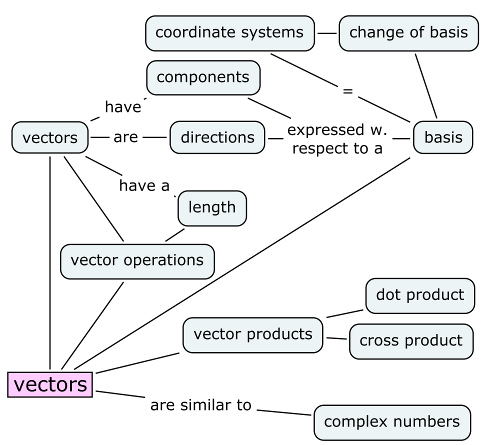



**Figure 2.1:** This figure illustrates the new concepts related to vectors. As you can see, there is quite a bit of new vocabulary to learn, but don't be phased—all these terms are just fancy ways of talking about arrows.

# Chapter 3

# Intro to linear algebra

The first two chapters of this book reviewed some core ideas of mathematics and basic notions about vectors that you may have been exposed to while learning physics. We're now done with the prerequisites so we can start the main matter: linear algebra.

## 3.1 Introduction

Linear algebra is the math of vectors and matrices. An $n$-dimensional vector $\vec{v}$ is an array of $n$ numbers. 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 need to specify the values for its three components $v_1$, $v_2$, and $v_3$.

A matrix $M \in \mathbb{R}^{m\times n}$ is an table of numbers with $m$ rows and $n$ columns. Consider as an example the following $3 \times 3$ matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \in \begin{bmatrix} \mathbb{R} & \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} & \mathbb{R} \\ \mathbb{R} & \mathbb{R} & \mathbb{R} \end{bmatrix} \equiv \mathbb{R}^{3\times 3}.$$

To specify the matrix $A$ we need to specify the values of its nine components $a_{11}$, $a_{12}$, ... , $a_{33}$.

In the remainder of the book we'll learn about the mathematical operations we can perform on vectors and matrices and their applications. Many problems in science, business, and technology can be described in terms of vectors and matrices so it is important for you to understand how to work with these things.

### Context

To illustrate what is 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 on 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 you know how to use these operations to evaluate math expressions.

You should also be familiar with *functions* that take real numbers as inputs and give real numbers as outputs, $f : \mathbb{R} \to \mathbb{R}$. Recall that, by definition, the *inverse function* $f^{-1}$ *undoes* the effect of $f$. If you are given $f(x)$ and you 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 $g^{-1}(x) = x^2$.

Vectors $\vec{v} \in \mathbb{R}^n$ and matrices $A \in \mathbb{R}^{m\times n}$ are the new objects of study in linear algebra. Our first step will be to define the basic operations which we can perform on them.

### Vector operations

The operations we can perform on vectors are

- addition (denoted $+$)
- subtraction, the inverse of addition (denoted $-$)
- scaling
- dot product (denoted $\cdot$)
- cross product (denoted $\times$)

We'll discuss each of these vector operations in Section **??**, but I expect you are already familiar with vectors from Chapter 2.

### Matrix operations

The mathematical operations defined for matrices $A$ and $B$ are the constant $\alpha$ are

- addition (denoted $A + B$)
- subtraction, the inverse of addition (denoted $A - B$)
- scaling (denoted implicitly $\alpha A$)
- matrix product (denoted implicitly $AB$)
- matrix inverse (denoted $A^{-1}$)
- trace (denoted $\mathrm{Tr}(A)$)
- determinant (denoted $\det(A)$ or $|A|$)

We will formally define each of these operations in Section 3.3 and you'll learn about the various computational, geometrical, and theoretical considerations associated with these matrix operations in the remainder of the book. For now, we'll focus on one important special case of the matrix product: the matrix-vector product $A\vec{x}$.

### Matrix-vector product

The matrix-vector product $A\vec{x}$ produces a linear combination of the columns of the matrix $A$ with coefficients $\vec{x}$. For example, consider the product of a $3 \times 2$ matrix $A$ and a $2 \times 1$ vector $\vec{x}$. The result of the product $A\vec{x}$ is the $3 \times 1$ vector $\vec{y}$ obtained as follows:

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

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

The key thing to observe in the above formula is the interpretation of the matrix-vector product in the "column picture": $\vec{y} = A\vec{x} = x_1A_{[:,1]} + x_2A_{[:,2]}$, where $A_{[:,1]}$ and $A_{[:,2]}$ are the first and second columns of $A$.

### Linear combinations as matrix products

Consider now some set of vectors $\{\vec{e}_1, \vec{e}_2\}$ and a third vector $\vec{y}$ which 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 above linear combination as the following matrix-vector product:

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

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 $\vec{y}$.

### Matrices as vector functions

Okay my dear readers, we have now reached the key notion in the study of linear algebra. One could even say the *main idea*.

I know you are ready to handle it because you are familiar with functions of a real variable $f : \mathbb{R} \to \mathbb{R}$ and you just saw the definition of the matrix-vector product in which the variables were chosen to subliminally remind you of the standard convention for calling the function input $x$ and the function output $y = f(x)$. Without further ado, I present to you: vector functions, also known as *linear transformations*.

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 $\vec{y} = T_A(\vec{x})$ for the linear transformation $T_A$ applied to the vector $\vec{x}$, we can write $\vec{y} = A\vec{x}$. 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 $A$.

When the matrix $A \in \mathbb{R}^{n\times n}$ is invertible, there exists an inverse matrix $A^{-1}$ which *undoes* the effect of $A$ to give back the original input vector:

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

For example, the transformation which multiplies the first components of input vectors by 3 and multiplies the second components by 5 is described by the matrix

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

Its inverse 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 has the effect of undoing what $A$ did.

Things get a little more complicated when matrices *mix* the different components of the input vector as in the following 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 $B(\vec{x}) \equiv B\vec{x}$ using the definition 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 the multiplication by $B$:

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

We will discuss matrix inverses and how to compute them in more detail later. For now it is important that you know they exist. By definition, the inverse matrix $A^{-1}$ *undoes* the effects of the matrix $A$:

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

The cumulative effect of applying $A^{-1}$ after $A$ is the *identity matrix* $\mathbb{1}$, which has ones on the diagonal and zeros everywhere else. Note that $\mathbb{1}\vec{x} = \vec{x}$ for any vector $\vec{x}$.

## What lies ahead?

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 range). The linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a "vector function" that takes $n$-vectors as inputs and produces $m$-vectors as outputs. If the vector 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 $A_T$. The matrix $A_T$ is a particular "implementation" of the abstract linear transformation $T$. **The coefficients of the matrix $A_T$ depend on the choice of basis for the input space and a the basis for the output space**.

Equivalently, every matrix $A \in \mathbb{R}^{m\times n}$ corresponds to some linear transformation $T_A : \mathbb{R}^n \to \mathbb{R}^m$. What does $T_A$ do? We define the action of $T_A$ on input $\vec{x}$ as the matrix-vector product $A\vec{x}$.

Given the equivalence between matrices and linear transformation, we can reinterpret the statement "linear algebra is about vectors and matrices" as saying "linear algebra is about vectors and linear transformations." If high school math is about numbers and functions then linear algebra is about vectors and vector functions. The action of a function on a number is similar to the action of a linear transformation (matrix) on a vector:

| | | |
|---|---|---|
| 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 $A^{-1}$ |
| zeros of $f$ | $\Leftrightarrow$ | null space of $T_A\equiv$ null space of $A\equiv\mathcal{N}(A)$ |
| image of $f$ | $\Leftrightarrow$ | image of $T_A\equiv$ column space of $A\equiv\mathcal{C}(A)$ |

The above table of correspondences serves as a roadmap for the rest of the material in the book. There are several new concepts, but not too many. You can do this!

A good strategy is to try adapting your existing knowledge about functions to the world of linear transformations. For example, the zeros of a function $f(x)$ are the set of inputs for which the function's output is zero. Similarly, the *nulls space* of a linear transformation $T$ is is the set of inputs that $T$ sends to the zero vector. It's the same concept really, we're just upgrading. If you ever get confused by some calligraphic letter like $\mathcal{N}$, $\mathcal{C}$, or $\mathcal{R}$, you can come back and review this table.

Recall what we said in Chapter 1 about knowing about functions can be useful for modelling the real world. Linear algebra is like a "vector upgrade" to your real-world modelling skills; you'll learn how to model complex relationships between multivariable inputs and multivariable outputs.

To build "modelling skills" you must first develop you geometrical intuition about lines, planes, vectors, bases, linear transformations, vector spaces, vector subspaces, etc. You'll also need to learn new computational techniques and develop new ways of thinking. Using these new ways of thinking and the computational tools you'll be able to apply linear algebra techniques to many areas of application. Let's now look in a little more detail at what lies ahead in the book.

### Computational linear algebra

The first steps toward understanding linear algebra will be a little tedious. In Chapter 4 you'll develop basic skills for manipulating vectors and matrices. Matrices and vectors have many components and performing operations on them involves a lot of 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 about the more advanced material.

The good news is that, except on your homework assignments and on your final exam, you won't have to do matrix algebra by hand. It is much more convenient to use a computer for large matrix calculations. For small instances, like $4\times 4$ matrices, you should be able to perform all the matrix algebra operations with pen and paper. The more you develop your matrix algebra skills, the deeper you'll be able to go into the advanced material.

### Geometrical linear algebra

So far we described vectors and matrices as arrays of numbers. This is fine for the purpose of doing *algebra* on vectors and matrices, but it is not sufficient to understand their geometrical properties. The components of a vector $\vec{v}\in\mathbb{R}^n$ can be thought of as measuring distances 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 geometrical interpretation of each of the algebraic operations on vectors and matrices.

Consider some unit length 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 $\theta$ 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 we mean that 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 5.2 we'll learn how to calculate all kinds of projections in using the dot product.

As another example of the geometrical aspects of vector operations, consider the following situation. Suppose I gave you two vectors $\vec{u}$ and $\vec{v}$ and I 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 5.1 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 intercept. There is an algebraic procedure called *Gauss–Jordan elimination* that you can use to find the solution.

The determinant of a matrix also has a geometrical interpretation (Section 4.4). 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*—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 a set of vectors has this property.

It is important that you try to *visualize* every new concept you learn about. You should 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 intuitions you build about vectors in two and three dimensions is applicable to vectors with more dimensions.

### Theoretical linear algebra

The most interesting aspect of linear algebra is that it teaches you to reason about vectors and matrices in a very abstract way. By thinking abstractly, you will be able to extend your geometrical intuition of two and three-dimensional problems to problems in higher dimensions. A lot of *knowledge buzz* awaits you as you learn about new mathematical ideas and develop new ways of thinking.

You're no doubt familiar with the normal coordinate system made up of two orthogonal axes: the $x$-axis and the $y$-axis. A vector $\vec{v}$ can be specified in terms of its coordinates $(v_x, v_y)$ with respect to these axes, that is, we can write any vector $\vec{v} \in \mathbb{R}^2$ as $\vec{v} = v_x\hat{\imath} + v_y\hat{\jmath}$, where $\hat{\imath}$ and $\hat{\jmath}$ are unit vectors that point along the $x$-axis and the $y$-axis. It turns out that 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 write all vectors $\vec{v} \in \mathbb{R}^2$ as a linear combination 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 $\{\hat{e}_1, \hat{e}_2\}$ basis. 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.

The eigenvalues and eigenvectors of matrices (Section **??**) allow us to describe the actions of matrices in a very natural way. The set of eigenvectors of a matrix is a special set of input vectors for which the action of the matrix is described as a *scaling*. When a matrix is multiplied by one of its eigenvectors the output is a vector in the same direction scaled by a constant, which we call an eigenvalue. Thinking of matrices in term of their eigenvalues and eigenvectors is a very powerful technique for describing their properties.

In the above text I explained that computing the product between a matrix and a vector $A\vec{x} = \vec{y}$ can be thought of as a linear vector function, with input $\vec{x}$ and output $\vec{y}$. Any linear transformation (Section 6.1) can be represented (Section 6.2) 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 implementing some linear transformation (vector function): $T_A \colon \mathbb{R}^n \to \mathbb{R}^m$. 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 $T_A$).

Part of what makes linear algebra so powerful is that linear algebra techniques can be applied to all kinds of "vector-like" objects. The abstract concept of a vector space (Section 7.3) 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 $P_2(x)$ 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 can be expressed as linear equations in multiple unknown quantities. You can solve for $n$ unknowns simultaneously if you have a set of $n$ linear equations. To solve this system of equations, you can use basic techniques such as substitution and subtraction to eliminte the variables one by one (see Section **??**), but the procedure will be slow and tedious when there are many equations and many unknowns. If the system of equations is linear, then it can be expressed as an *augmented matrix* build from the coefficients in the equations. You can then use the Gauss–Jordan elimination algorithm to solve for the $n$ unknowns (Section 4.1). The key benefit of augmented matrix approach is it allows you to focus on the coefficients and not worry about the variable names. This saves a lot of time when you have to solve many unknowns. Another approach for solving $n$ linear equations in $n$ unknowns is to express the system of equations as a matrix equation

(Section 4.2) and then solve the matrix equation by computing the matrix inverse (Section 4.5).

In Section 7.6 you'll learn how to *decompose* a matrix into a product of simpler matrices in various ways. 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 formalized as the search for an eigenvector of a matrix. The matrix in question contains the information about all the hyperlinks that exist between webpages. The eigenvector we are looking for corresponds to a vector which tells you the relative importance of each page. So when I tell you eigenvectors are *valuable information*, I am not kidding: a 300-billion dollar company started out as an eigenvector idea.

The techniques of linear algebra find application 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 8. As a special bonus for readers in this kind of stuff, there is also a short introduction to quantum mechanics in Appendix B; if you have a good grasp of linear algebra, you can understand matrix quantum mechanics at no additional mental cost.

### What next?

But let's not get ahead of ourselves and bring geometry, vector spaces, algorithms, and the applications of linear algebra all into the mix. Let's start with the basics. If linear algebra is about vectors and matrices, then we better define vectors and matrices precisely and describe the math operations we can do on them.

## 3.2 Review of vector operations

In Chapter 2 we described vectors from a practical point of view. Vectors are useful for describing directional quantities like forces and velocities in physics.

In this section we'll describe vectors more abstractly—as math objects. The first thing to do after one defines a new mathematical object is to specify its properties and the operations that we can perform on them.

### 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:

$$\begin{aligned}
\alpha\vec{u} &= (\alpha u_1, \alpha u_2, \alpha u_3) \\
\vec{u} + \vec{v} &= (u_1 + v_1, u_2 + v_2, u_3 + v_3) \\
\vec{u} - \vec{v} &= (u_1 - v_1, u_2 - v_2, u_3 - v_3) \\
||\vec{u}|| &= \sqrt{u_1^2 + u_2^2 + u_3^2} \\
\vec{u} \cdot \vec{v} &= u_1 v_1 + u_2 v_2 + u_3 v_3 \\
\vec{u} \times \vec{v} &= (u_2 v_3 - u_3 v_2,\ u_3 v_1 - u_1 v_3,\ u_1 v_2 - u_2 v_1)
\end{aligned}$$

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

## Notation

The set of real numbers is denoted $\mathbb{R}$. An $n$-dimensional real vector consists of $n$ real numbers slapped together in a bracket. We denote the set of 3-dimensional vectors as $(\mathbb{R}, \mathbb{R}, \mathbb{R}) \equiv \mathbb{R}^3$. Similarly, the set of $n$-dimensional real vectors is denoted $\mathbb{R}^n$.

## Addition and subtraction

Addition and subtraction take two vectors as inputs and produce another vector as output:

$$+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \qquad \text{and} \qquad - : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n.$$

Addition and subtraction are performed component wise:

$$\vec{w} = \vec{u} \pm \vec{v} \qquad \Leftrightarrow \qquad w_i = u_i \pm v_i, \quad \forall i \in [1, \ldots, n].$$

## Scaling by a constant

Scaling a vector by a constant is an operation that has the signature:

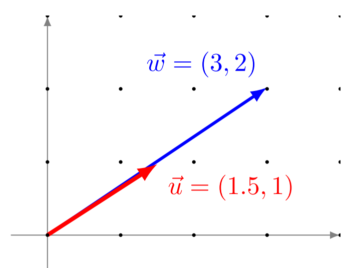
$$\text{scalar-mult} : \mathbb{R} \times \mathbb{R}^n \ \to \ \mathbb{R}^n.$$

There is no symbol to denote scalar multiplication—we just write the scaling factor in front of the vector and it is implicit that we are multiplying the two.

The scaling factor $\alpha$ multiplying the vector $\vec{u}$ is equivalent to $\alpha$ multiplying each component of the vector:

$$\vec{w} = \alpha \vec{u} \qquad \Leftrightarrow \qquad w_i = \alpha u_i.$$

For example, choosing $\alpha = 2$ we obtain the vector $\vec{w} = 2\vec{u}$ which is two times longer than the vector $\vec{u}$:




$$\vec{w} = (w_1, w_2, w_3) = (2u_1, 2u_2, 2u_3) = 2(u_1, u_2, u_3) = 2\vec{u}.$$

## Vector multiplication

The **dot product** is defined as follows:

$$\cdot : \mathbb{R}^n \times \mathbb{R}^n \ \to \mathbb{R}, \qquad \vec{u} \cdot \vec{v} \equiv \sum_{i=1}^{n} u_i v_i.$$

The dot product is defined for vectors of any dimension. As long as two vectors have the same length, we can "zip" through them computing the sum of the products of their corresponding entries.

The dot product is the key tool for projections, decompositions, and calculating orthogonality. It is also known as the *scalar product* or the *inner product*. Intuitively, applying the dot product to two vectors produces a scalar number which carries information about *how similar* the two vectors are. Orthogonal vectors are not similar at all: no part of one vector goes in the same direction as the other vector so their dot product is zero. For example: $\hat{\imath} \cdot \hat{\jmath} = 0$. Another notation for the inner product is $\langle u | v \rangle \equiv \vec{u} \cdot \vec{v}$.

The **cross product** is defined as follows:

$$\times : \mathbb{R}^3 \times \mathbb{R}^3 \ \to \mathbb{R}^3, \qquad \vec{w} = \vec{u} \times \vec{v} \quad \Leftrightarrow \quad \begin{array}{rcl} w_1 & = & u_2 v_3 - u_3 v_2, \\ w_2 & = & u_3 v_1 - u_1 v_3, \\ w_3 & = & u_1 v_2 - u_2 v_1. \end{array}$$

The cross product, or *vector product* as it is sometimes called, is an operator which returns a vector that is perpendicular to both of the input vectors. For example: $\hat{\imath} \times \hat{\jmath} = \hat{k}$. Note the cross product is only defined for 3-dimensional vectors.

## Length of a vector

The length of the vector $\vec{u} \in \mathbb{R}^n$ is computed as follows:

$$\|\vec{u}\| = \sqrt{u_1^2 + u_2^2 + \cdots + u_n^2} = \sqrt{\vec{u} \cdot \vec{u}}.$$

The length of a vector is a nonnegative number that describes the extent of the vector in space. The notion of length is an extension of Pythagoras' formula for the length of the hypotenuse of a triangle given the lengths of its two sides to $n$ dimensions. The length of a vector is sometimes called the *magnitude* or the *norm* of a vector. The length of a vector $\vec{u}$ is denoted $\|\vec{u}\|$ or $|\vec{u}|_2$ and sometimes simply $u$.

There are many mathematical concepts that correspond to the intuitive notion of length. The formula above computes the *Euclidian length* (or *Euclidian norm*) of the vector. Another name for the Euclidian length is the $\ell^2$-norm (pronounced *ell-two norm*[1]).

Note the length of a vector can be computed as the square root of the dot product of the vector with itself: $\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}$. Indeed, there is a deep mathematical norms norms and inner products.

## Unit vector

Given a vector $\vec{v}$ of any length, we can build a unit vector in the same direction by dividing $\vec{v}$ by its length:

$$\hat{v} = \frac{\vec{v}}{||\vec{v}||}\,.$$

Unit vectors are useful in many contexts. When we want to specify a direction in space, we use a unit vector in that direction.

## Projection

Okay, pop-quiz time! Let's see if you remembered anything from Chapter 2. Suppose I give you a direction $\hat{d}$ and some vector $\vec{v}$ and ask you how much of $\vec{v}$ is in the $\hat{d}$-direction? To find the answer, compute the dot product:

$$v_d = \hat{d} \cdot \vec{v} \equiv \|\hat{d}\|\|\vec{v}\| \cos\theta = 1\|\vec{v}\| \cos\theta,$$

where $\theta$ is the angle between $\vec{v}$ and $\hat{d}$. This formula is used in physics to compute $x$-components of forces: $F_x = \vec{F} \cdot \hat{\imath} = \|\vec{F}\| \cos\theta$.

Define the *projection* of a vector $\vec{v}$ in the $\hat{d}$ direction as follows:

$$\Pi_{\hat{d}}(\vec{v}) = v_d \hat{d} = (\hat{d} \cdot \vec{v})\hat{d}.$$

If the direction is specified by a vector $\vec{d}$ that is not of unit length then the projection formula becomes:

$$\Pi_{\vec{d}}(\vec{v}) = \left( \frac{\vec{d} \cdot \vec{v}}{\|\vec{d}\|^2} \right) \vec{d}.$$

[1] The name $\ell^2$-*norm* is because each of the coefficient is squared in the sum, and then take the square root to compute the length. An example of another norm is the $\ell^4$-norm which is defined as the fourth root of the sum of the components raised to the power four: $|\vec{u}|_4 \equiv \sqrt[4]{u_1^4 + u_2^4 + u_3^4}$.

The division by the length squared transforms the two appearances of the vector $\vec{d}$ into the unit vectors $\hat{d}$ needed for the projection formula:

$$\Pi_{\hat{d}}(\vec{v}) \;=\; \underbrace{(\vec{v}\cdot\hat{d})}_{\|\vec{v}\|\cos\theta}\,\hat{d} \;=\; \left(\vec{v}\cdot\frac{\vec{d}}{\|\vec{d}\|}\right)\frac{\vec{d}}{\|\vec{d}\|} \;=\; \left(\frac{\vec{v}\cdot\vec{d}}{\|\vec{d}\|^2}\right)\vec{d} \;=\; \Pi_{\vec{d}}(\vec{v}).$$

### Discussion

This section was a review of the properties of $n$-dimensional vectors. These are simply ordered tuples (lists) of $n$ coefficients. It is important to think of vectors as whole mathematical objects and not as coefficients. Sure, all the vector operations boil down to manipulations of the coefficients in the end, but vectors are most useful (and best understood) if you think of them as one *thing* that has components rather than focussing on the components.[2]

### Links

[ Example of a non-Euclidian distance formula (points on a sphere) ]
http://www.movable-type.co.uk/scripts/latlong.html

## 3.3 Matrix operations

A matrix is a two-dimensional array (a table) of numbers. Consider the $m$ by $n$ matrix $A \in \mathbb{R}^{m\times n}$. What are the mathematical operations we can do on this matrix?

We denote the matrix as a whole $A$ and refer to its individual entries as $a_{ij}$, where $a_{ij}$ is the entry in the $i^{\text{th}}$ row and the $j^{\text{th}}$ column of $A$.

### Addition and subtraction

The matrix addition and subtraction operations take two matrices as inputs and produce a matrix of the same size as output:

$$+ : \mathbb{R}^{m\times n}\times\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n} \quad \text{and} \quad - : \mathbb{R}^{m\times n}\times\mathbb{R}^{m\times n}\to\mathbb{R}^{m\times n}.$$

Addition and subtraction are performed component wise:

$$C = A \pm B \quad\Leftrightarrow\quad c_{ij} = a_{ij} \pm b_{ij},\ \forall i \in [1,\ldots,m], j\in[1,\ldots,n].$$

Written explicitly for $3\times 3$ matrices, addition is

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & a_{13}+b_{13} \\ a_{21}+b_{21} & a_{22}+b_{22} & a_{23}+b_{23} \\ a_{31}+b_{31} & a_{32}+b_{32} & a_{33}+b_{33} \end{bmatrix}.$$

### Multiplication by a constant

Given a number $\alpha$ and a matrix $A$, we can *scale* $A$ by $\alpha$ as follows:

$$\alpha A = \alpha \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \\ \alpha a_{31} & \alpha a_{32} & \alpha a_{33} \end{bmatrix}.$$

[2]Vector components are Māyā. Do not try to hold on to them. A vector is something in the real world. The components of the vector are only a representation of this vector with respect to some coordinate system. Change the basis and the coordinates will change.

## Matrix-vector multiplication

The matrix-vector product between matrix $A \in \mathbb{R}^{m\times n}$ and a vector $\vec{v} \in \mathbb{R}^n$ results in $m$-dimensional vector as output.

$$\text{matrix-vector product} : \mathbb{R}^{m\times n} \times \mathbb{R}^n \to \mathbb{R}^m$$

$$\vec{w} = A\vec{v} \qquad \Leftrightarrow \qquad w_i = \sum_{j=1}^{n} a_{ij}v_j, \quad \forall i \in [1,\ldots,m].$$

For example, the product of a $3\times 3$ matrix $A$ and the $3\times 1$ column vector $\vec{v}$ results in a $3\times 1$ output:

$$\begin{aligned} A\vec{v} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} &= v_1 \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} + v_2 \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} + v_3 \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix} \\ &= \begin{bmatrix} (a_{11}, a_{12}, a_{13}) \cdot \vec{v} \\ (a_{21}, a_{22}, a_{23}) \cdot \vec{v} \\ (a_{31}, a_{32}, a_{33}) \cdot \vec{v} \end{bmatrix} \\ &= \begin{bmatrix} a_{11}v_1 + a_{12}v_2 + a_{13}v_3 \\ a_{21}v_1 + a_{22}v_2 + a_{23}v_3 \\ a_{31}v_1 + a_{32}v_2 + a_{33}v_3 \end{bmatrix} \qquad \in \mathbb{R}^{3\times 1}. \end{aligned}$$

Note the two different ways to understand the matrix-vector product.[3]

In the "column picture" the multiplication of the matrix $A$ by the vector $\vec{v}$ is a **linear combination of the columns of the matrix**: $A\vec{v} = v_1 A_{[:,1]} + v_2 A_{[:,2]} + v_3 A_{[:,3]}$, where $A_{[:,1]}$, $A_{[:,2]}$ and $A_{[:,3]}$ are the first, second and third columns of $A$.

In the "row picture," multiplication of the matrix $A$ by the vector $\vec{x}$ produces a column vector with coefficients equal to the **dot products of the rows of the matrix** with the vector $\vec{v}$.

## Matrix-matrix multiplication

The matrix multiplication $AB$ of matrices $A \in \mathbb{R}^{m\times \ell}$ and $B \in \mathbb{R}^{\ell\times n}$ consists of computing the dot product between each the rows of $A$ and each the columns of $B$:

$$\text{matrix-product} : \mathbb{R}^{m\times \ell} \times \mathbb{R}^{\ell\times n} \to \mathbb{R}^{m\times n}$$

$$C = AB \qquad \Leftrightarrow \qquad c_{ij} = \sum_{k=1}^{\ell} a_{ik}b_{kj}, \forall i \in [1,\ldots,m], j \in [1,\ldots,n].$$

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

## Transpose

The transpose matrix $A^{\mathsf{T}}$ is defined by the formula $a^{\mathsf{T}}_{ij} = a_{ji}$. In other words, we obtain the transpose by "filipping" the matrix through its diagonal:

$$^{\mathsf{T}} : \mathbb{R}^{m\times n} \to \mathbb{R}^{n\times m},$$

$$\begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \\ \beta_1 & \beta_2 & \beta_3 \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} \alpha_1 & \beta_1 \\ \alpha_2 & \beta_2 \\ \alpha_3 & \beta_3 \end{bmatrix}.$$

Note entries on the diagonal are not affected by the transpose.

[3] For more info see the video of Prof. Strang's MIT lecture: `bit.ly/1ayRcrj`

**Properties of transpose operation**

$$
\begin{aligned}
(A+B)^{\mathsf{T}} &= A^{\mathsf{T}} + B^{\mathsf{T}} \\
(AB)^{\mathsf{T}} &= B^{\mathsf{T}} A^{\mathsf{T}} \\
(ABC)^{\mathsf{T}} &= C^{\mathsf{T}} B^{\mathsf{T}} A^{\mathsf{T}} \\
(A^{\mathsf{T}})^{-1} &= (A^{-1})^{\mathsf{T}}
\end{aligned}
$$

## Vectors as matrices

Vectors are special types of matrices. You can treat a vector $\vec{v} \in \mathbb{R}^n$ either as a *column vector* ($n \times 1$ matrix) or as a *row vector* ($1 \times n$ matrix).

**Inner product**

Recall the definition of the *dot product* or *inner product* for vectors:

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

The dot product of two $n$-vectors $\vec{u}$ and $\vec{v}$ is $\vec{u} \cdot \vec{v} \equiv \sum_{i=1}^{n} u_i v_i$.

If we think of these vectors as *column* vectors then we can write the dot product in terms of the matrix transpose operation $^{\mathsf{T}}$ and the standard rules of matrix multiplication:

$$\vec{u} \cdot \vec{v} = \vec{u}^{\mathsf{T}} \vec{v} = \begin{bmatrix} u_1 & u_2 & u_3 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = u_1 v_1 + u_2 v_2 + u_3 v_3.$$

The dot product for vectors is really a special case of matrix multiplication. Alternately, we could say that matrix multiplication is defined in terms of the dot product.

**Outer product**

Consider again two *column* vectors $\vec{u}$ and $\vec{v}$ ($n \times 1$ matrices). We obtain the inner product if we put the transpose on the *first* vector: $\vec{u}^{\mathsf{T}} \vec{v} \equiv \vec{u} \cdot \vec{v}$. If instead we put the transpose on the *second* vector, we'll obtain the outer product of $\vec{u}$ and $\vec{v}$. The outer product operation takes two vectors as inputs and produces a matrix as output:

$$\text{outer-product} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^{n \times n}.$$

For example, the outer product of two vectors in $\mathbb{R}^3$ is

$$\vec{u}\vec{v}^{\mathsf{T}} = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} \begin{bmatrix} v_1 & v_2 & v_3 \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & u_1 v_3 \\ u_2 v_1 & u_2 v_2 & u_2 v_3 \\ u_3 v_1 & u_3 v_2 & u_3 v_3 \end{bmatrix} \quad \in \mathbb{R}^{3 \times 3}.$$

The matrix-matrix product between a $3 \times 1$ matrix and a $1 \times 3$ matrix results in a $3 \times 3$ matrix.

The outer product is used to build *projection matrices*. For example, the matrix which corresponds to the projection onto the $x$-axis is given by $M_x = \hat{\imath}\hat{\imath}^{\mathsf{T}} \in \mathbb{R}^{3 \times 3}$. The $x$-projection of any vector $\vec{v}$ can be computed as a matrix-vector product: $M_x \vec{v} = \hat{\imath}\hat{\imath}^{\mathsf{T}} \vec{v} = \hat{\imath}(\hat{\imath} \cdot \vec{v}) = v_x \hat{\imath}$.

## Matrix inverse

The inverse matrix $A^{-1}$ has the property that $AA^{-1} = \mathbb{1} = A^{-1}A$, where $\mathbb{1}$ is the *identity matrix* which obeys $\mathbb{1}\vec{v} = \vec{v}$ for all vectors $\vec{v}$. The inverse matrix $A^{-1}$ has the effect of *undoing* whatever $A$ did.

The cumulative effect of multiplying by $A$ and $A^{-1}$ is equivalent to the identity transformation:

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

We can think of "finding the inverse" $\textrm{inv}(A) = A^{-1}$ as an operation of the form:

$$\textrm{inv} : \mathbb{R}^{n\times n} \to \mathbb{R}^{n\times n}.$$

Note that only *invertible* matrices have an inverse.

**Properties of inverse operation**

$$\begin{aligned}
(A+B)^{-1} &= A^{-1} + B^{-1} \\
(AB)^{-1} &= B^{-1}A^{-1} \\
(ABC)^{-1} &= C^{-1}B^{-1}A^{-1} \\
(A^{\mathsf{T}})^{-1} &= (A^{-1})^{\mathsf{T}}
\end{aligned}$$

The matrix inverse plays the role of "division by the matrix A" in matrix equations. We'll discuss matrix equations in Section 4.2.

## Trace

The *trace* of an $n \times n$ matrix is the sum of the $n$ values on its diagonal:

$$\textrm{Tr} : \mathbb{R}^{n\times n} \to \mathbb{R}, \qquad \textrm{Tr}[A] \equiv \sum_{i=1}^{n} a_{ii}.$$

**Properties of trace operation**

$$\begin{aligned}
\textrm{Tr}[\alpha A + \beta B] &= \alpha\textrm{Tr}[A] + \beta\textrm{Tr}[B] && \text{(linear property)} \\
\textrm{Tr}[AB] &= \textrm{Tr}[BA] \\
\textrm{Tr}[ABC] &= \textrm{Tr}[CAB] = \textrm{Tr}[BCA] && \text{(cyclic property)} \\
\textrm{Tr}[A] &= \sum_{i=1}^{n} \lambda_i && \text{where } \{\lambda_i\} \equiv \textrm{eig}(A) \text{ are the eigenvalues} \\
\textrm{Tr}\left[A^{\mathsf{T}}\right] &= \textrm{Tr}[A]
\end{aligned}$$

## Determinant

The determinant of a matrix is a calculation which involves all the coefficients of the matrix and the output of which a single real number:

$$\det : \mathbb{R}^{n\times n} \to \mathbb{R}.$$

The determinant describes the relative geometry of the vectors that make up the matrix. More specifically, the determinant of a matrix $A$ tells you the *volume* of a box with sides given by rows of $A$.

The determinant of a $2 \times 2$ matrix is

$$\det(A) = \det\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

The quantity $ad - bc$ corresponds to the area of the parallelogram formed by the vectors $(a, b)$ and $(c, d)$. Observe that if the rows of $A$ point in the same direction $(a, b) = \alpha(c, d)$ for some $\alpha \in \mathbb{R}$, then the area of the parallelogram will be zero. If the determinant of a matrix is non-zero then the rows the matrix are linearly independent.

**Properties of determinants**

$$
\begin{aligned}
\det(AB) &= \det(A)\det(B) \\
\det(A) &= \prod_{i=1}^{n} \lambda_i \qquad \text{where } \{\lambda_i\} \equiv \operatorname{eig}(A) \text{ are the eigenvalues} \\
\det\left(A^{\mathsf{T}}\right) &= \det(A) \\
\det\left(A^{-1}\right) &= \frac{1}{\det(A)}
\end{aligned}
$$

### Discussion

In the remainder of this book you'll learn about various algebraic and geometric interpretations of the matrix operations defined above. Understanding vector and matrix operations is essential for understanding the more advanced theoretical topics and the applications of linear algebra.

Thus far, we've defined two of the main actors in linear algebra (vectors and matrices), but the cast would not be complete until we introduce *linearity*. Linearity is the main "theme" that runs through all the topics in linear algebra.

## 3.4 Linearity

What is linearity? Why do we need an entire course to learn about linearity? Consider the following arbitrary function that contains terms with different *powers* of the input variable $x$:

$$f(x) \;=\; \frac{a}{x} \;+\; b \;+\; \underbrace{mx}_{\text{linear term}} \;+\; qx^2 \;+\; cx^3.$$

The term $mx$ is the only *linear* term in this expression—it contains $x$ to the first power. The term $b$ is *constant* and all other terms are *nonlinear*.

### Introduction

A single-variable function takes as input a real number $x$ and outputs a real number $y$. The signature of this class of functions is

$$f\colon \mathbb{R} \to \mathbb{R}.$$

The most general *linear function* from $\mathbb{R}$ to $\mathbb{R}$ looks like this:

$$y \equiv f(x) = mx,$$

where $m \in \mathbb{R}$ is called the *coefficient* of $x$. The action of a linear function is to multiply the input by the constant $m$. This is not too complicated, right?

**Example of composition of linear functions** Given the linear functions $f(x) = 2x$ and $g(y) = 3y$, what is the equation of the function $h(x) \equiv g \circ f\,(x) = g(f(x))$? The composition of the functions $f(x) = 2x$ and $g(y) = 3y$ is the function $h(x) = g(f(x)) = 3(2x) = 6x$. Note the composition of two linear functions is also a linear function. The coefficient of $h$ is equal to the product of the coefficients of the coefficients of $f$ and $g$.

## Definition

A function is *linear* if, for any two inputs $x_1$ and $x_2$ and constants $\alpha$ and $\beta$, the following equation is true:

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

Linear functions map linear combination of inputs to the same linear combination of outputs.

## Lines are not linear functions!

Consider the equation of a line:

$$l(x) = mx + b,$$

where the constant $m$ corresponds to the slope of the line and the constant $b \equiv f(0)$ is the $y$-intercept of the line. A *line* $l(x) = mx + b$ with $b \neq 0$ is *not* a linear function. This is a bit weird, but if you don't trust me, you can check for yourself:

$$l(\alpha x_1+\beta x_2) = m(\alpha x_1+\beta x_2)+b \neq m(\alpha x_1)+b+m(\beta x_2)+b = \alpha l(x_1)+\beta l(x_2).$$

A function with a linear part plus a constant term is called an *affine transformation*. These are cool too, but a bit off topic since the focus of our attention here is on *linear* functions.

## Multivariable functions

The study of linear algebra is the study of *all* things linear. In particular we'll learn how to work with functions that take multiple variables as inputs. Consider the set of functions that take pairs of real numbers as inputs and produce real numbers as outputs:

$$f\colon \mathbb{R} \times \mathbb{R} \to \mathbb{R}.$$

The most general linear function of two variables is

$$f(x, y) = m_x x + m_y y.$$

You can think of $m_x$ as the $x$-slope and $m_y$ as the $y$-slope of the function. We say $m_x$ is the $x$-coefficient of and $m_y$ the $y$-coefficient in the linear expression $m_x x + m_y y$.

## Linear expressions

A *linear expression* in the variables $x_1$, $x_2$, and $x_3$ has the form:

$$a_1x_1 + a_2x_2 + a_3x_3,$$

where $a_1$, $a_2$, and $a_3$ are arbitrary constants. Note the new terminology used "`expr` is linear in $v$" to refer to an expressions in which the variable $v$ appears only raised to the first power. The expression $\frac{1}{a}x_1 + b^6x_2 + \sqrt{c}\,x_3$ contains nonlinear factors ($\frac{1}{a}$, $b^6$, and $\sqrt{c}$), but is a linear expression in the variables $x_1$, $x_2$, and $x_3$.

## Linear equations

A linear equation in the variables $x_1$, $x_2$, and $x_3$ has the form

$$a_1x_1 + a_2x_2 + a_3x_3 = c.$$

This equation is linear because it contains no nonlinear terms in $x_i$.

**Example** Linear equations are very versatile. Suppose you know the following equation describes some real-world phenomenon:

$$4k - 2m + 8p = 10,$$

where $k$, $m$, and $p$ correspond to three variables of interest. You can think of this equation as describing the variable $m$ as a function of the variables $k$ and $p$ and rewrite the equation as:

$$m(k, p) = 2k + 4p - 5.$$

Using this function you can predict the value of $m$ given the knowledge of the quantities $k$ and $p$.

Another option would be to think of $k$ as a function of $m$ and $p$: $k(m, p) = \frac{10}{4} + \frac{m}{2} - 2p$. This model would be useful if you know the quantities $m$ and $p$ and you want to predict the value of the variable $k$.

## Applications

### Geometrical interpretation of linear equations

The linear equation in $x$ and $y$,

$$ax + by = c, \qquad b \neq 0,$$

corresponds to the equation of a line $y(x) = mx + y_0$ in the Cartesian plane. The slope of this line is $m = \frac{-a}{b}$ and its $y$-intercept is $y_0 = \frac{c}{b}$. The special case when $b = 0$ corresponds to a vertical line with equation $x = \frac{c}{a}$.

The most general linear equation in $x$, $y$, and $z$,

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

corresponds to the equation of a plane in a three-dimensional space. Assuming $c \neq 0$, we can rewrite this equation so that $z$ (the "height") is a function of the coordinates $x$ and $y$: $z(x, y) = z_0 + m_x x + m_y y$. The slope of the plane in the $x$-direction is $m_x = -\frac{a}{c}$ and $m_y = -\frac{b}{c}$ in the $y$-direction. The $z$-intercept of the plane is $z_0 = \frac{d}{c}$.

### First-order approximations

When we use a linear function as a mathematical model for a nonlinear real-world input-output process, we say the function represents a *linear model* or a *first-order approximation* for the process. Let's analyze in a little more detail what that means and see why linear models are so popular in science.

In calculus, we learn that functions can be represented as infinite Taylor series:

$$f(x) = \text{taylor}(f(x)) = a_0 + a_1 t + a_2 t^2 + a_3 t^3 + \cdots = \sum_{n=0}^{\infty} a_n x^n,$$

where the coefficient $a_n$ depends on the $n^{\text{th}}$ derivative of $f(x)$. The Taylor series is only equal to the function $f(x)$ if infinitely many terms in the series are calculated. If we sum together only a finite number of terms in the series, we obtain a *Taylor series approximation*. The first-order Taylor series approximation to $f(x)$ is

$$f(x) \approx \text{taylor}_1(f(x)) = a_0 + a_1 x = f(0) + f'(0)x.$$

The above equation describes the best approximation to $f(x)$ near $x = 0$, by a line of the form $l(x) = mx+b$. To build a linear model $f(x)$

of a real-world process, it is sufficient to measure two parameters: the initial value $b \equiv f(0)$ and the rate of change $m \equiv f'(0)$.

The reason why scientists use linear models routinely is because they allow for easy *parametrization.* To build a linear model, the first step is to establish the initial value $f(0)$ by inputting $x = 0$ to the process and seeing what comes out. Next, we vary the input by some amount $\Delta x$ and observe the resulting change in the output $\Delta f$. The rate of change parameter is equal to the change in the output divided by the change in the input $m = \frac{\Delta f}{\Delta x}$. Thus, in two simple steps we can obtain the parameters of a linear model. In contrast, finding the parametrization of nonlinear models is a more complicated task.

For a function $F(x, y, z)$ that takes many variables as inputs, the first-order Taylor series approximation is

$$F(x, y, z) \approx b + m_x x + m_y y + m_z z.$$

Except for the constant term, the function has the form of a linear expression. The "first-order approximation" to a function of $n$ variables $F(x_1, x_2, \ldots, x_n)$ has the form $b + m_1 x_1 + m_2 x_2 + \cdots + m_n x_n$.

As in the single-variable case, finding the parametrization of a multivariable linear model is a straightforward task. Suppose we want to model some complicated real-world phenomenon that has $n$ input variables. We proceed as follows. First, we input zero to obtain the initial value $F(0, \ldots, 0) \equiv b$. Next, we go through each of the input variables one-by-one and measure how a small change in each input $\Delta x_i$ affects the output $\Delta f$. The rate of change with respect to the input $x_i$ is $m_i = \frac{\Delta f}{\Delta x_i}$. By combining the knowledge of the initial value $b$ and the "slopes" with respect to each input parameter, we'll obtain a complete linear model of the phenomenon.

## Discussion

In the next three chapters of this book, we'll learn about new mathematical objets and mathematical operations. Linear algebra is the study of objects like vectors, matrices, linear transformations, vector spaces, etc. Most of the new mathematical operations we'll perform on these objects will linear: $f(\alpha\mathbf{obj}_1 + \beta\mathbf{obj}_2) = \alpha f(\mathbf{obj}_1) + \beta f(\mathbf{obj}_2)$. Linearity is the core assumption of linear algebra.

Our journey of all things linear begins with the computational aspects of linear algebra. In Chapter 4 we'll learn how to efficiently solve large systems of linear equations, practice computing matrix products, learn about the matrix determinant, and how to compute the matrix inverse.

# Chapter 4

# Computational linear algebra

This chapter covers the computational aspects of performing matrix calculations. Suppose we're given a huge matrix $A \in \mathbb{R}^{n\times n}$ with $n > 1000$. Hidden behind the innocent-looking mathematical notation of the matrix inverse $A^{-1}$, the matrix product $A^2$, and the matrix determinant $|A|$, there are monster computations involving all the $1000\times 1000 = 1$ million entries of the matrix $A$. Millions of arithmetic operations must be performed!

Okay, calm down. I won't make you do millions of arithmetic operations. To learn algebra, it is sufficient to know how to compute things with $3\times 3$ and $4\times 4$ matrices. Even for such moderately sized matrices, computing products, inverses, and determinants by hand are serious computational task. If you have to write a linear algebra final exam, you need to make sure you can do these calculations quickly. Even if you don't have an exam, it's important to practice doing matrix operations by hand to get a "feel" of what they're like.

**Gauss–Jordan elimination** Suppose we're trying to solve a system of two linear equations in two unknowns $x$ and $y$:

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

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

$$\begin{array}{rcrcl} ax & + & by & = & c, \\ (d+\alpha a)x & + & (e+\alpha b)y & = & f+\alpha c. \end{array}$$

This is called a *row operation*: we added $\alpha$-times the first row to the second row. Row operations change the coefficients in the linear equations but leave the solutions 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 4.3 we'll "unpack" this formula and learn about its intuitive interpretation: 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 $|A|$, is an operation that provides us with 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 all these aspects of determinants in Section 4.4.

**Matrix inverse** In Section 4.5 we'll build upon our knowledge of Gauss–Jordan elimination, matrix products, and determinants to derive three different procedures for computing the matrix inverse $A^{-1}$.

## 4.1 Reduced row echelon form

In this section we'll learn how to solve systems of linear equations using the *Gauss–Jordan elimination* procedure. A system of equations can be represented as a matrix of coefficients. The Gauss–Jordan elimination procedure converts any matrix into its *reduced row echelon form* (RREF). We can easily read off the solution of the system of equations from the RREF.

This section requires your full-on caffeinated attention because the procedures you'll learn are somewhat tedious. Gauss–Jordan elimination involves a lot of repetitive mathematical manipulations of arrays of numbers. It is important for you to suffer through the steps and verify each step presented below on your own with pen and paper. You shouldn't trust me—always verify the steps!

### Solving equations

Suppose you're asked to solve the following system of equations:

$$\begin{aligned} 1x_1 + 2x_2 &= 5, \\ 3x_1 + 9x_2 &= 21. \end{aligned}$$

The standard approach is to use substitution or subtraction tricks to combine the equations and find the values of the two unknowns $x_1$ and $x_2$. We learned about these tricks in Section **??**.

Observe that the *names* of the two unknowns are irrelevant to the solution of these equations. Indeed, the solution $(x_1, x_2)$ to the above equations would be the same as the solution $(s, t)$ in the following system of equations:

$$\begin{aligned} 1s + 2t &= 5, \\ 3s + 9t &= 21. \end{aligned}$$

The important parts of a system of linear equations are the *coefficients* in front of the variables and the numbers in the column of constants on the right-hand side of each equation.

### Augmented matrix

The above system of linear equations can be written as an *augmented matrix*:

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 3 & 9 & 21 \end{array}\right].$$

The first column corresponds to the coefficients of the first variable, the second column is for the second variable, and the last column corresponds to the constants of the right-hand side. It is customary to draw a vertical line where the equal signs in the equations would normally appear. This line helps to distinguish the coefficients of the equations from the column of constants on the right-hand side.

Once we have the augmented matrix, we can use *row operations* on its entries to simplify it. In the last step, we use the correspondence between the augmented matrix and the systems of linear equations to read off the solution.

After simplification by row operations, the above augmented matrix will look like this:

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

This augmented matrix corresponds to the following system of equations:

$$\begin{aligned} x_1 &= 1, \\ x_2 &= 2. \end{aligned}$$

This is a *trivial* system of equations; there is nothing left to solve! We can simply read off the solutions $x_1 = 1$ and $x_2 = 2$. This simple example illustrates the general idea of the Gauss–Jordan elimination procedure for solving systems of equations by manipulating an augmented matrix.

The augmented matrix approach to solving linear equations is very convenient for solving equations with many variables.

## 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:

1. Add a multiple of one row to another row
2. Swap two rows
3. Multiply a row by a constant

Let's trace the sequence of row operations we would need to solve the system of linear equations which we described above.

- We start with the augmented matrix:

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 3 & 9 & 21 \end{array}\right].$$

- As a first step, we'll eliminate the first variable in the second row. We do this by subtracting three times the first row from the second row:

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 3 & 6 \end{array}\right].$$

  We can denote this row operation: $R_2 \leftarrow R_2 - 3R_1$.
- To simplify the second row we divide it by three: $R_2 \leftarrow \frac{1}{3}R_2$:

$$\left[\begin{array}{cc|c} 1 & 2 & 5 \\ 0 & 1 & 2 \end{array}\right].$$

- The final step is to eliminate the second variable from the first row. We do this by subtracting two times the second row from the first row, $R_1 \leftarrow R_1 - 2R_2$:

$$\left[\begin{array}{cc|c} 1 & 0 & 1 \\ 0 & 1 & 2 \end{array}\right].$$

  We can now read off the solution: $x_1 = 1$ and $x_2 = 2$.

The procedure we used to find simplify the augmented matrix and get the solution was not random. We followed the Gauss–Jordan elimination algorithm to bring the matrix into its *reduced row echelon form*.

The *reduced row echelon form* (RREF) is in some sense the simplest form for a matrix. Each row contains a *leading one*, also known as a *pivot*. The pivot of each column is used to eliminate the numbers below and above it in the same column. The end result of this procedure is the reduced row echelon form:

$$\left[\begin{array}{cccc|c} 1 & 0 & * & 0 & * \\ 0 & 1 & * & 0 & * \\ 0 & 0 & 0 & 1 & * \end{array}\right].$$

The asterisks $*$ denote arbitrary numbers that could not be eliminated because no leading one is present in these columns.

## Definitions

- The *solution* to a system of linear equations in the variables $x_1, x_2$ is the set of values $\{(x_1, x_2)\}$ 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 the entries below the leading ones are zero. This 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 which contains the pivots from the rows above them.

## Gauss–Jordan elimination algorithm

Forward phase (left to right):

1. Obtain a pivot (leading one) in the leftmost column.
2. Subtract this row from all rows below this one to get zeros below in the entire column.
3. Look for a leading one in the next column and repeat.

Backward phase (right to left):

1. Find the rightmost pivot and use it to eliminate all the numbers above it in the column.
2. Move one step to the left and repeat.

**Example** We're asked to solve the following system of equations:

$$\begin{aligned} 1x + 2y + 3z &= 14, \\ 2x + 5y + 6z &= 30, \\ -1x + 2y + 3z &= 12. \end{aligned}$$

**E4.2** Find the solutions to the systems of equations that corresponds to the following augmented matrices:

$$(1)\left[\begin{array}{cc|c} -1 & -2 & -2 \\ 3 & 3 & 0 \end{array}\right],\quad (2)\left[\begin{array}{ccc|c} 1 & -1 & -2 & 1 \\ -2 & 3 & 3 & -1 \\ -1 & 0 & 1 & 2 \end{array}\right],\quad (3)\left[\begin{array}{ccc|c} 2 & -2 & 3 & 2 \\ 1 & -2 & -1 & 0 \\ -2 & 2 & 2 & 1 \end{array}\right].$$

**4.2** (1) $(-2, 2)$, (2) $(-4, -1, -2)$, (3) $(\frac{-2}{5}, \frac{-1}{2}, \frac{3}{5})$.

In this section you learned a practical computational algorithm for solving systems of equations by a using row operations on an augmented matrix build from the coefficients of the equations. In the next section we'll increase the level of abstraction: if we "zoom out" one level we can see the entire system of equations as a matrix equation $A\vec{x} = \vec{b}$, and solve the problem using in one step: $\vec{x} = A^{-1}\vec{b}$.

## 4.2 Matrix equations

The problem of solving a system of linear equations can be expressed as a matrix equation, which can be solved using the matrix inverse. Consider the following system of linear equations:

$$\begin{aligned} x_1 + 2x_2 &= 5, \\ 3x_2 + 9x_2 &= 21. \end{aligned}$$

We can rewrite this system of equations using the matrix-vector product:

$$\begin{bmatrix} 1 & 2 \\ 3 & 9 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 5 \\ 21 \end{bmatrix},$$

or more compactly as

$$A\vec{x} \;=\; \vec{b},$$

where $A$ is a $2\times2$ matrix, $\vec{x}$ is the vector of unknowns (a $2\times1$ matrix), and $\vec{b}$ is a vector of constants (a $2 \times 1$ matrix).

We can solve for $\vec{x}$ in this matrix equation by multiplying both sides of the equation by the inverse $A^{-1}$:

$$A^{-1}A\vec{x} = \mathbb{1}\vec{x} = \vec{x} = A^{-1}\vec{b}.$$

Thus, to solve a system of linear equations, we can find the inverse of the matrix of coefficients and then compute the product:

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = A^{-1}\vec{b} = \begin{bmatrix} 3 & -\frac{2}{3} \\ -1 & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 5 \\ 21 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

The computational "cost" of finding $A^{-1}$ is equivalent to the computational cost of brining an augmented array $[\,A \mid \vec{b}\,]$ to reduced row echelon form, so it's not like we get the solution for free. Nevertheless, expressing the system of equations as $A\vec{x} = \vec{b}$ and its solution as $\vec{x} = A^{-1}\vec{b}$ useful level of abstraciton which prevents us from having to deal with all dozens individual coefficients. The same symbolic expression $\vec{x} = A^{-1}\vec{b}$ applies no matter whether $A$ is a $2 \times 2$ matrix or a $1000 \times 1000$ matrix.

## Introduction

We'll no go into an important discussion about matrix equations and how they differ from regular equations with numbers. If $a$, $b$, and $c$ are three numbers, and I told you to *solve* for $a$ in the equation

$$ab = c,$$

then you would know to tell me that the answer is $a = c/b = c\frac{1}{b} = \frac{1}{b}c$ and that would be the end of it.

Now suppose that $A$, $B$, and $C$ are matrices and you want to solve for $A$ in the matrix equation

$$AB = C.$$

The naive answer $A = C/B$ is not allowed: so far, we have defined a matrix *multiplication* and matrix *inversion*, but not matrix *division*. Instead of division, we must do a multiplication by $B^{-1}$, which plays the role of the "divide by $B$" operation since the product of $B$ and $B^{-1}$ gives the identity matrix:

$$BB^{-1} = \mathbb{1}, \qquad B^{-1}B = \mathbb{1}.$$

When applying the inverse matrix $B^{-1}$ to the equation, we must specify whether we are multiplying from the left or from the right because matrix multiplication is not commutative. What do you think is the right answer for $A$ in the above equations? Is it this one $A = CB^{-1}$ or this one $A = B^{-1}C$?

To solve matrix equations we employ the same technique as we used to solve equations in Chapter 1: undoing the operations that stand in the way of the unknown. Recall that we must always **do the same thing to both sides** when manipulating equations for them to remain true.

With matrix equations it's the same story all over again, but there are two new things you need to keep in mind:

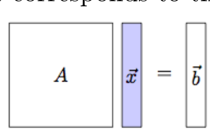
- The order in which matrices are multiplied matters because matrix multiplication is not a commutative operation $AB \neq BA$. The expressions $ABC$ and $BAC$ are different despite the fact that they are the product of the same three matrices.
- When performing operations on matrix equations you can act either *from the left* or *from the right* on the equation.

The best way to get used to the peculiarities of matrix equations is to look at example calculations. Don't worry, there won't be anything too mathematically demanding in this section: we'll just look at pictures.

## Matrix times vector

Suppose we want to solve the equation $A\vec{x} = \vec{b}$, in which an $n \times n$ matrix $A$ multiples the vector $\vec{x}$ to produce a vector $\vec{b}$. Recall that we can think of vectors as "tall and skinny" $n \times 1$ matrices.

The picture that corresponds to the equation $A\vec{x} = \vec{b}$ is




.

Assuming $A$ is invertible, can multiply by the inverse $A^{-1}$ on the left of both sides of the equation:

$$A^{-1} \; A \; \vec{x} = A^{-1} \; \vec{b} \; .$$

By definition, $A^{-1}$ times its inverse $A$ is equal to the identity matrix $\mathbb{1}$, which is a diagonal matrix with ones on the diagonal and zeros everywhere else:

$$I \; \vec{x} = A^{-1} \; \vec{b} \; .$$

Any vector times the identity matrix remains unchanged so

$$\vec{x} = A^{-1} \; \vec{b} \; ,$$

which is our final answer.

Note that the question "Solve for $\vec{x}$ in $A\vec{x} = \vec{b}$" can sometimes be asked in situations where the matrix $A$ is not invertible. If the system of equations is under-specified ($A$ is wider than it is tall), then there will be a whole subspace of acceptable solutions $\vec{x}$. We learned about infinite solution (lines and planes) in the previous section.

If the system is over-specified ($A$ is taller than it is wide) there will be no solution. Nevertheless we might be interested in finding the *best fit* vector $\vec{x}$ such that $A\vec{x} \approx \vec{b}$. One way to quantify the notion of "best fit solution" is called the least-squares solution where we look for the $\vec{x}_*$ such that square of the length of the error is minimized $\vec{x}^* = \text{argmin}_{\vec{x}} e(\vec{x}) = \|A\vec{x} - \vec{b}\|^2$. Such approximate solutions are of great practical importance in much of science. We'll describe how to compute *least-squares fit* solutions in Section 8.3.

## Matrix times matrix

Let's continue looking at some more matrix equations. Suppose we want to solve for $A$ in the equation $AB = C$:

$$A \; B = C \; .$$

To isolate $A$, we multiply by $B^{-1}$ *from the right* on both sides:

$$A \; B \; B^{-1} = C \; B^{-1} \; .$$

When $B^{-1}$ hits $B$ they cancel ($BB^{-1} = \mathbb{1}$) and we obtain the answer:

$$A = C \; B^{-1} \; .$$

## Matrix times matrix variation

What if we want to solve for $B$ in the same equation $AB = C$?

$$A \; B = C \; .$$

# Chapter 5

# Geometrical aspects of linear algebra

In the introduction of the book, I stated several times that understanding the boring and tedious matrix computation techniques like the Gauss–Jordan elimination procedure will allow you to learn about lots of *cool stuff*. I was not lying.

In this section we'll discuss geometrical objects like infinite lines, infinite planes, and vector spaces. We'll use what we learned about vectors to develop an intuitive understanding of these geometrical objects. We'll see how vector and matrix operations can be used to perform geometric calculations like projections and distance computations.

Developing geometrical intuition about the problems in linear algebra is very important, because this is the only knowledge in this course that will stay with you in the long term. Years from now, you may have forgotten the details of the Gauss–Jordan elimination procedure, but you'll still remember that the solution to 3 linear equations in 3 variables corresponds to the intersection of 3 planes in $\mathbb{R}^3$.

## 5.1 Lines and planes

We'll now learn about *points*, *lines*, and *planes*. The purpose of this section is to help you understand these geometrical objets, both in terms of the equations that describe them as well as to visualize what they look like.

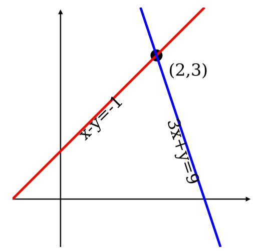
### Concepts

- $p = (p_x, p_y, p_z)$: 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 $\ell$ is a one-dimensional space defined by:
    - ▷ $\ell : \{p_o + t\,\vec{v}, t \in \mathbb{R}\}$: a *parametric equation* of a line with direction vector $\vec{v}$ passing through the point $p_o$
    - ▷ $\ell : \left\{ \frac{x - p_{ox}}{v_x} = \frac{y - p_{oy}}{v_y} = \frac{z - p_{oz}}{v_z} \right\}$: a *symmetric equation*
- An infinite plane $P$ is a two-dimensional space defined by:
    - ▷ $P : \{Ax + By + Cz = D\}$: a *general equation* of a plane
    - ▷ $P : \{\vec{n} \cdot [(x, y, z) - p_o] = 0\}$: a *geometric equation* of the plane that contains point $p_o$ and has normal vector $\hat{n}$
    - ▷ $P : \{p_o + s\,\vec{v} + t\,\vec{w},\ s, t \in \mathbb{R}\}$: a *parametric equation*
- $d(a, b)$: the shortest *distance* between geometric objects $a$ and $b$

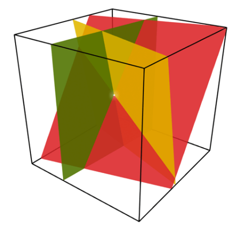
## Points

We can specify a point in $\mathbb{R}^3$ by its coordinates $p = (p_x, p_y, p_z)$, which is similar to how we specify vectors. In fact, the two notions are equivalent: we can either talk about the destination point $p$ or the vector $\vec{p}$ that takes us from the origin to the point $p$. By this equivalence, it makes sense to add vectors and points. For example, $\vec{d} = q - p$ denotes the displacement vector that takes the point $p$ to the point $q$.

We can also specify a point as the intersection of two lines. For an example in $\mathbb{R}^2$, let's define $p = (p_x, p_y)$ to be the intersection of the lines $x - y = -1$ and $3x + y = 9$. We must solve the two equations in parallel to find the point $p$. In other words, we are looking for a point which lies on both lines. We can use the standard techniques for solving equations to find the answer. The intersection point is $p = (2, 3)$.




In three dimensions, a point can also be specified as the intersection of three planes. This is precisely what is going on when we are solving equations of the form



$$\begin{aligned} A_1x + B_1y + C_1z &= D_1, \\ A_2x + B_2y + C_2z &= D_2, \\ A_3x + B_3y + C_3z &= D_3. \end{aligned}$$

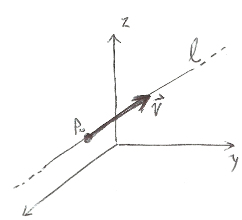
To solve this system of equations, we must find the point $(x, y, z)$ that satisfies all three equation, which means this point is contained in all three planes.

**Example 1** Find where the lines $x + 2y = 5$ and $3x + 9y = 21$ intersect. To find the intersection point we must solve these equations simultaneously: we must find the point $(x, y)$ that lies on both lines. The answer is the point $p = (1, 2)$.

## Lines

A line $\ell$ is a one-dimensional space that is infinitely long. There are many ways to specify the equation of a line in space.

The *parametric equation* of a line is obtained as follows. Given a direction vector $\vec{v}$ and some point $p_o$ on the line, we can define the line as:




$$\ell: \; \{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z) = p_o + t\vec{v}, t \in \mathbb{R}\}.$$

We say the line is *parametrized* by the variable $t$. The line consists of all the points $(x, y, z)$ which can be reached starting from the point $p_o$ and adding any multiple of the direction vector $\vec{v}$.

The *symmetric equation* is an equivalent way for describing a line that does not require an explicit parametrization. Consider the equations that correspond to each of the coordinates in the parametric equation of the line:

$$x = p_{ox} + t\,v_x, \quad y = p_{oy} + t\,v_y, \quad z = p_{oz} + t\,v_z.$$

When we solve for $t$ in these equations and equate the results, we obtain the *symmetric equation* for a line:

$$\ell: \; \left\{ \frac{x - p_{ox}}{v_x} = \frac{y - p_{oy}}{v_y} = \frac{z - p_{oz}}{v_z} \right\},$$

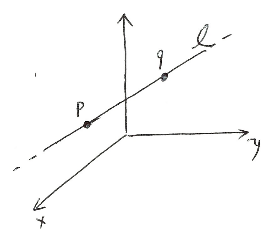
in which the parameter $t$ does not appear. The symmetric equation specifies the line as the relationships between the $x$, $y$, and $z$ coordinates that holds for all points on the line.

You're probably most familiar with the symmetric equation of lines in $\mathbb{R}^2$ when there is no $z$ variable. For non-vertical lines in $\mathbb{R}^2$ ($v_x \neq 0$), we can think of $y$ as being a function of $x$ and write the equation of the line in the equivalent form:

$$\frac{x - p_{ox}}{v_x} = \frac{y - p_{oy}}{v_y} \qquad \Rightarrow \qquad y(x) = mx + b,$$

where $m = \frac{v_y}{v_x}$ and $b = p_{oy} - \frac{v_y}{v_x} p_{ox}$. The equation $m = \frac{v_y}{v_x}$ makes sense intuitively: the slope of a line $m$ corresponds to how much the line "moves" the $y$-direction divided by how much the line "moves" in the $x$-direction.

Another way to describe a line is to specify two points that are part of the line. The equation of a line that contains the points $p$ and $q$ can be obtained as follows:



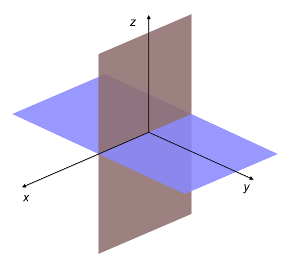

$$\ell : \ \{\vec{x} = p + t\,(q - p), \ t \in \mathbb{R}\},$$

where $(q - p)$ plays the role of the direction vector $\vec{v}$ for this line.

To understand this formula, recall that any vector in the same direction as the line can be used as the direction vector for this line. Therefore, we can build a parametric equation of the line using the vector that points from $p$ to $q$ as the direction vector.

### Lines as plane intersections

In three dimensions, the intersection of two non-parallel planes forms a line. For example, the intersection of the $xy$-plane $P_{xy}$ : $\{(x, y, z) \in \mathbb{R}^3 \mid z = 0\}$ and the $xz$-plane $P_{xz}$ : $\{(x, y, z) \in \mathbb{R}^3 \mid y = 0\}$, is the $x$-axis: $\{(x, y, z) \in \mathbb{R}^3 \mid (0, 0, 0) + (1, 0, 0)t, \ t \in \mathbb{R}\}$. For this simple case, we can imagine the two planes (use your hands) and visually realize that they intersect at a line that is the $x$-axis. Wouldn't it be nice if there was a general procedure for finding the line of intersection of two planes?




You already know such a procedure! The line of intersection between the planes $A_1x + B_1y + C_1z = D_1$ and $A_2x + B_2y + C_2z = D_2$ is the solution of the following set of linear equations:

$$\begin{aligned} A_1x + B_1y + C_1z &= D_1, \\ A_2x + B_2y + C_2z &= D_2. \end{aligned}$$

**Example 2** Find a parametric equation of the line that passes through the points $p = (1, 1, 1)$ and $q = (2, 3, 4)$. What is the symmetric equation of this line?

Using the direction vector $\vec{v} = q - p = (1, 2, 3)$ and the point $p$ on the line, we can write a parametric equation for this line as $\{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z) = (1, 1, 1) + t(1, 2, 3), t \in \mathbb{R}\}$. Note that a parametric equation using the direction vector $(-1, -2, -3)$ would be equally valid: $\{(1, 1, 1) + t(-1, -2, -3), t \in \mathbb{R}\}$. The symmetric equation of this line is $\frac{x-1}{1} = \frac{y-1}{2} = \frac{z-1}{3}$.

**Example 3** Find the intersection of the planes $0x + 0y + 1z = 0$ and $0x + 1y + 1z = 0$. We follow the standard Gauss–Jordan elimination procedure: construct an augmented matrix, perform two row operations (denoted $\sim$), obtain the RREF, and interpret the solution:

$$\left[\begin{array}{ccc|c} 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right] \sim \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

The first column is a free variable $t \in \mathbb{R}$. The solution is the line

$$\left\{ \begin{array}{l} x = t \\ y = 0 \ , \quad \forall t \in \mathbb{R} \\ z = 0 \end{array} \right\} = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + t \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \forall t \in \mathbb{R} \right\},$$

which corresponds to the $x$-axis.

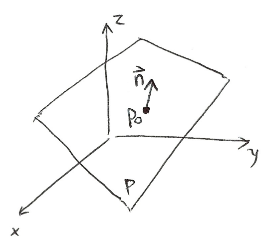
## Planes

A plane $P$ in $\mathbb{R}^3$ is a two-dimensional space with infinite extent. In general, we specify a plane as a constraint satisfied by all points in the plane:

$$P: \ \{(x, y, z) \in \mathbb{R}^3 \mid Ax + By + Cz = D\}.$$

The plane $P$ is the set of all points $(x, y, z) \in \mathbb{R}^3$ that satisfy the equation $Ax + By + Cz = D$. The equation $Ax + By + Cz = D$ is called the *general equation* of the plane. This definition represents the algebraic view of planes, which is useful for algebraic calculations with planes.

There is an equally useful geometric view of planes. A plane can be specified by a *normal vector* $\vec{n}$ and some point $p_{\text{o}}$ in the plane. The normal vector $\vec{n}$ is perpendicular to the plane: it sticks out at right angles to the plane like the normal force between two surfaces. All the points in the plane $P$ can be obtained starting from the point $p_{\text{o}}$ and moving in a direction orthogonal to the normal vector $\vec{n}$. The formula in compact notation is




$$P: \quad \vec{n} \cdot [(x, y, z) - p_{\text{o}}] = 0.$$

Recall that the dot product of two vectors is zero if and only if these vectors are orthogonal. In the above equation, the expression $[(x, y, z) - p_{\text{o}}]$ forms an arbitrary vector with one endpoint at $p_{\text{o}}$. From all these vectors we select *only* those that are perpendicular to $\vec{n}$ and thus we obtain all the points of the plane.

The geometric equation $\vec{n} \cdot [(x, y, z) - p_{\text{o}}] = 0$ is equivalent to the general equation $Ax + By + Cz = D$. Calculating the dot product we find $A = n_x, B = n_y, C = n_z$, and $D = \vec{n} \cdot p_{\text{o}} = n_x p_{\text{o}x} + n_y p_{\text{o}y} + n_y p_{\text{o}z}$.

Observe that scaling the general equation of a plane by a constant factor does not change the plane: the equations $Ax + By + Cz = D$ and $\alpha Ax + \alpha By + \alpha Cz = \alpha D$ define the same plane. Similarly the geometric equations $\vec{n} \cdot [(x, y, z) - p_{\text{o}}] = 0$ and $\alpha\vec{n} \cdot [(x, y, z) - p_{\text{o}}] = 0$ define the same plane.

We can also give a parametric description of a plane $P$. Suppose we know a point $p_{\text{o}}$ in the plane and two linearly independent vectors $\vec{v}$ and $\vec{w}$ that lie inside the plane, then a parametric equation for the plane is

$$P: \ \{(x, y, z) \in \mathbb{R}^3 \mid (x, y, z) = p_{\text{o}} + s\,\vec{v} + t\,\vec{w}, \ s, t \in \mathbb{R}\}.$$

Since a plane is a two-dimensional space, we need two parameters ($s$ and $t$) to describe the location of arbitrary points in the plane.

# Chapter 6

# Linear transformations

Linear transformations are one of the central ideas in linear algebra. Understanding linear transformations is the cornerstone that will connect and unite all the seemingly unrelated concepts which we studied until now.

We previously introduced linear transformations informally as a type of "vector functions." In this chapter, we'll introduce linear transformations more formally, describe their properties, and discuss their power and limitations.

In Section 6.2, 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.

In Section 6.3 we'll discuss the relation between bases and matrix representations. We'll learn how the choice of basis for the input and output spaces determines the coefficients of matrix representations. The same linear transformations corresponds to different matrix representations depending on the choice of basis for the input and output spaces.

Finally, in Section 6.4 we'll discuss and characterize the class of *invertible linear transformations*. This section will serve to combine several different topics we have discussed so far in the book and solidify your understanding of linear transformations, their matrix representations, and the fundamental vector spaces we learned about previously.

## 6.1 Linear transformations

In this section we'll study functions that take vectors as inputs and produce vectors as outputs. A function $T$ that takes $n$-dimensional vectors as inputs and produces $m$-dimensional vectors as outputs is denoted $T\colon \mathbb{R}^n \to \mathbb{R}^m$.

We'll restrict our attention to the class of *linear transformations*, which includes most of the useful transformations of analytical geometry: stretchings, projections, reflections, rotations, and combinations of these. Linear transformations are used to describe and model many real-world phenomena in physics, chemistry, biology, and computer science so it's worth spending some time to understand the theory behind them.

### Concepts

Linear transformation are mappings between *vector inputs* and *vector outputs*. We'll use following concepts to describe the input and output spaces:

- $V$: the input vector space of $T$
- $W$: the output 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. We use $T(\vec{v}) = \vec{w}$ to denote the $T$ acting on $\vec{v}$ to produce $\vec{w}$.

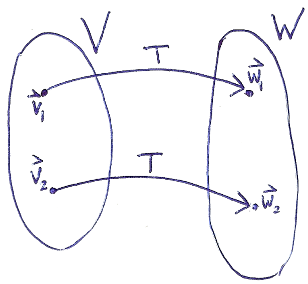



**Figure 6.1:** An illustration of the linear transformation $T : V \to W$.

- $\mathrm{Im}(T)$: 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

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

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

- $\mathrm{Null}(T)$: The *null space* of the linear transformation $T$. This is the set of vectors that get mapped to the zero vector by $T$. Mathematically we write:

$$\mathrm{Null}(T) \equiv \{\vec{v} \in V \mid T(\vec{v}) = \vec{0}\},$$

  and we have $\mathrm{Null}(T) \subseteq V$. The null space is the vector equivalent of the *roots* of a function, the values of $x$ where $f(x) = 0$.

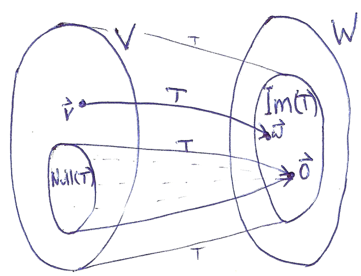



**Figure 6.2:** Illustration of two key properties of a linear transformation $T : V \to W$, its null space $\mathrm{Null}(T) \subseteq V$ and its image space $\mathrm{Im}(T) \subseteq W$.

**Example** Consider the linear transformation $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by the equation $T(x, y) = (x, y, x+y)$. Applying $T$ to the input vector $(1, 0)$ produces the output vector $(1, 0, 1 + 0) = (1, 0, 1)$. Applying $T$ to the input vector $(3, 4)$ produces the output vector $(3, 4, 7)$.

The null space of $T$ contains only the zero vector $\mathrm{Null}(T) = \{\vec{0}\}$. The image space of $T$ is a two-dimensional subspace of the output space $\mathbb{R}^3$, $\mathrm{Im}(T) = \mathrm{span}\{(1, 0, 1), (0, 1, 1)\} \subseteq \mathbb{R}^3$.

## Matrix representations

Given bases for the input and the output spaces, we can represent the action of a linear transformation as a matrix 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 $T$:
$$\vec{w} = T(\vec{v}) \qquad \Leftrightarrow \qquad \vec{w} = M_T\vec{v}.$$
  If we want to be precise, we denote the matrix representation as ${}_{B_W}[M_T]_{B_V}$ to show the input and output bases.
- $\mathcal{C}(M_T)$: the *column space* of a matrix $M_T$
- $\mathcal{N}(M_T)$: the *null space* a matrix $M_T$

## Properties of linear transformation

Before we get into to the details, I'd like to highlight the feature of linear transformations which makes them suitable for modelling a wide range of phenomena in science, engineering, business, and computing.

### Linearity

The fundamental property of linear transformations is—you guessed it—their *linearity.* If $\vec{v}_1$ and $\vec{v}_2$ are two input vectors and $\alpha$ and $\beta$ are two constants then:

$$T(\alpha\vec{v}_1 + \beta\vec{v}_2) = \alpha T(\vec{v}_1) + \beta T(\vec{v}_2) = \alpha\vec{w}_1 + \beta\vec{w}_2,$$

where $\vec{w}_1 = T(\vec{v}_1)$ and $\vec{w}_2 = T(\vec{v}_2)$.

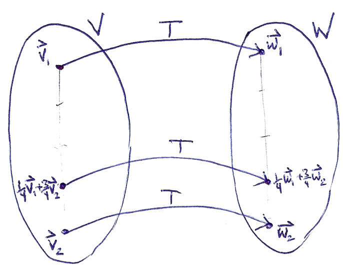



**Figure 6.3:** Linear transformations map linear combination of inputs to the same linear combination of outputs. The figure illustrates how the linear combination $\frac{1}{4}\vec{v}_1 + \frac{3}{4}\vec{v}_2$ gets mapped to $\frac{1}{4}\vec{w}_1 + \frac{3}{4}\vec{w}_2$.

Linear transformations map linear combination of inputs to the same linear combination of outputs. If you know the outputs of $T$ for the inputs $\vec{v}_1$ and $\vec{v}_2$, then you can deduce the output $T$ for any linear combination of the vectors $\vec{v}_1$ and $\vec{v}_2$ by computing the appropriate linear combination of the known outputs.

### Linear transformations as black boxes

Suppose someone gives you a *black box* which implements the linear transformation $T$. You're not allowed to look inside the box to see how $T$ acts, but you're allowed to *probe* the transformation by choosing various input vectors and observing what comes out.

Assume the linear transformation $T$ is of the form $T\colon \mathbb{R}^n \to \mathbb{R}^m$. It turns out that probing this transformation with $n$ linearly independent input vectors and observing the outputs is sufficient to characterize it completely!

To see why this is true, consider a basis $\{\vec{e}_1, \vec{e}_2, \ldots, \vec{e}_n\}$ for the $n$-dimensional input space $V = \mathbb{R}^n$. In order to characterize $T$, all we have to do is input each of the $n$ basis vectors $\vec{e}_i$ into the black box that implements $T$ and record the $T(\vec{e}_i)$ that comes out.

Any input vector $\vec{v}$ can be written as a linear combination of the basis vectors:

$$\vec{v} = v_1\vec{e}_1 + v_2\vec{e}_2 + \cdots + v_n\vec{e}_n.$$

Using these observations and the linearity of $T$, we can predict the output of $T$ for this vector:

$$T(\vec{v}) = v_1T(\vec{e}_1) + v_2T(\vec{e}_2) + \cdots + v_nT(\vec{e}_n).$$

This *black box* model is used in many areas of science and is one of the most important ideas in linear algebra. The transformation $T$ could be the description of a chemical process, an electrical circuit, or some phenomenon in biology. So long as we know that $T$ is (or can be approximated by) a linear transformation, we can obtain a complete description for it by *probing* it with a small number of inputs. This is in contrast to non-linear transformations which correspond to arbitrarily complex input-output relationships and require significantly more *probing* in order to be characterized.

## Input and output spaces

Consider the linear transformation $T$ from $n$-vectors to $m$-vectors:

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

The *domain* of the transformation $T$ is $\mathbb{R}^n$ and the *codomain* is $\mathbb{R}^m$.

The *image space* $\text{Im}(T)$ consists of all possible outputs of the transformation $T$. In general $\text{Im}(T)$ is a subset of the output space, $\text{Im}(T) \subseteq \mathbb{R}^m$. A linear transformation $T$ whose image space is equal to its codomain ($\text{Im}(T) = \mathbb{R}^m$) is called *surjective* or *onto*. Recall that a function is surjective is one that covers the entire output set.

The *null space* of $T$ is the subspace of the domain $\mathbb{R}^n$ that is mapped to the zero vector by $T$: $\text{Null}(T) \equiv \{\vec{v} \in \mathbb{R}^n \mid T(\vec{v}) = \vec{0}\}$. A linear transformations which has a trivial null space $\text{Null}(T) = \{\vec{0}\}$ is called *injective*. Injective transformations map different inputs to different outputs.

If a linear transformation $T$ is both injective and surjective it is called *bijective*. In this case, $T$ is a *one-to-one correspondence* between the input vector space and the output vector space.

Note the terminology we use here to characterize linear transformations (injective, surjective, and bijective) is the same as the terminology we used to characterize functions in Section **??**. Indeed, linear transformations are simply vector functions so it makes sense to use the same terminology.

The concepts of image space and null space of a linear transformation are illustrated in Figure 6.4.

**Observation** The dimensions of the input space and the output space of a bijective linear transformation must be the same. Indeed, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is bijective then it is both injective and surjective. Since $T$ is injective, the output space must be larger or equal to the inputs space $m \geq n$. Since $T$ is surjective, the input space must be at least as large as the output space $n \geq m$. Combining these observations, we find that if $T : \mathbb{R}^n \to \mathbb{R}^m$ is bijective then $m = n$.

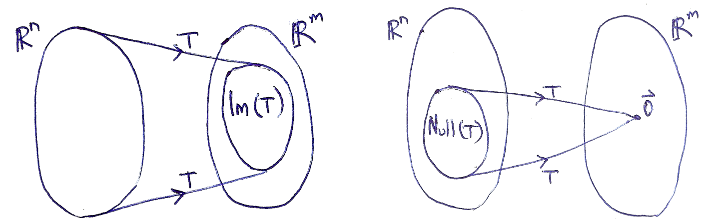



**Figure 6.4:** Pictorial representations of the image space $\text{Im}(T)$ and the null space $\text{Null}(T)$ of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$. The image space is the set of all possible outputs of $T$. The null space is the set of inputs that $T$ maps to the zero vector.

**Example 2** Consider the linear transformation $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by the equation $T(x, y, z) = (x, z)$. Find the null space and the image space of $T$. Is $T$ injective? Is $T$ surjective?

The action of $T$ is to kill the $y$-components of inputs. Any vector which has only a $y$-component will be sent to the zero vector. We have $\text{Null}(T) = \text{span}\{(0, 1, 0)\}$. The image space is $\text{Im}(T) = \mathbb{R}^2$. The transformation $T$ is not injective. To see an explicit example proving $T$ is not injective observe that $T(0, 1, 0) = T(0, 2, 0)$ but $(0, 1, 0) \neq (0, 2, 0)$. Since $\text{Im}(T)$ is equal to the codomain $\mathbb{R}^2$, $T$ is surjective.

## Linear transformations as matrix multiplications

There is an important relationship between linear transformations and matrices. If you fix a basis for the input vector space and a basis for the output vector space, a linear transformation $T(\vec{v}) = \vec{w}$ can be represented as matrix multiplication $M_T\vec{v} = \vec{w}$ for some matrix $M_T$.

We have the following equivalence:

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

Using this equivalence, we can re-interpret several properties of matrices as properties of linear transformations. The equivalence is useful in the other direction too since it allows us to use the language of linear transformations to talk about the properties of matrices.

The idea of representing the action of a linear transformation as a matrix product is extremely important since it allows us to transform the *abstract* description of what the transformation $T$ does into the *practical* description: "take the input vector $\vec{v}$ and multiply it on the left by a matrix $M_T$."

**Example 3** We'll now illustrate the "linear transformation $\Leftrightarrow$ matrix" equivalence with an example. Define $\Pi_{P_{xy}}$ to be the *orthogonal projection* onto the $xy$-plane $P_{xy}$. In words, the action of this projection is simply to "kill" the $z$-component of the input vector. The matrix that corresponds to this projection is

$$T\left(\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}\right) = \begin{bmatrix} v_x \\ v_y \\ 0 \end{bmatrix} \qquad \Leftrightarrow \qquad M_T\vec{v} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \begin{bmatrix} v_x \\ v_y \\ 0 \end{bmatrix}.$$

### Finding the matrix

In order to find the matrix representation of a the transformation $T\colon \mathbb{R}^n \to \mathbb{R}^m$, it is sufficient to probe $T$ with the $n$ vectors in the standard basis for $\mathbb{R}^n$:

$$\hat{e}_1 \equiv \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \hat{e}_2 \equiv \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots, \quad \hat{e}_n \equiv \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

To obtain $M_T$, we combine the outputs $T(\hat{e}_1)$, $T(\hat{e}_2)$, ..., $T(\hat{e}_n)$ as the *columns* of a matrix:

$$M_T = \begin{bmatrix} | & | & & | \\ T(\hat{e}_1) & T(\hat{e}_2) & \ldots & T(\hat{e}_n) \\ | & | & & | \end{bmatrix}.$$

Observe that the matrix constructed in this way has the right dimensions $m \times n$. We have $M_T \in \mathbb{R}^{m\times n}$ since we used $n$ "probe vectors" and since each of the outputs of $T$ are $m$-dimensional column vectors.

To help you visualize this new "column thing," let's analyze what happens when we compute the product $M_T\hat{e}_2$. The probe vector $\hat{e}_2 \equiv (0, 1, 0, \ldots, 0)^{\mathsf{T}}$ will "select" only the second column from $M_T$ and thus we'll obtain the correct output: $M_T\hat{e}_2 = T(\hat{e}_2)$. Similarly, applying $M_T$ to the other basis vectors selects each of the columns of $M_T$.

Any input vector can be written as a linear combination of the standard basis vectors $\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2 + \cdots + v_n\hat{e}_n$. Therefore, by linearity, we can compute the output $T(\vec{v})$:

$$\begin{aligned} T(\vec{v}) &= v_1T(\hat{e}_1) + v_2T(\hat{e}_2) + \cdots + v_nT(\hat{e}_n) \\ &= v_1 \begin{bmatrix} | \\ T(\hat{e}_1) \\ | \end{bmatrix} + v_2 \begin{bmatrix} | \\ T(\hat{e}_2) \\ | \end{bmatrix} + \cdots + v_n \begin{bmatrix} | \\ T(\hat{e}_n) \\ | \end{bmatrix} \\ &= \begin{bmatrix} | & | & & | \\ T(\hat{e}_1) & T(\hat{e}_2) & \ldots & T(\hat{e}_n) \\ | & | & & | \end{bmatrix} \begin{bmatrix} | \\ \vec{v} \\ | \end{bmatrix} \\ &= M_T\vec{v}. \end{aligned}$$

**Input and output spaces**

We can identify some correspondences between the properties of a linear transformation $T$ and the properties of the matrix $M_T$ which implements it.

The outputs of the linear transformation $T$ consist of all possible linear combinations of the columns of the matrix $M_T$. Thus, we can identify the *image space* of the linear transformation $T$ with the *column space* of the matrix $M_T$:

$$\operatorname{Im}(T) = \{\vec{w} \in W \mid \vec{w} = T(\vec{v}), \text{ for some } \vec{v} \in V\} = \mathcal{C}(M_T).$$

There is also an equivalence between the null space of the linear transformation $T$ and the null space of the matrix $M_T$:

$$\operatorname{Null}(T) \equiv \{\vec{v} \in \mathbb{R}^n | T(\vec{v}) = \vec{0}\} = \{\vec{v} \in \mathbb{R}^n | M_T\vec{v} = \vec{0}\} \equiv \mathcal{N}(M_T).$$

The null space of a matrix $\mathcal{N}(M_T)$ consists of all vectors that are orthogonal to the rows of the matrix $M_T$. The vectors in the null space of $M_T$ have a zero dot product with each of the rows of $M_T$. This orthogonality can also be phrased in the opposite direction. Any vector in the *row space* $\mathcal{R}(M_T)$ of the matrix is orthogonal to the null space $\mathcal{N}(M_T)$ of the matrix.

These observation allows us to decompose the domain of the transformation $T$ as the *orthogonal sum* of the null space and the row space of the matrix $M_T$:

$$\mathbb{R}^n \;=\; \mathcal{N}(M_T) \oplus \mathcal{R}(M_T).$$

# Chapter 7

# Theoretical linear algebra

Thus far in the book we focused on the practical aspects of linear algebra. We learned about vector and matrix operations. We learned how to solve systems of linear equations using row operations. We learned about linear transformations and their matrix representations. With this toolset under your belt, you can handle the computational aspects of linear algebra and you're now ready to learn about the more abstract, theoretical concepts.

In math, using abstraction is looking for what mathematical objects have in common. Seeing the parallels between different mathematical structures helps us understand them better. In this chapter we'll extend what we know about the vector space $\mathbb{R}^n$ to the study of abstract vector spaces of vector-like mathematical objects (Section 7.3). We'll learn how to discuss linear independence, find bases, and count dimensions. In Section 7.4 we'll also define an abstract inner product, and use it to define the concepts of orthogonality and length for abstract vectors. We'll also learn about the Gram–Schmidt orthogonalization procedure for distilling a high quality orthonormal basis from any set of linearly independent vectors (Section 7.5).

Another task we'll work on in this chapter is to develop a taxonomy for the different types of matrices (Section 7.2) according to their properties and applications. We'll also learn about matrix decompositions—techniques for splitting matrices as the product of simpler matrices. In Section 7.1 we'll learn about the decomposition of matrices into their "own" basis and "own" values, which is called *eigendecomposition*. Later, we'll learn about a several other decompositions in Section 7.6).

In the last section of this chapter (Section 7.7) we'll learn about vectors and matrices with complex coefficients. This section will also serve as a review of everything we've learned in this book so be sure to read it even if complex numbers are not required for your course.

Throughout the chapter, we won't be concerned so much with calculations but with mind expansion. Let's get right to it.

## 7.1 Eigenvalues and eigenvectors

The set of eigenvectors of a matrix is a special set of input vectors for which the action of the matrix is described as a simple *scaling*. Recall the observations we made in Section 6.2 about how linear transformations act differently on different input spaces. You're also familiar with the special case of the "zero eigenspace," which is called the *null space* of a matrix.

Decomposing a matrix in terms of its eigenvalues and its eigenvectors gives valuable insights into the properties of the matrix. Also, certain matrix calculations, like computing the power of the matrix, become much easier when we use the *eigendecomposition* of the matrix. For example, suppose you are given a square matrix $A$ and you want to compute $A^7$. To make this example more concrete, let's use the matrix

$$A = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}.$$

We want to compute

$$A^7 = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}\begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}\begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}\begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}\begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}\begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}\begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}.$$

That's a lot of matrix multiplications. You'll have to multiply and add entries for a while! Imagine how many times you would have to multiply the matrix if I had asked you to find $A^{17}$ or $A^{77}$ instead?

Let's be smart about this. Every matrix corresponds to some linear operation. This means that it is a legitimate question to ask "what does the matrix $A$ do?" and once we figure out what it does, we can compute $A^{77}$ by simply doing what $A$ does 77 times.

The best way to see what a matrix does is to look inside of it and see what it is made of. What is its *natural basis* (own basis) and what are *its values* (own values). The matrix $A$ can be written as the product of three matrices:

$$A = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}}_{Q}\underbrace{\begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}}_{\Lambda}\underbrace{\begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & -\frac{1}{5} \end{bmatrix}}_{Q^{-1}} = Q\Lambda Q^{-1}.$$

You can multiply these three matrices together and you'll obtain $A$. Note that the "middle matrix" $\Lambda$ (the capital Greek letter *lambda*) has entries only on the diagonal. The diagonal matrix $\Lambda$ is sandwiched between the matrix $Q$ on the left and $Q^{-1}$ (the inverse of $Q$) on the right. This way of writing $A$ allows us to compute $A^7$ in a very civilized manner:

$$\begin{aligned} A^7 &= AAAAAAA \\ &= Q\Lambda \underbrace{Q^{-1}Q}_{\mathbb{1}}\Lambda\underbrace{Q^{-1}Q}_{\mathbb{1}}\Lambda\underbrace{Q^{-1}Q}_{\mathbb{1}}\Lambda\underbrace{Q^{-1}Q}_{\mathbb{1}}\Lambda\underbrace{Q^{-1}Q}_{\mathbb{1}}\Lambda\underbrace{Q^{-1}Q}_{\mathbb{1}}\Lambda Q^{-1} \\ &= Q\Lambda\mathbb{1}\Lambda\mathbb{1}\Lambda\mathbb{1}\Lambda\mathbb{1}\Lambda\mathbb{1}\Lambda\mathbb{1}\Lambda Q^{-1} \\ &= Q\Lambda\Lambda\Lambda\Lambda\Lambda\Lambda\Lambda Q^{-1} \\ &= Q\Lambda^7 Q^{-1}. \end{aligned}$$

Since the matrix $\Lambda$ is diagonal, it's easy to compute its seventh power:

$$\Lambda^7 = \begin{bmatrix} 5 & 0 \\ 0 & 10 \end{bmatrix}^7 = \begin{bmatrix} 5^7 & 0 \\ 0 & 10^7 \end{bmatrix} = \begin{bmatrix} 78125 & 0 \\ 0 & 10000000 \end{bmatrix}.$$

Thus we can express our calculation of $A^7$ as

$$A^7 = \begin{bmatrix} 9 & -2 \\ -2 & 6 \end{bmatrix}^7 = \underbrace{\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}}_{Q}\begin{bmatrix} 78125 & 0 \\ 0 & 10000000 \end{bmatrix}\underbrace{\begin{bmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & -\frac{1}{5} \end{bmatrix}}_{Q^{-1}}.$$

We still have to multiply these three matrices together, but we have brought down the work from 6 matrix multiplications down to just two.

The answer is

$$A^7 = Q\Lambda^7 Q^{-1} = \begin{bmatrix} 8015625 & -3968750 \\ -3968750 & 2062500 \end{bmatrix}.$$

Using this technique, we can compute $A^{17}$ just as easily:

$$A^{17} = Q\Lambda^{17}Q^{-1} = \begin{bmatrix} 80000152587890625 & -39999694824218750 \\ -39999694824218750 & 20000610351562500 \end{bmatrix}.$$

We could even compute $A^{777} = Q\Lambda^{777}Q^{-1}$ if we wanted to! I hope by now you get the point: if you look at $A$ in the right *basis*, repeated multiplication only involves computing the powers of its *eigenvalues*, which is much simpler than carrying out matrix multiplications.

## Definitions

- $A$: an $n \times n$ square matrix. We denote the entries of $A$ 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 $\Lambda$ 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 $\Lambda$ corresponds to the matrix $A$ expressed in its eigenbasis.
- $Q$: a matrix whose columns are the 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

The fundamental eigenvalue equation is

$$A\vec{e}_\lambda = \lambda\vec{e}_\lambda,$$

where $\lambda$ is an eigenvalue and $\vec{e}_\lambda$ is an eigenvector of the matrix $A$. When we multiply $A$ on of its eigenvectors $\vec{e}_\lambda$, the result is the same vector scaled by the constant $\lambda$.

To find the eigenvalues of a matrix we start from the eigenvalue equation $A\vec{e}_\lambda = \lambda\vec{e}_\lambda$, insert the identity $\mathbb{1}$, and rewrite it as a null-space problem:

$$A\vec{e}_\lambda = \lambda\mathbb{1}\vec{e}_\lambda \qquad \Rightarrow \qquad (A - \lambda\mathbb{1})\,\vec{e}_\lambda = \vec{0}.$$

This equation has a solution whenever $|A - \lambda\mathbb{1}| = 0$. The eigenvalues of $A \in \mathbb{R}^{n\times n}$, denoted $(\lambda_1, \lambda_2, \ldots, \lambda_n)$, are the roots of the *characteristic polynomial*:

$$p(\lambda) = \det(A - \lambda\mathbb{1}) \equiv |A - \lambda\mathbb{1}| = 0.$$

# Chapter 8

# Applications

In this chapter we'll learn about a bunch of cool stuff you can do with linear algebra. I'm sure you'll be amazed by reading this chapter—certainly I was amazed while writing it—amazed with how many different areas of science depend on linear algebra.

Don't worry if you're not able to follow all the details in each situation. I'm taking a shotgun approach here and covering topics from many different areas in the hope to hit one some that are of interest to you.

## 8.1 Solving systems of equations

Systems of linear equations come up all over the place. No matter what you choose to study in the future, your *reduced row echelon*-computing techniques will come in handy.

### Input–output models in economics

Suppose you run a small country and you want to make an economic production plan for the coming year. Your job as the leader of the country is to choose the production rates: $x_e$ units of electric power, $x_w$ units of wood, and $x_a$ units of aluminum to satisfy the external demand for these commodities $\vec{d} = (d_e, d_w, d_a)$. The problem is complicated by the interdependence of the industries. For example, it takes 0.3 units of electricity to produce each unit of aluminum so you cannot simply choose the electric production to match the demand ($x_e = d_e$); you must account for internal need for electric power.

In the real world, your decisions about which industry to sponsor and the production rates would be guided by which industry gave you the biggest kickback the previous year. Let's ignore reality for a moment though, and assume you're an honest leader interested in using math to do what is right for the country instead of abusing his/her position of power like a blood thirsty leech.

You can use linear algebra to account for the internal production demands and choose the appropriate production rates. Suppose we can model that the interdependence between the industries is described by the following system of equations:

$$\begin{aligned} x_e &= 25 + 0.05x_w + 0.3x_a, \\ x_w &= 10 + 0.01x_e + 0.01x_a, \\ x_a &= 14 + 0.1x_e. \end{aligned}$$

The first equation indicates that the electric production $x_e$ must be chosen to satisfy the external demand of 25 units plus an additional 0.05 units for each unit of wood produced (electricity needed for saw

# End matter

## Links

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

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

[ Wikibook on the subject (for additional reading) ]
http://en.wikibooks.org/wiki/Linear_Algebra

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

[ Problems and solutions ]
http://en.wikibooks.org/wiki/Linear_Algebra

[ Proofs ]
http://www.proofwiki.org/wiki/Category:Linear_Transformations
http://www.proofwiki.org/wiki/Category:Linear_Algebra
http://www.proofwiki.org/wiki/Category:Matrix_Algebra

[ Free books ]
Peeyush Chandra, A. K. Lal, G. Santhanam , V. raghavendra. *Notes on Linear Algebra*, July 2011
http://home.iitk.ac.in/~arlal/book/nptel/mth102/book.html

[ List of applications of linear algebra ]
http://aix1.uottawa.ca/~jkhoury/app.htm

[ Lots of good examples here ]
http://isites.harvard.edu/fs/docs/icb.topic1011412.files/applications.pdf

[ Thirty-three Miniatures: Mathematical and Algorithmic Applications of Linear Algebra ]
http://kam.mff.cuni.cz/~matousek/stml-53-matousek-1.pdf

# Appendix A

# Notation

This appendix contains a summary of the notation used in this book.

## Math notation

| Expression | Read as | Used to |
|---|---|---|
| $a, b, x, y$ | | denote variables |
| $=$ | is equal to | indicate two expressions are equal in value |
| $\equiv$ | is defined as | define a variable in terms of an expression |
| $a+b$ | $a$ plus $b$ | combine lengths |
| $a-b$ | $a$ minus $b$ | find the difference in length |
| $a \times b \equiv ab$ | $a$ times $b$ | find the area of a rectangle |
| $a^2 \equiv aa$ | $a$ squared | find the area of a square of side length $a$ |
| $a^3 \equiv aaa$ | $a$ cubed | find the volume of a cube of side length $a$ |
| $a^n$ | $a$ exponent $n$ | denote $a$ multiplied by itself $n$ times |
| $\sqrt{a} \equiv a^{\frac{1}{2}}$ | square root of $a$ | find the side length of a square of area $a$ |
| $\sqrt[3]{a} \equiv a^{\frac{1}{3}}$ | cube root of $a$ | find the side of a cube with volume $a$ |
| $a/b \equiv \frac{a}{b}$ | $a$ divided by $b$ | denote parts of a whole |
| $a^{-1} \equiv \frac{1}{a}$ | one over $a$ | denotes division by $a$ |
| $f(x)$ | $f$ of $x$ | denote the output of the function $f$ applied to the input $x$ |
| $f^{-1}$ | $f$ inverse | denote the inverse function of $f(x)$<br>if $f(x) = y$, then $f^{-1}(y) = x$ |
| $e^x$ | $e$ to the $x$ | denote the exponential function base $e$ |
| $\ln(x)$ | natural log of $x$ | logarithm base $e$ |
| $a^x$ | $a$ to the $x$ | denote the exponential function base $a$ |
| $\log_a(x)$ | log base $a$ of $x$ | logarithm base $a$ |
| $\theta, \phi$ | *theta*, *phi* | denote angles |
| $\sin, \cos, \tan$ | sin, cos, tan | obtain trigonometric ratios |
| $\%$ | percent | denote proportions of a total $a\% \equiv \frac{a}{100}$ |

## Set notation

You don't need a lot of fancy notation to understand mathematics. It really helps, though, if you know a little bit of set notation.

| Symbol | Read as | Denotes |
|---|---|---|
| $\{\ \dots\ \}$ | the set ... | define a sets |
| $\vert$ | such that | describe or restrict the elements of a set |
| $\mathbb{N}$ | the naturals | the set $\mathbb{N} \equiv \{0, 1, 2, 3, \dots\}$ |
| $\mathbb{Z}$ | 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 |
| $\mathbb{A}$ | | the set of algebraic numbers |
| $\mathbb{R}$ | | the set of real numbers |
| $\mathbb{C}$ | | the set of complex numbers |
| | | |
| $\subset$ | subset | one set 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 |

# Vectors notation

| Expression | Denotes |
|---|---|
| $\vec{v}$ | 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 $\vec{v}$ |
| $\theta$ | angle the vector $\vec{v}$ makes with the $x$-axis |
| $\hat{v} \equiv \frac{\vec{v}}{\\|\vec{v}\\|}$ | unit length 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 $\vec{v}$ |

# Matrix notation

| Expression | Denotes |
|---|---|
| $A$ | a matrix |
| $a_{ij}$ | entry in the $i^{\text{th}}$ row, $j^{\text{th}}$ column of $A$ |
| $A^{-1}$ | matrix inverse |
| $\mathbb{1}$ | the identity matrix |
| $\vert A\vert$ | determinant of $A$, also denoted $\det(A)$ |
| $AB$ | matrix product |
| $A \sim A'$ | matrix $A'$ obtained from $A$ by *row operations* |

# Appendix B

# Intro to quantum physics

The fundamental principles of quantum mechanics are simple enough to be explained in the space available on the back of an envelope, but to truly understand the implications of these principles takes years of training and effort.

## what is quantum mechanics?

- computer science view
  - ▷ bits
    $x = 0$ *or* $1$ (two level systems)
  - ▷ qubits
    $|x\rangle = \alpha|0\rangle + \beta|1\rangle$ (vectors in $\mathbb{C}^2$)

- physics view
  - ▷ quantum mechanics *is* useful calculations
  - ▷ pauli matrices, Hamiltonians, ground state, entanglement, non-locality, uncertainty
  - ▷ position, energy, momentum, spin, polarization

- philosophy view
  - ▷ ultimate nature of reality

- psychology view?

## simple view

- Quantum mechanics has two "weird" aspects
- Presented on next slides in simplified form

- Think of this as a 5 minute quantum mechanics primer...

### two worlds

- theoretical (quantum) reality and practical (classical) reality

- Some object are assumed to live in the quantum reality
  $|x\rangle, |y\rangle, \ldots$
- We live in the classical reality $i$, $j$, ...
- Classical objects remain classical forever
- Quantum objects can become classical through measurement:
  $position(|x\rangle) = p_x$
  $momentum(|x\rangle) = m_x$
  $spin(|x\rangle) = s_x$
  ...

- Measurement changes the system
  not because of "magic" – just because of size of systems

### interference

- classically:
  apple + apple = 2 apples
  rock + rock = 2 rocks
  $1 + 1 = 2$

- quantum world:
  $|x\rangle + |y\rangle = ?$
  adding vectors remember ...
  The whole thing is just a model so we can't really know...
- quantum as seen through measurement:
  $abs(|x\rangle) = 1, \quad abs(|y\rangle) = 1$
  $abs(|x\rangle + |y\rangle) = 2$ is possible
  but also
  $abs(|x\rangle + |y\rangle) = 0$ !
- Interference occurs when two things annihilate each other in the quantum reality — it is very real

### applications

- What can you do with quantum?
- Quantum cryptography uses keys in the quantum world in order to achieve better security
- Quantum computation performs computations in the quantum world which can sometimes solve hard problems fast
- Quantum error correction is necessary to make computation work
- Quantum network theory deals with multiple senders and receivers of quantum information
- Applications to physics problems

## B.1 Introduction to quantum information

The use of quantum systems for information processing tasks is no more mysterious than the use of digital technology for information processing. The use of an *analog to digital converter* (ADC) to transform an analog signal to a digital representation and the use of a *digital to analog converter* (DAC) to transform from the digital world back into the analog world are similar to the *state preparation* and the *measurement* steps used in quantum information science. The *digital world* is sought after because of the computational, storage and communication benefits associated with manipulation of discrete systems instead of continuous signals. Similarly, there are benefits associated with using the *quantum world* (Hilbert space) in certain computation problems. The use of digital and quantum technology can therefore both be seen operationally as a black box process with information encoding, processing and readout steps.

The focus of this thesis is the study of *quantum* aspects of communication which are relevant for *classical communication* tasks. In order to make the presentation more self-contained, we will present below a brief introduction to the subject which describes how quantum systems are represented, how information can be encoded and how information can be read out.

### Quantum states

In order to describe the *state* of a quantum system $B$ we use a density matrix $\rho^B$ acting on a $d$-dimensional complex vector space $\mathcal{H}^B$ (Hilbert space). To be a density matrix, the operator $\rho^B$ has to be Hermitian, positive semidefinite and have unit trace. We denote the set of density matrices on a Hilbert space $\mathcal{H}^B$ as $\mathcal{D}(\mathcal{H}^B)$.

A common choice of basis for $\mathcal{H}^B$ is the standard basis $\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$:

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

which is also known as the *computational* basis.

In two dimensions, another common basis is the *Hadamard* basis:

$$|+\rangle \equiv \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle, \tag{B.1}$$

$$|-\rangle \equiv \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle. \tag{B.2}$$

## B.2 Dirac notation

### Vectors

Let $\alpha, \beta \in \mathbb{C}$ and $\vec{v}, \vec{w}, \vec{a}, \vec{b}$ be vectors in $\mathbb{C}^d$, then the Dirac notation for them is as follows:

$$\vec{v} \equiv |v\rangle \qquad \text{(called a “ket”)}$$

The standard basis $\mathcal{B}_Z = \{\vec{e}_0, \vec{e}_1, \ldots, \vec{e}_{d-1}\}$ for $\mathbb{C}^d$ is denoted:

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

For a *qubit* ($d = 2$) we have:

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \equiv |0\rangle, |1\rangle$$
$$[1,0],\ [0,1] \equiv \langle 0|, \langle 1|$$
$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \equiv \alpha|0\rangle + \beta|1\rangle$$

The dagger operation ($\dagger$) is transpose ($^\intercal$) + complex conjugate ($\bar{\ }$):

$$\overline{(\vec{v}^\intercal)} = (\overline{\vec{v}})^\intercal = \vec{v}^\dagger \equiv \langle v| = |v\rangle^\dagger \qquad (\langle v| \text{ is called a “bra”})$$
$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix}^\dagger = [\overline{\alpha}, \overline{\beta}] \equiv \overline{\alpha}\langle 0| + \overline{\beta}\langle 1| = (\alpha|0\rangle + \beta|1\rangle)^\dagger$$

In physics, we use an inner product $(.,.)$ with $\dagger$ on the first entry:

$$\sum_{i=0}^{d-1} \overline{a_i} b_i \equiv (\vec{a}, \vec{b}) \equiv \vec{a}^\dagger \vec{b} \equiv \langle a\rangle b$$
$$(\vec{v}, \alpha\vec{a} + \beta\vec{b}) \equiv \alpha\vec{v}^\dagger\vec{a} + \beta\vec{v}^\dagger\vec{b} \equiv \alpha\langle v\rangle a + \beta\langle v\rangle b = \langle v|\big(\alpha|a\rangle + \beta|b\rangle\big)$$
$$(\alpha\vec{a}+\beta\vec{b}, \vec{w}) \equiv \overline{\alpha}\vec{a}^\dagger\vec{w}+\overline{\beta}\vec{b}^\dagger\vec{w} \equiv \overline{\alpha}\langle a\rangle w+\overline{\beta}\langle b\rangle w = \big(\overline{\alpha}\langle a|+\overline{\beta}\langle b|\big)|w\rangle$$
$$\|\vec{v}\| = \sqrt{\vec{v}^\dagger\vec{v}} \equiv \sqrt{\langle v\rangle v} = ||v\rangle|$$

## B.3 Change of basis

Let $\vec{v} = [v_0, v_1, \ldots, v_{d-1}]^\intercal$, where $v_i \in \mathbb{C}$ are the coefficients of $\vec{v}$ with respect to the standard basis $\mathcal{B}_Z = \{\vec{e}_i\}_{i=0,\ldots,d-1}$. We can calculate each coefficient using the inner product:

$$[\underbrace{0,\ldots,}_{i}1,0,\ldots]\begin{bmatrix} v_0 \\ v_1 \\ \vdots \\ v_{d-1} \end{bmatrix} = \vec{e}_i^\dagger\vec{v} \equiv v_i = \langle i\rangle v$$
$$\begin{aligned}\vec{v} = [v_0, v_1, \ldots, v_{d-1}]^\intercal &\equiv v_0|0\rangle + v_1|1\rangle + \cdots + v_{d-1}|d-1\rangle \\ &= \langle 0\rangle v|0\rangle+\langle 1\rangle v|1\rangle + \cdot + \langle d-1\rangle v|d-1\rangle\end{aligned}$$

The last expression explicitly shows that the basis $\mathcal{B}_Z = \{|i\rangle\}_{i=0,\ldots,d-1}$ was used. This comes in handy when using a different choice of basis like the Hadamard basis for example $\mathcal{B}_X = \{\vec{h}_0, \vec{h}_1\}$ when $d = 2$:

$$\frac{1}{\sqrt{2}}\vec{e}_0 + \frac{1}{\sqrt{2}}\vec{e}_1 = \vec{h}_0 \equiv |+\rangle = \frac{1}{\sqrt{2}}|0\rangle + \frac{1}{\sqrt{2}}|1\rangle$$
$$\frac{1}{\sqrt{2}}\vec{e}_0 - \frac{1}{\sqrt{2}}\vec{e}_1 = \vec{h}_1 \equiv |-\rangle = \frac{1}{\sqrt{2}}|0\rangle - \frac{1}{\sqrt{2}}|1\rangle$$

To get the coefficients of $\vec{v}$ with respect to the Hadamard basis:

$$\begin{aligned}\vec{v} = [v_+, v_-]^\intercal_{\mathcal{B}_X} &\equiv v_+|+\rangle + v_-|-\rangle \\ &= \langle +\rangle v|+\rangle + \langle -\rangle v|-\rangle.\end{aligned}$$

In the bra-ket notation, the “$X+$ coefficient” w.r.t. $\mathcal{B}_X$ is $v_+ \equiv \langle +\rangle v$ and the “$X-$ coefficient” w.r.t. $\mathcal{B}_X$ is $v_- \equiv \langle -\rangle v$.

Given a vector $\vec{v} = [v_0, v_1]$ expressed in terms of its coefficients with respect to the standard basis $\mathcal{B}_Z$, performing the change of basis operation is very intuitive in ket notation:

$$\begin{aligned}
\vec{v} &\equiv [v_+, v_-]^{\mathsf{T}}_{\mathcal{B}_X} \\
&\equiv \langle + \rangle v|+\rangle + \langle - \rangle v|-\rangle \\
&= \langle +|\big(v_0|0\rangle + v_1|1\rangle\big)|+\rangle + \langle -|\big(v_0|0\rangle + v_1|1\rangle\big)|-\rangle \\
&= \big(v_0\langle +||0\rangle + v_1\langle +||1\rangle\big)|+\rangle + \big(v_0\langle -||0\rangle + v_1\langle -||1\rangle\big)|-\rangle \\
&= \underbrace{\frac{1}{\sqrt{2}}(v_0 + v_1)}_{v_+}|+\rangle + \underbrace{\frac{1}{\sqrt{2}}(v_0 - v_1)}_{v_-}|-\rangle.
\end{aligned}$$

## Matrices

Consider a linear transformation $T : \mathbb{C}^n \to \mathbb{C}^n$. It can be represented by a matrix ${}_{\mathcal{B}_\ell}[A_T]_{\mathcal{B}_r}$ with respect to a "left basis" and a "right basis". The coefficients of the matrix are computed with respect to these bases.

Consider for example a transformation on qubits $T : \mathbb{C}^2 \to \mathbb{C}^2$, which be expressed as a matrix $A_T$ with respect to the standard basis as follows:

$$\begin{aligned}
{}_{\mathcal{B}_Z}\begin{bmatrix} a_{00} & a_{01} \\ a_{10} & a_{11} \end{bmatrix}_{\mathcal{B}_Z} &\equiv {}_{\mathcal{B}_Z}[A_T]_{\mathcal{B}_Z} \\
&\equiv a_{00}|0\rangle\langle 0| + a_{01}|0\rangle\langle 1| + a_{10}|1\rangle\langle 0| + a_{11}|1\rangle\langle 1| \\
&= {}_{\mathcal{B}_Z}\begin{bmatrix} \langle 0|A|0\rangle & \langle 0|A|1\rangle \\ \langle 1|A|0\rangle & \langle 1|A|1\rangle \end{bmatrix}_{\mathcal{B}_Z}
\end{aligned}$$

Observe that in the Dirac notation we the basis appears explicitly. In a different basis, the same transformation $T$ will correspond to a different matrix $A_T$:

$${}_{\mathcal{B}_X}\begin{bmatrix} a_{++} & a_{+-} \\ a_{-+} & a_{--} \end{bmatrix}_{\mathcal{B}_X} \equiv A_T \equiv {}_{\mathcal{B}_X}\begin{bmatrix} \langle +|A|+\rangle & \langle +|A|-\rangle \\ \langle -|A|+\rangle & \langle -|A|-\rangle \end{bmatrix}_{\mathcal{B}_X}.$$

# B.4 Superoperators = operators on matrices

Kraus opertors Choi-Jamilkowski isomorphism

How can we describe probability distributions over quantum states?

We know about classical probability distributions. R.V. $X \sim p_X$ defined over a finite set $\mathcal{X}$

$$p_X(x) \equiv \Pr\{X = x\}.$$

Define $\mathcal{P}(\mathcal{X})$ to be the set of all probability distributions over $\mathcal{X}$

$$\mathcal{P}(\mathcal{X}) \equiv \left\{ p_X \in \mathbb{R}^{|\mathcal{X}|} \middle| p_X(a) \geq 0 \ \forall a \in \mathcal{X}, \sum_a p_X(a) = 1 \right\}$$

A probabilistic mixture of quantum states is denoted $\{p_X(x), |\psi_x\rangle\}$. The probability of the system to be in state $|\psi_x\rangle$ is given by $p_X(x)$. Is there a better way to represent such mixtures of quantum states?

# B.5 Quantum states

**E1** 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 length vector in state space.

The *global phase* doesn't matter:

$$|\phi\rangle \equiv e^{i\theta}|\phi\rangle \quad \forall\theta \in \mathbb{R}.$$

- Unit length has a probabilistic interpretation (to follow).
- Instead of a unit vector we can think of quantum state as a *ray* in state space, i.e., the length doesn't matter. The ray that contains $|\phi\rangle \in \mathcal{H}$ is $\{\lambda|\phi\rangle \mid \lambda \in \mathbb{C}\}$.

## The qubit

- Classical bit: $b \in \{0, 1\}$.
- Quantum bit or qubit is a unit vector in a two dimensional Hilbert space $\mathcal{H}^2 \cong \mathbb{C}^2$.
- The standard basis for $\mathcal{H}^2$:

$$|0\rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad |1\rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- Qubit:

$$\begin{aligned} |\psi\rangle &= \alpha|0\rangle + \beta|1\rangle, && \text{(B.3)} \\ &\quad \alpha \in \mathbb{R}, && \text{(global phase doesn't matter)} \\ &\quad \beta \in \mathbb{C}, \\ &\quad |\alpha|^2 + |\beta|^2 = 1 \end{aligned}$$

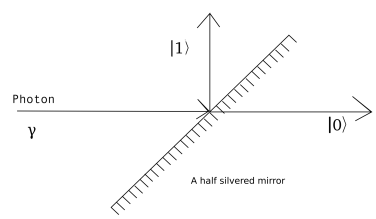



**Figure B.1:** A photon encounters a half silvered mirror. The state after the mirror is $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$.

- WT1: Quantum superposition the photon is both in $|0\rangle$ and $|1\rangle$.
- Observe that $\mathbb{C}^2$ is 4 dimensional (four degrees of freedom: real and imaginary parts of $\alpha$ and $\beta$).
- But a qubit is 2 d.f.

$$4 \text{ d.f.} - \alpha \text{ real} - \{\||\psi\rangle\| = 1\} = 2 \text{ d.f.}$$

### Bloch sphere

Observe that

$$\begin{aligned} |\psi\rangle &= \alpha|0\rangle + \beta|1\rangle \\ &= \alpha|0\rangle + |\beta|e^{i\varphi}|1\rangle \\ &= \cos(\theta/2)\,|0\rangle + \sin(\theta/2)\,e^{i\varphi}|1\rangle \end{aligned}$$

Thus we can identify a qubit by the two angles:

$$0 \leq \theta \leq \pi, \qquad 0 \leq \varphi \leq 2\pi.$$

These are the angles on the Bloch sphere.

- Don't trust the geometry of the Bloch spehre — $|0\rangle$ and $|1\rangle$ are orthogonal in Hilbert space, but they appear anitpodal on the Block sphere.

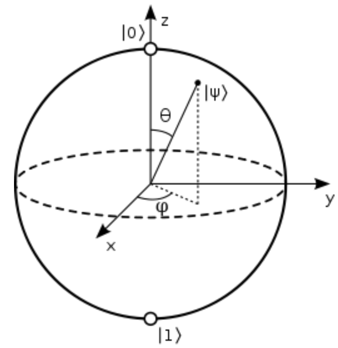



**Figure B.2:** The Bloch sphere is a good way to represent quantum states.

## State preparation

$$\text{x} \longrightarrow \boxed{\text{state preparation}} \longrightarrow |x\rangle$$

## Quantum operations

**E2** Time evolution of an isolated quantum system is unitary. If the state at time $t_1$ is $|\phi_1\rangle$ and at time $t_2$ is $|\phi_2\rangle$ then $\exists$ unitary $U$ such that $|\phi_2\rangle = U|\phi_1\rangle$.

$$|\psi\rangle \longrightarrow \boxed{\text{U}} \longrightarrow |\psi'\rangle$$

Unitarity:

$$\begin{aligned} U^\dagger U &= UU^\dagger = \mathbb{I} \\ U^\dagger &= \bar{U}^\top \quad \text{complex conjugate transpose} \end{aligned}$$

Note that unitaries are invertible.

- Unitarity ensures state remain unit length.
- Before $U$ we have: $\||\varphi\rangle\|^2 = \langle\varphi|\varphi\rangle = 1$.
- After $U$ has been applied we have: $\|U|\varphi\rangle\|^2 = \langle\varphi|U^\dagger U|\varphi\rangle = \langle\varphi|I|\varphi\rangle = \langle\varphi|\varphi\rangle = 1$.

**E1** The $Z$ operator is defined by its action on the standard basis:

$$Z|0\rangle = |0\rangle, \qquad Z|1\rangle = -1|1\rangle.$$

By linearity we have:

$$Z(\alpha|0\rangle + \beta|1\rangle) = \alpha|0\rangle - \beta|1\rangle.$$

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} = |0\rangle\langle 0| - |1\rangle\langle 1|.$$

**E2** The $Z$ operator is defined by its action on the standard basis:

$$X|0\rangle = |1\rangle, \qquad X|1\rangle = |0\rangle.$$

By linearity we have:

$$X(\alpha|0\rangle + \beta|1\rangle) = \beta|0\rangle + \alpha|1\rangle.$$

$$X = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = |0\rangle\langle 1| + |1\rangle\langle 0|.$$

**E3** The Hadamard operator:

$$H = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

$$H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \equiv |+\rangle, \tag{B.4}$$

$$H|1\rangle = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle) \equiv |-\rangle. \tag{B.5}$$

## Measurement

**E3** Measurement is modelled by a collection of operators $\{M_j\}$ acting on the state space of the system being measured. The operators obey:

$$\sum_j M_j^\dagger M_j = \mathbb{I}$$

The index $j$ labels measurement outcomes. If our pre-measurement state is $|\phi\rangle$, then the probability of getting $j$ is

$$p(j|\phi) = \langle\phi|M_j^\dagger M_j|\phi\rangle$$

The post measurement state for outcome $j$ is:

$$\frac{M_j|\phi\rangle}{\sqrt{\langle\phi|M_j^\dagger M_j|\phi\rangle}}.$$

- WT2: Measurement is not passive – measuring changes the state of the system!

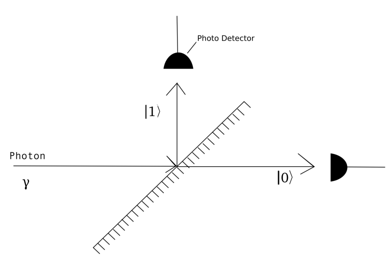



**Figure B.3:** Photodetectors are modelled by projectors

**E4** In Figure $B.3$ a state vector $|\phi\rangle = \alpha|0\rangle + \beta|1\rangle$ is being measured with photo detectors modelled as projectors given by

$$M_0 = |0\rangle\langle 0| \qquad \sum_j M_j^\dagger M_j = \mathbb{I}$$
$$M_1 = |1\rangle\langle 1|$$

$$\begin{aligned} p(0|\phi) &= \langle\phi| \quad |0\rangle\langle 0| \quad |\phi\rangle \\ &= (\bar{\alpha}\langle 0| + \bar{\beta}\langle 1|)|0\rangle\langle 0|(\alpha|0\rangle + \beta|1\rangle) \\ &= \bar{\alpha}\alpha = |\alpha|^2 \\ p(1|\phi) &= |\beta|^2 \end{aligned}$$

## More on measurements

- Projective measurement if $M_j^2 = M_j$.

- Complete projective measurement if $M_j^2 = M_j$ and $\mathrm{Tr}M_j = 1, \forall j$ (rank one projectors).
- Born's rule: Probability is the square of the modulus of the overlap: $p(j|\phi) = \|\langle j|\phi\rangle\|^2 = \langle\phi|M_j^\dagger M_j|\phi\rangle$.
- General measurements: Specify a set of positive semidefinite operators (Hermitian with non-negative eigenvalues) $\{E_j\}$ such that:

  $$p(j|\phi) = \langle\phi|E_j|\phi\rangle.$$

  This is called a positive-operator valued measurement (POVM).

## B.6 Composite quantum systems

Classically, if we have bits $b_1 \in \{0,1\}$, $b_2 \in \{0,1\}$ then we can concatenate them to obtain a bit string $b_1 b_2 \in \{0,1\}^2$.

**E4** The state space of a composite quantum system is the tensor product of the state spaces of the individual systems. If you have systems $1, 2, \ldots, n$ in states $|\varphi_1\rangle, |\varphi_2\rangle, \cdots, |\varphi_n\rangle$, then the state of the composite system is $|\varphi_1\rangle \otimes |\varphi_2\rangle \otimes \cdots \otimes |\varphi_n\rangle$.

So if we have

$$|\varphi_1\rangle \in \mathcal{H}_1, \qquad |\varphi_2\rangle \in \mathcal{H}_2$$

then

$$|\varphi_1\rangle \otimes |\varphi_2\rangle \in \mathcal{H}_1 \otimes \mathcal{H}_2$$

Often we drop the tensor symbol:

$$|0\rangle \otimes |1\rangle = |0\rangle|1\rangle = |01\rangle$$

A basis for $\mathcal{H}_1 \otimes \mathcal{H}_2$ is:

$$\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}.$$

**E5** Suppose that we are given

$$|\varphi_1\rangle = \alpha_1|0\rangle_1 + \beta_1|1\rangle_1, \qquad |\varphi_2\rangle = \alpha_2|0\rangle_2 + \beta_2|1\rangle_2,$$

and we want to find $|\varphi_{12}\rangle = |\varphi_1\rangle \otimes |\varphi_2\rangle$.

$$\begin{aligned}
|\varphi_{12}\rangle &= |\varphi_1\rangle \otimes |\varphi_2\rangle && \text{(B.6)}\\
&= (\alpha_1|0\rangle_1 + \beta_1|1\rangle_1) \otimes (\alpha_2|0\rangle_2 + \beta_2|1\rangle_2) && \text{(B.7)}\\
&= \alpha_1\alpha_2|0\rangle_1|0\rangle_2 + \beta_1\alpha_2|1\rangle_1|0\rangle_2 + \alpha_1\beta_2|0\rangle_1|1\rangle_2 + \beta_1\beta_2|1\rangle_1|1\rangle_2 && \text{(B.8)}\\
&= \alpha_1\alpha_2|00\rangle + \beta_1\alpha_2|10\rangle + \alpha_1\beta_2|01\rangle + \beta_1\beta_2|11\rangle. && \text{(B.9)}
\end{aligned}$$

**E6** Consider now the Einstein-Podolsky-Rosen (EPR) state:

$$|\Phi_+\rangle := \frac{1}{\sqrt{2}}(|00\rangle_A + |11\rangle_B)$$

Can you write $|\Phi_+\rangle$ as $|\varphi_1\rangle \otimes |\varphi_2\rangle$ for some states $|\varphi_1\rangle$ and $|\varphi_2\rangle$ ?

- The EPR pair is non-local — it cannot be adequately described as the tensor product of two local states.

## B.7 Lasers demonstration

### Polarizing filter example

$$\text{unpolarized light} \rightarrow \boxed{H} \rightarrow \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \text{horizontally polarized light}$$

$$\text{unpolarized light} \rightarrow \boxed{V} \rightarrow \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \text{vertically polarized light}$$

#### Two lenses

$$\underbrace{\text{light} \longrightarrow \underbrace{\boxed{H}}_{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}}_{\text{state preparation}} \longrightarrow \begin{bmatrix} 1 \\ 0 \end{bmatrix} \longrightarrow \underbrace{\underbrace{\boxed{V}}_{\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}} \longrightarrow \emptyset}_{\text{measurement}}$$

#### Three lenses

$$\underbrace{\text{light} \rightarrow \underbrace{\boxed{H}}_{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}}}_{\text{state preparation}} \rightarrow p = 1, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \rightarrow \underbrace{\underbrace{\boxed{D}}_{\frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}}}_{\text{measurement 1}} \rightarrow p = \frac{1}{2}, \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \rightarrow \underbrace{\underbrace{\boxed{V}}_{\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}}}_{\text{measurement 2}} \rightarrow p = \frac{1}{4}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

- Note that experiment shown actually can perfectly well be explained with classical theory of electromagnetic waves so it cannot be used as a demonstration of quantum mechanics.
- However, the "Stern-Gerlach experiment" which uses the spin of electrons for the same demonstration is based on the same reasoning.

**E7** When we measure a quantum system $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ in the basis $\{|0\rangle, |1\rangle\}$, we are asking a question of it:

> Are you $|0\rangle$ or are you $|1\rangle$ ?

example:

$$\begin{aligned} \Pr\{\text{measure } \psi = 0\} &= \langle\psi| \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} |\psi\rangle \\ &= \operatorname{Tr}\left\{ |\psi\rangle\langle\psi| \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} = \operatorname{Tr}\left\{ \underbrace{\begin{bmatrix} |\alpha|^2 & \beta^*\alpha \\ \alpha^*\beta & |\beta|^2 \end{bmatrix}}_{\text{density matrix}} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \right\} \\ &= |\alpha|^2 \end{aligned}$$

## B.8 Quantum information theory

By merging the ideas of classical information theory and quantum mechanics, a new field is born. The classic question of in information transmission and information compression must be revisited to see whether we can take advantage of the properties of quantum mechanics in our encoding and decoding procedures.

Note that quantum information theory is not a special type of

> "Why can't you quantum people give us an error model?"
> —An communications engineer

The reason is that collective quantum measurements on $n$ outputs of a quantum channel is a more powerful procedure than performing individual measurements of each channel output. Thus, quantum information theory has a raison-d'etre because the quantum detection strategies are more powerful than classical decoding strategies.

## B.9 Quantum computing

Certain tasks can be done faster on a quantum computer than on a classical computer. What is a quantum computer you ask?

It's a little complicated to define what a "quantum computer" should be able to do, so we could say that it can store quantum registers. I.e. the information is encoded in a quantum state, and can be manipulated as a quantum state.

Deutsch's algorithm

Deutsch–Jozsa algorithm