Vectors and Matrices

Adam Kelly

January 22, 2021

This set of notes is a work-in-progress account of the course 'Vectors and Matrices', originally lectured by Dr. Jonathan Evans in Michaelmas 2020 at Cambridge. These notes are not a transcription of the lectures, but they do roughly follow what was lectured (in content and in structure).

These notes are my own view of what was taught, and should be somewhat of a superset of what was actually taught. I frequently provide different explanations, proofs, examples, and so on in areas where I feel they are helpful. Because of this, this work is likely to contain errors, which you may assume are my own. If you spot any or have any other feedback, I can be contacted at ak2316@cam.ac.uk.

Contents

0 Introduction

Vectors and Matrices covers topics in both algebra and geometry, and the way in which they relate to one another. The course uses approaches that are quite varied in their nature (can be abstract or more concrete, conceptual or more computational, and so on). You will need to be able to fluently switch between these approaches.

The course assumes that you are vaguely familiar with Euclidean and coordinate geometry, along with the idea of geometric transformations.

§0.1 Course Structure

This course is divided into a number of chapters.

1. *Complex Numbers*

   This chapter takes the point of view of thinking of points in the plane as pairs of real numbers, and defining 'multiplication' on it to turn it into the complex numbers.

2. *Vectors in Three Dimensions*

   Here, we will recap on the relationship between three dimensional vectors and some of their geometrical applications, and we will discuss things like the dot and cross product. Towards the end of that discussion, we will introduce the 'index notation', a powerful and helpful notation for dealing with vectors. We will also introduce the 'summation convention', which is also incredibly useful.

3. *Vectors in a General Setting*

   This chapter will discuss what vectors are in general, and different ways of looking at them. We will be particularly concerned with vectors in $\mathbb{R}^n$ and $\mathbb{C}^n$, that is, vectors who's entries are in $\mathbb{R}$ and $\mathbb{C}$ respectively.

4. *Matrices and Linear Maps*

   Picking up on the idea of generalizing vectors, this chapter will consider the idea of a 'linear map', an abstraction of matrices.

5. *Determinants and Inverses*

   This chapter will detail how to define and compute determinants of general $n \times n$ matrices. This will take two points of view, in that we need to be able to compute them but we also must understand what they mean. The relation between determinants and finding inverses of matrices will also be considered.

6. *Eigenvalues and Eigenvectors*

   This chapter also involves both geometry and algebra. The core question of this chapter is: given a linear map or matrix, what does it act on in a very straightforward way?

7. *Changing Basis, Canonical Forms and Symmetries*

In this final chapter, we will consider a set of far reaching results by trying to describing an arbitrary linear map. These ideas are far reaching, and immensely useful.

§0.2 Differences to the Lecture Course

This set of notes may diverge slightly from the lectures. If this occurs, I will attempt to describe the differences in this section. For now, while these notes are incomplete, I will leave it up to the reader to check themselves what is included or missing.

1 Complex Numbers

The complex numbers arose through the study of polynomials, but there is hardly an area of modern mathematics where they are not of use, with applications stretching from number theory to geometry to quantum mechanics.

§1.1 Defining the Complex Numbers

We are going to start right at the beginning, though some familiarity with complex numbers is assumed. We will construct the set of complex numbers, denoted by $\mathbb{C}$, from the real numbers $\mathbb{R}$ by adjoining an element i with the property that $i^2 = -1$.

Definition 1.1.1 (Complex Numbers)
A **complex number** is a number $z \in \mathbb{C}$ of the form $z = x + yi$ with $x, y \in \mathbb{R}$ such that $i^2 = -1$.

We call $x = \mathrm{Re}(z)$ the **real part**, and $y = \mathrm{Im}(z)$ the **imaginary part** of z.

The reals are a subset of the complex numbers, as for $x \in \mathbb{R}$, we have $x + 0i \in \mathbb{C}$.

§1.1.1 Basic Operations

With the complex numbers defined, we have to come up with some things to do with them. We can add, subtract, multiply and divide complex numbers in a sensible way, and indeed they form a *field* (we will elaborate on this later).

Definition 1.1.2 (Basic Operations in $\mathbb{C}$)
Let $z_1 = x_1 + y_1 i$ and $z_2 = x_2 + y_2 i$ be two complex numbers. Then we define **addition** and **multiplication** as

$$\begin{aligned} z_1 + z_2 &= (x_1 + x_2) + (y_1 + y_2)i, \\ z_1 \times z_2 &= (x_1 x_2 - y_1 y_2) + (x_1 y_2 + x_2 y_1)i. \end{aligned}$$

We can see from this definition that the operation of multiplication respects our original requirement that $i^2 = -1$. From these definitions we can also immediately define **subtraction** and **division** as the inverse operations of these, with

$$\begin{aligned} z_1 - z_2 &= (x_1 - x_2) + (y_1 - y_2)i \\ \frac{z_1}{z_2} &= \frac{x_1 + y_1 i}{x_2 + y_2 i} \cdot \frac{x_2 - y_2 i}{x_2 - y_2 i} = \left(\frac{x_1 y_1 + x_2 y_2}{x_2^2 + y_2^2}\right) + \left(\frac{x_2 y_1 - x_1 y_2}{x_2^2 + y_2^2}\right) i. \end{aligned}$$

From the definitions of addition and multiplication, it is clear that both operations are commutative and associative. Indeed, as we also have inverse operations, the complex numbers form groups.

Proposition 1.1.3 ($\mathbb{C}$ is a Group)

$\mathbb{C}$ with the operation $+$ is an abelian group with identity 0, and $\mathbb{C}\backslash\{0\}$ with the operation $\times$ is an abelian group with identity 1.

Proof Sketch. Check definitions and verify that $+$ and $\times$ are indeed associative. □

Another useful operation is the *complex conjugate*.

Definition 1.1.4 (Complex Conjugation)

For a complex number $z = x + iy$, we define its **complex conjugate** to be $\overline{z} = z^* = x - iy$.

Complex conjugation allows us to extract the real and imaginary parts of a complex number, with

$$\operatorname{Re}(z) = \frac{z+\overline{z}}{2}, \quad \text{and} \quad \operatorname{Im}(z) = \frac{z-\overline{z}}{2i}.$$

For complex numbers z_1 and z_2, we have $\overline{(\overline{z})} = z$, $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$ and $\overline{z_1 z_2} = \overline{z_1} \cdot \overline{z_2}$.

Aside: The History of Complex Numbers

The complex numbers arose through the study of polynomials, and their serious study really began in the study of the *cubic* polynomials of the form[1]

$$x^3 - 3px - 2q = 0.$$

The sixteenth century mathematicians del Ferro, Tartaglia and Cardano discovered that this equation could be solved explicitly, with the result being published in Cardano's 1545 work *Ars Magna* (where he mentioned del Ferro as a first author, and noted that Tartaglia had independently found the same result). The solution they had obtained was

$$x = \sqrt[3]{q + \sqrt{q^2 - p^3}} + \sqrt[3]{q - \sqrt{q^2 - p^3}}.$$

For example, for the cubic $x^3 - 12x - 20$ we obtain the solution $x = \sqrt[3]{16} + \sqrt[3]{2}$.

Some 30 years later, Rafael Bombelli discussed one particular cubic in his work *l'Algebra*. He looked at $x^3 - 15x - 4 = 0$. One might observe (through mere trial and error) that $x = 4$ is a solution, but using Cardano's formula, we obtain the solution

$$x = \sqrt[3]{2 + \sqrt{-121}} + \sqrt[3]{2 - \sqrt{-121}}.$$

Indeed, Bombelli noticed this and gave an explanation for how these two solutions were essentially the same, even with the presence of the rather puzzling term $\sqrt{-121}$. His explanation essentially began by noting that if we wrote $\sqrt{-121} = 11i$, then we could find that

$$\sqrt[3]{2 + \sqrt{-121}} = 2 + ni$$
$$\sqrt[3]{2 - \sqrt{-121}} = 2 - ni,$$

[1] This form of a cubic equation is a 'depressed', and any general cubic can be expressed in this form.

and adding would then result in $2 + ni + 2 - ni = 4$, our other solution.

Inherent in this explanation was an expectation that this i would obey many of the same rules of arithmetic that we would expect of, say, a real number. After this work, contributions from mathematicians such as Descartes, Wallis, de Moivre, Euler, Wessel, Argand, Hamilton, Gauss, Cauchy and many others began to build up a rich theory of the complex numbers, later developing the field of complex analysis, all centered on this construction of some i with the property that $i^2 = -1$.

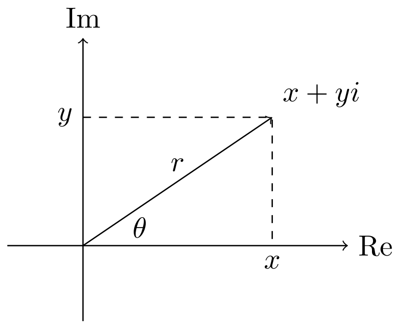
Looking again at the solutions to polynomials it turns out that the complex numbers are sufficient to solve *every* polynomial equation. This result is known as the 'Fundamental Theorem of Algebra', and was proved in the 18th century. We will use it frequently in this course, but you will have to wait until a later course to see a proof.

Theorem 1.1.5 (Fundamental Theorem of Algebra)

Let $p(z)$ be a polynomial of degree $n \geq 1$ with complex coefficients. Then $p(z) = 0$ has precisely n (not necessarily distinct) complex roots, counted with multiplicity.

§1.1.2 The Argand Diagram

There is a geometric way to interpret complex numbers and the way that they operate. We can plot a complex number $z = x + yi$ on the **Argand diagram** or **complex plane** by associating it with a point (x, y). We call the x axis the 'real' axis, and the y axis the 'imaginary' axis.



From the complex number written $z = x + yi$, we can plot it in rectangular coordinates. However, it can also be useful to plot it using polar coordinates. With this in mind, we introduce some useful definitions.

Definition 1.1.6 (Modulus)

For a complex number $z = x + yi$, we define its **modulus** to be $r = |z|$ such that $r \geq 0$ and $r^2 = |z|^2 = x^2 + y^2$.

Indeed, this definition matches the distance from the origin of the Argand diagram to the point (x, y).

Definition 1.1.7 (Argument)

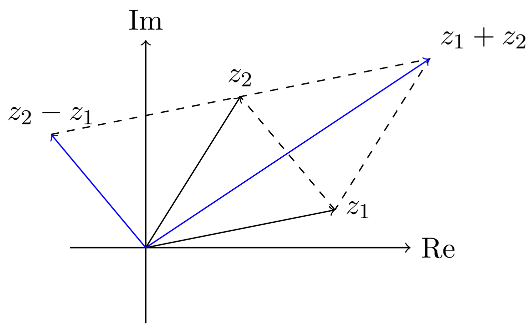
For a non-zero complex number z, we define its **argument** to be $\theta = \arg(z)$, a real number such that $z = r(\cos\theta + i\sin\theta)$. We say this is the **polar form** of z.

From this, we get that for $z = x + yi$,

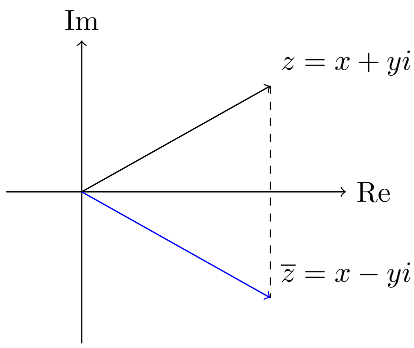
$$\cos\theta = \frac{x}{\sqrt{x^2+y^2}}, \quad \sin\theta = \frac{y}{\sqrt{x^2+y^2}}, \quad \text{and} \quad \tan\theta = \frac{y}{z}.$$

An important point to note is that $\arg(z)$ is only defined up to multiples of 2π, as mapping $\theta \mapsto \theta + 2\pi$ leaves z unchanged. So, to make θ unique (which is often useful), we typically restrict its range. For example, if we restrict $-\pi < \theta \leq \pi$, we get the **principle value** of θ.

Using the Argand diagram, there is a geometric way to define addition, multiplication and complex conjugation. Addition and subtraction correspond to parallelograms constructed as shown.



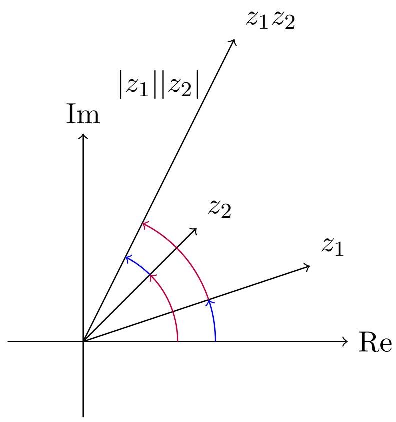
Complex conjugation corresponds to reflection in the real axis.



Multiplication is slightly more complex. If we have two complex numbers in polar form, $z_1 = r_1(\cos\theta_1 + \sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + \sin\theta_2)$, then $z_1 z_2 = r_1 r_2(\cos\theta_1 + \theta_2 + \sin\theta_1 + \theta_2)$. This gives the following.

Proposition 1.1.8 (Moduli Multiply and Arguments Add)
For complex numbers z_1 and z_2, we have $\arg(z_1 z_2) = \arg(z_1) + \arg(z_2)$ and $|z_1 z_2| = |z_1| \cdot |z_2|$.

This is shown in the diagram below.



To see how useful a geometric perspective can be, consider the following result.

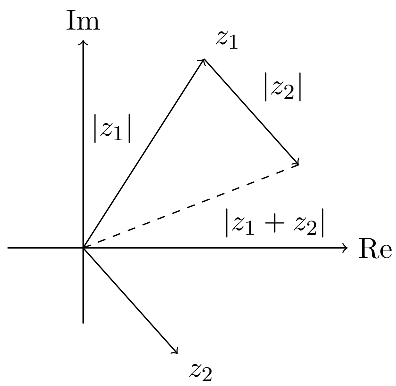
Proposition 1.1.9 (Triangle Inequality)
For complex numbers z_1 and z_2, $|z_1 + z_2| \leq |z_1| + |z_2|$.

Algebraic Proof. We have $|z_1+z_2|^2 = (z_1+z_2)\overline{(z_1+z_2)} = |z_1|^2 + z_1\overline{z_2} + \overline{z_1}z_2 + |z_2|^2$, so it suffices to show that $z_1\overline{z_2} + \overline{z_1}z_2 \leq 2|z_1||z_2|$. To show this, we have

$$\begin{aligned} z_1\overline{z_2} + \overline{z_1}z_2 &\leq 2|z_1||z_2| \\ \iff \frac{z_1\overline{z_2} + \overline{z_1\overline{z_2}}}{2} &\leq |z_1||\overline{z_2}| \\ \iff \operatorname{Re}(z_1\overline{z_2}) &\leq |z_1\overline{z_2}|, \end{aligned}$$

which is true. □

Now put everything on an Argand diagram and stare at it for a minute.



Geometric Proof. The shortest distance between 0 and $z_1 + z_2$ is $|z_1 + z_2|$. This is a side length of the triangle at 0, z_1 and $z_1 + z_2$, and this must be shorter than the other two sides of the triangle, which has length $|z_1| + |z_2|$. This gives the triangle inequality. □

Alternate forms of the triangle inequality are

$$|z_2 - z_1| \geq |z_2| - |z_1|, \quad \text{and} \quad |z_2 - z_1| \geq |z_1| - |z_2|.$$

One of these tends to be trivially true, however, due to one side being negative. Still, we can combine these into a single statement

$$|z_2 - z_1| \geq ||z_2| - |z_1||.$$

We will end our discussion with the following theorem which will repeatedly appear.

Theorem 1.1.10 (De Moivre's Theorem)
For $n \in \mathbb{Z}$ and $\theta \in \mathbb{R}$, we have

$$(\cos\theta + i\sin\theta)^n = \cos n\theta + i\sin n\theta,$$

where $z^0 = 1$ for non-zero z.

Proof Sketch. Use induction and angle sum identities. A different proof will be given after exponential functions are defined. □

§1.2 Exponential and Trigonometric Functions

We define the functions exp, cos and sin on the complex numbers by

$$\begin{aligned}
\exp(z) &= \sum_{n=0}^{\infty} \frac{1}{n!} z^n = e^z \\
\cos(z) &= \frac{1}{2}(e^z + e^{-iz}) = 1 - \frac{1}{2!}z^2 + \frac{1}{4!}z^4 - \cdots \\
\sin(z) &= \frac{1}{2i}(e^z - e^{-iz}) = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \cdots
\end{aligned}$$

These functions are defined by **power series**, which will be explored more in Analysis. They converge for all z, and can be manipulated and rearranged as you may expect.

Aside: Power Series (Preview of Analysis I)

A **power series** about a point x_0 is a sum of the form

$$\sum_{n=0}^{\infty} a_n (x - x_0)^n,$$

where $a_n \in \mathbb{C}$. A power series is said to **converge** at a point x if the finite sums $\sum_{n=0}^{N} a_n(x - x_0)^n$ tend to a limit as $N \to \infty$. With each power series there is a **radius of convergence** ρ such that if $|x - x_0| < \rho$, then the power series converges absolutely. If $|x - x_0| > \rho$, then the series does not converge. If $|x - x_0| = \rho$, then it may or may not converge. In the functions defined above, the radius of convergence is infinity.

We can manipulate power series as you might expect. Notably, we can differentiate term by term (and the radius of convergence will stay the same), and we can add and multiply power series if we are in the radius of convergence of both.

Power series will be discussed in detail in Analysis I.

These definitions reduce to familiar definitions when $z \in \mathbb{R}$. Indeed, if we differentiate the series, we get

$$\frac{\mathrm{d}}{\mathrm{d}x}\exp(x) = \exp(x), \quad \frac{\mathrm{d}}{\mathrm{d}x}\cos(x) = -\sin(x), \quad \frac{\mathrm{d}}{\mathrm{d}x}\sin(x) = \cos(x),$$

and $\exp(0) = 1$, $\cos(0) = 1$, $\sin(0) = 0$, which together uniquely defines these functions over the reals. However, the behavior over $\mathbb{C}$ is somewhat richer.

Proposition 1.2.1 (Properties of the Complex Exponential)
For $z = x + yi$,

(i) $e^z = e^x(\cos y + i \sin y)$.

(ii) e^z can take on all values in $\mathbb{C}$ except 0.

(iii) $e^z = 1$ if and only if $z = 2n\pi i$ for $n \in \mathbb{Z}$.

Proof. (i) $e^{x+iy} = e^x e^{iy}$, and $e^{iy} = \cos y + i \sin y$.

(ii) $|e^z| = |e^x|$, which takes on all positive real values, and $\arg(e^z) = y$, which takes on all real values.

(iii) $e^z = 1 \iff e^x = 1$ and $\cos y = 1$, $\sin y = 0$ which gives $x = 0$ and $y = 2n\pi$ as required. □

Returning to polar form, we can write

$$z = r(\cos\theta + i\sin\theta) = re^{i\theta},$$

where $r = |z|$ and $\theta = \arg(z)$. Then De Moivre's theorem follows immediately from $(e^{i\theta})^n = e^{in\theta}$.

§1.3 Roots of Unity

An important application of these results are *roots of unity*, which answers the question 'find all z such that $z^n = 1$'.

Definition 1.3.1 (Roots of Unity)
An **nth root of unity** is a complex number z such that $z^n = 1$.

Proposition 1.3.2 (Finding Roots of Unity)
The nth roots of unity are

$$z = e^{2k\pi i/n} = \cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n},$$

where $k \in \{0, 1, \dots, n-1\}$.

Proof. Writing z in polar form, these are easy to find. We have $z = re^{i\theta}$, then

$$z^n = r^n e^{in\theta} = 1,$$

so $r = 1$, and $\theta = \frac{2k\pi}{n}$ for some $k \in \mathbb{Z}$. This gives n distinct solutions. □

To get a feel for this, we can plot the cases for $n = 2, 3, 4$ and 5 on an Argand diagram. This hints at a geometric interpretation of the roots of unity.

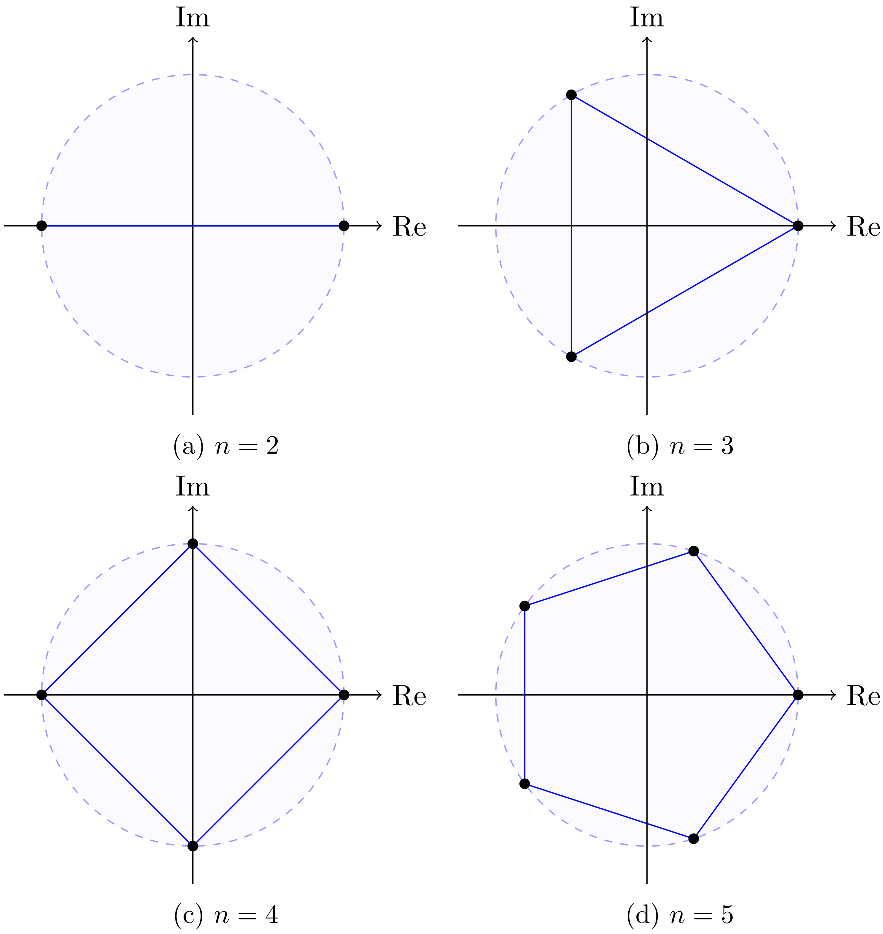


Figure 1.1: nth roots of unity on an Argand diagram

Proposition 1.3.3 (Geometric Interpretation of Roots of Unity)
The nth roots of unity correspond to the vertices of a regular n-gon, inscribed within the unit circle with a vertex at 1.

Proof. Follows directly from roots of unity written in polar form. □

§1.4 Logarithms and Complex Powers

Over $\mathbb{R}$, the exponential function is one to one. This means that we can quite easily define the inverse function, the logarithm, on the positive reals (the image of the exponential over $\mathbb{R}$).

This is not true over $\mathbb{C}$, as the complex exponential function is many to one. For example, $e^{3\pi i/2} = e^{7\pi i/2} = -i$. However, we do know that if we have $e^{i\theta} = e^{i\gamma}$, then θ and γ differ by a multiple of 2π[2]. Thus we can define a *multi-valued* logarithm function.

§1.4.1 The Complex Logarithm

To see how to define the logarithm over $\mathbb{C}$, let z be any non-zero complex number. Then if we let $\theta = \arg(z)$ with $-\pi < \theta \leq \pi$, then we can write

$$z = re^{i\theta} = e^{\log r} e^{i\theta + 2n\pi i} = e^{\log r + i(\theta + 2n\pi)i},$$

for any $n \in \mathbb{Z}$.

Definition 1.4.1 (Complex Logarithm)

For a complex number z, we define the **complex logarithm** $\log z$ to be the multi-valued function

$$\log z = \log r + i(\theta + 2n\pi),$$

where $r = |z|$, $\theta = \arg(z)$ with $-\pi < \theta \leq \pi$ and $n \in \mathbb{Z}$.

Example 1.4.2 (Calculating Complex Logarithms)

Let $z = \frac{5}{2} + \frac{5\sqrt{3}i}{2}$. We will find $\log z$. We have $|z| = 5$, and the principle value is $\arg(z) = \pi/3$. Thus we have

$$\log z = \log 5 + i(\pi/3 + 2n\pi), \quad n \in \mathbb{Z}.$$

Sometimes we will want single-valued behavior from the complex logarithm, and we can define the **principle logarithm** so that we fix $n = 0$. A warning thou h: if this is done, not all of the familiar properties of the logarithm still hold.[3]

§1.4.2 Complex Powers

The multi-valued nature of the logarithm carries into the notion of taking powers of complex numbers. For real numbers, we can define $x^a = e^{a \ln x}$ for $x > 0$. We can do the same for complex numbers.

Definition 1.4.3 (Complex Powers)

We define **complex powers** for $z, \alpha \in C$ with $z \neq 0$ to be the multi-valued

$$z^\alpha = e^{\alpha \log z}.$$

This is not *always* multi valued. For example, if $\alpha \in \mathbb{Z}$, then z^α is unique. If $\alpha \in \mathbb{Q}$, then there will be only finitely many values. However, in most cases there will be infinitely many values.

[2] This comes from the argument of a complex number only being defined up to a multiple of 2π

[3] When taking the principle logarithm, we are really taking a *branch* of the multi-valued logarithm function. This corresponds with taking a slice of the Riemann surface that $\log z$ defines. This gives some weird behavior as we move past certain sets of points (singular points) which breaks many of the identities that we use frequently for the real valued logarithm.

Example 1.4.4 (Calculating i^i)
We have

$$\begin{aligned} i^i &= e^{i \log i} \\ &= e^{i \log 1 + i(\pi/2 + 2n\pi)} \\ &= e^{-(2n+1/2)\pi}, \end{aligned}$$

for any $n \in \mathbb{Z}$. Notably $i^i \in \mathbb{R}$.

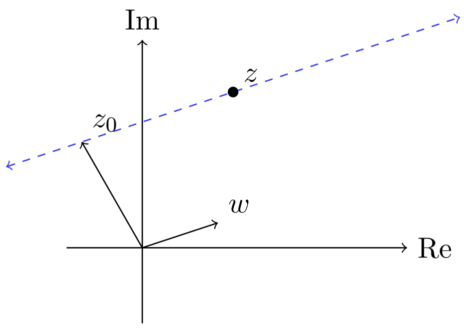
§1.5 Transformations, Lines and Circles

Consider the following transformations on $\mathbb{C}$.

- $z \mapsto z + \alpha$, translation by $\alpha \in \mathbb{C}$.
- $z \mapsto \lambda z$, scaling by $\lambda \in \mathbb{R}$,
- $z \mapsto e^{i\theta} z$, rotation by $\theta \in \mathbb{R}$.
- $z \mapsto \overline{z}$, reflection in the real axis.
- $z \mapsto \frac{1}{z}$, inversion.

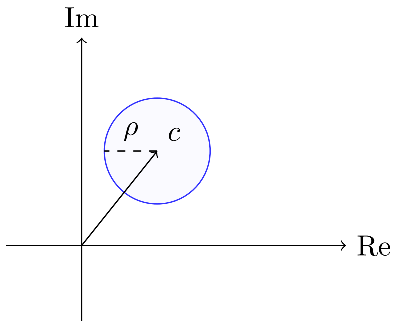
Using these transformations, we can think geometrically to do things like find equations of various geometric objects.

First we will find the general point on a line in $\mathbb{C}$ through z_0, parallel to $w \in \mathbb{C}$ where $w \neq 0$.



We can write $z = z_0 + \lambda w$ for $\lambda \in \mathbb{R}$. To eliminate λ, we can take the conjugate: $\overline{z} = \overline{z_0} + \lambda \overline{w}$, then $wz - w\overline{z} = \overline{w} z_0 + w\overline{z_0}$.

Similarly, we can find the general point on a circle with a center $c \in \mathbb{C}$ and positive real radius ρ.



A general point on this circle can be written $z = c + \rho e^{i\alpha}$, $\alpha \in \mathbb{R}$. To get rid of α, note that we could instead write $|z - c| = \rho$.

We can unify these two forms in the following way.

Proposition 1.5.1 (Circles and Lines in $\mathbb{C}$)
The equation of any line, and of any circle, may be written respectively as

$$Bz + \bar{B}\bar{z} + C = 0 \quad \text{and} \quad z\bar{z} + \bar{B}z + B\bar{z} + C = 0$$

for some complex B and real C.

In particular, this is a line in the direction iB or (if $|B|^2 \geq C$), a circle with center $-B$ and radius $\sqrt{|B|^2 - C}$.

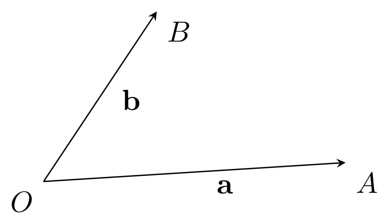
Proof Sketch. Do algebra □

2 Vectors in Three Dimensions

Fundamentally, vectors are mathematical objects that we can add together and take multiples of, in both cases obtaining another vector. We will consider vectors in a general setting later on, but in this chapter we will consider vectors in $\mathbb{R}^3$, three dimensional space. Specifically, we will take a geometric approach, thinking of vectors as position vectors and using Euclidean notions of points, lines, planes, etc.

We begin by choosing a point O as the origin. Then points A and B have position vectors

$$\mathbf{a} = \overrightarrow{OA}, \qquad \mathbf{b} = \overrightarrow{OB}.$$



The vectors have length $|\mathbf{a}| = |\overrightarrow{OA}|$, the distance between O and A. We also let $\mathbf{0}$ denote the position vector of O.

§2.1 Vector Addition and Scalar Multiplication

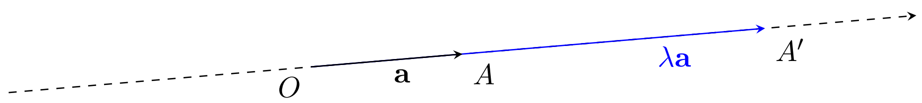
The most important operations on vectors (in $\mathbb{R}^3$ and in general) is vector addition and scalar multiplication. We define these geometrically as follows.

Definition 2.1.1 (Scalar Multiplication in $\mathbb{R}^3$)

GIven $\mathbf{a}$, the position vector for a point A, and a scalar $\lambda \in \mathbb{R}$, we define **scalar multiplication** $\lambda\mathbf{a}$ to be the position vector for the point A' on the line OA such that

$$|\lambda\mathbf{a}| = |OA'| = |\lambda||\mathbf{a}|,$$

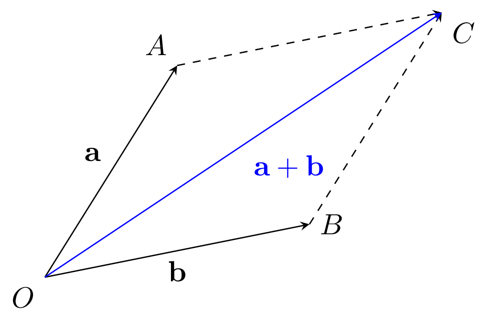
with the vector being in the direction $\overrightarrow{OA}$ if λ is positive, and in the opposite direction otherwise.



If two vectors are scalar multiples of each other, then we say that they are **parallel**. Specifically, we write $\mathbf{a} \parallel \mathbf{b}$ if and only if $\mathbf{a} = \lambda\mathbf{b}$ or $\mathbf{b} = \lambda\mathbf{a}$ for some $\lambda \in \mathbb{R}$. Notably, $\mathbf{a} \parallel \mathbf{0}$ for all vectors $\mathbf{a}$.

Definition 2.1.2 (Vector Addition in $\mathbb{R}^3$)

Given $\mathbf{a}$ and $\mathbf{b}$, position vectors of points A and B, if $\mathbf{a} \nparallel \mathbf{b}$, we construct the parallelogram $OACB$, and define **vector addition** $\mathbf{a} + \mathbf{b} = \mathbf{c}$.



If $\mathbf{a} \parallel \mathbf{b}$, then writing $\mathbf{a} = \alpha\mathbf{u}$ and $\mathbf{b} = \beta\mathbf{u}$ where $\mathbf{u}$ is a unit vector, then $\mathbf{a}+\mathbf{b} = (\alpha+\beta)\mathbf{u}$.

Given some set of vectors $\mathbf{v_1}$, $\mathbf{v_2}$, $\dots$, $\mathbf{v_k}$, we can form a **linear combination**

$$\lambda_1\mathbf{v_1} + \lambda_2\mathbf{v_2} + \cdots + \lambda_k\mathbf{v_k},$$

where $\lambda_i \in \mathbb{R}$. With this in mind, we can consider the set of all vectors that can be formed as a linear combination of some vectors.

Definition 2.1.3 (Span)
For vectors $\mathbf{v_1}$, $\mathbf{v_2}$, $\dots$, $\mathbf{v_k}$, we define their **span** to be the set

$$\operatorname{span}\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_k}\} = \{\lambda_1\mathbf{v_1} + \lambda_2\mathbf{v_2} + \cdots + \lambda_k\mathbf{v_k} \mid \lambda_i \in \mathbb{R}\}.$$

If two vectors $\mathbf{a} \not\parallel \mathbf{b}$, then $\operatorname{span}\{\mathbf{a}, \mathbf{b}\}$ is a plane through OAB.

Vector addition and scalar multiplication obey some basic properties that you should keep in mind, as they will be used again when we define vectors in a general sense.

Proposition 2.1.4 (Basic Properties of Vector Operations)
For any vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$,

(i) Vectors along with vector addition form an abelian group.

(ii) $\lambda(\mathbf{a}+\mathbf{b}) = \lambda\mathbf{a} + \lambda\mathbf{b}$.

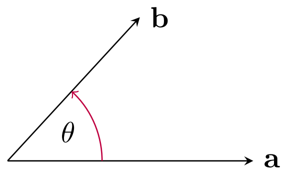
(iii) $(\lambda+\mu)\mathbf{a} = \lambda\mathbf{a} + \mu\mathbf{a}$.

(iv) $\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}$.

Proof Sketch. Check geometric definitions. Checking associativity of vector addition will require the construction of a parallelepiped. □

§2.2 The Dot Product

We saw in the previous section how to add vectors, and how to multiply a vector by a scalar. In the next two sections, we will consider how to multiply a vector and a vector. We will first define the *scalar* or *dot product*. Consider two vectors $\mathbf{a}$ and $\mathbf{b}$, with the angle between then being θ as shown.



Definition 2.2.1 (Dot Product)

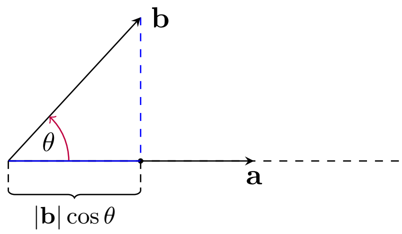
For two vectors $\mathbf{a}$ and $\mathbf{b}$, and θ the angle between them as shown, we define the **scalar** or **dot product** as

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}||\mathbf{b}| \cos\theta.$$

[Note that θ is defined unless $\mathbf{a}$ or $\mathbf{b}$ is zero, but then $|\mathbf{a}| = 0$ or $|\mathbf{b}| = 0$ and $\mathbf{a} \cdot \mathbf{b} = 0$.]

The scalar product encodes an angle condition, and gives us a dot for perpendicularity. We say vectors $\mathbf{a}$ and $\mathbf{b}$ are **orthogonal** or **perpendicular** if and only if $\mathbf{a} \cdot \mathbf{b} = 0$. This corresponds with $\theta = \pm\pi/2$, or with either $|\mathbf{a}| = 0$ or $|\mathbf{b}| = 0$.

There is a geometric interpretation of the dot product. Fundamentally, the dot product is a *projection*. For $\mathbf{a} \neq \mathbf{0}$, the quantity $|\mathbf{b}| \cos\theta$ is the component of $\mathbf{b}$ when projected along the direction of $\mathbf{a}$.



Proposition 2.2.2 (Basic Properties of the Dot Product)

For any vectors $\mathbf{a}$ and $\mathbf{b}$,

(i) $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$.

(ii) $\mathbf{a} \cdot \mathbf{a} = |a|^2 \geq$, and $\mathbf{a} \cdot \mathbf{a} = 0$ if and only if $\mathbf{a} = \mathbf{0}$.

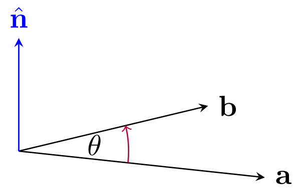
(iii) $(\lambda\mathbf{a}) \cdot \mathbf{b} = \lambda(\mathbf{a} \cdot \mathbf{b}) = \mathbf{a} \cdot (\lambda\mathbf{b})$.

(iv) $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$.

Proof. Properties (i) to (iii) follow directly from the definition of the dot product. Property (iv) follows from the geometric interpretation of the dot product shown above. □

§2.3 The Cross Product

The next operation only works with vectors in three dimensions, and is the *vector* or *cross product*. As before, consider two vectors $\mathbf{a}$ and $\mathbf{b}$, and let θ be measured as shown with respect to $\hat{\mathbf{n}}$, a unit normal to the plane spanned by $\mathbf{a}$ and $\mathbf{b}$.



Definition 2.3.1 (Cross Product)

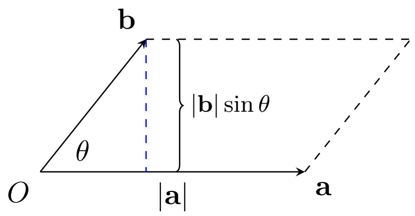
For two vectors $\mathbf{a}$ and $\mathbf{b}$ with angle θ measured as shown, we define the **vector** or **cross product** as

$$\mathbf{a} \times \mathbf{b} = |\mathbf{a}||\mathbf{b}| \sin\theta \hat{\mathbf{n}},$$

where $\hat{\mathbf{n}}$ is a unit vector perpendicular to the plane spanned by $\mathbf{a}$ and $\mathbf{b}$.

[$\hat{\mathbf{n}}$ is defined up to a sign if $\mathbf{a} \nparallel \mathbf{b}$, but changing the sign changes θ to $2\pi - \theta$, and thus leaves $\mathbf{a} \times \mathbf{b}$ unchanged. When $\hat{\mathbf{n}}$ is not defined ($\mathbf{a} \parallel \mathbf{b}$ or θ not defined), $\mathbf{a} \times \mathbf{b} = \mathbf{0}$.]

The geometric interpretation of the cross product is that $|\mathbf{a} \times \mathbf{b}|$ is the area of the parallelogram as shown.



The direction of $\mathbf{a} \times \mathbf{b}$ gives the orientation of this parallelogram in space.

For an alternate geometric interpretation, fix a vector $\mathbf{a}$ and consider some vector $\mathbf{x}$ with $\mathbf{x} \perp \mathbf{a}$. Then the transformation $\mathbf{x} \longmapsto \mathbf{a} \times \mathbf{x}$ corresponds with scaling $\mathbf{x}$ by $\mathbf{a}$ and rotating by $\pi/2$ in the plane perpendicular to $\mathbf{a}$.

Proposition 2.3.2 (Basic Properties of the Cross Product)

For vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$,

(i) $\mathbf{a} \times \mathbf{b} = -(\mathbf{b} \times \mathbf{a})$.

(ii) $(\lambda\mathbf{a}) \times \mathbf{b} = \lambda(\mathbf{a}) \times \mathbf{b})$.

(iii) $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$.

(iv) $\mathbf{a} \times \mathbf{b} = \mathbf{0}$ if and only if $\mathbf{a} \perp \mathbf{b}$.

(v) $\mathbf{a} \times \mathbf{b} \perp \mathbf{a}, \mathbf{b}$, so $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = \mathbf{b} \cdot (\mathbf{a} \times \mathbf{b}) = 0$.

Proof Sketch. Check definitions. □

§2.4 Orthonormal Bases and Components

So far we have considered vectors in a purely geometric sense. But just as a coordinate system can be used to reason about geometry, basis (and by extension components) can be used to reason about vectors. We will look at bases again in a more general setting later on, but for now we will focus on vectors in three dimensions.

When we use a cartesian coordinate system, we pick a set of axes and using them, we can describe the position of any point. In a similar way, a *basis* is a set of vectors that span the space. That is, any vector can be written as a linear combination of vectors in the set. However we also have an additional requirement: this linear combination must be unique.

Definition 2.4.1 (Basis – Informal)

We say that a set of vectors is a **basis** if any vector can be uniquely written as a linear combination of vectors in the set.

Before we continue, it's helpful when working with vectors in three dimensions (and in other settings) to pick basis vectors in a certain way. We introduce the following definition.

Definition 2.4.2 (Orthonormal)

We say that a set of vectors $\{\mathbf{v_1}, \mathbf{v_2}, \dots, \mathbf{v_n}\}$ is **orthonormal** if they are all unit vectors and are orthogonal to each other. That is, if $|\mathbf{v_i}| = 1$ and

$$\mathbf{v_i} \cdot \mathbf{v_j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

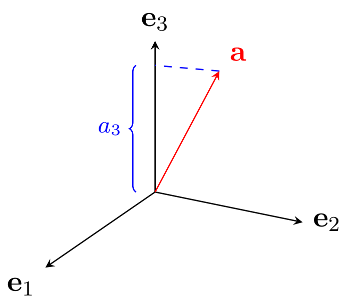
for all $1 \leq i, j \leq n$.

Now choose vectors $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ that are **orthonormal**. Then the set $\{\mathbf{e}_i\}$ is a basis, so for any vector $\mathbf{a}$, we can write

$$\mathbf{a} = \sum_i a_i \mathbf{e}_i = a_1\mathbf{e}_1 + a_2\mathbf{e}_2 + a_3\mathbf{e}_3,$$

and each coefficient is uniquely defined by $a_i = \mathbf{e}_i \cdot \mathbf{a}$. Note that this only works because of our orthonormal condition.

Recalling that dot products correspond to projections, we can see that this really is analogous to a cartesian coordinate system.



Once we have specified the basis vectors $\{\mathbf{e}_i\}$, we can then identify the vector $\mathbf{a}$ by its **components**, writing

$$\mathbf{a} = (a_1, a_2, a_3), \qquad \text{or} \qquad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

which is a **row** or **column** vector respectively[1].

An advantage of writing vectors in this form is that we can work directly with the components of vectors, rather than using a purely geometric approach.

§2.4.1 Dot Product in Components

For two vectors $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$, their dot product is

$$\mathbf{a} \cdot \mathbf{b} = (a_1, a_2, a_3) \cdot (b_1, b_2, b_3) = a_1 b_1 + a_2 b_2 + a_3 b_3.$$

This comes directly from writing $\mathbf{a}$ and $\mathbf{b}$ in terms of $\mathbf{e}_i$, and using the orthonormality condition.

We can use this to directly calculate the length of the vector, with

$$\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2 = a_1^2 + a_2^2 + a_3^2.$$

§2.4.2 Cross Product in Components

The cross product is orientation dependent, so it helps to choose the basis in a specific way. Usually, we pick a **right-handed** bases, which is one that satisfies

$$\begin{aligned}
\mathbf{e}_1 \times \mathbf{e}_2 &= \mathbf{e}_3 \\
\mathbf{e}_2 \times \mathbf{e}_3 &= \mathbf{e}_1 \\
\mathbf{e}_3 \times \mathbf{e}_1 &= \mathbf{e}_2,
\end{aligned}$$

and note that just one of these implies all of the others. Then we can calculate the cross product of two vectors in component form.

$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= (a_1 \mathbf{e}_1 + a_2 \mathbf{e}_2 + a_3 \mathbf{e}_3) \times (b_1 \mathbf{e}_1 + b_2 \mathbf{e}_2 + b_3 \mathbf{e}_3) \\
&= (a_2 b_3 - a_3 b_2)\mathbf{e}_1 + (a_3 b_1 - a_1 b_3)\mathbf{e}_2 + (a_1 b_2 - a_2 b_1)\mathbf{e}_3.
\end{aligned}$$

Example 2.4.3 (Computing a Cross Product)

The following is an example of computing a cross product using vector components.

$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix} \times \begin{pmatrix} 7 \\ -3 \\ 5 \end{pmatrix} \\
&= \begin{pmatrix} 0 \times 5 - (-3) \times (-1) \\ 7 \times (-1) - 2 \times 5 \\ 2 \times (-3) - 7 \times 0 \end{pmatrix} \\
&= \begin{pmatrix} -3 \\ -17 \\ -6 \end{pmatrix}.
\end{aligned}$$

Later on in this chapter, we will develop some more tools for working efficiently with components.

[1] The difference will be important when we move onto matrices.

§2.5 Triple Products

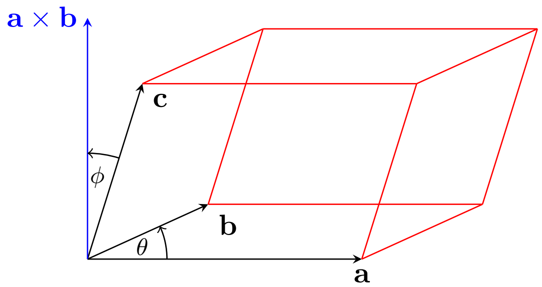
§2.5.1 Scalar Triple Product

Given three vectors, we define the *scalar triple product* as follows.

Definition 2.5.1 (Scalar Triple Product)
Let $\mathbf{a}, \mathbf{b}, \mathbf{c}$ be vectors. The **scalar triple product** of $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is

$$\begin{aligned}[\mathbf{a}, \mathbf{b}, \mathbf{c}] &= \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \mathbf{b} \cdot (\mathbf{c} \times \mathbf{a}) = \mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) \\ &= -\mathbf{a} \cdot (\mathbf{c} \times \mathbf{b}) = -\mathbf{b} \cdot (\mathbf{a} \times \mathbf{c}) = -\mathbf{c} \cdot (\mathbf{b} \times \mathbf{a}).\end{aligned}$$

The most natural interpretation of the scalar triple product is as the 'signed' volume of the parallelepiped formed by $\mathbf{a}, \mathbf{b}, \mathbf{c}$.



In the sketch above (where we assume $\theta, \phi < \pi/2$), we have

$$|\mathbf{c}||\mathbf{a} \times \mathbf{b}| \cos\phi = \underbrace{|a||b| \sin\theta}_{\text{base}} \cdot \underbrace{|c| \cos\phi}_{\text{height}},$$

which is the volume as shown.

We do have a sign, and indeed we have that $\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) > 0$ implies that $\mathbf{a}, \mathbf{b}, \mathbf{c}$ is a right handed set. Also $\mathbf{c} \cdot (\mathbf{a} \times \mathbf{b}) = 0$ if and only if $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are coplanar (corresponding with the parallelepiped having zero volume).

We can compute the scalar triple product using components. For $\mathbf{a} = (a_1, a_2, a_3)$, $\mathbf{b} = (b_1, b_2, b_3)$, $\mathbf{c} = (c_1, c_2, c_3)$, then we have

$$\begin{aligned}\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) &= a_1 b_2 c_3 - a_1 b_3 c_2 \\ &+ a_2 b_3 c_1 - a_2 b_1 c_3 \\ &+ a_3 b_1 c_2 - a_3 b_2 c_1 \\ &= \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix},\end{aligned}$$

which is a determinant (we shall study these in Chapter 5).

§2.5.2 Vector Triple Product

Another slightly less used product on three vectors is the *vector triple product*.

Definition 2.5.2 (Vector Triple Product)
For vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, we define the **vector triple product** to be

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

We will prove this identity later on, but note that this product is not associative. As the cross product is anticommutative, we do have

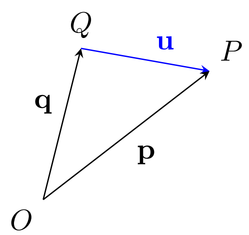
$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = -(\mathbf{b} \times \mathbf{c}) \times \mathbf{a}.$$

There's not really a sensible geometric interpretation of the vector triple product, but it is used in other vector formulas as it can shorten certain expressions.

§2.6 Lines, Planes & Vector Equations

We defined vectors as position vectors from an origin O, but with the introduction of vector addition, we can also use vectors to describe the displacement between two points,

$$\mathbf{u} = \overrightarrow{QP} = \mathbf{p} - \mathbf{q}.$$



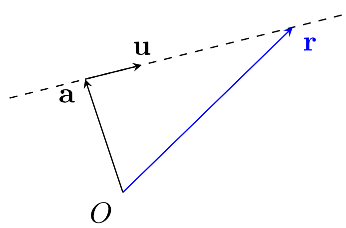
With this description of displacement vectors, we can define geometric objects such as lines and planes using equations involving vectors.

§2.6.1 Lines

A general point on a line through $\mathbf{a}$ in the direction of $\mathbf{u} \neq \mathbf{0}$ can be written as

$$\mathbf{r} = \mathbf{a} + \lambda\mathbf{u}, \qquad \lambda \in \mathbb{R}.$$

This is known as the **parametric form** of the line. Note the similarity between this equation and the one for a line in the complex plane.



To eliminate the parameter from this equation, we can take the cross product with $\mathbf{u}$

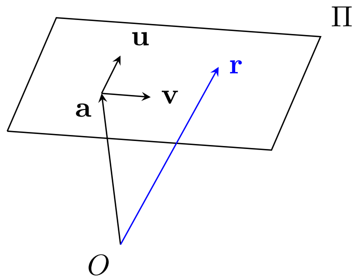
$$\begin{aligned} \mathbf{u} \times \mathbf{r} &= \mathbf{u} \times (\mathbf{a} + \lambda\mathbf{u}) \\ &= \mathbf{u} \times \mathbf{a} \\ \implies \mathbf{u} \times (\mathbf{r} - \mathbf{a}) &= \mathbf{0}. \end{aligned}$$

With this in mind, consider also the related equation $\mathbf{u} \times \mathbf{r} = \mathbf{c}$ with $\mathbf{u}, \mathbf{c} \neq \mathbf{0}$. From this, $\mathbf{u} \cdot (\mathbf{u} \times \mathbf{r}) = \mathbf{u} \cdot \mathbf{c} = \mathbf{0}$, so if $\mathbf{u} \cdot \mathbf{c} \neq 0$, this is inconsistent and there is no solutions. Otherwise, if $\mathbf{u} \cdot \mathbf{c} = \mathbf{0}$, then $\mathbf{u} \times (\mathbf{u} \times \mathbf{c}) = -|\mathbf{u}|^2\mathbf{c}$, so $\mathbf{a} = \frac{-1}{|\mathbf{u}|^2}(\mathbf{u} \times \mathbf{c})$ is one solution. Then $\mathbf{r} = \mathbf{a} + \lambda\mathbf{u}$ is the general solution.

§2.6.2 Planes

A general point on a plane Π thorough $\mathbf{a}$ with directions $\mathbf{u}, \mathbf{v}$ in the plane ($\mathbf{u} \not\parallel \mathbf{v}$) is

$$\mathbf{r} = \mathbf{a} + \lambda\mathbf{u} + \mu\mathbf{v}, \qquad \lambda, \mu \in \mathbb{R}.$$



An alternative form comes from considering a normal to the plane, $\mathbf{n} = \mathbf{u} \times \mathbf{v}$ ($\neq \mathbf{0}$). With this, we get

$$\begin{aligned} \mathbf{n} \cdot \mathbf{r} &= \mathbf{n} \cdot \mathbf{a} \\ \implies \mathbf{n} \cdot (\mathbf{r} - \mathbf{a}) &= 0. \end{aligned}$$

For a geometric interpretation of this, note that $\mathbf{r} - \mathbf{a}$ is the displacement in the direction of the plane,and thus should be perpendicular to $\mathbf{n}$, a normal to the plane.

§2.6.3 Other Vector Equations

To handel other vector equations, a general strategy is to manipulate equations and dot/cross with constant vectors.

Example 2.6.1 (Vector Equation of a Sphere)
We will attempt to find a geometric interpretation of the vector equation

$$|\mathbf{r}|^2 + \mathbf{r} \cdot \mathbf{a} = k,$$

with $\mathbf{a}$, k constant.

Here, we can 'complete the square' to get

$$\left|\mathbf{r} + \frac{1}{2}\mathbf{a}\right|^2 = k + \frac{1}{4}|\mathbf{a}|^2,$$

which is a sphere with center $-\frac{1}{2}\mathbf{a}$ and radius $(k + \frac{1}{4}|\mathbf{a}|^2)^{1/2}$, if $k > -\frac{1}{4}|\mathbf{a}|^2$. If this is not the case, then there is no solutions.

Example 2.6.2 (Arbitrary Vector Equation)

We will solve the vector equation

$$\mathbf{r} + \mathbf{a} \times (\mathbf{b} \times \mathbf{r}) = \mathbf{c},$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}$ are given.

We will use the 'dot and cross stuff' approach. We use the vector triple product identity to rewrite this as

$$\mathbf{r} + (\mathbf{a} \cdot \mathbf{r})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{r} = \mathbf{c}. \tag{$\dagger$}$$

Then dotting with $\mathbf{a}$ we get

$$\begin{aligned} \mathbf{a} \cdot \mathbf{r} + (\mathbf{a} \cdot \mathbf{r})(\mathbf{b} \cdot \mathbf{a}) - (\mathbf{a} \cdot \mathbf{b})(\mathbf{a} \cdot \mathbf{r}) &= \mathbf{c} \cdot \mathbf{a} \\ \implies \mathbf{a} \cdot \mathbf{r} &= \mathbf{c} \cdot \mathbf{a}. \end{aligned} \tag{$\dagger\dagger$}$$

Taking a dot product has 'lost information', so we substitution ($\dagger\dagger$) into ($\dagger$) to get

$$(1 - \mathbf{a} \cdot \mathbf{b})\mathbf{r} = \mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b}.$$

If $\mathbf{a} \cdot \mathbf{b} \neq 1$, then we can rearrange to get our solution

$$\mathbf{r} = \frac{1}{1 - \mathbf{a} \cdot \mathbf{b}}(\mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b}),$$

which is a point.

If $\mathbf{a} \cdot \mathbf{b} = 1$, then $\mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b} \neq \mathbf{0}$ would cause the equation to be inconsistent. So if $\mathbf{c} - (\mathbf{a} \cdot \mathbf{c})\mathbf{b} = \mathbf{0}$, then ($\dagger$) can be rewritten

$$(\mathbf{a} \cdot \mathbf{r} - \mathbf{a} \cdot \mathbf{c})\mathbf{b} = \mathbf{0},$$

of which the set of solutions is given by ($\dagger\dagger$), a plane.

§2.7 Index Notation & The Summation Convention

Returning to the component approach that we began to develop earlier, we will now introduce some machinery to deal with algebraically with components in an easier way.

This section needs to be rewritten.

§2.7.1 Components, δ and ε

Begin by writing vectors $\mathbf{a}$, $\mathbf{b}$, $\dots$ in terms of components a_i, b_i with respect to an orthonormal, right-handed basis $\{\mathbf{e}_i\}$. In the following discussion, the indices i, j, k, l, p, q, $\dots$ will take on values $1, 2, 3$[2].

Using indices, we can take some vector $\mathbf{c} = \alpha\mathbf{a} + \beta\mathbf{b}$ and express the components as $c_i = [\alpha\mathbf{a} + \beta\mathbf{b}]_i = \alpha a_i + \beta b_i$, for $i = 1, 2, 3$ (a **free variable**). Using the same notation,

[2] This corresponds to the number of basis vectors, and the approach used will naturally generalize to indices taking on more values.

we can write things like

$$\mathbf{a} \cdot \mathbf{b} = \sum_i a_i b_i,$$

and

$$\begin{aligned} &\mathbf{x} = \mathbf{a} + (\mathbf{b} \cdot \mathbf{c})\mathbf{d} \\ \iff &x_j = a_j + \left(\sum_k b_k c_k\right) d_j, \qquad \text{for } j = 1, 2, 3. \end{aligned}$$

Definition 2.7.1 (Kronecker Delta Symbol)
We define the **Kronecker delta** symbol to be

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

With this definition, we get $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, which encodes our orthonormality condition from before. Then

$$\begin{aligned} \mathbf{a} \cdot \mathbf{b} &= \left(\sum_i a_i \mathbf{e}_i\right) \left(\sum_i b_j \mathbf{e}_j\right) \\ &= \sum_{ij} a_i b_j \mathbf{e}_i \cdot \mathbf{e}_j \\ &= \sum_{ij} a_i b_j \delta_{ij} \\ &= \sum_i a_i b_i. \end{aligned}$$

Definition 2.7.2 (Levi-Civita Symbol)
We define the **Levi-Civita symbol** such that

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i\ j\ k) \text{ is an even permutation of } (1\ 2\ 3), \\ -1 & \text{if } (i\ j\ k) \text{ is an odd permutation of } (1\ 2\ 3), \\ 0 & \text{otherwise} \end{cases}$$

For example,

$$\begin{aligned} \varepsilon_{123} = \varepsilon_{231} = \varepsilon_{312} &= +1, \\ \varepsilon_{321} = \varepsilon_{213} = \varepsilon_{132} &= -1, \\ \varepsilon_{ijk} = 0 \text{ if any of } i, j, k \text{ repeat.} \end{aligned}$$

The Levi-Civita symbol is **totally antisymetric**, in that exchangin a pair of indices produces a change in sign.

With this definition, in our right-handed orthonormal basis we have

$$\mathbf{e}_i \times \mathbf{e}_j = \sum_k \varepsilon_{ijk} \mathbf{e}_k,$$

for example

$$\mathbf{e}_2 \times \mathbf{e}_1 = \sum_k \varepsilon_{21k}\mathbf{e}_k = \varepsilon_{213}\mathbf{e}_3 = -\mathbf{e}_3.$$

So we can compute the cross product of two vectors as

$$\begin{aligned}
\mathbf{a} \times \mathbf{b} &= \left(\sum_i a_i \mathbf{e}_i\right) \times \left(\sum_j b_j \mathbf{e}_j\right) \\
&= \sum_{ij} a_i b_j \mathbf{e}_i \times \mathbf{e}_j \\
&= \sum_{ijk} a_i b_j \varepsilon_{ijk} \mathbf{e}_k \\
\iff \quad (\mathbf{a} \times \mathbf{b})_k &= \sum_{ij} \varepsilon_{ijk} a_i b_j.
\end{aligned}$$

For example, $(\mathbf{a} \times \mathbf{b})_3 = \sum_{ij} \varepsilon_{ij3} a_i b_j = \varepsilon_{123} a_1 b_2 + \varepsilon_{213} a_2 b_1 = a_1 b_2 - a_2 b_1$, as we had before.

§2.7.2 Summation Convention

With components and index notation, indices that appear twice in a given term are usually summed over. In the **summation convention**, we omit the sigma ($\sum$) signs for repeated indices, and the sums are implicitly understood.

Example 2.7.3 (Examples of the Summation Convention)

The following all use the summation convention.

(i) $a_i \delta_{ij} = a_1 \delta_{1j} + a_2 \delta_{2j} + a_3 \delta_{3j} = a_j$.

(ii) $\mathbf{a} \cdot \mathbf{b} = a_i b_i$.

(iii) $(\mathbf{a} \times \mathbf{b})_i = \varepsilon_{ijk} a_j b_k$.

(iv) $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \varepsilon_{ijk} a_i b_j c_k$.

(v) $\delta_{ii} = \delta_{11} + \delta_{22} + \delta_{33} = 3$.

(vi) $[(\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}]_i = (\mathbf{a} \cdot \mathbf{c}) b_i + (\mathbf{a} \cdot \mathbf{b}) c_i = a_j c_j b_i - a_k b_k c_i$.

We can explicitly define the rules of the summation convention as follows.

Definition 2.7.4 (Summation Convention Rules)

When using the summation convention, we have the following rules.

(i) An index occurring exactly once in any given term must appear once in every term of an equation, and it can take any value – a **free index**.

(ii) Any index occurring exactly *twice* in a given term is summed over – a **repeated**, **contracted** or **dummy index**.

(iii) No index can occur more than twice in a given term.

Let's have a look at applying the summation convention to prove the vector triple product identity we stated earlier.

Proposition 2.7.5 (Vector Triple Product Identity)
For vectors $\mathbf{a}, \mathbf{b}, \mathbf{c}$, we have

$$\mathbf{a} \times (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}.$$

Proof. We have

$$\begin{aligned} [\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_i &= \varepsilon_{ijk} a_j (\mathbf{b} \times \mathbf{c})_k \\ &= \varepsilon_{ijk} a_j \varepsilon_{kpq} b_p c_q \\ &= (\varepsilon_{ijk} \varepsilon_{pqk}) a_j b_p c_q, \end{aligned}$$

now $\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$ (we will prove this later on), hence

$$\begin{aligned} [\mathbf{a} \times (\mathbf{b} \times \mathbf{c})]_i &= (\delta_{ip}\delta_{jq}) a_j b_p c_q - (\delta_{iq}\delta_{jp}) a_j b_p c_q \\ &= [(\mathbf{a} \cdot \mathbf{c})\mathbf{b} - (\mathbf{a} \cdot \mathbf{b})\mathbf{c}]_i, \end{aligned}$$

as required. □

§2.7.3 $\varepsilon\varepsilon$ Identities

We will finish this section by looking at $\varepsilon\varepsilon$ identities, which can be used in various computations involving δ and ε symbols.

- $\varepsilon_{ijk}\varepsilon_{pqk} = \delta_{ip}\delta_{jk} - \delta_{iq}\delta_{jp} = \varepsilon_{kij}\varepsilon_{kpq}$.

  To check this, note that RHS and LHS are antisymmetric under $i \leftrightarrow j$, $p \leftrightarrow q$, so both vanish if i & j or p & q take the same value. Now it suffices to check examples $i = p = 1$, $j = q = 2$, where LHS and RHS are both $+1$, or $i = q = 1$, $j = p = 2$, where LHS and RHS are both -1. All other index choices giving non-zero results work similarity.

- $\varepsilon_{ijk}\varepsilon_{pjk} = 2\delta_{ip}$.

  This comes from the above identity, where the LHS becomes $\delta_{ip}\delta_{jj} - \delta_{ij}\delta_{jp} = 3\delta_{ip} - \delta_{ip} = 2\delta_{ip}$.

- $\varepsilon_{ijk}\varepsilon_{ijk} = 6$.

  This follows directly from the previous result, with $i = p$. Then $2\delta_{ii} = 6$.

3 Vectors in a General Setting, $\mathbb{R}^n$ and $\mathbb{C}^n$

In the previous chapter, we studied vectors as they occur when considering position vectors in three dimensions. However, vectors are more general objects, which we will see in this chapter. First, we will consider how to immediately generalise vectors from $\mathbb{R}^3$ (as in the previous chapter) to $\mathbb{R}^n$.

§3.1 Vectors in $\mathbb{R}^n$

By regarding vectors as sets of components, it's easy to generalize from 3 to n dimensions.

Definition 3.1.1 ($\mathbb{R}^n$ Space)
We define $\mathbb{R}^n$ to be the set of n-tuples of real numbers,

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) \mid x_i \in \mathbb{R}\}.$$

We also define **vector addition** and **scalar mulitplication** as

$$\begin{aligned} (x_1, \dots, x_n) + (y_1, \dots, y_n) &= (x_1 + y_1, \dots, x_n + y_n), \\ \text{and} \quad \lambda(x_1, \dots, x_n) &= (\lambda x_1, \dots, \lambda x_n). \end{aligned}$$

As before, we can take arbitrary linear combinations and we also get a notion of parallelism, where $\mathbf{x} \parallel \mathbf{y}$ if and only if $\mathbf{x} = \lambda\mathbf{y}$ or $\mathbf{y} = \lambda\mathbf{x}$ for $\lambda \in \mathbb{R}$.

We also still have a dot product, but in more general settings, such operations are normally referred to as *inner products*.

Definition 3.1.2 (Inner Product)
The inner product on $\mathbb{R}^n$ is defined by

$$\mathbf{x} \cdot \mathbf{y} = \sum_i x_i y_i = x_1 y_1 + x_2 y_2 + \cdot + x_n y_n.$$

Proposition 3.1.3 (Properties of the Inner Product)
In $\mathbb{R}^n$, the inner product satisfies

(i) *Symmetric.* $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$.

(ii) *Bilinear.* $(\lambda\mathbf{x} + \lambda'\mathbf{x}') \cdot \mathbf{y} = \lambda\mathbf{x} \cdot \mathbf{y} + \lambda'\mathbf{x}' \cdot \mathbf{y}$.

(iii) *Positive Definite.* $\mathbf{x} \cdot \mathbf{x} = \sum_i x_i^2 \geq 0$, and is equal 0 if and only if $\mathbf{x} = \mathbf{0}$.

Also, if we return back to vectors in three-dimensions, we had an equivalent definition of the dot product as

$$\mathbf{x} \cdot \mathbf{y} = |\mathbf{x}||\mathbf{y}| \cos\theta,$$

where θ is the angle between the vectors. Using this as a definition of the inner product in $\mathbb{R}^n$, we get a definition for the angle between vectors.

When we look at vector spaces in general, we will define the inner product in general to be any operation that satisfies these properties.

We also have the notions of length and orthogonality, much the same as in three-dimensions.

Definition 3.1.4 (Length/Norm)
For $\mathbf{x} \in \mathbb{R}^n$, we define its **length** or **norm** to be $|\mathbf{x}| \geq 0$ with $|\mathbf{x}|^2 = \mathbf{x} \cdot \mathbf{x}$.

Definition 3.1.5 (Orthogonality)
We say two vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ are **orthogonal** if and only if $\mathbf{x} \cdot \mathbf{y} = 0$.

When working in $\mathbb{R}^n$, we typically use a set of basis vectors known as the **standard basis**. They are

$$\begin{aligned} \mathbf{e}_1 &= (1, 0, 0, \dots, 0), \\ \mathbf{e}_2 &= (0, 1, 0, \dots, 0), \\ &\vdots \\ \mathbf{e}_n &= (0, 0, 0, \dots, 1). \end{aligned}$$

Then we have

$$\mathbf{x} = \sum_i x_i \mathbf{e}_i,$$

and $\mathbf{e}_i \cdot \mathbf{e}_j = \delta_{ij}$, so thus basis is orthonormal.

§3.1.1 Cauchy-Schwarz and the Triangle Inequality

We will now state an inequality about the inner product of two vectors in $\mathbb{R}^n$.

Proposition 3.1.6 (Cauchy-Schwarz Inequality)
For all vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have

$$|\mathbf{x} \cdot \mathbf{y}| \leq |\mathbf{x}| \cdot |\mathbf{y}|,$$

with equality if and only if $\mathbf{x} \parallel \mathbf{y}$.

Proof. If $\mathbf{y} = \mathbf{0}$, then we are done. Otherwise, consider

$$\begin{aligned} |\mathbf{x} - \lambda \mathbf{y}|^2 &= (\mathbf{x} - \lambda \mathbf{y}) \cdot (\mathbf{x} - \lambda \mathbf{y}) \\ &= |\mathbf{x}|^2 - 2\lambda \mathbf{x} \cdot \mathbf{y} + \lambda^2 |\mathbf{y}|^2 \geq 0 \end{aligned}$$

for any $\lambda \in \mathbb{R}$. This is a real quadratic in λ with at most one real root. So it's discriminant is

$$(-2\mathbf{x} \cdot \mathbf{y})^2 - 4|\mathbf{x}|^2 |\mathbf{y}|^2 \leq 0,$$

so equality holds if and only if the discriminant is zero, that is, if $\mathbf{x} = \lambda\mathbf{y}$, as required. □

As vectors in $\mathbb{R}^n$ are just sets of components, we can state this inequality in terms of real numbers. For a sequence of real numbers (a_i) and (b_i), we have

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right),$$

where we have equality if and only if $\frac{a_i}{b_i}$ is a constant for all i where $a_i b_i \neq 0$.

We also have the triangle inequality, similar to the one stated in Chapter 1.

Proposition 3.1.7 (Triangle Inequality)
For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$,

$$|\mathbf{x} + \mathbf{y}| \leq |\mathbf{x}| + |\mathbf{y}|$$

Proof. We have

$$\begin{aligned} |\mathbf{x} + \mathbf{y}|^2 &= |\mathbf{x}|^2 + 2\mathbf{x} \cdot \mathbf{y} + |\mathbf{y}|^2 \\ &\leq |\mathbf{x}|^2 + 2|\mathbf{x}| \cdot |\mathbf{y}| + |\mathbf{y}|^2 \\ &= (|\mathbf{x}| + |\mathbf{y}|)^2, \end{aligned}$$

as required. □

§3.1.2 A comment about $\mathbb{R}^n$ and $\mathbb{R}^3$

The definition we gave for the inner product on $\mathbb{R}^n$ can be written as

$$\mathbf{a} \cdot \mathbf{b} = \delta_{ij} a_i b_j,$$

for $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ where we use the summation convention. For $n = 3$, this matches with the compact definition of the dot product we gave previously.

However, in three dimensions we also had a compact definition of the cross product with

$$(\mathbf{a} \times \mathbf{b})_i = \varepsilon_{ijk} a_j b_k.$$

In $\mathbb{R}^n$ however, we have $\varepsilon_{ij\dots k}$ where there is n indices (we will look at this again in Chapter 5). This is totally antisymmetric, so we cannot use ε to define a vector-valued product for any n except 3.

In $\mathbb{R}^2$ however, we have ε_{ij} where $\varepsilon_{12} = -\varepsilon_{21} = 1$, and we can define an additional scalar product

$$[\mathbf{a}, \mathbf{b}] = \varepsilon_{ij} a_i b_j = a_1 b_2 - a_2 b_1.$$

This corresponds with the signed area of a parallelogram, and $|[\mathbf{a}, \mathbf{b}]| = |\mathbf{a}||\mathbf{b}| \sin\theta$. Compare this with $[\mathbf{a}, \mathbf{b}, \mathbf{c}] = \varepsilon_{ijk} a_i b_j c_k$, the signed volume of the parallelepiped in three dimensions we defined earlier.

§3.2 Vector Spaces

We will now go one step further and define what a vector is in a much more general sense. This will be done by defining the properties that vectors have, rather than what they are themselves.

Definition 3.2.1 (Vector Space)
Let V be a set of objects called **vectors** with operations $\mathbf{v} + \mathbf{w} \in V$ defined for all $\mathbf{v}, \mathbf{w} \in V$, and $\lambda\mathbf{v} \in V$ for all $\mathbf{v} \in V$ and $\lambda \in F$, where F is some field.

Then V is a **vector space** if

(i) V with $+$ is an abelian group.

(ii) $\lambda(\mathbf{v} + \mathbf{w}) = \lambda\mathbf{v} + \lambda\mathbf{w}$.

(iii) $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$.

(iv) $\lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}$.

(v) $1\mathbf{v} = \mathbf{v}$.

In this course, we will deal only with the fields $\mathbb{R}$ and $\mathbb{C}$. If $F = \mathbb{R}$, then we call V a **real vector space**, and if $F = \mathbb{C}$, then we call it a **complex vector space**.

Example 3.2.2 (Example of a Real Vector Space)
Let V be the set of all functions $f : [0, 1] \to \mathbb{R}$ where $f(0) = f(1) = 0$ and f is smooth. Then V is a real vector space.

Indeed we have

$$(f + g)(x) = f(x) + g(x),$$
$$(\lambda f)(x) = \lambda(f(x)).$$

The other properties can be checked as required.

We can also define a *subspace* in the natural way.

Definition 3.2.3 (Subspace)
A **subspace** of a vector space V is a subset $U \subseteq V$ that is also a vector space.

Note that a non-empty subset is a subspace if and only if $\mathbf{v}, \mathbf{w} \in U$ implies $\lambda\mathbf{v} + \mu\mathbf{w} \in U$ where λ, μ are some arbitrary scalars.

For any vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$, their span

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_r\} = \{\lambda_1\mathbf{v}_1 + \cdots + \lambda_r\mathbf{v}_r \mid \lambda_i \in \mathbb{R} \text{ or } \mathbb{C}\}.$$

is a subspace. Also note that V and $\{\mathbf{0}\}$ are both subspaces of a vector space V.

§3.2.1 Linear Dependance & Independence

We will now think about whether sets of vectors are independent.

Definition 3.2.4 (Linear Dependance/Independence)

For $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r \in V$, some vector space, consider the **linear relation**

$$\lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \cdots + \lambda_r \mathbf{v}_r = \mathbf{0} \qquad (\dagger)$$

If ($\dagger$) implies that $\lambda_i = 0$ for all i, then we these vectors form a **linearly independent** set. If ($\dagger$) holds with some $\lambda_i \neq 0$, then they form a **linearly dependant** set.

Example 3.2.5 (Example of a Linearly Dependent Set)

In $\mathbb{R}^3$, the set $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 2 \end{pmatrix} \right\}$ is linearly dependent, since

$$0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} - 1 \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which is a non-trivial linear relation.

Any set containing $\mathbf{0}$ is linearly dependent, since we can $\lambda \mathbf{0} = \mathbf{0}$ for any λ.

In $\mathbb{R}^3$, the set of vectors $\{\mathbf{a}, \mathbf{b}, \mathbf{c}\}$ are linearly independent if $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) \neq \mathbf{0}$, since if $\alpha \mathbf{a} + \beta \mathbf{b} + \gamma \mathbf{c} = \mathbf{0}$, then dotting with $\mathbf{b} \times \mathbf{c}$ gives $\alpha \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = 0 \implies \alpha = 0$, and β, γ similarly.

§3.2.2 Inner Products

A real vector space can have an additional structure that we can define by axioms.

Definition 3.2.6 (Inner Product)

For a real vector space V and $\mathbf{v}, \mathbf{w} \in V$, an **inner product** denoted by

$$\mathbf{v} \cdot \mathbf{w} \quad \text{or} \quad (\mathbf{v}, \mathbf{w}) \in \mathbb{R}$$

is a map that is symmetric, bilinear, and positive definite.

If these conditions hold, then we we get things like the Cauchy-Schwartz inequality and a norm as by the previous argument.

Example 3.2.7 (Inner Product on a Real Vector Space)

Consider the real vector space of functions $V = \{f : [0, 1] \to \mathbb{R} \mid f \text{ smooth, and} f(0) = f(1) = 0\}$. We can define the inner product as

$$(f, g) = \int_0^1 f(x) g(x) \, \mathrm{d}x,$$

which satisfies the conditions specified above. Then we have

$$||(f, g)|| \leq ||f|| \cdot ||g||,$$

that is,

$$\left|\int_0^1 f(x)g(x)\,\mathrm{d}x\right| \leq \left(\int_0^1 f(x)^2\,\mathrm{d}x\right)^{1/2}\left(\int_0^1 g(x)^2\,\mathrm{d}x\right)^{1/2}.$$

We can also say something about linear independence from an inner product.

Lemma 3.2.8

In any real vector space V with an inner product, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are non-zero and orthogonal vectors, then they are linearly independent.

Proof. If $\sum_i \alpha_i \mathbf{v}_i = \mathbf{0}$, then $(\mathbf{v}_j, \sum_i \alpha_i \mathbf{v}_i) = 0$ for a fixed j, so $\alpha_j(\mathbf{v}_j, \mathbf{v}_j) = 0$, and thus $\alpha_j = 0$. □

§3.3 Basis and Dimension

We can generalize the notion of a basis to an arbitrary vector space.

Definition 3.3.1 (Basis)

For a vector space V, a **basis** is a set $B = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ such that

(i) B spans V, that is, any $\mathbf{v} \in V$ can be written $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e}_i$.

(ii) B is linearly independent.

The linear independence implies that the coefficients v_i in (i) are unique. We say that v_i are the components of $\mathbf{v}$ with respect to the basis B.

Example 3.3.2 (Standard Basis)

The standard basis for $\mathbb{R}^n$ consists of the vectors

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad \mathbf{e}_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

This is a basis since it satisfies the definition as $\mathbf{x} \in \mathbb{R}^n$, we have

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n,$$

and $\mathbf{x} = \mathbf{0}$ if and only if $x_1 = x_2 = \cdots = x_n = 0$.

Of course in the example above we just used one particular base. For example in $\mathbb{R}^2$ another basis consists of $\{(1,0),(1,1)\}$.

Theorem 3.3.3 (The Basis Theorem)

If $\{\mathbf{e}_1, \dots, \mathbf{e}_m\}$ and $\{\mathbf{f}_1, \dots, \mathbf{f}_m\}$ are both bases, then $n = m$.

Proof. We can find coefficients A_{ai} and B_{ia} such that $\mathbf{f}_a = \sum_i A_{ai}\mathbf{e}_i$ and $\mathbf{e}_i = \sum_a B_{ia}\mathbf{f}_a$. So we have that

$$\begin{aligned}\mathbf{f}_a &= \sum_i A_{ai}\left(\sum_b B_{ib}\mathbf{f}_b\right)\\ &= \sum_b\left(\sum_i A_{ai}B_{ib}\right)\mathbf{f}_b,\end{aligned}$$

but coefficients with respect to a basis are unique, hence

$$\sum_i A_{ai}B_{ib} = \delta_{ab}.$$

Similarly, $\sum_a B_{ia}A_{aj} = \delta_{ij}$. Finally,

$$\begin{aligned}\sum_{ia} A_{ai}B_{ia} &= \sum_a \delta_{aa} = m\\ &= \sum_i \delta_i i = n,\end{aligned}$$

hence $n = m$. □

This theorem allows us to state the following definition.

Definition 3.3.4 (Definition)

The **dimension** of a vector space V, written $\dim V$, is the size of its basis.

Note that $\mathbb{R}^n$ has dimension n.

The step sin the proof of the basis theorem are within the scope of the course, but you will not be expected to prove this theorem without prompts (it is non-examinable in the schedules). The same applies with the following.

Proposition 3.3.5

Let V be a vector space with finite subsets $Y = \{\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m\}$ that spans V, and $X = \{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ that is linearly independent. Then $k \leq n \leq m$ where $n = \dim V$. Also

(i) A basis can be formed from subsets of Y,

(ii) A basis can be obtained from X by adding additional vectors (from Y if necessary).

Proof. (i) If Y is linearly independent, then it's a basis and $m = n = \dim V$. If

Y is not linearly independent, then

$$\sum_{i=1}^{m} \lambda_i \mathbf{w}_i = \mathbf{0}, \quad \text{with some } \lambda_i \neq 0.$$

Take $\lambda_m \neq 0$ without loss of generality, then $\mathbf{w}_m = -\frac{1}{\lambda_m} \sum_{i=1}^{m-1} \mathbf{w}_i$, so $\operatorname{span} Y = \operatorname{span} Y'$, with $Y' = \{\mathbf{w}_1, \dots, \mathbf{w}_{m-1}\}$. This can be repeated until a basis is obtained.

(ii) If X spans V then it's a basis and $k = n = \dim V$. If not, then there exists $\mathbf{u}_{k+1} \in V$ not in $\operatorname{span} X$. But then $\sum_{i=1}^{k+1} \mu_i \mathbf{u}_i = \mathbf{0}$ implies $\mu_{k+1} = 0$, thus $X' = \{\mathbf{u}_1, \dots, \mathbf{u}_{k+1}\}$ is linearly independent. Furthermore, we can choose $\mathbf{u}_{k+1} \in Y$ as otherwise $Y \subseteq \operatorname{span} X \implies \operatorname{span} Y \subseteq \operatorname{span} X \implies \operatorname{span} X = V$. We can repeat this process until we obtain a basis (and this must stop since Y is a finite set). $\square$

Not every vector space has a finite dimension, but in this course we will deal only with vector spaces that do, save for the occasional example.

Example 3.3.6 (Example of an Infinite Dimensional Vector Space)
Consider the vector space

$$V = \{f : [0,1] \to \mathbb{R} \mid \text{f smooth}, f(0) = f(1) = 0\}.$$

Then consider $S_n(x) = \sqrt{2}\sin(n\pi x)$, $n = 1, 2, \dots$. These belong to V and

$$(S_n, S_m) = 2\int_0^1 \sin(n\pi x)\sin(m\pi x)\, \mathrm{d}x = \delta_{mn},$$

so they are all linearly independent. But we have infinitely many of them, so V is infinite dimensional.

§3.4 Vectors in $\mathbb{C}^n$

So far we have looked at real vector spaces, and particularly at R^n. We can define a similar vector space over the complex numbers.

Definition 3.4.1 ($\mathbb{C}^n$ Space)
We define $\mathbb{C}^n$ to be the set of n-tuples of complex numbers,

$$\mathbb{C}^n = \{(z_1, z_2, \dots, z_n) \mid z_j \in \mathbb{C}\}.$$

We also define **addition** and scalar multiplication with

$$\begin{aligned} &(z_1, \dots, z_n) + (w_1, \dots, w_n) = (z_1 + w_1, \dots, z_n + w_n) \\ \text{and} \quad &\lambda (z_1, \dots, z_n) = (\lambda z_1, \dots, \lambda z_n) \end{aligned}$$

Note that we could have used real scalars and have gotten a real vector space, but having a complex vector space is more natural.

The distinction between real and complex vector spaces is important. For example, if we have $\mathbf{z} = (z_1, \ldots, z_n) \in \mathbb{C}^n$, with $z_j = x_j + iy_j$ then

$$\mathbf{z} = \sum_j x_j \mathbf{e}_j + \sum_j y_i \mathbf{f}_i$$

is a real linear combination, and thus $\{\mathbf{e}_1, \ldots, \mathbf{e}_n, \mathbf{f}_1, \ldots, \mathbf{f}_n\}$ is a basis for $\mathbb{C}^n$ as a real vector space, and it has dimension $2n$. However, taking C^n as defined before, it is of dimension n which is clearly different.

From now on we will view $\mathbb{C}^n$ as a complex vector space unless we say otherwise.

§3.4.1 Complex Inner Product

We can define an inner product on $\mathbb{C}^n$.

Definition 3.4.2 (Complex Inner Product)
For $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$, we define the inner product to be

$$(\mathbf{z}, \mathbf{w}) = \sum_j \overline{z_j} w_j = \overline{z_1} w_1 + \cdots + \overline{z_n} w_n.$$

Proposition 3.4.3 (Properties of the Complex Inner Product)
The complex inner product satisfies

(i) *Hermitian.* $(\mathbf{w}, \mathbf{z}) = \overline{(\mathbf{z}, \mathbf{w})}$.

(ii) *Linear/anti-linear.* $(\mathbf{z}, \lambda\mathbf{w} + \lambda'\mathbf{w}') = \lambda(\mathbf{z}, \mathbf{w}) + \lambda'(\mathbf{z}, \mathbf{w}')$ and $(\lambda\mathbf{z} + \lambda'\mathbf{z}', \mathbf{w}) = \overline{\lambda}(\mathbf{z}, \mathbf{w}) + \overline{\lambda'}(\mathbf{z}', \mathbf{w})$.

(iii) *Positive definite.* $(\mathbf{z}, \mathbf{z}) = \sum_j |z_j|^2$, which is real, non-negative and zero if and only if $\mathbf{z} = \mathbf{0}$.

Proof Sketch. Check definitions. □

We also define *length* or *norm* of $\mathbf{z}$ to be $|\mathbf{z}| \geq 0$ with $|\mathbf{z}|^2 = (\mathbf{z}, \mathbf{z})$.

We can also say that $\mathbf{z}, \mathbf{w} \in \mathbb{C}^n$ are orthogonal if $(\mathbf{z}, \mathbf{w}) = 0$. With this, note that the standard basis for $\mathbb{C}^n$ is orthonormal.

4 Matrices and Linear Maps

5 Determinants and Inverses

6 Eigenvalues and Eigenvectors

7 Changing Basis, Canonical Forms and Symmetries