CHAPTER 0

INTRODUCTION

In this chapter we introduce key concepts that will be used in later chapters. For this reason, unlike other chapters it contains many statements, sometimes given without thorough explanations or reasoning. While all of these statements are grounded in deep ideas and can be formulated in a rigorous manner, it is advised to first get an intuitive understanding of the ideas before diving into their more formal construction.

Note 0.1 In case you are already familiar with the topics

It is recommended for readers who are familiar with the topics to at least gloss over this chapter and make sure they know and understand all the concepts presented here.

!

0.1 MATHEMATICAL SYMBOLS AND SETS

0.1.1 Logical Statements and their Truth Value

We start our discussion with the simplest mathematical concept: a **proposition**. A proposition is simply a statement that might be either **true** or **false**.

Example 0.1 Truth of propositions

- $3 > 1$ (**true**)
- $-2 = 5 - 7$ (**true**)
- $7 < 5$ (**false**)
- The radius of the earth is bigger than that of the moon. (**true**)
- The word 'House' starts with the letter 'G'. (**false**)

We can group together propositions using **logical operators**. Two of the most common logical operators are **AND** and **OR**.

The **AND** operator returns a **true** statement only if **both** the statements it groups are themselves **true**, otherwise it returns **false**.

Example 0.2 The AND operator

- $2+4=6$ is **true**, $4-2=2$ is **true**. $(2+4=6$ **AND** $4-2=2)$ is therefore **true**.
- $2+4=6$ is **true**, $2>6$ is **false**. $(2+4=6$ **AND** $2>6)$ is therefore **false**.
- $\frac{10}{2}=1$ is **false**, $2^4=16$ is **true**. $\left(\frac{10}{2}=1 \textbf{ AND } 2^4=16\right)$ is therefore **false**.
- $7<5$ is **false**, $10+2=13$ is **false**. $(7<5$ **AND** $10+2=13)$ is therefore **false**.

The **OR** operator returns **true** if **at least** one of the statements it groups is true.

Example 0.3 The OR operator

- $2+4=6$ is **true**, $4-2=2$ is **true**. $(2+4=6$ **OR** $4-2=2)$ is therefore **true**.
- $2+4=6$ is **true**, $2>6$ is **false**. $(2+4=6$ **OR** $2>6)$ is therefore **true**.
- $\frac{10}{2}=1$ is **false**, $2^4=16$ is **true**. $\left(\frac{10}{2}=1 \textbf{ OR } 2^4=16\right)$ is therefore **true**.
- $7<5$ is **false**, $10+2=13$ is **false**. $(7<5$ **OR** $10+2=13)$ is therefore **false**.

Table 0.1 The truth table for the operators **AND** and **OR**.

A	B	A AND B	A OR B
true	**true**	**true**	**true**
true	**false**	**false**	**true**
false	**true**	**false**	**true**
false	**false**	**false**	**false**

Table 0.2 Common mathematical notations used in this Book.

Symbol	In words
$\neg a$	**not** a
$a \wedge b$	a **and** b
$a \vee b$	a **or** b
$a \Rightarrow b$	a **implies** b
$a \Leftrightarrow b$	a **is equivalent to** b
$\forall x$	**For all** x (...)
$\exists x$	**There exists** x **such that** (...)
$a := b$	a **is defined to be** b
$a \equiv b$	a **is equivalent to** b

The behaviour of both operators can be summarized using a **truth table** (see Table 0.1 below).

When writing, it is convenient to use **notations** to represent operators: the **AND** operator is denoted by $\wedge$, while the **OR** operator is denoted by $\vee$.

Example 0.4 Using the notations for AND and OR

$$\underbrace{(2+2=5)}_{\text{false}} \wedge \underbrace{(1-1=0)}_{\text{true}} \Rightarrow \textbf{false}$$

$$\underbrace{(2+2=5)}_{\text{false}} \vee \underbrace{(1-1=0)}_{\text{true}} \Rightarrow \textbf{true}$$

0.1.2 Common mathematical notations

Several more common mathematical notations are given in Table 0.2.

The notation $\Rightarrow$ need a bit of clarification: implication means that we can directly derive a proposition from another proposition. For example, if $x = 3$ then $x > 2$. The opposite implication can be a **false** statement, i.e. for the example above $x > 2$ does not imply $x = 3$ (denoted as $x > 2 \not\Rightarrow x = 3$). Sometimes implication is expressed by using the word *if*: in the above example $x > 2$ if $x = 3$, but the other way around is not **true**.

We say that two propositions are **equivalent** when they imply each other. For example:

$x = 2$ implies that $\frac{x}{2} = 1$, while $\frac{x}{2} = 1$ implies that $x = 2$. We can write this as

$$\frac{x}{2} = 1 \Leftrightarrow x = 2.$$

Instead of the word *equivalent*, the phrase *if and only if* (sometimes shortened to **iff**) is commonly used, e.g.

$$x = 2 \text{ iff } \frac{x}{2} = 1.$$

0.1.3 Sets and subsets

The concept of **sets** is perhaps one of the most basic ideas in modern mathematics. Much of the material covered in this book will be built upon sets and their properties. However, as with the rest of the material presented here - our description of sets will not be thorough nor precise.

For our purposes, a set is a collection of **elements**. These elements can be any concept - be it physical (a chair, a bicycle, a tapir) or abstract (a number, an idea). However, we will consider only sets comprised of numbers. Sets can have finite of infinite number of elements in them.

We denote sets by using curly brackets, and if the number of elements in them is not too big - we display the elements, separated by commas, inside the brackets. In other cases we can express the sets as a sentence or a mathematical proposition.

Example 0.5 Simple sets

$$\{1, 2, 3, 4\} \qquad \left\{-4, \frac{3}{7}, 0, \pi, 0.13, -2.5, \frac{e}{3}, 2^{-\pi}\right\} \qquad \{\text{all even numbers}\}$$

Sets have two important properties:

0.1. Elements in a set do not repeat. i.e each element is unique.

0.2. The order of elements in a set does not matter.

Example 0.6 Important set properties

Examples demonstrating the two aforementioned important properties of sets:

0.1. The following is not a proper set:

$$\{1, 1, 0, 1, 0, 0, -1, 0, 0, -1, -1, 1\}$$

0.2. The following sets are all identical:

$$\{1, 2, 3, 4\} \qquad \{1, 3, 2, 4\} \qquad \{3, 4, 1, 2\} \qquad \{1, 3, 2, 4\} \qquad \{4, 3, 2, 1\}$$

Sets can be denoted using **conditions**, with the symbol | representing the phrase ”such that”.

Example 0.7 Defining a set using a condition

he following set contains all the odd whole numbers between 0 and 10, including both:

$$\{0 < x < 10 \mid x \text{ is an odd number}\}.$$

The definition of this set can be read as

all numbers x that are bigger than 0 *and are smaller than* 10, *such that x is odd.*

(note that the requirement of x to be an odd number means that it is necessarily a whole number as well)

This set can be written explicitly as

$$\{1, 3, 5, 7, 9\}.$$

Sets are usually denoted with an uppercase Latin letter $(A, B, C, \dots)$, while their elements are denoted as lowercase letters $(a, b, \alpha, \phi, \dots)$. When we want to denote that an element belongs to a set we use the following symbol: $\in$. Conversely, $\notin$ is used to denote that an element *does not* belong to a set.

Example 0.8 Elements in sets

For the two sets

$$A = \{1, 2, 5, 7\}, \quad B = \{\text{even numbers}\},$$

all the following propositions are **true**:

$$\begin{aligned} &1 \in A, \quad 2 \in A, \quad 5 \in A, \quad 7 \in A, \\ &2 \in B, \quad 1 \notin B, \quad 5 \notin B, \quad 7 \notin B. \end{aligned}$$

The number of elements in a set, also called its **cardinality** is denoted using two vertical bars (similar to the way absolute values are denoted).

Example 0.9 Cardinality

For $S = \left\{-3, 0, -2, 7, 1, \frac{1}{2}, 5\right\}$, $|S| = 7$.

An important special set is the **empty set**, which is the set containing no elements. It is denoted by $\emptyset$, and has the unique property that $|\emptyset| = 0$.

0.1.4 Intersection, union, difference and complement sets

Two sets are equal if they both contain the exact same elements and only these elements, i.e.

$$A = B \Longleftrightarrow x \in A \Leftrightarrow x \in B. \tag{0.1.1}$$

This proposition reads 'The sets A and B are equal *if and only if* any element x in A is also in B, and any element x in B is also in A'. When all the elements of a set B are also elements of another set A, we say that B is a **subset** of A, and we denote that as $B \subset A$. In mathematical notation, we write

$$B \subset A \Leftrightarrow \forall x \in B, x \in A. \tag{0.1.2}$$

i.e. B is a subset of A **iff** the following is true: any element in B is also an element in A.

Note 0.2 (not so) Surprising properties of subsets

The definition of a subset (Equation 0.1.2) gives rise to two interesting properties:

- The empty set $\emptyset$ is a subset of any set.
- Any set is a subset of itself.

Note 0.3 The uniqueness of $\emptyset$

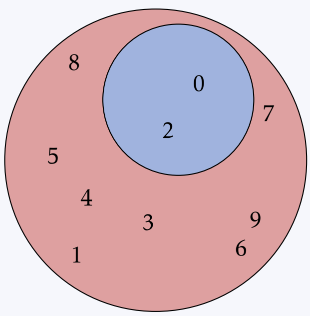
There is only a single empty set, as any set that has no elements is equivalent to any other set with no elements (i.e. they have the same elements). Due to the way subsets are defined, the empty set is a subset of any set (including itself!).

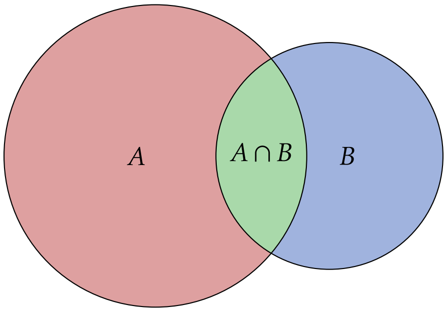
Of course, since we have a definition for a subset, the opposite concept also exists: if B is a subset of A, then we say that A is a **superset** of B.

A very useful way of illustrating the relationship between two or more sets is by using **Venn diagrams**, where sets are represented by circles (or other 2D shapes).

Example 0.10 Subsets and Venn diagrams

A Venn diagram depicting the set $B = \{0, 2\}$ as a subset of $A = \{0, 1, 2, 3, 4, \ldots, 9\}$:





If for two sets A, B both $A \subset B$ and $B \subset A$, then $A = B$. We can write this fact as a mathematical proposition:

$$(A \subset B) \wedge (B \subset A) \Leftrightarrow A = B. \tag{0.1.3}$$

The **intersection** of two sets A and B, denoted $A \cup B$, is the set of all elements x such that $x \in A$ **AND** $x \in B$:

$$A \cup B = \{x \mid x \in A \wedge x \in B\}. \tag{0.1.4}$$

Example 0.11 Intersection of sets

The intersection of the sets $A = \{1,2,3,4\}$ and $B = \{3,4,5,6\}$ is the set $A \cup B = \{3,4\}$. The intersection of the sets $C = \{0,1,2,6,7\}$ and $D = \{3,9,-4,5\}$ is the empty set $\emptyset$, since no element is in both sets.

The following Venn diagram depicts the intersection of two sets (the green area):

Note 0.4 Disjoint sets

When the intersection of two sets is the empty set, we say that the set is **disjoint**.

The **union** of two sets (denoted using the symbol $\cup$) is the set composed of all the elements that belong to any of the sets, including elements that are in both sets:

$$A \cup B = \{x \mid x \in A \vee x \in B\}. \tag{0.1.5}$$

Example 0.12 Union of sets

The union of the sets $A = \{1,2,3,4\}$ and $B = \{3,4,5,6\}$ is the set $A \cup B = \{1,2,3,4,5,6\}$.
The union of the sets $C = \{0,1,2,6,7\}$ and $D = \{3,9,-4,5\}$ is the set $C \cup D = \{0,1,2,3,-4,5,6,7,9\}$.

The following Venn diagram depicts the union of two sets (the purple area):

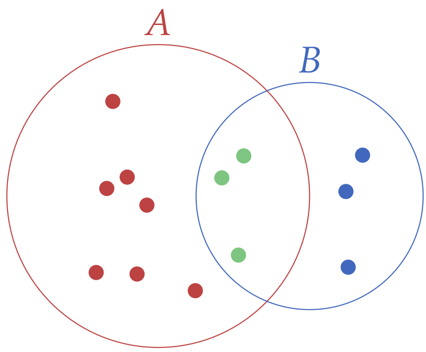
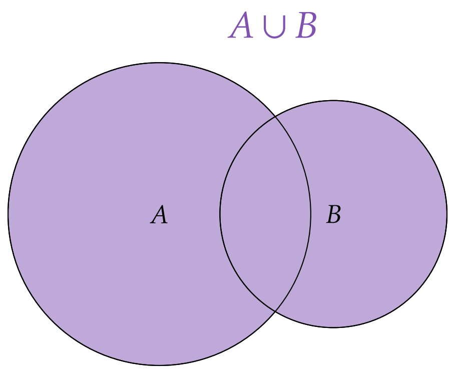


Figure 0.1 Counting the number of elements in the union of two sets: A has 10 elements (red + green dots), while B has 6 elements (blue + green dots). If we count both we get 16 elements, but this counts the joint elements (green dots) twice. Therefore we should subtract the number of joint points, and get that there are only 13 elements in the union.



Naively, the number of elements of a union $A \cup B$ is simply the sum of the number of elements in A and the number of elements in B. However, this naive approach might count the elements in both sets twice: once for A and once for B (see Figure 0.1) - this is exactly the set $A \cap B$. We therefore subtract the number of elements in $A \cap B$ and get

$$|A \cup B| = |A| + |B| - |A \cap B|. \tag{0.1.6}$$

When two sets A, B are disjoint, then $|A \cap B| = 0$, and so $|A \cup B| = |A| + |B|$.

The definitions of intersections and unions can be easily extended to any whole number of sets.

Example 0.13 Intersection and union of 3 sets

The intersection of 3 sets $A = \{1, 2, 3, 4, 5\}$, $B = \{-2, -1, 0, 1, 2\}$ and $C = \{2, 3, 4, 5, 6\}$ is the set of all elements that are in A and in B and in C, i.e. the set $A \cap B \cap C = \{2\}$. The union of these sets is the set of all elements that are in either of the sets, i.e. $A \cup B \cup C = \{-2, -1, 0, 1, 2, 3, 4, 5, 6\}$.

The most general definition of an intersection of n sets (where n is a whole number), which we will call $A_1, A_2, A_3, \cdots, A_n$ is

$$A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n = \{x \mid (x \in A_1) \wedge (x \in A_2) \wedge (x \in A_3) \wedge \cdots \wedge (x \in A_n)\}. \tag{0.1.7}$$

The left hand side of Equation 0.1.7 can be written as

$$A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i. \tag{0.1.8}$$

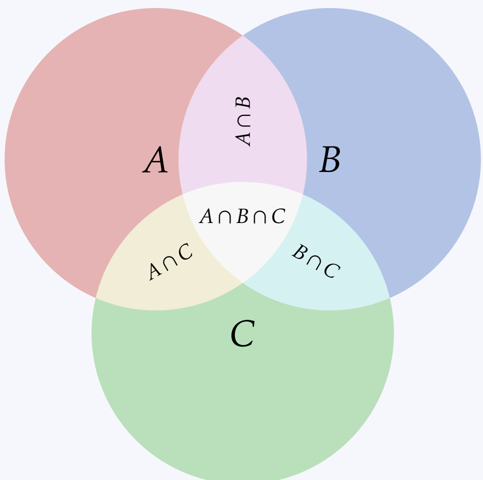
(clarifying the notation? i.e. indexing, etc.)

Similarly, the union of n different sets is defined as

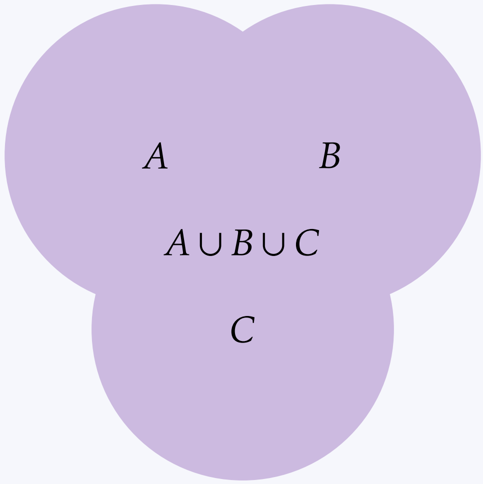
$$\begin{aligned} \bigcup_{i=1}^{n} A_i &= A_1 \cup A_2 \cup A_3 \cup \cdots \cup A_n \\ &= \{x \mid (x \in A_1) \vee (x \in A_2) \vee (x \in A_3) \vee \cdots \vee (x \in A_n)\}. \end{aligned} \tag{0.1.9}$$

Example 0.14 Venn diagrams: intersection and union of 3 sets

The following Venn diagram shows all possible intersections between three sets:



...and the following Venn diagram depicts the union of the same three sets:



The **difference** of two sets A and B (written $A-B$ or $A\setminus B$) is, in a sense, the opposite of their intersection: it is the set of all elements in A that are not in B. Note that $A-B$ doesn't necessarily equal $B-A$, i.e. it is not **commutative**.

Example 0.15 Difference of two sets

Given the two sets $A=\{1,2,3,4,5\}$ and $B=\{3,4,5,6,7,8,9\}$,

$$\begin{aligned} A-B &= \{1,2\}, \\ B-A &= \{6,7,8,9\}. \end{aligned}$$

(note how in this case $A-B\neq B-A$)

Given a set A and a subset of A, $B\subset A$, we can define the **complement** of B in relation to A (notation: $B^{\complement}$) as all the elements in A that are not in B. As the name suggest, the elements of $B^{\complement}$ complete B: $B\cup B^{\complement}=A$.

Example 0.16 Complement of a set

Given the set $A=\mathbb{Z}$ and $B=\{x\in\mathbb{Z}\mid x \text{ is odd}\}$, the complement $B^{\complement}$ in relation to A is the set of all even numbers. The reason is that any integer number can be either odd (in which case it belongs in B) or even (in which case it belongs in $B^{\complement}$).

Given a set A with $|A|$ elements - how many different subsets does it have? We'll start by looking at a practical example: $A=\{1,2,3\}$. We can imminently see that any set which contains just one of the elements of A is a subset of A, i.e. $\{1\},\{2\},\{3\}$ are all subsets of A. In addition, any set which contains only two elements from A is a subset of A, i.e. $\{1,2\},\{1,3\},\{2,3\}$. Of course, we must not forget the empty set and A itself - both subsets of A (see Note 0.2). Thus altogether A has 8 subsets:

$$\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}.$$

Generally, any set A with $|A|$ elements has $2^{|A|}$ different subsets. The set of all these subsets is called the **power set** of A, and is denoted as $P(A)$.

Example 0.17 Power set

The power set of $A = \{1,2,3\}$ is

$$P(A) = \{\emptyset, \{1\}, \{2\}, \{3\}, \{1,2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

0.1.5 Important number sets

It is now time to introduce some important number sets. We begin with the simplest of these sets: the **natural numbers**, denoted by $\mathbb{N}$. These are the numbers $1, 2, 3, 4, \dots$. Adding the opposites to the natural numbers and adding 0 to the set yields the **integers**, denoted by $\mathbb{Z}$. Loosely speaking, we can define the integers as

$$\mathbb{Z} = \{0, \pm 1, \pm 2, \pm 3, \pm 4, \dots\}. \tag{0.1.10}$$

This makes the integers a superset of the natural numbers, i.e.

$$\mathbb{N} \subset \mathbb{Z}. \tag{0.1.11}$$

One can think of the integers as all the number needed for solving an equation of the form $a + x = b$, where a and b are integers themselves, and x is an unknown. No matter which integer values we put in a and b, the unknown x will always be an integer as well (whether it be positive, negative or zero depends on the values of a and b). However, when one wishes to solve an equation of the sort $ax = b$, the integers are not longer sufficient: for example, if $a = 2$ and $b = 1$, then x is not an integer.

To solve $ax = b$ (where $a, b \in \mathbb{Z}$) we must introduce the **rational numbers**: numbers with values that are ratios of two integers. We denote the set of rational numbers with the symbol $\mathbb{Q}$, and write

$$\mathbb{Q} = \left\{ \frac{a}{b} \;\middle|\; a, b \in \mathbb{Z} \wedge b \neq 0 \right\}. \tag{0.1.12}$$

(TBW: discuss briefly why $b \neq 0$)

For some combinations of a and b the ratio $\frac{a}{b}$ is an integer. For example: $\frac{3}{1}$, $\frac{8}{4}$, $\frac{-2}{2}$. This makes the integers a subset of the rational numbers, i.e.

$$\mathbb{Z} \subset \mathbb{Q}. \tag{0.1.13}$$

About 2500 years ago it was discovered that some numbers are not rational (and thus also not integers). The most famous example is the number $\sqrt{2}$ - there are not two integers a, b such that $\frac{a}{b} = \sqrt{2}$. We call some of these numbers **algebraic numbers** (denoted by $\mathbb{A}$), and what makes them special is that they are solutions to **polynomial equations**, which we will not define yet (see section xxx). Instead, here is an example for a 2nd order polynomial equation (called a **quadratic equation**):

$$x^2 - 2x - 1 = 0. \tag{0.1.14}$$

Similar to what we saw before, the rational numbers are a subset of the algebraic numbers, i.e.

$$\mathbb{Q} \subset \mathbb{A}. \tag{0.1.15}$$

The algebraic numbers together with other non-rational numbers, such as π and e, form the set of **real numbers**, denoted as $\mathbb{R}$. The definition of real numbers is way beyond the scope of this book, but it is important to understand that the progression we used so far still holds, i.e.

$$\mathbb{A} \subset \mathbb{R}. \tag{0.1.16}$$

The final set of numbers we will touch upon here is the set of **complex numbers**, denoted $\mathbb{C}$, which we can define as

$$\mathbb{C} = \left\{ a + \mathrm{i}b \mid a, b \in \mathbb{R},\ \mathrm{i} = \sqrt{-1} \right\}. \tag{0.1.17}$$

When $b = 0$, Equation 0.1.17 becomes just a single real number - and so

$$\mathbb{R} \subset \mathbb{C}. \tag{0.1.18}$$

(Section 0.6 is dedicated to complex numbers)

Equations 0.1.11-0.1.18 can be merged together to the following single equation:

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{A} \subset \mathbb{R} \subset \mathbb{C}. \tag{0.1.19}$$

There are more advanced constructions that generalize the complex numbers (i.e. create supersets of the complex number set). These include **quaternions** and **Clifford algebras**. However, as stated before, we will not consider them in this book.

0.1.6 Intervals on the real number line

An important concept that is easily defined over the set $\mathbb{R}$ is an **interval**. A **closed interval** $[a, b]$ is a subset of $\mathbb{R}$ which is defined as

$$[a, b] = \{x \in \mathbb{R} \mid a \leq x \leq b\}. \tag{0.1.20}$$

An **open interval** (a, b) is a subset of $\mathbb{R}$ which is defined as

$$(a, b) = \{x \in \mathbb{R} \mid a < x < b\}. \tag{0.1.21}$$

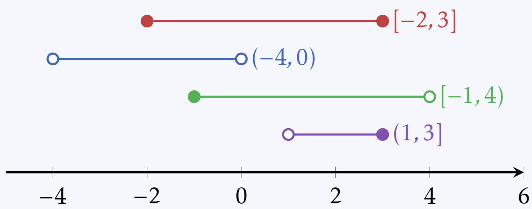
The difference between closed and open intervals is the inclusion and exclusion, respectively, of the edge point: in a closed interval the points a, b are included, while they are not included in an open interval. Of course, we can also create **half open intervals**, i.e.

$$
\begin{aligned}
[a, b) &= \{x \in \mathbb{R} \mid a \leq x < b\}, \\
(a, b] &= \{x \in \mathbb{R} \mid a < x \leq b\}, \qquad (0.1.22)
\end{aligned}
$$

where the first interval includes a but not b, and the second interval includes b but not a.

Example 0.18 Intervals

Intervals can be drawn as colored line segments on top of the real number line:



Note how a full point denotes a closed edge, while an empty point denotes an open edge.

In some cases, it is necessary to use intervals that are infinite in one side, i.e. the left or the right edge are at infinity. In these cases, we use the symbol ∞ to denote infinity, and always keep the interval open at that end:

$$
\begin{aligned}
(-\infty, b) &= \{x \mid x < b\}, \\
(-\infty, b] &= \{x \mid x \leq b\}, \\
(a, \infty) &= \{x \mid x > a\}, \\
[a, \infty) &= \{x \mid x \geq a\}. \qquad (0.1.23)
\end{aligned}
$$

0.1.7 Cartesian Products

The **Cartesian product** of two sets A, B (denoted $A \times B$) is the set of all possible **ordered** pairs, where the first component is an element of A and the second component is an element of B:

$$A \times B = \{(a, b) \mid a \in A,\ b \in B\}. \qquad (0.1.24)$$

Example 0.19 Cartesian products

Consider $A = \{1, 2, 3\}$, $B = \{x, y\}$. Then

$$A \times B = \{(1, x),\ (1, y),\ (2, x),\ (2, y),\ (3, x),\ (3, y)\}$$

The concept of 'ordered pairs' is paramount: if we reverse the order of the elements in a pair the result might not be in the Cartesian product. We therefore say that the Cartesian product is **not commutative**.

Example 0.20 Non-commutativity of the Cartesian product

The elements $(x,1)$, $(y,1)$, $(x,2)$ and so on **are not** in the Cartesian product $A\times B$ as defined in the previous example, since in each one of the pairs the first element is from B and the second element is from A.

The number of elements in a Cartesian product is the product of the number of elements in each of the sets it is composed of, i.e.

$$|A\times B| = |A|\cdot|B|. \tag{0.1.25}$$

Example 0.21 Number of elements in a Cartesian product

The Cartesian product described in the previous two examples has in total $3\cdot 2=6$ elements, as seen in **??**.

As with intersections and unions, the definition of a Cartesian product can be expanded into any natural number of sets:

$$A_1\times A_2\times\cdots\times A_n = \{(x_1,x_2,\dots,x_n)\mid x_1\in A_1,\ x_2\in A_2,\ \dots,x_n\in A_n\}. \tag{0.1.26}$$

Example 0.22 Cartesian product of three sets

The Cartesian product of the sets $A=\{1,2,3\}$, $B=\{x,y\}$, $C=\{\alpha,\beta\}$ is

$$\begin{aligned} A\times B\times C = \{&(1,x,\alpha),\ (1,x,\beta),\ (1,y,\alpha),\ (1,y,\beta)\\ &(2,x,\alpha),\ (2,x,\beta),\ (2,y,\alpha),\ (2,y,\beta)\\ &(3,x,\alpha),\ (3,x,\beta),\ (3,y,\alpha),\ (3,y,\beta)\}. \end{aligned}$$

A special case of Cartesian products are those products for which all the sets composing them are the same set. We denote these as the respective integer power, for example the Cartesian product $\mathbb{R}\times\mathbb{R}$ is denoted as $\mathbb{R}^2$, the Cartesian product $\mathbb{R}\times\mathbb{R}\times\mathbb{R}$ is denoted as $\mathbb{R}^3$, etc.

Specifically, the Cartesian product $\mathbb{R}^2$ can be interpreted as the two-dimensional **Euclidean space**, which is the space used to draw graphs in one-dimensional calculus and shapes in two-dimensional analytical geometry. We will explore this idea (and higher dimensional spaces) in more details in upcoming chapters.

0.2 RELATIONS AND FUNCTIONS

0.2.1 Basics

The Cartesian product of two sets can be viewed as describing all possible connections between the elements of the first set to the elements of the second set, and thus any subset of a Cartesian product forms a specific **relation** between the sets.

Example 0.23 Relations as subsets of Cartesian products

Given the following two sets:

$$A = \{1, 2, 3, 4\}, \ B = \{\alpha, \beta, \gamma\},$$

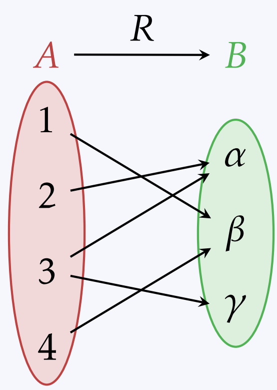
then

$$\begin{aligned} A \times B = \{&(1,\alpha), \ (1,\beta), \ (1,\gamma), \\ &(2,\alpha), \ (2,\beta), \ (2,\gamma), \\ &(3,\alpha), \ (3,\beta), \ (3,\gamma), \\ &(4,\alpha), \ (4,\beta), \ (4,\gamma)\}. \end{aligned}$$

We can choose the following pairs to form a subset of $A \times B$:

$$R = \{(1,\beta), \ (2,\alpha), \ (3,\alpha), \ (3,\beta), (4,\gamma)\}.$$

R is thus a relation between A and B. We can graphically illustrate R as follows:



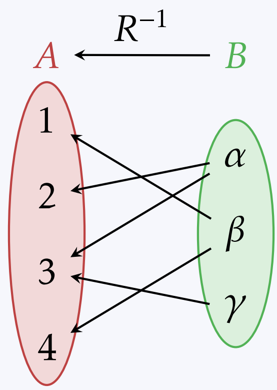
Relations can be inverted by reversing the order of each of its pairs.

Example 0.24 Inverse relation

The inverse relation to the relation in Example 0.23 is

$$R^{-1} = \{(\beta,1),\ (\alpha,2),\ (\alpha,3),\ (\beta,3), (\gamma,4)\}.$$

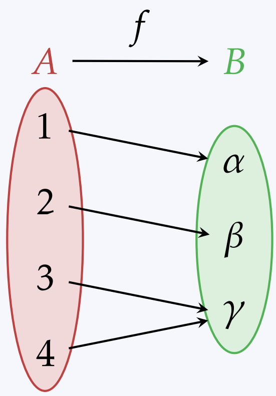
Graphically:

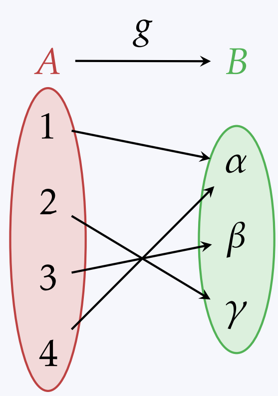


A **function** f from a set A to a set B is a relation for which any element in A is connected to a single element in B.

Example 0.25 Functions

The following are two functions from the set A to the set B defined in Example 0.23:





The pairs making up f are $(1,\alpha)$, $(2,\beta)$, $(3,\gamma)$ and $(4,\gamma)$, and the pairs making up g are $(1,\alpha)$, $(2,\gamma)$, $(3,\beta)$ and $(4,\alpha)$.

Note 0.5 Relations which are not functions

Note that the relation in Example 0.23 is **not** a function, since the element $3 \in A$ is connected to more than one element in B, namely α and γ.

Different names are used in some branches of mathematics to describe functions, such as **maps** and **transformations**. Barring context, they all mean the same thing.

A common way to denote that a function f is connecting elements in A to elements in B is

$$f: A \to B. \tag{0.2.1}$$

A is called the **domain** of f, and B its **image**. In this book and many other sources, the following notation is used: $f(x) = y$, which means that when we apply the function f to an element $x \in A$, the result is the element it is connected to, i.e. $y \in B$. We write this as $x \mapsto y$ (the special symbol $\mapsto$ is called a **mapping notation**).

Example 0.26 Value $\mapsto$ value notation for functions

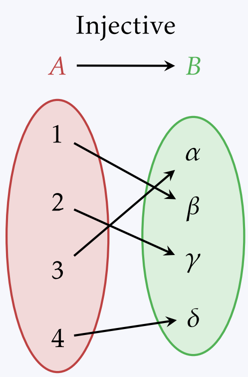
For the functions f, g as defined in Example 0.25:

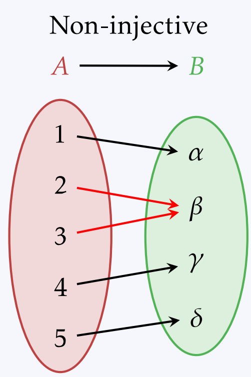
$$f(1) = \alpha,\ f(2) = \beta,\ f(3) = f(4) = \gamma.$$
$$g(1) = g(4) = \alpha,\ g(2) = \gamma,\ g(3) = \beta.$$

0.2.2 Injective, surjective and bijective functions

A function is **injective** if each of the elements in its **image** is connected to by at most a single element in its **domain**. An injective function is also known as an **injection**.

Example 0.27 Injective function





The function on the right in non-injective because the element $\beta \in B$ is connected to by two elements in A (2 and 3, red arrows).

A function is **surjective** if every element in its image is connected to by at least a single element in its domain (see Example 0.28). As with injective functions, a surjective function is also known as a **surjection**. A non surjective function can be made into surjective function by excluding from its image any element that is not connected to by any element from its domain (see Example 0.29).

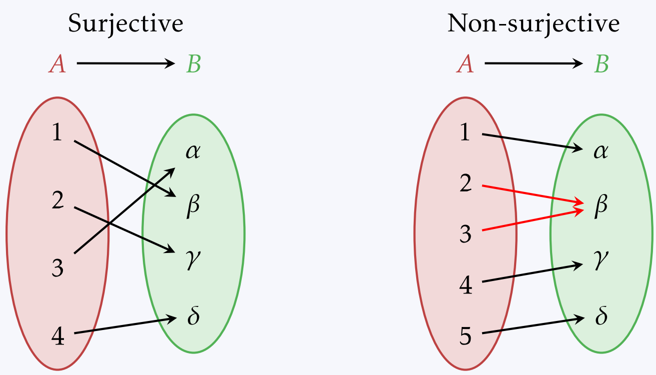
A function $f : A \to B$ that is both surjective and bijective is called a **bijective function** (also a **bijection**). All elements in the image of a bijection are connected to by exactly a single element in its domain. This means that the direction of the connections can be flipped, yielding the **inverse** of the original function (denoted f^{-1}).

The reason only bijective functions have inverses is as follows: Given a function $f : A \to B$,

- if f is non-injective, then there is at least one element $y_1 \in B$ which is connected to by at least two elements from A. We can name these elements x_1 and x_2. When inverted, $f^{-1} : B \to A$ has an element $y_1 \in B$ (note that for f^{-1}, B is its domain), which is connected to two or more elements in A, the image of f^{-1}. These are of course x_1, x_2. This fact disqualifies f^{-1} from being a function.
- If f is non-surjective, then there exists at least one element $y_2 \in B$ that is not connected to by any element from A. When inverted, y_2 in the domain B of f^{-1} is not connected to any element in its image A. This fact disqualifies f^{-1} from being a function.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

Example 0.28 Surjective function



Example 0.29 Making a non-surjective function into a surjection

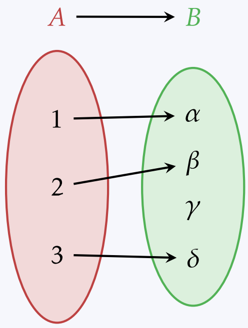
Given the two sets $A = \{1, 2, 3, 4\}$ and $B = \{\alpha, \beta, \gamma, \delta\}$, the following non-surjective function $f : A \to B$ is defined:

$$f = \{(1, \alpha),\ (2, \beta),\ (3, \gamma),\ (4, \gamma)\}.$$

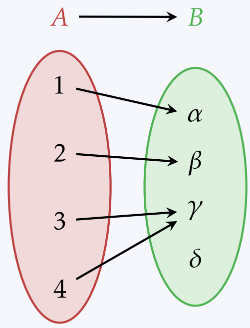
By removing δ from B, the function f becomes surjective (though it remains non-injective).

Example 0.30 Cross examples

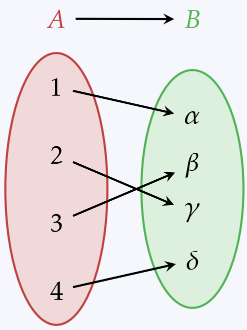
Injective, non surjective



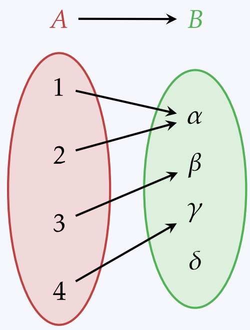
Injective, non surjective



Injective and surjective



Neither injective nor surjective



Note 0.6 Other names for bijections

Bijections are also called **one-to-one correspondences** and **invertible functions**.

0.2.3 Real functions

In suitable cases, a function is defined via a general mapping rule. This should be very familiar to anyone who learned mathematics in high school, where many times functions are defined this way, e.g.

$$f(x) = x^2 + 3x - 4. \tag{0.2.2}$$

In mapping notation we can write Equation 0.2.2 as $f : x \mapsto x^2 + 3x - 4$. In high school mathematics, both the domain and image of such functions is $\mathbb{R}$, although it is almost never specified explicitly. Such functions are commonly referred to as **real functions**, a convention used in this book as well.

Example 0.31 Functions defined using a mapping rule

The following are real functions:

$$f_1(x) = 2x^2 - 5, \quad f_2(x) = \sin\left(\frac{x}{3}\right), \quad f_3(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{\sigma^2}}.$$

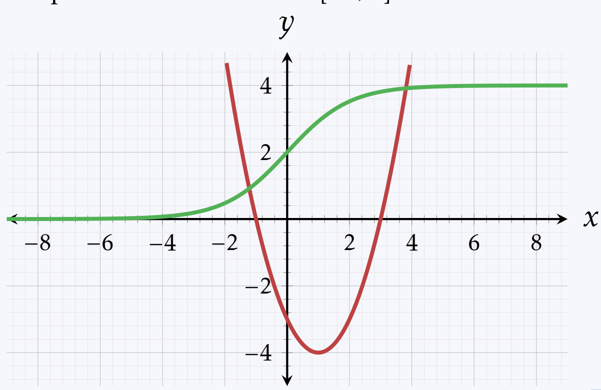
Note that these functions can also be defined using different sets, for example $f_1 : \mathbb{N} \to \mathbb{Z}, \quad f_2 : \mathbb{N} \to [-1, 1], \quad$ etc.

Real functions can be easily plotted in a **Cartesian coordinate system** by drawing all the points $(x, f(x))$ (i.e. all the points (x, y), where $x, y \in \mathbb{R}$ and $x \mapsto y$). We call these points the **graph** of f over $\mathbb{R}$.

Example 0.32 Graphs of real functions

The following two functions are plotted on the domain $[-9, 9]$:

- $f(x) = x^2 - 2x - 3$,
- $g(x) = 4e^x / (e^x + 1)$.

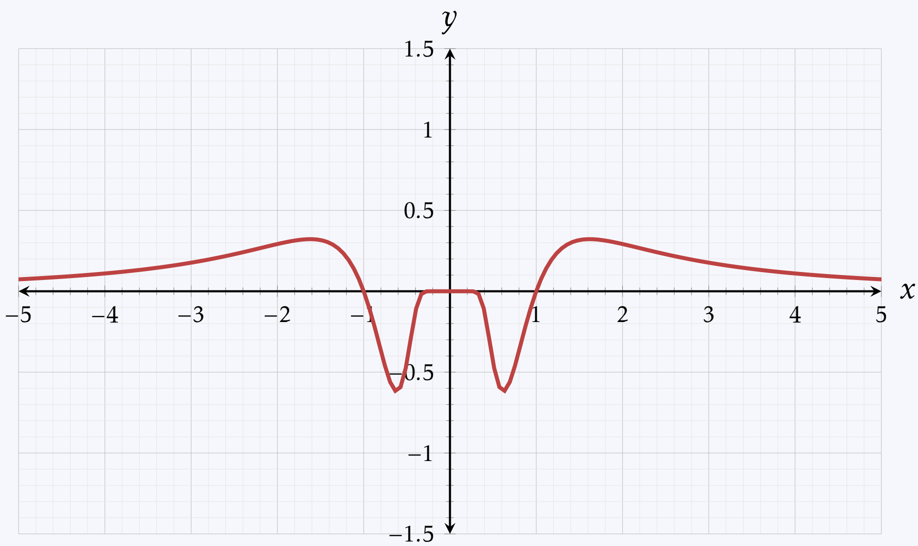


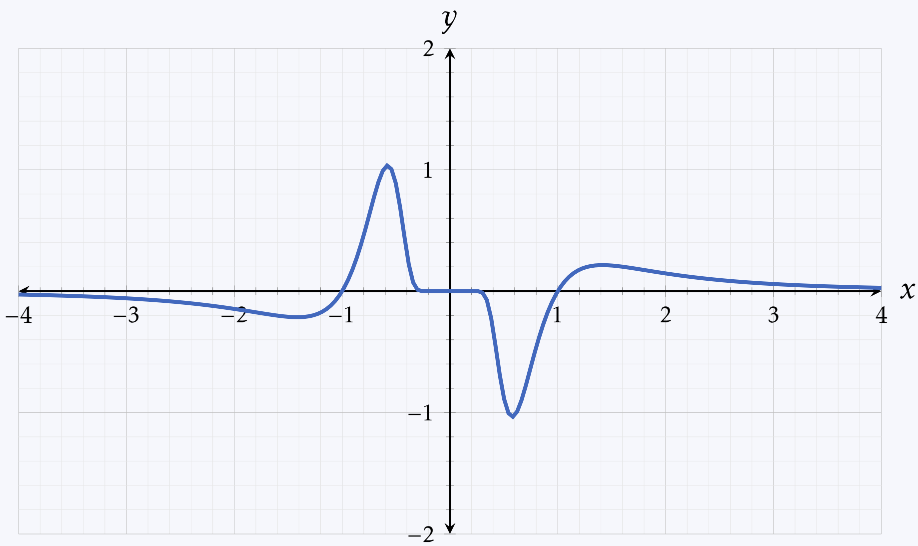
In Example 0.32, the function $g(x)$ always increases in value from left to right. Let's give this notion a more formal tone: a function f is said to be **increasing** on an interval I if for any $x_1, x_2 \in I$, if $x_2 > x_1$ then $f(x_2) > f(x_1)$. We can similarily define the idea of **decreasing** on an interval.

A property of some functions which is visually easy to depict is symmetry. A real function f is said to be **symmetric** if for any $x \in \mathbb{R}$, $f(-x) = f(x)$. This essentially means that the y-axis mirrors the function's plot. If for any $x \in \mathbb{R}$, $f(-x) = -f(x)$, we say that the function is **anti-symmetric**. A function can be neither, but there's only a single function which is both: the zero function, i.e. $f(x) = 0$.

Example 0.33 Symmetric and anti-symmetric functions

In the following graphs, the function on the top is symmetric, while the function on the bottom is anti-symmetric:





(injections/surjections of real functions?)

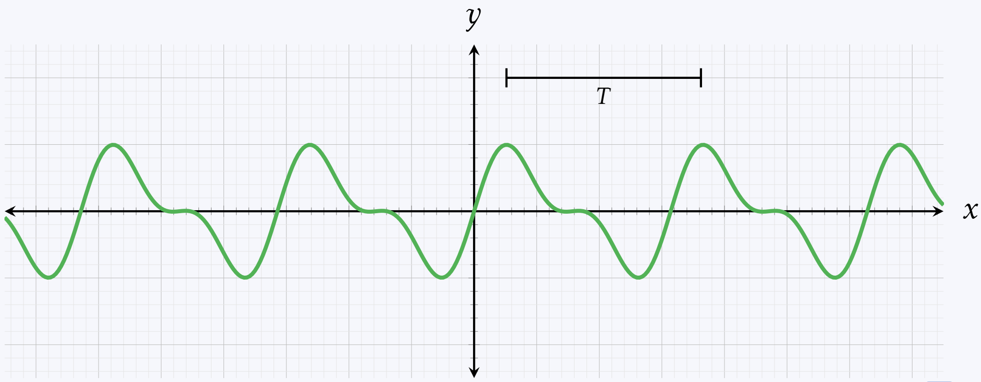
A real function is said to be **periodic** if it repeats its values exactly over and over with increasing x. In more precise terms we define a real function f to be periodic if for any integer value k,

$$f(x+kT) = f(x). \tag{0.2.3}$$

where $T = [a,b]$ is a finite interval of $\mathbb{R}$ which we call the **period** of the function.

Example 0.34 A periodic function

The following graph depicts a periodic function f, with its period T shown. Notice that for any x $f(x+T) = f(x)$, i.e. you can move the period measure left and right along the x-axis and the values of $f(x)$ in both its edges would always be the equal.



Two additional measures that arise from a period T are the **frequency** $f = \frac{1}{T}$, and the **angular frequency** $\omega = 2\pi f = \frac{2\pi}{T}$. We will use these measures later in the book.

Note 0.7 Units of period and frequency

In a periodic function such as the one in the above example, the units for the period are the same one used for the horizontal axis, while the units of both frequency and angular frequency are both 1 over the unit used for the horizontal axis/period. For example, if the unit of the horizontal axis is that of seconds, then the frequency units are 1/seconds, i.e. Hertz (SI symbol: Hz).

0.2.4 Composition of functions

Functions can be **composed** together, generating new functions. Given two functions $f: A \to B$ and $g: B \to C$, their composition is denoted as $f \circ g$. For the composition to be well defined, the **image** of f must be the same as the **domain** of g, and the resulting composition would have A as its domain and C as its image, i.e. $f \circ g: A \to C$.

Example 0.35 Composition of functions

Consider the functions

$$f(x) = x^2, \quad g(x) = \sin(x).$$

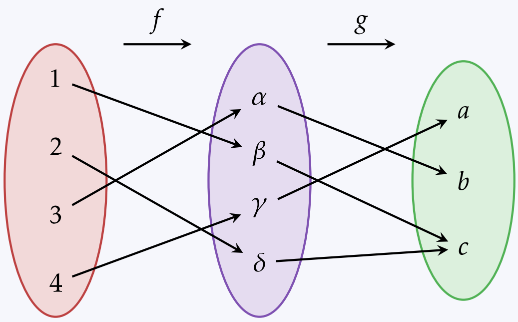
Using these functions, the two possible compositions are

- $f \circ g = f(g(x)) = [\sin(x)]^2$, and
- $g \circ f = g(f(x)) = \sin\left(x^2\right)$.

Example 0.36 Graphical representation of function composition

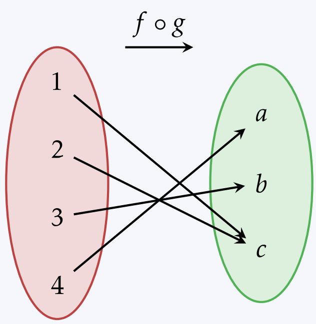
A graphical representation of composing two functions:

$$f : \{1, 2, 3, 4\} \to \{\alpha, \beta, \gamma, \delta\}, \quad g : \{\alpha, \beta, \gamma, \delta\} \to \{a, b, c\}.$$



The composition results in the following function

$$f \circ g : \{1, 2, 3, 4\} \to \{a, b, c, \}.$$



0.3 POLYNOMIAL FUNCTIONS

A very useful family of real functions can be derived using only three fundamental operations: addition, multiplication and exponentiation: the (real) **polynomial functions**. These are functions of the form

$$P_n(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n, \tag{0.3.1}$$

where $a_0, a_1, \ldots, a_n$ are real numbers called the **coefficients** of the polynomial function. Note that $a_n \neq 0$, i.e. the **degree** of the polynomial function is the index of the highest non-zero coefficient (and thus the highest power in the expression). We also call this the **order** of the polynomial function.

Example 0.37 Polynomial

The following is a polynomial function of degree $n = 6$:

$$P(x) = 4 + 2x - 3x^2 + 7x^4 - x^5 + 3x^6.$$

Breaking down this polynomial to its constituent terms:

$$\begin{array}{ccccccc}
P(x) = & 4 & +2x & -3x^2 & +7x^4 & -x^5 & +3x^6 \\
 & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\
 & a_0 = 4 & a_1 = 2 & a_2 = -3 & a_4 = 7 & a_5 = -1 & a_6 = 3
\end{array}$$

Note that a_3 is missing from the polynomial function (i.e. there is no x^3 term). This means that $a_3 = 0$.

A shorthand way to write the general form of a polynomial function is by using the **summation notation**:

$$P(x) = \sum_{k=0}^{n} a_k x^k. \tag{0.3.2}$$

This notation, called the **Capital-sigma notation**, essentially represents addition of n elements (in the case shown here), each with its own **index of summation**, in this case i. The most general form of the summation notation is

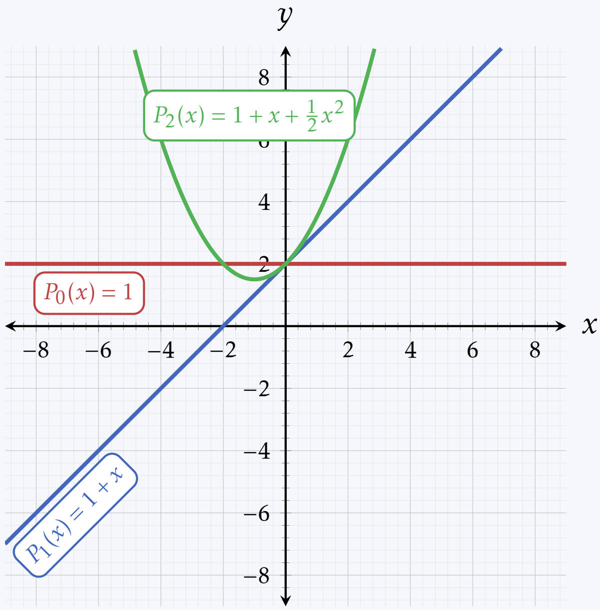
$$\sum_{i=k}^{n} a_i = a_k + a_{k+1} + a_{k+2} + \cdots + a_{n-1} + a_n, \tag{0.3.3}$$

i.e. the notation tells us to add those elements a_i for which $k \geq i \geq n$. Note that in the case of Equation 0.3.2, when $k = 0$, $x^k = x^0 = 1$ and the first term of the polynomial function has no x power (i.e. it is simply a_0), and when $k = 1$, $x^k = x^1 = x$ and thus the second term is $a_1 x$. We will encounter the summation notation in more details later in the book.

In the special case $n = 0$, i.e. when $P(x) = a_0$, the function is constant. When $n = 1$ the function $P(x) = a_0 + a_1 x$ is a line, and when $n = 2$, $P(x) = a_0 + a_1 x + a_2 x^2$ is a quadratic function.

Example 0.38 Polynomial functions for $n = 0, 1, 2$

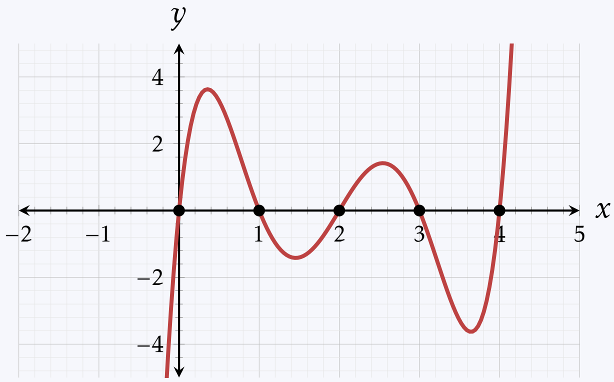
The following graphs represent the polynomial functions of degrees $n = 0, 1, 2$ with coefficients $a_0 = 2$, $a_1 = 1$, $a_2 = \frac{1}{2}$:



The values $x \in \mathbb{R}$ for which $P(x) = 0$ are called the **roots** (also: **zeros**) of the polynomial function.

Example 0.39 Roots of a polynomial function

The polynomial function $P(x) = 24x - 50x^2 + 35x^3 - 10x^4 + x^5$ has the following 5 roots: $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$. In the following graph of $P(x)$ the roots are shown as black dots.



The maximum number of **real** roots of a polynomial function with degree $n \geq 1$ is n, e.g. a polynomial of degree $n = 4$ has at most 4 real roots. This statement is a consequence of a very important theorem called **the fundamental theorem of algebra**, which due to its importance we will mention here without proof:

Theorem 0.1 The fundamental theorem of algebra

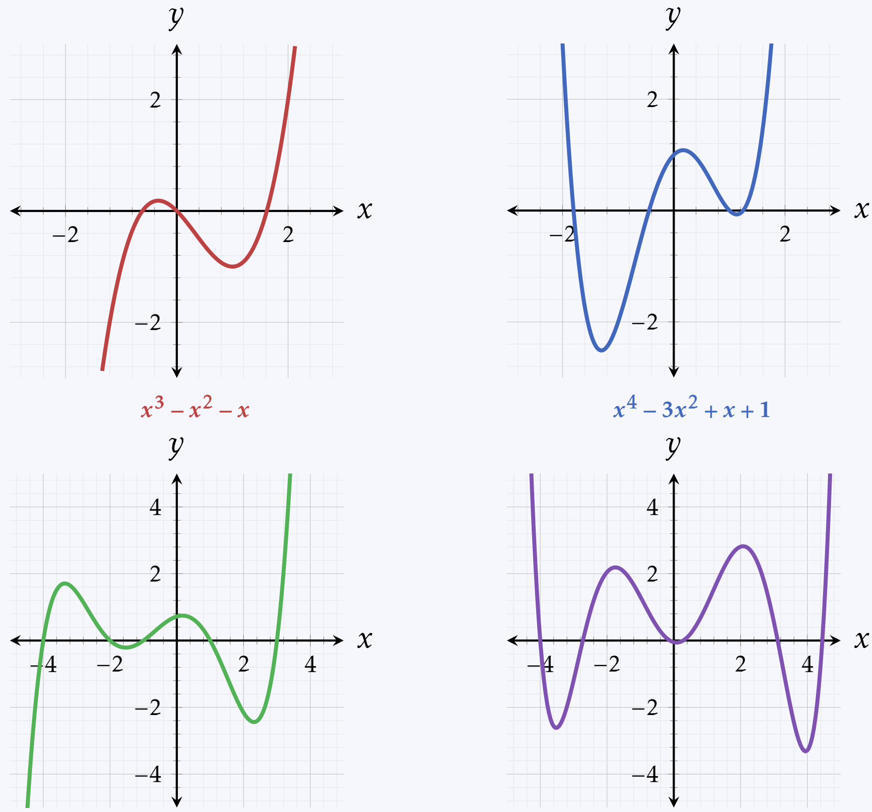
For any $n \geq 1$, the polynomial function $P(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_n z^n$, where $a_0, a_1, a_2, \ldots, a_n$ are all **complex numbers** and $a_n \neq 0$, has n complex roots.

Given a polynomial function $P(x)$ with n roots $r_1,\ r_2,\ \cdots,\ r_n$, the function can be written as a product of terms of the form $x - r_i$ (up to a constant), e.g. the polynomial function of degree $n = 3$ with roots $-1, 1, 2$ can be written as

$$P(x) = (x+1)(x-1)(x-2) = x^3 - 2x^2 - x + 2. \tag{0.3.4}$$

Example 0.40 Higher order polynomial functions

The following are the graphs of high-order polynomial functions ($n = 3, 4, 5, 6$):



$x^3 - x^2 - x$; $x^4 - 3x^2 + x + 1$; $\frac{3}{100}\left(x^5 + 3x^4 - 11x^3 - 27x^2 + 10x + 24\right)$; $\frac{1}{100}\left(x^6 - x^5 - 26x^4 + 15x^3 + 150x^2 - 25x - 5\right)$

As can be seen in **??**, the maximal number of 'bends' in a polynomial function of order n is $n-1$ (i.e. one less than the order of the function).

We will continue to explore polynomial functions in more details in future chapters.

0.4 EXPONENTIAL AND LOGARITHMIC FUNCTIONS

In the previous section we dealt with functions composed of integer powers of x. We will now shortly focus on functions where x is in the power itself and their inverse functions.

An **exponential function**, or simply an **exponential**, is a real function of the type

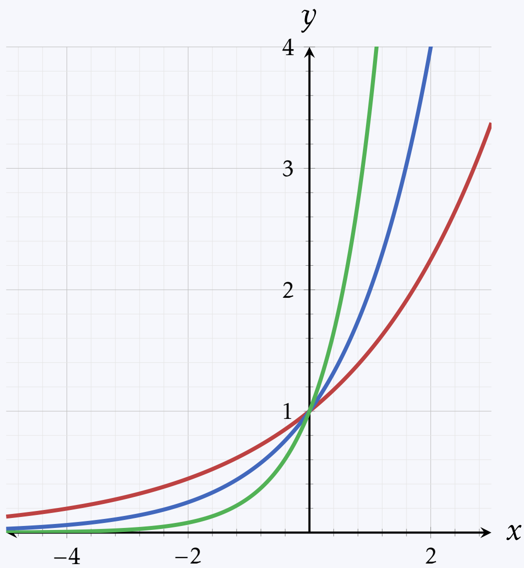
$$f(x) = b^x, \tag{0.4.1}$$

where $b > 0$ is called the **base** of the exponentiation, and x the exponential. All exponents, regardless of base, are always positive. In addition, all exponents pass through the point $(0,1)$ since $b^0 = 1$ for any real positive number, and through the point $(1,b)$

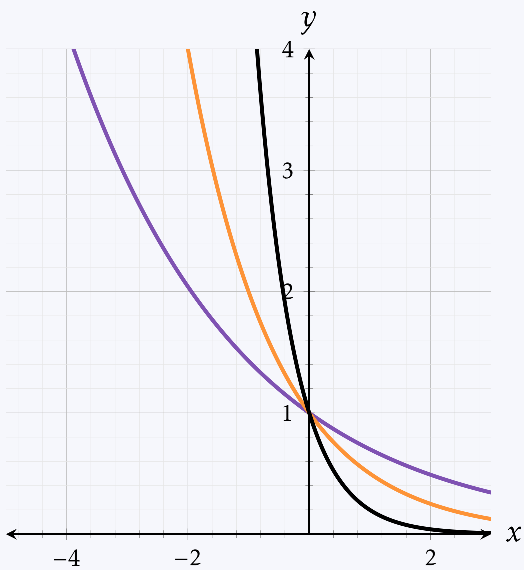
since $b^1 = b$. When $b > 1$ the function is increasing on $\mathbb{R}$, while for $b < 1$ the function is descending on $\mathbb{R}$.

Example 0.41 Exponential functions

The following are graphs of the exponential functions $\mathbf{1.5^x}$, 2^x and $\mathbf{3.5^x}$:



And the following are graphs of the exponential functions $\mathbf{0.7^x}$, $\mathbf{0.5^x}$ and $\mathbf{0.2^x}$:



As a reminder, the following are two well known properties of exponents: given a base

$b > 0$,

$$b^{-x} = \frac{1}{b^x}, \tag{0.4.2}$$
$$b^x b^y = b^{x+y}. \tag{0.4.3}$$

A special base for exponential functions is the real, non-algebraic number e. This number has many names, among them is **Euler's number**, but in the constant of exponentials it is known as the **natural base**. Its exact value is not entirely important for the moment: it is about 2.718, and in any case it is not possible to write it as there it has infinitely many digits after the period. It is very common across different fields of mathematics and science to write $\exp(x)$ instead of e^x.

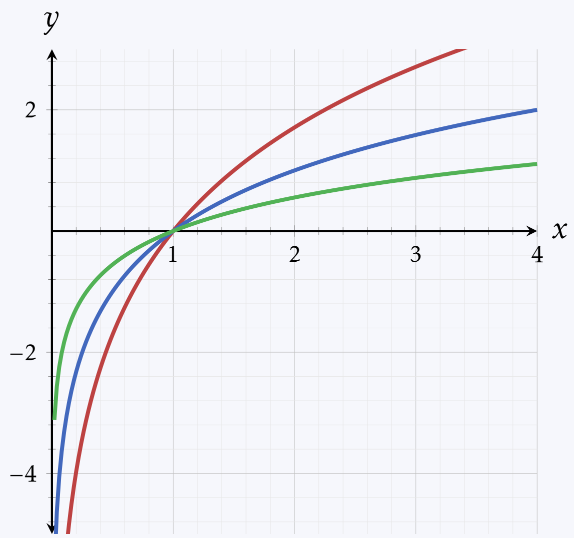
The inverse function to exponentials are the **logarithmic functions** (or simply **logarithms**), i.e. for any real $b > 0$, $b \neq 1$,

$$\log_b\left(b^x\right) = b^{\log_b(x)} = x. \tag{0.4.4}$$

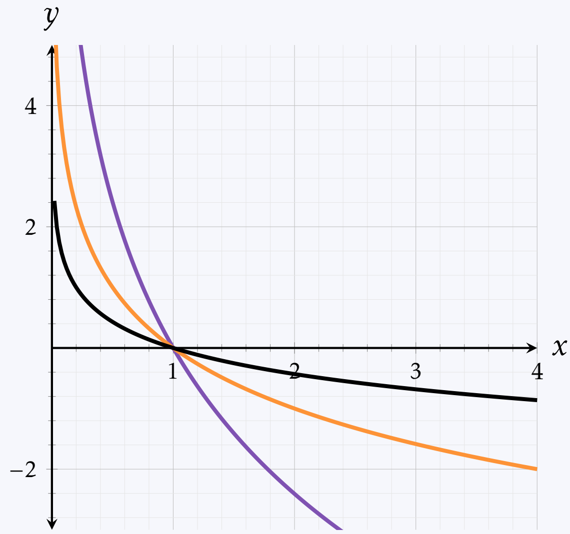
In essence, the logarithm in base b of a number x answers the question *"what is the number a for which $b^a = x$?"*. Being the inverses of exponential functions, all logarithms go through the point $(1,0)$, and each also passes through its own point $(b,1)$.

Example 0.42 Logarithmic functions

The following are graphs of the logarithmic functions $\log_{1.5}(x)$, $\log_2(x)$ and $\log_{3.5}(x)$:



...and the following are graphs of the exponential functions $\log_{0.75}(x)$, $\log_{0.5}(x)$ and $\log_{0.2}(x)$:



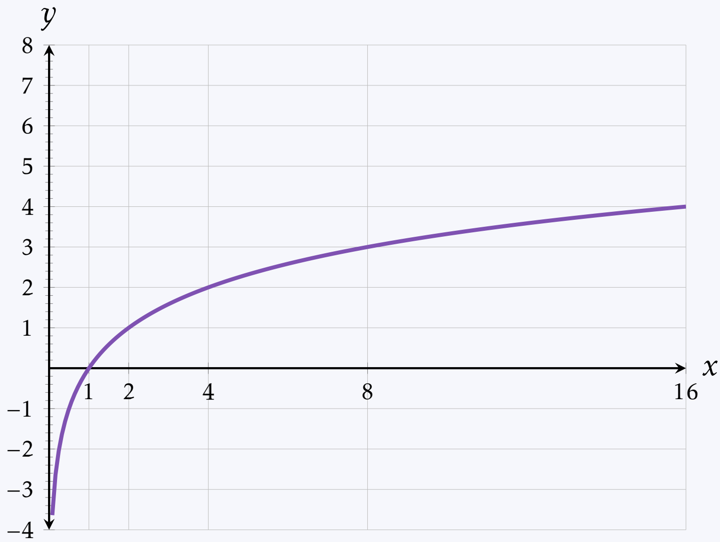
A useful property of logarithms is that they can help reduce ranges spanning several orders of magnitude to numbers humans can deal with. The easiest way to see this is using $b = 10$: $10^1 = 10$, and so $\log_{10}(10) = 1$. $10^2 = 100$, and so $\log_{10}(100) = 2$. $10^3 = 1000$, and so $\log_{10}(1000) = 3$, etc. The value of the logarithm goes by 1 for each raise in order of magnitude of its argument.

Therefore, if we have some measurement x which can hold values spanning several orders of magnitude (say $x \in [3, 1500000000]$), then it can sometimes be useful to use instead the logarithmic value of x (which in our case would span the range $\log_{10}(x) \in [0.477, 9.176]$). This is done in many fields of science, for example some definitions of entropy[1], acid dissociation constants[2], pH[3] and more.

Example 0.43 Logarithms as evaluating orders of magnitude

In the following graph of $\log_2(x)$, each increase by power of two in x (i.e. $x = 1, 2, 4, 8, 16, \dots$) yields only a single increase in y (i.e. $y = 0, 1, 2, 3, 4, \dots$). This shows how logarithms shift our perspective from absolute values to orders of magnitude.

[1] $S = k_B \log(\Omega)$
[2] $pK_a = -\log(K_{diss})$
[3] $pH = -\log([H^+])$



Using the definition of the logarithmic function $\log_b(x)$ (Equation 0.4.4) and the product rule for exponentials (Equation 0.4.3), a similar rule can be derived for logarithms. Let $x, y > 0$ and $b > 0,\ b \neq 1$ all be real numbers. We define

$$\log_b(x) = M,\ \log_b(y) = N, \tag{0.4.5}$$

which means

$$b^M = x,\ b^N = y. \tag{0.4.6}$$

From Equation 0.4.3 we know that

$$xy = b^M b^N = b^{M+N}, \tag{0.4.7}$$

and by re-applying the definition of logarithmic functions we get that

$$\log_b(xy) = M + N = \log_b(x) + \log_b(y). \tag{0.4.8}$$

Similarly to Equation 0.4.8, division yields subtraction:

$$\log_b\left(\frac{x}{y}\right) = \log_b(x) - \log_b(y). \tag{0.4.9}$$

Equations 0.4.8 and 0.4.9 reveal another valuable property of logarithms: they reduce multiplication to addition (and subsequently division to subtraction). While today this property doesn't seem very impressive, in pre-computers days it helped carrying on complicated calculations, using tables of pre-calculated logarithms (called simply **logarithm tables**) - a sight rarely seen today.

Taking one step forward in regards to reduction of operations, logarithms reduce powers to multiplication:

$$\log_b\left(x^k\right) = k \log_b(x). \tag{0.4.10}$$

for any $k \in \mathbb{R}$.

(TBW: proving this will be in the chapter questions to the reader)

Any logarithm $\log_b(x)$ can be expressed using another base, i.e. $\log_a(x)$ (where $a > 0$, $a \neq 1$) using the following formula:

$$\log_a(x) = \log_b(x) \cdot \log_a(b). \tag{0.4.11}$$

(TBW: proving this too will be a question to the reader)

Example 0.44 Changing logarithm base

Expressing $\log_4(x)$ in terms of $\log_2(x)$:

$$\log_4(x) = \log_2(x) \cdot \underbrace{\log_4(2)}_{=\frac{1}{2}} = \frac{1}{2}\log_2(x).$$

Much like with exponentials, the number e plays an important role when it comes to logarithms, for reasons that are discussed in the calculus chapter (ref). For now, we will just mention that $\log_e(x)$ gets a special notation: $\ln(x)$, which stands for **natural logarithm**. This notation is mainly used in applied mathematics and science, while in pure mathematics the notation is simply $\log(x)$, i.e. without mentioning the base [4].

For reason we will see in the calculus chapter, it is relatively simple to calculate both the exponential and logarithm in base e. Therefore, many operations in modern computations are actually done using these functions, for example calculating logarithms in other bases:

$$\log_b(x) = \frac{\ln(x)}{\ln(b)}. \tag{0.4.12}$$

Another operation commonly using both e^x and $\ln(x)$ is raising a real number a to a real power b: using the properties of both exponential and logarithmic functions, any such power can be expressed as

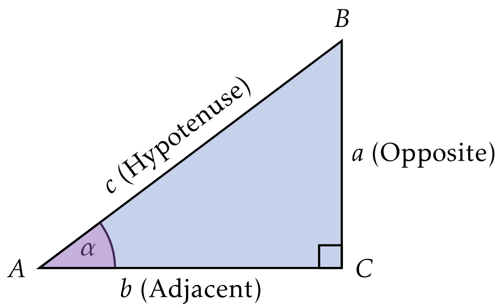
$$a^b = \mathrm{e}^{b\ln(a)}. \tag{0.4.13}$$

0.5 TRIGONOMETRIC FUNCTIONS

0.5.1 Basic Definitions

Consider a **right triangle** $\triangle ABC$ with sides a, b, and Hypotenuse c, where the angle $\angle ACB$ is $90°$, and the angle $\angle BAC$ is denoted as α:

[4] Depending on convention and context, this notation can refer to logarithm in any other base, most commonly $\log_{10}(x)$ and $\log_2(x)$.



We use the ratios between the three sides of the triangle to define three functions of α:

- The **sine** of the angle α is $\sin(\alpha) = \frac{a}{c}$,
- the **cosine** of the angle α is $\cos(\alpha) = \frac{b}{c}$, and
- the **tangent** of the angle α is $\tan(\alpha) = \frac{a}{b}$, which in turn is equal to $\frac{\sin(\alpha)}{\cos(\alpha)}$.

We can rearrange the above definitions:

$$\begin{aligned} a &= c\sin(\alpha), \\ b &= c\cos(\alpha). \end{aligned} \tag{0.5.1}$$

Normally, the Hypotenuse is the longest side of a right triangle. We will consider here the two edge cases where one of the sides a or b is equal to the Hypotenuse (and the other side is thus 0):

- if $a = c$ then $\alpha = 90°$,
- if $b = c$ then $\alpha = 0°$.

The possible length of a is therefore in the range $0 \leq a \leq c$, which means that $0 \leq \frac{a}{c} \leq 1$. Since $\sin(\alpha) = \frac{a}{c}$ this means that the image of $\sin(\alpha)$ is $[0,1]$. The same idea is also true for b, and therefore $[0,1]$ is the image of $\cos(\alpha)$ as well.

As a reminder, the **Pythagorean theorem**[5] states that for a right triangle with sides a, b and c,

$$a^2 + b^2 = c^2. \tag{0.5.2}$$

By substituting **Equation** 0.5.1 into the Pythagorean theorem we get

$$\begin{aligned} c^2 &= a^2 + b^2 \\ &= [c\sin(\alpha)]^2 + [c\cos(\alpha)]^2 \end{aligned}$$

[5] It's worth mentioning that no three positive integers a, b, and c satisfy the equation $a^n + b^n = c^n$ for any integer value of $n > 2$. This can be proven, however the proof is too large to fit in the footnotes.

$$= c^2 \sin^2(\alpha) + c^2 \cos^2(\alpha)$$
$$= c^2 \left[\sin^2(\alpha) + \cos^2(\alpha)\right],$$

and therefore

$$\sin^2(x) + \cos^2(x) = 1. \tag{0.5.3}$$

0.5.2 The Unit Circle

We defined $\sin(\alpha)$ and $\cos(\alpha)$ so far in way such that their domains are both $[0°, 90°]$, and their images are both $[0, 1]$. However, there is a simple way to extend these functions such that both their domains are $\mathbb{R}$, and both their images are $[-1, 1]$: by using a **unit circle**.

Figure 0.4 depicts a unit circle: it is simply a circle of radius $R = 1$, which is placed such that its center lies at the origin of a 2-dimensional axis system (i.e. at the point $\boldsymbol{O} = (0, 0)$. We then draw a line from $\boldsymbol{O}$ to a point $\boldsymbol{P} = (x, y)$ on the circumference of the unit circle. We call the angle between the line $\boldsymbol{OP}$ and the x-axis θ. We then draw another line, this time from the point $\boldsymbol{P}$ to a point $\boldsymbol{D}$ on the x-axis, such that $\boldsymbol{PD}$ is perpendicular to the x-axis.

The triangle $\triangle OPD$ is a right triangle. Therefore, we can use the trigonometric functions to calculate the coordinates of the point $\boldsymbol{P} = (x, y)$:

$$\begin{aligned} x &= R\cos(\theta) = \cos(\theta), \\ y &= R\sin(\theta) = \sin(\theta). \end{aligned} \tag{0.5.4}$$

We then define $\cos(\theta)$ and $\sin(\theta)$ as a function of θ:

$$\begin{aligned} \sin(\theta) &= y, \\ \cos(\theta) &= x. \end{aligned} \tag{0.5.5}$$

Using this definition, the angle θ can take any value between $0°$ and $360°$. In fact, the values of θ can be extended to any real number in degrees: any real value of degrees is equivalent to some value in the range $[0°, 360°]$, the first and most obvious example is that $360°$ is equivalent to $0°$. Similarly, $-30° \equiv 330°$, $-180° \equiv 180°$, $-90° \equiv 270°$, etc (see Figure 0.3). In fact, this property makes the trigonometric functions periodic, with a period $T = 360°$.

0.5.3 Radians

Using degrees to measure angles in a sphere creates an inconvenience: the domain and image of the trigonometric functions have different units. In order to measure both these magnitudes using the same unit we switch to measuring angles on a circle using **radians** instead of degrees. One radian equals the length of a single radius R of the circle (in the case of the unit circle this is always $R = 1$). We define an inner angle θ to equal one radian if the arc length it represents is equal to R (see Figure 0.5).

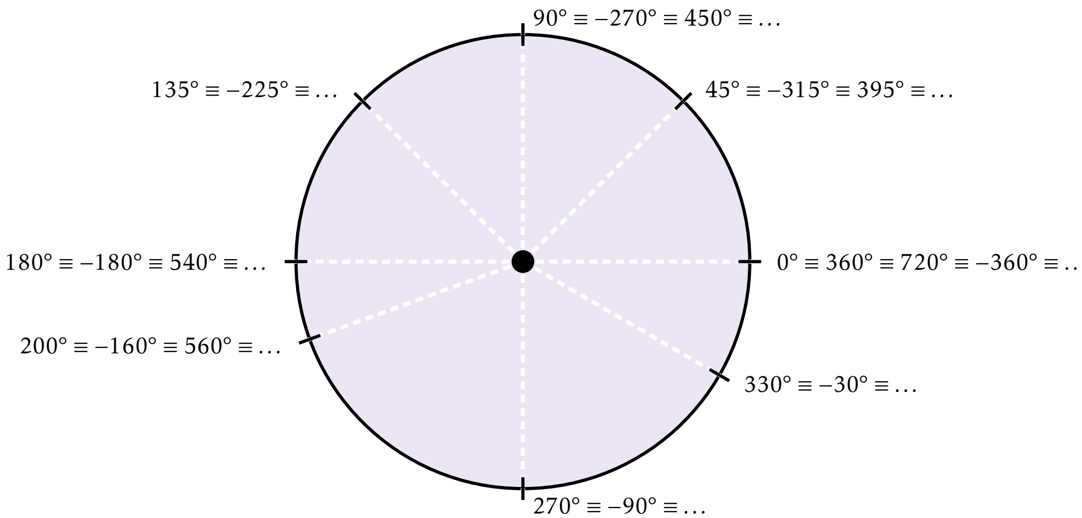


Figure 0.3 Angles equivalency on a circle.

Table 0.3 Common angles in radians, and their respective images for the three main trigonometric functions.

θ degrees	θ radians	$\sin(\theta)$	$\cos(\theta)$	$\tan(\theta)$
0°	0	0	1	0
30°	$\frac{\pi}{6}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{\sqrt{3}}$
45°	$\frac{\pi}{4}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	1
60°	$\frac{\pi}{3}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\sqrt{3}$
90°	$\frac{\pi}{2}$	1	0	undefined
180°	π	0	−1	0
270°	$\frac{3\pi}{2}$	−1	0	undefined
360°	2π	0	1	0

How much is a radian in degrees? The full circumference of any circle with radius R equals $2\pi R$, which means that a single radian R is equivalent to $\frac{180°}{\pi} \approx 57.3°$. Table 0.3 shows some common angles and their equivalent value in radians.

0.5.4 Graphs

As seen previously, the functions $\sin(x)$ and $\cos(x)$ are periodic, having both the period $T = 2\pi$. Their graphs are depicted in Figure 0.7. The value of $\sin(x)$ is always equal to that of $\cos\left(x - \frac{\pi}{2}\right)$: we say that the two functions have a **phase difference** of $\pi/2$. The graph of $\tan(x)$ is depicted in Figure 0.6.

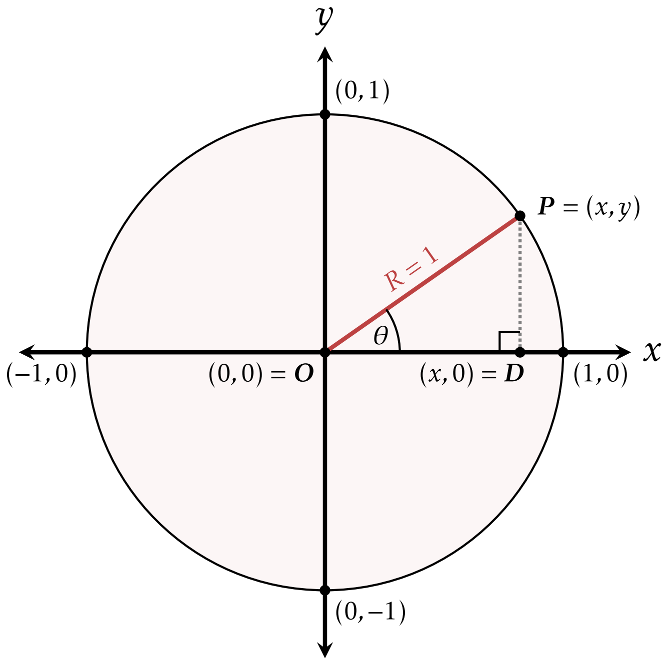


Figure 0.4 A unit circle with a point $\boldsymbol{P} = (x, y)$ on its circumference. The triangle $\triangle \boldsymbol{OPD}$ is a right triangle with sides $\boldsymbol{OD} = x$, $\boldsymbol{OP} = y$ and an angle θ opposing the side $\boldsymbol{DP}$.

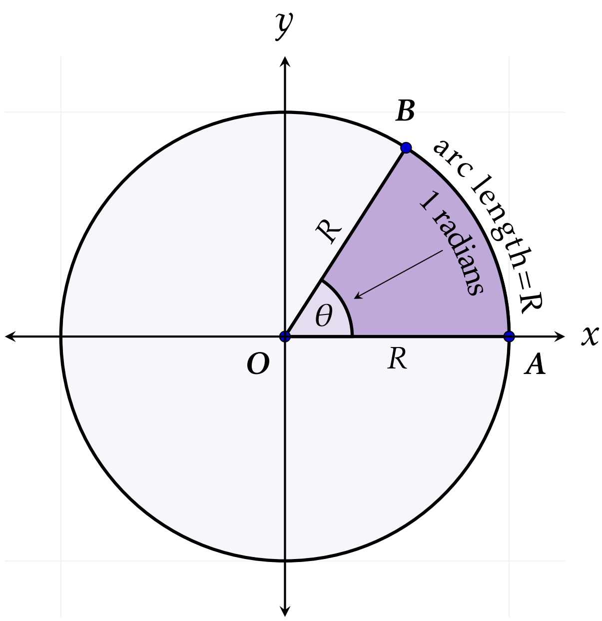


Figure 0.5 In this figure the arc $\boldsymbol{AB}$ has the same length of the radii $\boldsymbol{OA}$ and $\boldsymbol{OB}$ (all are equal to R), and therefore $\theta = 1$ radians.

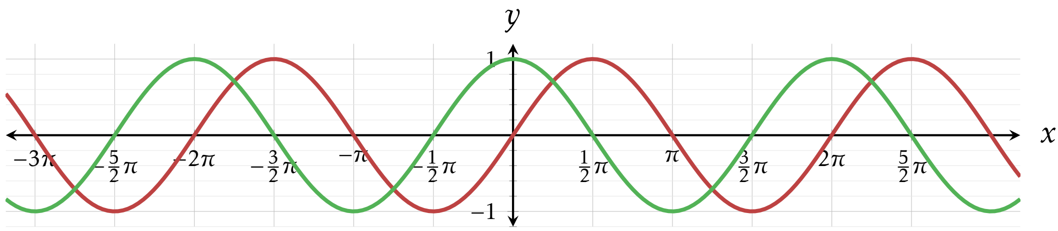


Figure 0.6 The graphs of sin (x) and cos (x) for $x \in [-10, 10]$. Note how the graph of cos (x) is "lagging" behind the graph of sin (x) by $\pi/2$.

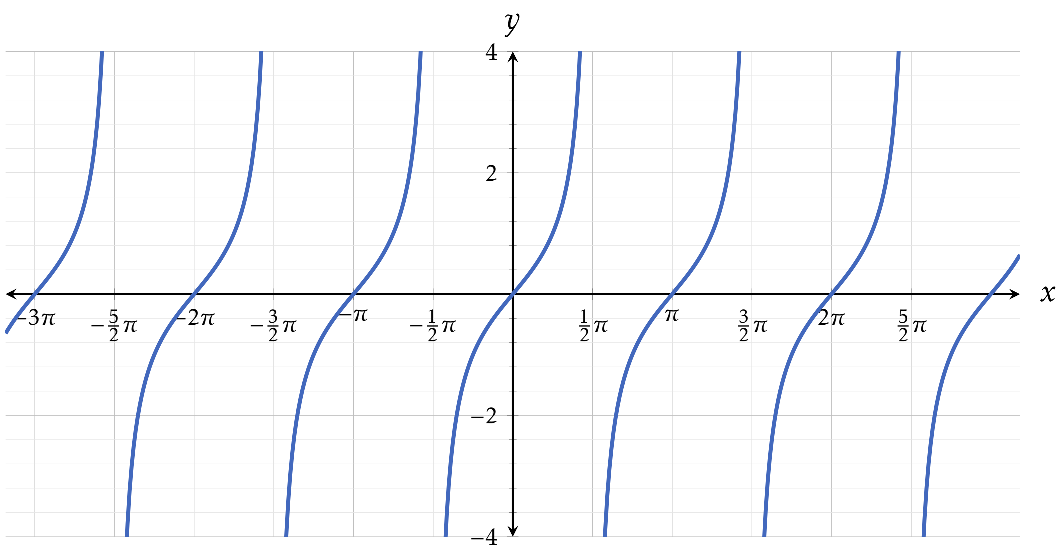


Figure 0.7 The graphs of $\tan(x)$ on the domain $[-3\pi, 3\pi]$.

0.5.5 Identities

The following are some useful facts and connections between trigonometric functions:

- Pythagorean identity:
$$\sin^2(\theta)+\cos^2(\theta)=1 \tag{0.5.6}$$

- Symmetry/Antisymmetry:
$$\sin(-\theta)=-\sin(\theta). \tag{0.5.7}$$
$$\cos(-\theta)=\cos(\theta). \tag{0.5.8}$$
$$\tan(-\theta)=-\tan(\theta). \tag{0.5.9}$$

- Tangent from sine and cosine:
$$\tan(\theta)=\frac{\sin(\theta)}{\cos(\theta)} \tag{0.5.10}$$

- Phase between sine and cosine:
$$\sin\left(\theta\pm\frac{\pi}{2}\right)=\pm\cos(\theta). \tag{0.5.11}$$
$$\cos\left(\theta\pm\frac{\pi}{2}\right)=\mp\sin(\theta). \tag{0.5.12}$$

- Half-period shift:
$$\sin(\theta+\pi)=-\sin(\theta). \tag{0.5.13}$$
$$\cos(\theta+\pi)=-\cos(\theta). \tag{0.5.14}$$

- Angle sum:
$$\sin(\alpha\pm\beta)=\sin(\alpha)\cos(\beta)\pm\cos(\alpha)\sin(\beta). \tag{0.5.15}$$
$$\cos(\alpha\pm\beta)=\cos(\alpha)\cos(\beta)\mp\sin(\alpha)\sin(\beta). \tag{0.5.16}$$

- Double angle:
$$\sin(2\theta)=2\sin(\theta)\cos(\theta). \tag{0.5.17}$$
$$\cos(2\theta)=1-2\sin^2(\theta). \tag{0.5.18}$$

- Half angle:
$$\sin\left(\frac{\theta}{2}\right)=\pm\sqrt{\frac{1-\cos(\theta)}{2}}. \tag{0.5.19}$$
$$\cos\left(\frac{\theta}{2}\right)=\pm\sqrt{\frac{1+\cos(\theta)}{2}}. \tag{0.5.20}$$
$$\tan\left(\frac{\theta}{2}\right)=\frac{\sin(\theta)}{1+\cos(\theta)}. \tag{0.5.21}$$

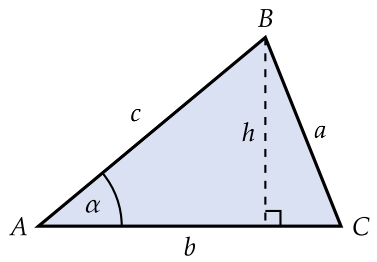


Figure 0.8 The area of a triangle using the side b as a base, and its corresponding height to the point B. The angle opposing the side A is marked as α.

- Product to sum:

$$\sin(\theta)\sin(\varphi) = \frac{1}{2}\left[\cos(\theta-\varphi) - \cos(\theta+\varphi)\right]. \tag{0.5.22}$$

$$\cos(\theta)\cos(\varphi) = \frac{1}{2}\left[\cos(\theta-\varphi) + \cos(\theta+\varphi)\right]. \tag{0.5.23}$$

$$\sin(\theta)\cos(\varphi) = \frac{1}{2}\left[\sin(\theta+\varphi) + \sin(\theta-\varphi)\right]. \tag{0.5.24}$$

$$\tan(\theta)\tan(\varphi) = \frac{\cos(\theta-\varphi) - \cos(\theta+\varphi)}{\cos(\theta-\varphi) + \cos(\theta+\varphi)}. \tag{0.5.25}$$

- Sum to product:

$$\sin(\theta) \pm \sin(\varphi) = 2\sin\left(\frac{\theta\pm\varphi}{2}\right)\cos\left(\frac{\theta\mp\varphi}{2}\right). \tag{0.5.26}$$

$$\cos(\theta) + \cos(\varphi) = 2\cos\left(\frac{\theta+\varphi}{2}\right)\cos\left(\frac{\theta-\varphi}{2}\right). \tag{0.5.27}$$

$$\cos(\theta) - \cos(\varphi) = -2\cos\left(\frac{\theta+\varphi}{2}\right)\cos\left(\frac{\theta-\varphi}{2}\right). \tag{0.5.28}$$

$$\tan(\theta) \pm \tan(\varphi) = \frac{\sin(\theta\pm\varphi)}{\cos(\theta)\cos(\varphi)}. \tag{0.5.29}$$

0.5.6 Useful theorems

The area S of a triangle $\triangle ABC$ can be calculated using the length L any side of the triangle (in this context called a **base**) and the height h to its opposing vertex (see Figure 0.8):

$$S = \frac{1}{2}Lh. \tag{0.5.30}$$

The triangle with sides cbh is a right triangle, c being its hypotenuse. We can therefore infer the size of h using α:

$$h = c\sin(\alpha). \tag{0.5.31}$$

Substituting this back to Equation 0.5.30 yields that the area of the triangle is

$$S = \frac{1}{2}bc\sin(\alpha). \tag{0.5.32}$$

There is nothing special about choosing the side b as a base: we can also use a or c for the calculation. This will yield, respectively,

$$S = \frac{1}{2}ac\sin(\gamma), \tag{0.5.33}$$

$$S = \frac{1}{2}ac\sin(\beta), \tag{0.5.34}$$

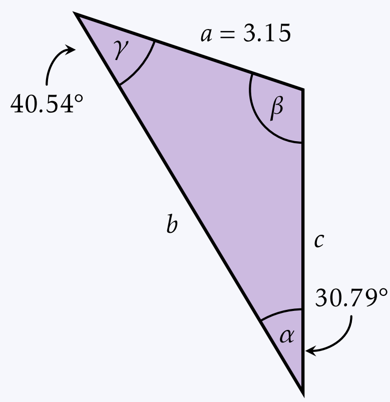
where β is the angle opposing b and γ is the angle opposing c. Since S is the same in all cases, we simply multiply each of the area equations by 2 and divide by abc, which yields

$$\frac{\sin(\alpha)}{a} = \frac{\sin(\beta)}{b} = \frac{\sin(\gamma)}{c}, \tag{0.5.35}$$

i.e. in a triangle, the ratio between any side and the sine of its opposing angle is always the same no matter which side we choose. This theorem is called the **law of sines**.

Example 0.45 Law of sines

Given the triangle $\triangle ABC$ below, what are β and b?



Since all angles in a triangle must add up to 180°,

$$\beta = 180° - 30.79° - 40.54° = 108.67°.$$

Using the law of sines,

$$b = \frac{a}{\sin(\alpha)} \cdot \sin(\beta) = \frac{3.15}{\sin(30.79°)} \cdot \sin(108.67°) \approx 5.83,$$

and

$$c = \frac{a}{\sin(\alpha)} \cdot \sin(\gamma) = \frac{3.15}{\sin(30.79°)} \cdot \sin(40.54°) \approx 4.$$

Note 0.8 Ambiguity of solutions

The above example reveals an issue that might arise due to the symmetrical nature of $\sin(x)$ around $x = \pi$ (180°): say we wanted to calculate β using the law of sines instead of by using $\beta = 180° - \alpha - \gamma$. In this case we would solve the equation

$$\frac{\sin(\alpha)}{a} = \frac{\sin(\beta)}{b},$$

which would result in $\beta = \arcsin\left(\frac{b\sin(\alpha)}{a}\right) = \arcsin(0.95)$. However, two angles can fit this requirement: the sines of 71.34° and 108.67° are both equal to 0.95! Therefore, we must be careful when using the law of sine and make sure we always choose values that make sense (e.g. such that all angles add up to 180°).

!

Of course, the sine function is not unique in having its own named "Law": another useful theorem is the so-called **law of cosines** (also **al-Kashi's theorem**). This theorem states that given a triangle with sides a, b, c and an angle γ opposing c,

$$c^2 = a^2 + b^2 - 2ab\cos(\gamma). \tag{0.5.36}$$

Much like the law of sines, the choice of angle does not matter, as long as we plug the correct sides to the equation: for α and β being the angles opposing to a and b respectively,

$$\begin{aligned} a^2 &= b^2 + c^2 - 2bc\cos(\alpha), \\ b^2 &= a^2 + c^2 - 2ac\cos(\beta). \end{aligned} \tag{0.5.37}$$

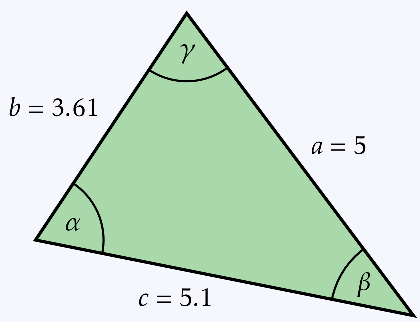
If the triangle in question is a right triangle then one of the angles is equal to 90°. Without loss of generality, let us assume that this is γ. Since $\cos(90°) = 0$ we get that in the case of a right triangle

$$c^2 = a^2 + b^2, \tag{0.5.38}$$

i.e. we retrieve back the Pythagorean theorem.

Example 0.46 Law of cosines

Calculate all angles in the following triangle:



Using the law of cosines:

$$\begin{aligned}
\cos(\gamma) &= \frac{c^2-b^2-a^2}{-2ab} = \frac{5.1^2-3.61^2-5^2}{-2\cdot 5\cdot 3.61} \approx 0.33302 \Rightarrow \gamma = 70.54°, \\
\cos(\beta) &= \frac{b^2-a^2-c^2}{-2ac} = \frac{3.61^2-5^2-5.1^2}{-2\cdot 5\cdot 5.1} \approx 0.74466 \Rightarrow \beta = 41.87°, \\
\cos(\alpha) &= \frac{a^2-b^2-c^2}{-2cb} = \frac{5^2-3.61^2-5.1^2}{-2\cdot 5.1\cdot 3.61} \approx 0.38135 \Rightarrow \alpha = 67.58°.
\end{aligned}$$

0.6 COMPLEX NUMBERS

0.6.1 Algebraic approach

Real numbers, while being extremely useful, are not complete - they can't solve all equations involving numbers. For example, the equation

$$x^2+1=0 \tag{0.6.1}$$

has no real solutions, since there can be no real number x such that $x^2=-1$. However, we can choose to define a new number, $\mathrm{i}=\sqrt{-1}$ and using it to build a new number system. This system is of course the set of complex numbers, $\mathbb{C}$. It is defined as the set of all z such that

$$z = a + \mathrm{i}b, \tag{0.6.2}$$

where $a,b\in\mathbb{R}$ and $\mathrm{i}=\sqrt{-1}$. We call a the **real component** of z or $\mathfrak{Re}(z)$, and b its **imaginary component** or $\mathfrak{Im}(z)$[6]. These numbers appear a lot all throughout the exact sciences (but especially in physics and engineering), so we must at the very least learn their basic properties.

It is not so obvious that we can add two different kinds of numbers together, but it works (the linear algebra chapter sheds more light on this idea). What is important is that we always keep these two parts separated. We see this when we add together two complex numbers z_1, z_2:

$$z = z_1+z_2 = (a_1+b_1\mathrm{i}) + (a_2+b_2\mathrm{i}) = (a_1+a_2)+(b_1+b_2)\mathrm{i}. \tag{0.6.3}$$

The real part of z is therefore a_1+b_1, and its imaginary part is b_1+b_2.

What happens when we multiply two complex numbers? Let's check:

$$\begin{aligned}
z = z_1z_2 &= (a_1+b_1\mathrm{i})(a_2+b_2\mathrm{i}) \\
&= a_1a_2 + \mathrm{i}a_1b_2 + \mathrm{i}a_2b_1 + \mathrm{i}^2b_1b_2 \\
&= a_1a_2 + \mathrm{i}a_1b_2 + \mathrm{i}a_2b_1 - b_1b_2
\end{aligned}$$

[6] There is nothing more "real" about real numbers than imaginary numbers, but unfortunately that's the terminology we're stuck with ¯_(ツ)_/¯

$$= (a_1 a_2 - b_1 b_2) + \mathrm{i}\,(a_1 b_2 + a_2 b_1). \tag{0.6.4}$$

We see that we can still separate the real part and imaginary part of the result. What happens in the case of two real numbers? For real numbers $b = 0$, and thus Equation 0.6.4 devolves to $z = a_1 a_2 \in \mathbb{R}$, which is exactly what we expect: multiplying two real numbers yields their product, which is a real number. Notice that this doesn't happen with purely imaginary numbers: multiplying together two imaginary numbers (i.e. numbers for which $a = 0$) results in a real number. Will get to understand why this happens very soon.

When discussing real numbers sometimes we like to refer to their *magnitude*, i.e. their absolute value. With complex numbers this is defined as

$$|z| = \sqrt{a^2 + b^2}, \tag{0.6.5}$$

i.e. in a sense, to get the magnitude of a complex number we imagine its two components as being perpendicular and calculate the length of the resulting hypotenuse (cf. the Pythagorean theorem). In fact, this is one very useful interpretation of complex numbers, which we will explore in depth in the next subsection.

A very important operation that can be applied to complex numbers is **conjugation**. The conjugate of a complex number $z = a + \mathrm{i}b$ is defined as

$$\bar{z} = a - \mathrm{i}b, \tag{0.6.6}$$

i.e. conjugating a number is simply negating its imaginary part. When we multiply a complex number by its own complex conjugate we get

$$z\bar{z} = (a + \mathrm{i}b)(a - \mathrm{i}b) = a^2 + \cancel{ab\mathrm{i}} - \cancel{ab\mathrm{i}} - b^2\mathrm{i}^2 = a^2 + b^2, \tag{0.6.7}$$

i.e. $z\bar{z} = |z|^2$. The inverse of a complex number can be expressed as

$$z^{-1} = \frac{\bar{z}}{|z|^2}. \tag{0.6.8}$$

0.6.2 Geometric approach

As alluded to in the previous subsection, we can interpret a complex number $z = a + \mathrm{i}b$ as two components in a 2-dimensional space (called the **complex plane**), in which the horizontal axis represents real components, and the vertical access represents imaginary components:

Drawing a line from z to a (on the real axis) creates a right triangle. We can then define θ to be the angle near the origin and r the length of the hypotenuse: We call r the **magnitude** of z, and θ its **argument**. The ranges for r and θ are, respectively, $[0, \infty)$ and $[0, 2\pi)$.

Using Equation 0.5.4 the real and imaginary components of z are

$$\begin{aligned} a &= r\cos(\theta), \\ b &= r\sin(\theta), \end{aligned} \tag{0.6.9}$$

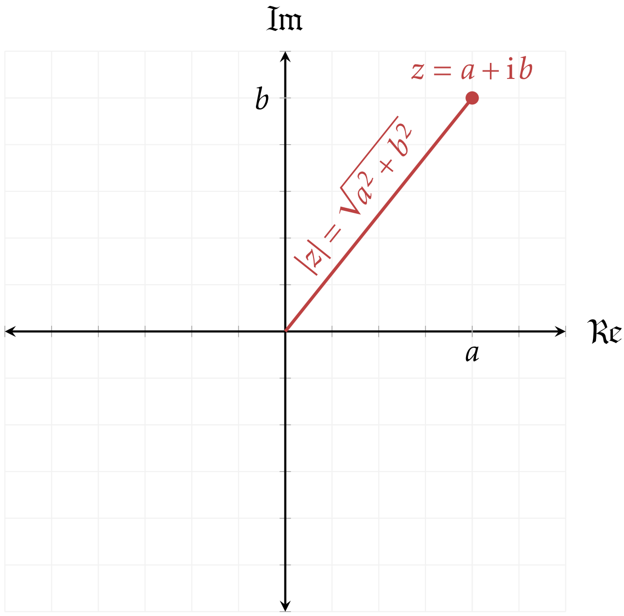


Figure 0.9 A complex number $z = a + ib$ shown on the complex plane: the horizontal and vertical axes represent the real and imaginary components, respectively.

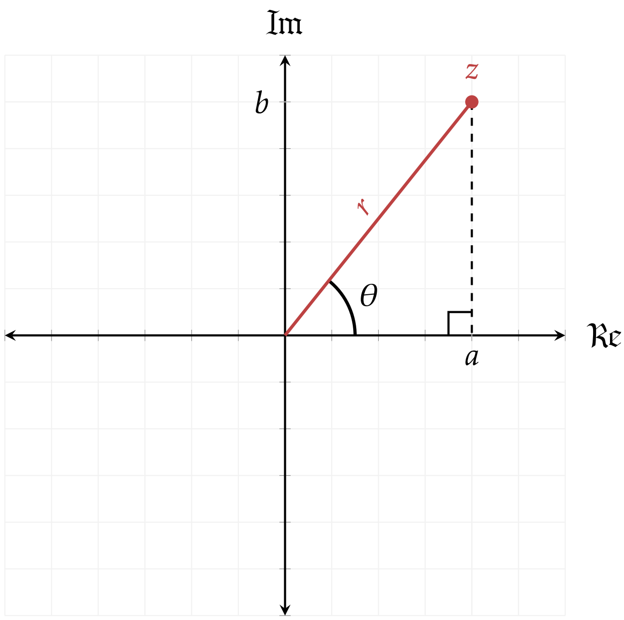


Figure 0.10 The same complex number z from Figure 0.9 shown with its polar components $r = |z| = \sqrt{a^2 + b^2}$ and $\theta = \arctan\left(\frac{b}{a}\right)$.

and z can be re-written as

$$z = r\left(\cos(\theta) + \mathrm{i}\sin(\theta)\right). \tag{0.6.10}$$

Which we call the **polar form** of z (contrasted with $z = a + \mathrm{i}b$ being the **Cartesian form** of z).

Inverting the relations in Equation 0.6.9 yields the relations

$$\begin{aligned} r &= a^2 + b^2, \\ \theta &= \arctan\left(\frac{b}{a}\right). \end{aligned} \tag{0.6.11}$$

Let's examine the same properties of complex numbers shown in Equations 0.6.3, 0.6.4 and 0.6.6, and verify that they work in the polar form of complex numbers. We start with addition (Equation 0.6.3):

$$\begin{aligned} z_1 + z_2 &= r_1\left[\cos(\theta_1) + \mathrm{i}\sin(\theta_1)\right] + r_2\left[\cos(\theta_2) + \mathrm{i}\sin(\theta_2)\right] \\ &= \underbrace{r_1\cos(\theta_1)}_{a_1} + \underbrace{r_2\cos(\theta_2)}_{a_2} + \mathrm{i}\,\underbrace{r_1\sin(\theta_1)}_{b_1} + \mathrm{i}\,\underbrace{r_2\sin(\theta_2)}_{b_2} \\ &= (a_1 + a_2) + \mathrm{i}\,(b_1 + b_2). \end{aligned} \tag{0.6.12}$$

We see that indeed, the polar form of complex numbers adheres to the addition rule in Equation 0.6.3. Next is the product rule:

$$\begin{aligned} z_1 z_2 &= r_1\left[\cos(\theta_1) + \mathrm{i}\sin(\theta_1)\right] \cdot r_2\left[\cos(\theta_2) + \mathrm{i}\sin(\theta_2)\right] \\ &= r_1 r_2\left[\cos(\theta_1)\cos(\theta_2) + \mathrm{i}\cos(\theta_1)\sin(\theta_2) + \mathrm{i}\sin(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)\right] \\ &= r_1\cos(\theta_1) r_2\cos(\theta_2) - r_1\sin(\theta_1) r_2\sin(\theta_2) + \mathrm{i}\left[r_1\cos(\theta_1) r_2\sin(\theta_2) + r_1\sin(\theta_1) r_2\cos(\theta_2)\right] \\ &= (a_1 a_2 - b_1 b_2) + \mathrm{i}\,(a_1 b_2 + a_2 b_1), \end{aligned} \tag{0.6.13}$$

which is indeed the result seen in Equation 0.6.4. We can also develop further the second row of Equation 0.6.13 using some trigonometry (specifically the trigonometric identities in Equation 0.5.25):

$$\begin{aligned} z_1 z_2 &= r_1 r_2\left[\cos(\theta_1)\cos(\theta_2) + \mathrm{i}\cos(\theta_1)\sin(\theta_2) + \mathrm{i}\sin(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2)\right] \\ &= r_1 r_2\left[\cos(\theta_1)\cos(\theta_2) - \sin(\theta_1)\sin(\theta_2) + i\left[\cos(\theta_1)\sin(\theta_2) + \sin(\theta_1)\cos(\theta_2)\right]\right] \\ &= r_1 r_2\left[\cos(\theta_1 + \theta_2) + \mathrm{i}\sin(\theta_1 + \theta_2)\right]. \end{aligned} \tag{0.6.14}$$

This is a very important result: it shows that multiplying a complex number z_1 by another complex number z_2 gives a complex number with magnitude $r_1 r_2$, i.e. the product of the magnitudes of the two complex numbers, and argument $\theta_1 + \theta_2$, i.e. the argument of z_1 rotated by the argument of z_2 (or vice-versa). We will consider this result in more detail soon.

In the polar form the complex conjugate of a number $z = r\left[\cos(\theta) + \mathrm{i}\sin(\theta)\right]$ can be brought about by substituting $-\theta$ into the arguments of the trigonometric functions:

$$\begin{aligned} \bar{z} &= r\left[\cos(-\theta) + \mathrm{i}\sin(-\theta)\right] \\ &= r\left[\cos(\theta) - \mathrm{i}\sin(\theta)\right] \end{aligned}$$

Table 0.4 Values of e^{ix} for some useful values of x (cf. Table 0.3 for the values of $\sin(\theta)$ and $\cos(\theta)$).

x	$\cos(x)$	$\sin(x)$	$z = e^{ix}$
$\frac{\pi}{2}$	0	1	i
π	-1	0	-1
$\frac{3\pi}{2}$	0	-1	$-i$
$\frac{\pi}{3}$	$\frac{1}{2}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}\left(1+i\sqrt{3}\right)$
$\frac{\pi}{4}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}(1+i)$
$\frac{\pi}{6}$	$\frac{\sqrt{3}}{2}$	$\frac{1}{2}$	$\frac{1}{2}\left(\sqrt{3}+i\right)$

$$
\begin{aligned}
&= r\cos(\theta) - ir\sin(\theta) \\
&= a - ib. \qquad (0.6.15)
\end{aligned}
$$

Lastly, let's show that Equation 0.6.7 can be derived in the polar form:

$$
\begin{aligned}
z\bar{z} &= r[\cos(\theta) + i\sin(\theta)] \cdot r[\cos(\theta) - i\sin(\theta)] \\
&= r^2\left[\cos^2(\theta) - \cancel{i\cos(\theta)\sin(\theta)} + \cancel{i\sin(\theta)\cos(\theta)} + \sin^2(\theta)\right] \\
&= r^2\left[\sin^2(\theta) + \cos^2(\theta)\right] \\
&= r^2 = a^2 + b^2. \qquad (0.6.16)
\end{aligned}
$$

In 1748 Leonhard Euler published his famous work *Introduction to analysis of the infinite*[7]. In it he introduced the following relation, called **Euler's formula**:

$$e^{ix} = \sin(x) + i\cos(x). \qquad (0.6.17)$$

Using Euler's formula a complex number z can be written as

$$z = re^{i\theta}. \qquad (0.6.18)$$

In Table 0.4 we can see some useful complex exponentials e^{ix}. Specifically, setting $x = \pi$ yields the famous **Eurler's identity**, considered by many to be one of the most beautiful equations in mathematics, as it binds together five important numbers, namely $0, 1, \pi, e$ and i:

$$e^{i\pi} + 1 = 0. \qquad (0.6.19)$$

Table 0.4 also shows us the integer behaviours of i:

$$i^2 = -1,\ i^3 = -i, i^4 = 1,\ i^5 = i,\ i^6 = -1,\ i^7 = -i,\ \ldots \qquad (0.6.20)$$

[7] Latin for **Introduction to the Analysis of the Infinite**.

0.6.3 Roots of complex numbers

What is the n-th order roots of a complex number z, i.e. $\sqrt[n]{z}$? An illuminating way to approach this problem is by looking at the polar form of z. As an example, we start with the number $z=1$ and find its 3rd order roots, i.e. all number w such that $w^3=1$ (spoiler alert: there are three such numbers).

Equation 0.6.14 taught us that complex numbers not only scale other numbers, but also rotate them: with real numbers the product $x\cdot y$ is equivalent to a scaling of x by y. With complex numbers the product has two components: its magnitude is the scale of x by y, and its argument is the argument of x rotated by the argument of y[8].

Example 0.47 Rotation using complex numbers

Let $z=2+2\mathrm{i}$ and $w=3\mathrm{i}$. Their polar forms are $z=2\sqrt{2}\left[\cos(\pi/4)+\mathrm{i}\sin(\pi/4)\right]$ and $w=3\left[\cos(\pi/2)+\mathrm{i}\sin(\pi/2)\right]$. Their product is

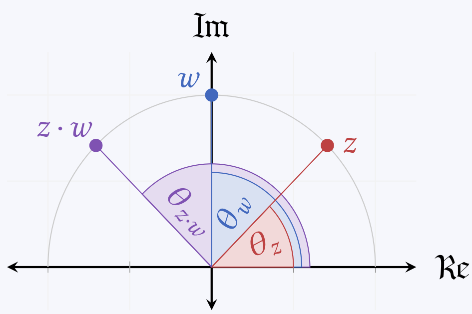
$$z\cdot w=(2+2\mathrm{i})\cdot 3\mathrm{i}=6\mathrm{i}+6\mathrm{i}^2=-6+6\mathrm{i},$$

which in polar form is $6\sqrt{2}\left[\cos(3\pi/4)+\mathrm{i}\sin(3\pi/4)\right]$. Note that

$$2\sqrt{2}\cdot 3=6\sqrt{2},\ \text{and}$$
$$\frac{\pi}{4}+\frac{\pi}{2}=\frac{3\pi}{4},$$

i.e. $z\cdot w$ has magnitude which is the **product** of the magnitudes of z and w, and an argument which is the **sum** of the arguments of z and w.

The figure below depict the arguments of the three numbers $z, w, z\cdot w$. Note that $\theta_{z\cdot w}=\theta_z+\theta_w$.

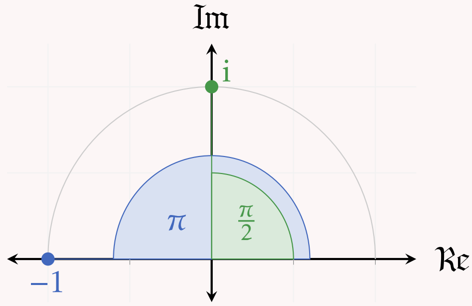


Note 0.9 $\mathrm{i}^2=-1$ from a geometric (polar) viewpoint

In polar coordinates i has magnitude 1 and argument $\frac{\pi}{2}$, i.e. multiplying by i is equivalent to rotation by $\frac{\pi}{2}$ (90°) counter clockwise. Therefore, multiplying i by

[8] Due to the commutativity of the complex product we can switch the order of z and w and get the same result.

itself, i.e. i^2, rotates i itself by $\frac{\pi}{2}$ counter clockwise, bringing it to -1.



In polar form $1 = \cos(0) + \mathrm{i}\sin(0)$. Finding the arguments of the cube roots of 1 is therefore done by answering the following question: what angles θ will equal 0 (or its equivalent angles $2\pi, 4\pi, 6\pi, \dots$) when multiplied by 3? The answer is very simple: the only possible solutions are

$$\begin{aligned} \theta_1 &= 0, \\ \theta_2 &= \frac{2\pi}{3}, \\ \theta_3 &= \frac{4\pi}{3}. \end{aligned} \tag{0.6.21}$$

(any other number in the range $[0, 2\pi)$ will give the same angles)

Therefore, the three cube roots of 1 are [9] (Figure 0.11)

$$\begin{aligned} z_1 &= \cos(\theta_1) + \mathrm{i}\sin(\theta_1) = \cos(0) + \mathrm{i}\sin(0) = 1, \\ z_2 &= \cos(\theta_2) + \mathrm{i}\sin(\theta_2) = \cos\left(\frac{2\pi}{3}\right) + \mathrm{i}\sin\left(\frac{2\pi}{3}\right) = -0.5 + \frac{\sqrt{3}}{2}\mathrm{i}, \\ z_3 &= \cos(\theta_3) + \mathrm{i}\sin(\theta_1) = \cos\left(\frac{4\pi}{3}\right) + \mathrm{i}\sin\left(\frac{4\pi}{3}\right) = -0.5 - \frac{\sqrt{3}}{2}\mathrm{i}. \end{aligned} \tag{0.6.22}$$

The n-th degree roots of 1 will follow the same pattern for $n \in \mathbb{N}$ (see Figure 0.12): their magnitude is always 1, and the argument of the k-th root is

$$\theta_k = \frac{2\pi}{n}k. \tag{0.6.23}$$

Finding the n-th degree roots of a general complex number $z = r[\cos(\theta) + \mathrm{i}\sin(\theta)]$ can be done in a similar fashion: all roots will have the magnitude $\sqrt[n]{r}$, and their argument θ_k will be such that multiplying it by n gives $\theta + 2\pi m$ for some integer value m, i.e.

$$\theta_k = \frac{\theta + 2\pi m}{n}. \tag{0.6.24}$$

[9] recall that $\sqrt[3]{1} = 1$, and therefore all roots have magnitude 1.

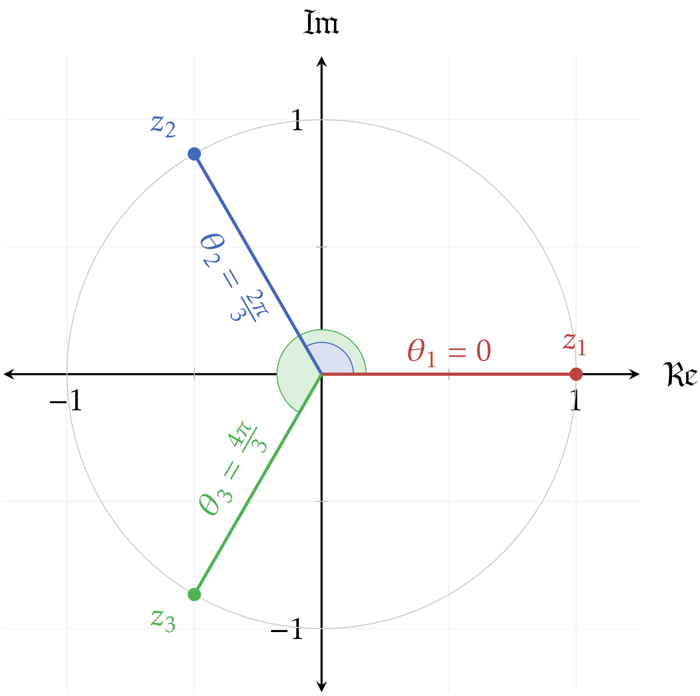


Figure 0.11 The three cube roots of $z = 1$.

0.7 EXERCISES

0.1. Write the following sets explicitly:

(i) $\{x \in \mathbb{N} \mid 1 < x \leq 7\}$

(ii) $\{x \in \mathbb{Z} \mid x < 5\}$

(iii) $\left\{x \in \mathbb{R} \mid x^2 = -1\right\}$

(iv) $\{x \in \mathbb{N} \wedge x \in \mathbb{Q}\}$

(v) $\left\{x \in \mathbb{R} \mid x^2 - 3x - 4 = 0\right\}$

(vi) $\{x \in \mathbb{R} \mid x < 5 \wedge x \geq 2\}$

0.2. Determine the relation between the sets:

(i) $A = \{1,\ 2,\ 3\},\ B = \{1,\ 2\}$

(ii) $A = \varnothing,\ B = \{2,\ -5,\ \pi\}$

(iii) $A = \mathbb{Z},\ B = \{\pm x \mid x \in \mathbb{N} \cup \{0\}\}$

(iv) $A = \left\{\pi, \mathrm{e}, \sqrt{2}\right\},\ B = \mathbb{Q}$

0.3. Write all elements in $S^2 \times W$, where $S = \{\alpha, \beta, \gamma\}$ and $W = \{x, y, z\}$. Find a condition that guarantees $S^2 \times W = W \times S^2$.

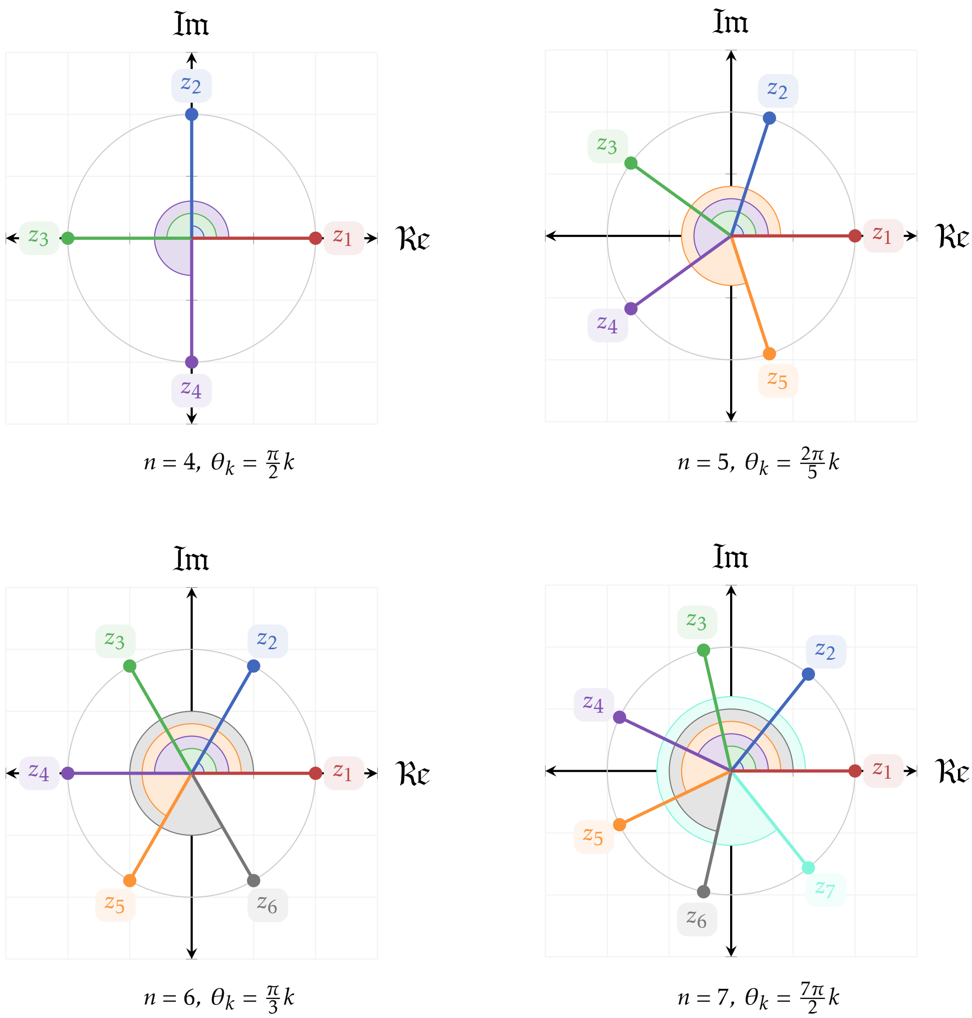


Figure 0.12 Complex n-th roots of $z = 1$ for $n = 4, 5, 6, 7$. Note that for all circles $r = 1$.

0.4. How many different injective functions $f:\{1,2\}\to\{1,2\}$ exist? How many injective functions $f:\{1,2,3\}\to\{1,2,3\}$ exist? How many inject functions $f:\{1,2,\ldots,n\}\to\{1,2,\ldots,n\}$ exist for a given $n\in\mathbb{N}$?

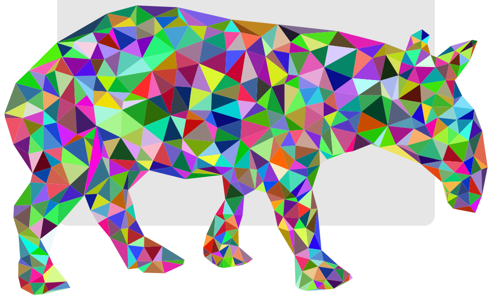
0.5. For each of the real functions below, find a set on which it is surjective (use a graphing calculator if you are not familiar with the shape of a function):

$$x^2,\ x^3-5,\ \mathrm{e}^{-x^2/2},\ \sin(x),\ \sin(x)+\cos(x),\ x\mathrm{e}^x.$$

0.6. Given two sets A,B such that $|B|=|A|-1$, can a bijective function $f:A\to B$ exist? Explain your answer.

0.7. MORE EXERCISES TO BE WRITTEN...

CHAPTER 1



LINEAR ALGEBRA

(INTUITIVE APPROACH)

Linear algebra is one of the most important and often used fields, both in theoretical and applied mathematics. It brings together the analysis of systems of linear equations and the analysis of linear functions (in this context usually called linear transformations), and is employed extensively in almost any modern mathematical field, e.g. approximation theory, vector analysis, signal analysis, error correction, 3-dimensional computer graphics and many, many more.

In this book, we divide our discussion of linear algebra into to chapters: the first (this chapter) deals with a wider, birds-eye view of the topic: it aims to give an intuitive understanding of the major ideas of the topic. For this reason, in this chapter we limit ourselves almost exclusively to discussing linear algebra using 2- and 3-dimensional analysis (and higher dimensions when relevant) using real numbers only. This allows us to first create an intuitive picture of what is linear algebra all about, and how to use correctly the tools it provides us with.

The next chapter takes the opposite approach: it builds all concepts from the ground-up, defining precisely (almost) all basic concepts and proving them rigorously, and only then using them to build the next steps. This approach has two major advantages: it guarantees that what we build has firm foundations and does not fall apart at any future point, and it also allows us to generalize the ideas constructed during the process to such extent that they can be used as foundation to build ever newer tools we can apply in a wide range of cases.

1.1 VECTORS

1.1.1 Basics

Vectors are the fundamental objects of linear algebra: the entire field revolves around manipulation of vectors. In this chapter we deal with the so-called **real vectors**, which can be be defined in a geometric way:

Definition 1.1 Real vectors

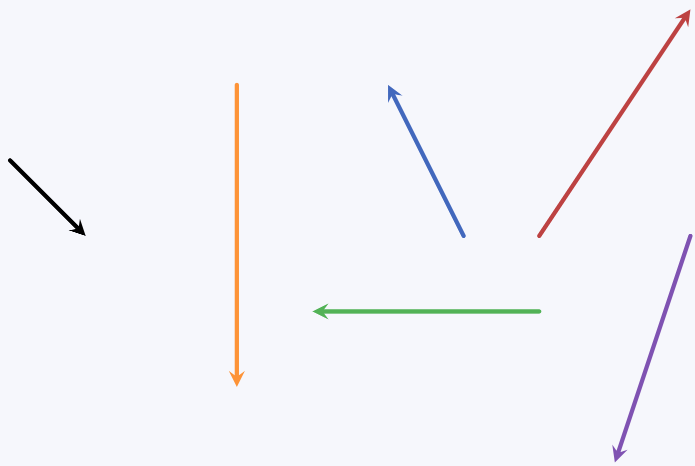
A *real vector* is an object with a **magnitude** (also called **norm**) and a **direction**.

π

In this chapter we refer to real vectors simply as *vectors*.

Example 1.1 Real vectors

The following are all vectors in 2-dimensional space depicted as arrows:



Vectors are usually denoted in one of the following ways:

- **Arrow above letter**: $\vec{u}, \vec{v}, \vec{x}, \vec{a}, \dots$

- **Bold letter**: $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}, \boldsymbol{a}, \ldots$

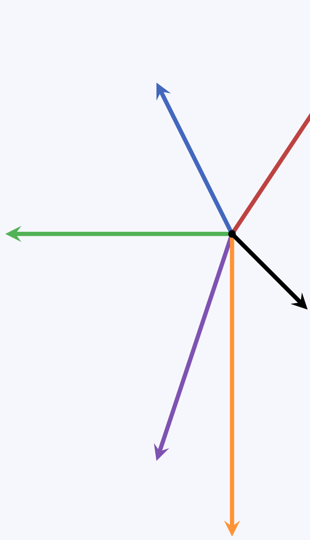
- **Bar below letter**: $\underline{u}, \underline{v}, \underline{x}, \underline{a}, \ldots$

In this book we use the first notation style, i.e. an arrow above the letter. In addition vectors will almost always be denoted using lowercase Lating script.

When discussing vectors in a single context, we always consider them starting at the same point, called the **origin**, and **translating** (moving) vectors around in space does not change their properties: only their norms and directions matter.

Example 1.2 Real vectors

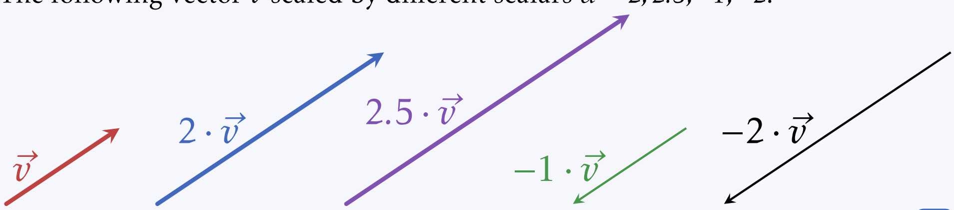
The vectors from the previous translated (moved) such that their origins all lie on the same point:



A vector can be scaled by a real number α: when this happens, its norm is multiplied by α while its direction stays the same. We call α a **scalar**.

Example 1.3 Scaling vectors

The following vector $\vec{v}$ scaled by different scalars $\alpha = 2, 2.5, -1, -2$:



Note 1.1 Negative scale

As can be seen in the example above, when scaling a vector by a negative amount its direction reverses. However, we consider two opposing direction (i.e. directions that are 180° apart) as being the same direction.

In this book we use the following notation for the norm of a vector $\vec{v}$: $\left\|\vec{v}\right\|$.

A vector $\vec{v}$ with norm $\left\|\vec{v}\right\| = 1$ is called a **unit vector**, and is usually denoted by replacing the arrow symbol by a hat symbol: $\hat{v}$. Any vector (except $\vec{0}$) can be scaled into a unit vector by scaling the vector by 1 over its own norm, i.e.

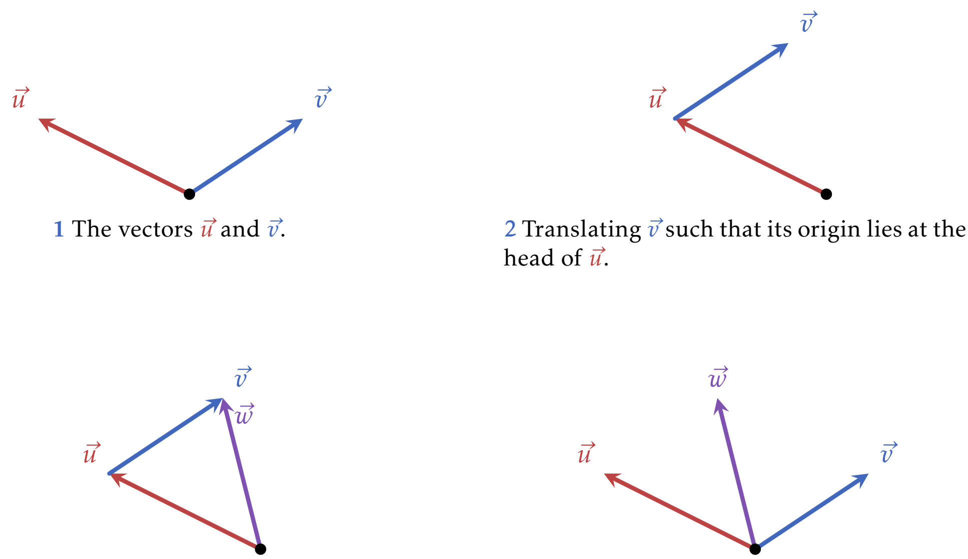
$$\hat{v} = \frac{1}{\left\|\vec{v}\right\|}\vec{v}. \tag{1.1.1}$$

The result of normalization is a vector of unit norm which points in the same direction of the original vector.

Two vectors can be added together to yield a third vector: $\vec{u} + \vec{v} = \vec{w}$. To find $\vec{w}$ we use the following procedure (depicted in Figure 1.1):

1.1. Move (translate) $\vec{v}$ such that its origin lies on the head of $\vec{u}$.

1.2. The vector $\vec{w}$ is the vector drawn from the origin of $\vec{u}$ to the head of $\vec{v}$.



1 The vectors $\vec{u}$ and $\vec{v}$.

2 Translating $\vec{v}$ such that its origin lies at the head of $\vec{u}$.

3 Drawing the vector $\vec{w}$ from the origin to the head of $\vec{v}$.

4 Showing all three vectors.

Figure 1.1 Vector addition.

The addition of vectors as depicted here is commutative, i.e. $\vec{u} + \vec{v} = \vec{v} + \vec{u}$. This can be seen by using the **parallogram law of vector addition** as depicted in Figure 1.2:

drawing the two vectors $\vec{u}, \vec{v}$ and their translated copies (each such that its origin lies on the other vector's head) results in a parallelogram.

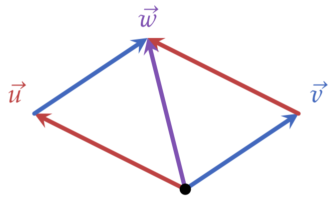


Figure 1.2 The parallogram law of vector addition.

An important vector is the **zero-vector**, denoted as $\vec{0}$. The zero-vector has a unique property: it is neutral in respect to vector addition, i.e. for any vector $\vec{v}$,

$$\vec{v} + \vec{0} = \vec{v}. \tag{1.1.2}$$

(we also say that $\vec{0}$ is the **additive identity** in respect to vectors.)

Any vector $\vec{v}$ always has an **opposite** vector, denoted $-\vec{v}$. The addition of a vector and its opposite always result in the zero-vector, i.e.

$$\vec{v} + (-\vec{v}) = \vec{0}. \tag{1.1.3}$$

1.1.2 Components

Vectors can be decomposed to their components, the number of which depends on the dimension of space we're using: 2-dimensional vectors can be decomposed into 2 components, 3-dimensional vectors can be decomposed into 3 components, etc. To decompose a vector, say $\vec{v}$, we first choose a coordinate system: the most commonly used system, and the one we will use for most of this chapter, is the Cartesian coordinate system. We place the vector in the coordinate system such that its origin lies at the origin of the system. We then draw a perpendicular line from its head to each of the axes in the system (see Figure 1.3), the point of interception on each axis is the component of the vector in that axis (we label these points v_x, v_y, v_z in the case of 2- or 3-dimensional spaces, and generally $v_1, v_2, v_3, \dots$). The vector can then be written as a column using these components:

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}. \tag{1.1.4}$$

Note 1.2 Order of components

The order of the components of a vector is important, and should always be consistent. In the case of 2- and 3-dimensional the order is always v_x, v_y, v_z.

!

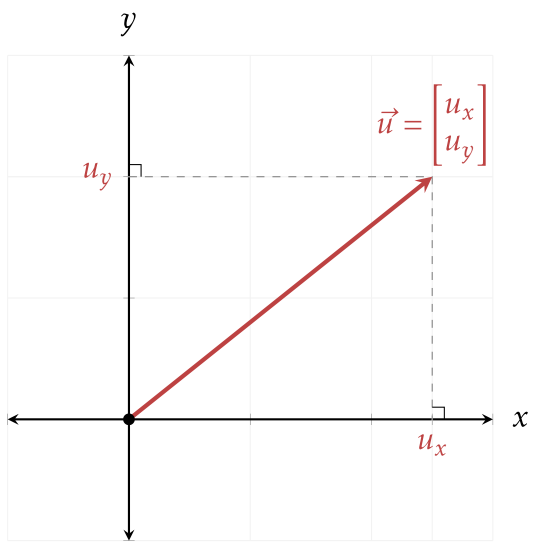
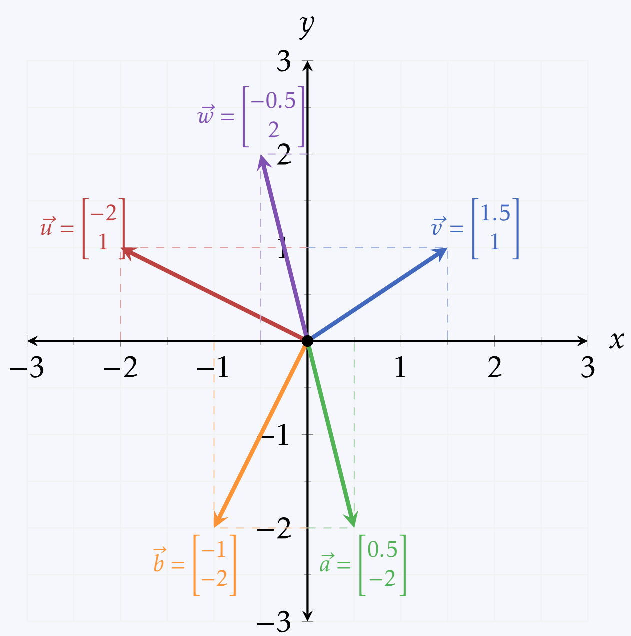


Figure 1.3 Placing a 2-dimensional vector $\vec{u}$ on the 2-dimensional Cartesian coordinate system, showing its x- and y-components.

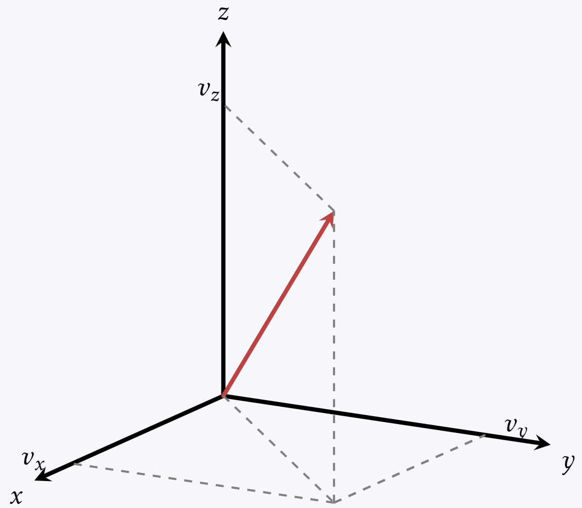
Example 1.4 Vector components in two dimensions

The following five 2-dimensional vectors are decomposed each into its x- and y-components:



Example 1.5 Vector components in three dimensions

The following 3-dimensional vector is decomposed into its x-, y- and z-components: (THIS NEEDS TO BE IMPROVED AND FINISHED)



The column form of a vector is essentially equivalent to an order list of n real numbers, i.e. $(v_1, v_2, \ldots, v_n)$. Why then are we using the column form and not the list form (mostly known as **row vectors**)? In fact, we could use either form - and even using both interchangeably - and with only minor adjusments the entire chapter would stay the same as it is now. However, there are some advantages of using only a single form, and consider the other form as a different object altogether. This idea will become clear in future chapters, when discussing **covariant vectors**, **contravarient vectors**, and **tensors**. For now, we stick with the column form of vectors to stay consistent with common notation.

However, the row form of vectors highlights the space in which they exist: n-dimensional vectors live in a space we call $\mathbb{R}^n$. Recall from Chapter 0 that the set $\mathbb{R}^n$ is a Cartesian product made up of n times the set of real numbers, i.e.

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n}. \tag{1.1.5}$$

Each member of this set is a list of n real numbers, and their order inside the list matters - very similar to vectors, be they in row or column form. For this reason, we refer to $\mathbb{R}^n$ as the space of n-dimensional real vectors. As mentioned, in this chapter we use $\mathbb{R}^2$ (the 2-dimensional real space) and $\mathbb{R}^3$ (the 3-dimensional real space) for most ideas and examples.

Looking at vectors in $\mathbb{R}^2$, it is rather straight-forward to calculate their norm: since the origin, the head of the vector and the point v_x form a right triangle (see Figure 1.4), we can use the Pythagorean theorem to calculate the norm of the vector, which is equal to the hypotenous of said triangle:

$$\left\|\vec{v}\right\| = \sqrt{v_x^2 + v_y^2}. \tag{1.1.6}$$

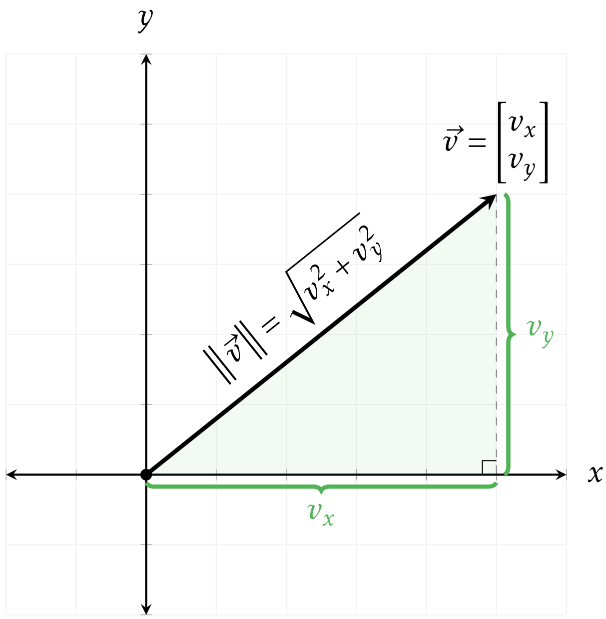


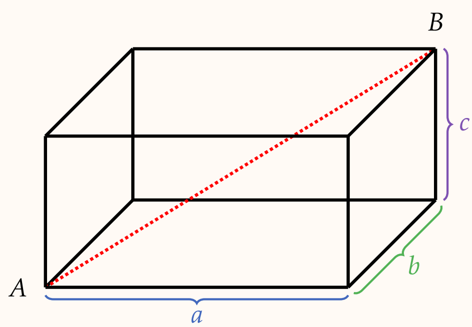
Figure 1.4 Calculating the norm of a 2-dimensional column vector.

In $\mathbb{R}^3$ the norm of a vector $\vec{v}$ is similarily

$$\left\|\vec{v}\right\| = \sqrt{v_x^2 + v_y^2 + v_z^2}. \tag{1.1.7}$$

Challange 1.1 Norm of a 3D vector

Show why Equation 1.1.7 is valid, by calculating the length AB in the following figure, depicting a box of sides $\boldsymbol{a}, \boldsymbol{b}$ and $\boldsymbol{c}$:



?

Generalizing the vector norms in $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$ yields the following form:

$$\left\|\vec{v}\right\| = \sqrt{v_1^2+v_2^2+v_3^2+\cdots+v_n^2} = \sqrt{\sum_{i=1}^{n} v_i^2}. \tag{1.1.8}$$

Note 1.3 Other norms

The norm shown here is called the 2-norm. There are other possible norm that can be defined, and are used in different situations, such as the 1-norm (also the called **taxicab norm**), general p-norm where $p \geq 1$ is a real number, the zero-norm, the max-norm, and many others. However, for the purpose of this chapter we use only the standard 2-norm, since it is the most useful for describing basic concepts of linear algebra and its uses.

!

Scaling a vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ by a real number α is done by multiplying each of its components by α, i.e.

$$\alpha\vec{v} = \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix}. \tag{1.1.9}$$

We can prove Equation 1.1.9 by directly calculating the norm of a scaled vector $\vec{w} = \alpha\vec{v}$:

Proof 1.1 Scaling a column vector

Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ and $\vec{w} = \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix}$, where $\alpha \in \mathbb{R}$. Then $\vec{w}$ has the following norm:

$$\begin{aligned}
\left\|\vec{w}\right\| &= \sqrt{\sum_{i=1}^{n} (\alpha v_i)^2} \\
&= \sqrt{(\alpha v_1)^2 + (\alpha v_2)^2 + \cdots + (\alpha v_1)^2} \\
&= \sqrt{\alpha^2 v_1^2 + \alpha^2 v_2^2 + \cdots + \alpha^2 v_n^2} \\
&= \sqrt{\alpha^2 \left(v_1^2 + v_2^2 + \cdots + v_n^2\right)} \\
&= \alpha\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} \\
&= \alpha\left\|\vec{v}\right\|.
\end{aligned}$$

This shows that indeed $\vec{w} = \alpha\vec{v}$.

QED

Another idea we can prove in column form is vector normalization (Equation 1.1.1), by showing that dividing each component of a vector by its norm gives a vector of unit norm:

Proof 1.2 Norm of a vector

Let $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$. Its norm is then $\left\|\vec{v}\right\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$. Scaling $\vec{v}$ by $\frac{1}{\|\vec{v}\|}$ yields

$$\hat{v} = \frac{1}{\left\|\vec{v}\right\|} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \frac{1}{\sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

The norm of $\hat{v}$ is therefore

$$\begin{aligned} \|\hat{v}\| &= \sqrt{\frac{v_1^2}{v_1^2 + v_2^2 + \cdots + v_n^2} + \frac{v_2^2}{v_1^2 + v_2^2 + \cdots + v_n^2} + \cdots + \frac{v_n^2}{v_1^2 + v_2^2 + \cdots + v_n^2}} \\ &= \sqrt{\frac{1}{v_1^2 + v_2^2 + \cdots + v_n^2}\left(v_1^2 + v_2^2 + \cdots + v_n^2\right)} \\ &= \sqrt{1} = 1, \end{aligned}$$

i.e. $\hat{v}$ is indeed a unit vector.

QED

Example 1.6 Normalizing a vector

Let's normalize the vector $\vec{v} = \begin{bmatrix} 0 \\ 4 \\ -3 \end{bmatrix}$. Its norm is

$$\left\|\vec{v}\right\| = \sqrt{0^2 + 4^2 + (-3)^2} = \sqrt{0 + 16 + 9} = \sqrt{25} = 5.$$

Therefore $\hat{v}$ (the normalized $\vec{v}$) is

$$\hat{v} = \begin{bmatrix} 0 \\ \frac{4}{5} \\ -\frac{3}{5} \end{bmatrix}.$$

By calculating the norm of $\hat{v}$ directly, we can see that it is indeed a unit vector:

$$\|\hat{v}\| = \sqrt{0^2 + \frac{4^2}{5^2} + \frac{3^2}{5^2}} = \sqrt{\frac{0^2 + 4^2 + 3^2}{5^2}} = \sqrt{\frac{16 + 9}{25}} = \sqrt{\frac{25}{25}} = \sqrt{1} = 1.$$

The addition of two column vectors $\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$ is done by adding their respective components together, i.e.

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}. \tag{1.1.10}$$

TBW: how this addition is the same as the one shown in Figure 1.1.

Note 1.4 No addition of vectors of different number of components!

Two vectors can only be added together if they have the same number of components. The addition of vectors with different number of components is undefined.

!

1.1.3 Linear combinations, spans and linear dependency

As seen above, scaling a vector by a scalar results in a vector that has the same number of dimensions as the original vector. The same is true for adding two vectors: both of them must be of the same dimension, and the result is also a vector of the same dimension. Therefore, any combination of scaling and addition of vectors results in a vector of the same dimension as the original vector(s). This kind of combination is called a **linear combination**.

Let's define linear combinations a little more formaly:

Definition 1.2 Linear combinations

A linear combination of n vectors $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ of the same dimension, using n scalars $\alpha_1, \alpha_2, \dots, \alpha_n$, is an expression of the form

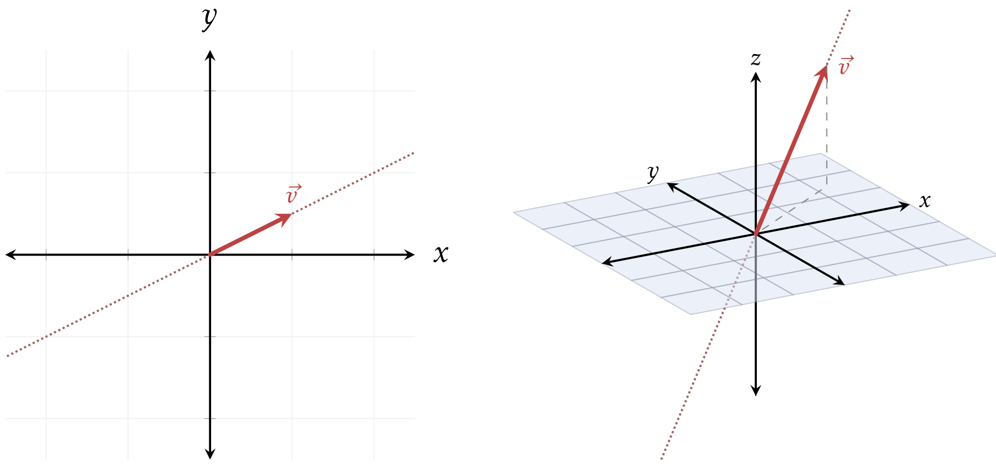
$$\vec{w} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n = \sum_{i=1}^{n} \alpha_i \vec{v}_i. \tag{1.1.11}$$

π

Linear combinations of real vectors have geometric meaningsc: we start with the set of all linear combinations of a single vector $\vec{v} \in \mathbb{R}^n$, i.e.

$$V = \{\alpha \vec{v} \mid \alpha \in \mathbb{R}\}. \tag{1.1.12}$$

The set V represents a line in the direction of $\vec{v}$ going through the origin (see Figure 1.5). The set V is itself a vector space of dimension 1, and as such a **subspace** of $\mathbb{R}^n$. We say that it is the **span** of the vector $\vec{v}$ (i.e. the vector $\vec{v}$ **spans** the subspace V).



1 $\mathbb{R}^2$ **2** $\mathbb{R}^3$

Figure 1.5 The span of a single vector $\vec{v}$, shown as a dashed line: in $\mathbb{R}^2$ (left) and $\mathbb{R}^3$ (right).

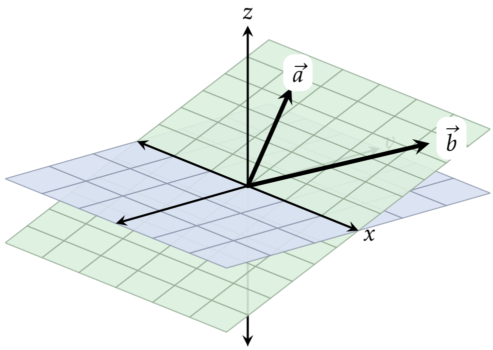


Figure 1.6 Two vectors $\vec{a}$ and $\vec{b}$ span a plane (colored green) in $\mathbb{R}^3$. The xy-plane (i.e. $z = 0$) is shown in blue for emphasis.

Similarly, the set of all linear combinations of two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ that are not scales of each other (i.e. there is no such $\alpha \in \mathbb{R}$ for which $\vec{v} = \alpha\vec{u}$),

$$V = \{\alpha\vec{u} + \beta\vec{v} \mid \alpha, \beta \in \mathbb{R}\}, \tag{1.1.13}$$

is a plane that goes through the origin (see Figure 1.6). Such vectors are also said to be **non-collinear**.

Example 1.7 Spanning $\mathbb{R}^2$ using two non-collinear vectors

Since any two non-collinear vectors span a 2-dimensional subspace of $\mathbb{R}^n$, in $\mathbb{R}^2$ this means that any vector $\vec{w}$ can be written as a linear combination of any two vectors $\vec{u}, \vec{v}$ that are not a scale of each other. For example, we can take the vector

$$\vec{w} = \begin{bmatrix} 7 \\ -1 \end{bmatrix},$$

and write it as a linear combination of any two non-collinear vectors, say

$$\vec{u} = \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \ \vec{v} = \begin{bmatrix} 0 \\ 5 \end{bmatrix}.$$

The equation which forces the relation is

$$\begin{bmatrix} 7 \\ -1 \end{bmatrix} = \alpha \begin{bmatrix} 2 \\ -3 \end{bmatrix} + \beta \begin{bmatrix} 0 \\ 5 \end{bmatrix},$$

and we should solve it for α and β. This is possible since the equation above is actually a system of two equations in two variables (namely α and β):

$$\begin{cases} 7 = 2\alpha, \\ -1 = -3\alpha + 5\beta. \end{cases}$$

The solution for the system is $\alpha = 3.5$ and $\beta = 1.9$, and therefore

$$\begin{bmatrix} 7 \\ -1 \end{bmatrix} = 3.5 \begin{bmatrix} 2 \\ -3 \end{bmatrix} + 1.9 \begin{bmatrix} 0 \\ 5 \end{bmatrix}.$$

As the reader, you should verify for yourself the above equation.

Generalizing the example above, any vector $\vec{w} = \begin{bmatrix} w_x \\ w_y \end{bmatrix}$ can be written as a linear combination of two vectors $\vec{u} = \begin{bmatrix} u_x \\ u_y \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_x \\ v_y \end{bmatrix}$, as long as $\vec{u}$ and $\vec{v}$ are non-collinear. Let's prove this:

Proof 1.3 $\mathbb{R}^2$ is spanned by any two non-collinear vectors in $\mathbb{R}^2$

Let $\vec{u}, \vec{v} \in \mathbb{R}^2$ be two non-collinear vectors. Their non-collinearity means that the equation

$$\vec{u} = \alpha \vec{v} \tag{1.1.14}$$

has no solution, i.e. the system

$$\begin{cases} u_x = \alpha v_x \\ u_y = \alpha v_y \end{cases} \tag{1.1.15}$$

has no solution. The system has solution only when $u_x v_y = u_y v_x$, and so the restriction is translated to the simple equation

$$u_x v_y \neq u_y v_x. \tag{1.1.16}$$

The system which defines $\vec{w}$ as a linear combination of $\vec{u}$ and $\vec{v}$ is

$$\begin{cases} w_x = \alpha u_x + \beta v_x \\ w_y = \alpha u_y + \beta v_y \end{cases} \tag{1.1.17}$$

Isolating α using the first equation yields

$$\alpha = \frac{w_x - \beta v_x}{u_x}, \tag{1.1.18}$$

and subtituting it into the second equation yields

$$\beta = \frac{w_y - \alpha u_y}{v_y} = \frac{w_y - \frac{w_x - \beta v_x}{u_x}}{v_y}, \tag{1.1.19}$$

which rearranges into

$$\beta = \frac{u_x w_y - u_y w_x}{u_x v_y - u_y v_x}, \tag{1.1.20}$$

and thus

$$\alpha = \frac{-v_x w_y + v_y w_x}{u_x v_y - u_y v_x}. \tag{1.1.21}$$

We can see that α and β exist iff $u_x v_y \neq u_y v_x$, which is guaranteed by Equation 1.1.16. Therefore, α and β always exist when $\vec{u}$ and $\vec{v}$ are non-collinear, and thus any vector in $\mathbb{R}^2$ can be written as a linear combination of any two non-collinear vectors in $\mathbb{R}^2$, i.e. any two non-collinear vectors in $\mathbb{R}^2$ span $\mathbb{R}^2$.

QED

Going a step further, any three vectors $\vec{u}, \vec{v}, \vec{w} \in \mathbb{R}^n$ that are not coplanar span a 3-dimensional subspace of $\mathbb{R}^n$ going through the origin. To generalize the notion of collinear and coplanar vectors to higher dimensions we introduct the concept of **linear dependency** of a set of vectors:

Definition 1.3 Linear dependent set of vectors

A set of n vectors

$$S = \{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\} \tag{1.1.22}$$

is said to be linearly dependent if there exist a linear combination

$$\alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n = \vec{0}, \tag{1.1.23}$$

and **at least** one the coefficients $\alpha_i \neq 0$.

π

The following examples shows that the definition above reduces to colinarity and coplanary in the case of 2 and 3 vectors:

Example 1.8 Linear dependency of 2 vectors

Let $\vec{u}$ and $\vec{v}$ be two linearly dependent vectors in $\mathbb{R}^n$. Then there exist a linear combination

$$\alpha\vec{u}+\beta\vec{v}=\vec{0},$$

with either $\alpha \neq 0$ or $\beta \neq 0$ (or both). We can look at the different possible cases:

- $\alpha \neq 0,\ \beta = 0$: in this case $\alpha\vec{u}=\vec{0}$, i.e. $\vec{u}=0$.
- $\alpha = 0,\ \beta \neq 0$: in this case $\beta\vec{v}=\vec{0}$, i.e. $\vec{v}=0$.
- $\alpha \neq 0,\ \beta \neq 0$: in this case we can rearrange the equation and get

  $$\vec{u}=-\frac{\beta}{\alpha}\vec{v},$$

  i.e. $\vec{u}$ and $\vec{v}$ are scales of each other and thus are collinear.

What we learn from this is that two vectors form a linearly dependent set if at least one of the is the zero vector, or if they are collinear.

Example 1.9 Linear dependency of 3 vectors

Now, let $\vec{u},\vec{v}$ and $\vec{w}$ be three linearly dependent vectors in $\mathbb{R}^n$. Then there exists a linear combination

$$\alpha\vec{u}+\beta\vec{v}+\gamma\vec{w}=\vec{0},$$

with either $\alpha \neq 0$ or $\beta \neq 0$ or $\gamma \neq 0$ or any combination where two of the coefficients are non-zero, or all of the coefficients are non-zero. Again, we look at all the possible cases:

- $\alpha \neq 0,\ \beta = \gamma = 0$: we get $\alpha\vec{u}=\vec{0}$, thus $\vec{u}=\vec{0}$.
- $\alpha = 0,\ \beta \neq 0,\ \gamma = 0$: we get $\beta\vec{v}=\vec{0}$, thus $\vec{v}=\vec{0}$.
- $\alpha = \beta = 0,\ \gamma \neq 0$: we get $\gamma\vec{w}=\vec{0}$, thus $\vec{w}=\vec{0}$.
- $\alpha \neq 0,\ \beta \neq 0, \gamma = 0$: we get that $\vec{u}$ and $\vec{v}$ are collinear, since this is exactly as the case for two linearly dependent vectors.
- $\alpha \neq 0,\ \beta = 0, \gamma \neq 0$: similar to the previous case, this time $\vec{u}$ and $\vec{w}$ are collinear.
- $\alpha = 0,\ \beta \neq 0, \gamma \neq 0$: similar to the previous case, this time $\vec{v}$ and $\vec{w}$ are collinear.
- $\alpha \neq 0,\ \beta \neq 0,\ \gamma \neq 0$: by rearranging we get

  $$\vec{w}=-\frac{1}{\gamma}\left(\alpha\vec{u}+\beta\vec{v}\right),$$

  i.e. $\vec{w}$ lies on the the plane spanned by $\vec{u}$ and $\vec{v}$. If we isolate $\vec{u}$ or $\vec{v}$ instead, we get the same result: the isolated vector is a lienar combination of the other two vectors, and thus lies on the plan spanned by these vectors.

From this example we learn that three vectors form a linearly dependent set if one or more of the vectors is the zero vector, or if any two vectors in the set are collinear, or if all three vectors are coplanar.

Just like the case of 2 and 3 vectors seen above, any set of $m \leq n$ vectors in $\mathbb{R}^n$ that are **not** linearly dependent span an m-dimensional subspace of $\mathbb{R}^n$ (which goes throught the origin) - i.e. any vector $\vec{v} \in \mathbb{R}^n$ can be written as a linear combination of these vectors. We call such a set a **basis set** of $\mathbb{R}^n$.

Example 1.10 Basis sets in n dimensions

The following three vectors are non coplanar (i.e. they are linearly independent), and thus form a basis set of $\mathbb{R}^3$:

$$B = \left\{ \begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -5 \end{bmatrix} \right\}.$$

This means that any vector in $\mathbb{R}^3$ can be written as a linear combination of these vectors. We can show this by writing a generic vector $\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in \mathbb{R}^3$ as a linear combination of the vectors:

$$\vec{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \alpha \begin{bmatrix} 0 \\ 4 \\ 5 \end{bmatrix} + \beta \begin{bmatrix} 4 \\ 2 \\ -2 \end{bmatrix} + \gamma \begin{bmatrix} 1 \\ 0 \\ -5 \end{bmatrix},$$

which can be expanded to the system of equations

$$\begin{cases} x = \cancel{0\alpha} + 4\beta + 1\gamma, \\ y = 4\alpha + 2\beta + \cancel{0\gamma}, \\ z = 5\alpha - 2\beta - 5\gamma. \end{cases}$$

The solution of the above system gives the coefficients of the linear combination to yield any vector in $\mathbb{R}^3$:

$$\begin{aligned} \alpha &= -\frac{5x}{31} + \frac{9y}{31} - \frac{z}{31}, \\ \beta &= \frac{10x}{31} - \frac{5y}{62} + \frac{2z}{31}, \\ \gamma &= -\frac{9x}{31} + \frac{10y}{31} - \frac{8z}{31}. \end{aligned}$$

For example, to yield the vector $\vec{v} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$ we sustitute $x = 1,\ y = -1,\ z = 0$ into the above solutions, and get that the following coefficients are needed:

$$\alpha = -\frac{28}{62},\ \beta = \frac{25}{62},\ \gamma = -\frac{38}{62},$$

i.e.

$$-\frac{28}{62}\begin{bmatrix}0\\4\\5\end{bmatrix}+\frac{25}{62}\begin{bmatrix}4\\2\\-2\end{bmatrix}-\frac{38}{62}\begin{bmatrix}1\\0\\-5\end{bmatrix}=\begin{bmatrix}1\\-1\\0\end{bmatrix}.$$

(you, the reader, should verify this!)

Having described basis sets in somewhat general terms, we can now define them a bit more precisely:

Definition 1.4 Basis sets

Let B be a **linearly independent set** of vectors in $\mathbb{R}^n$. If any vector $\vec{v} \in \mathbb{R}^{\ltimes}$ can be written as a linear combination of the vectors in B, then B is called a basis set of $\mathbb{R}^n$. The **dimension** of B is the number of vectors in B.

The dimension of a basis set B of $\mathbb{R}^n$ is always n. In fact, in a later chapter we will see that the dimension of a vector space is defined by the dimension of its basis sets, i.e. given a vector space V and a basis set $B \subseteq V$, the dimension of V is equal to $|B|$, or mathematically

$$\dim(V) = |B|. \tag{1.1.24}$$

It can be easily shown that any set of vectors in $\mathbb{R}^n$ which has more than n vectors must be a linearly dependent set:

Proof 1.4 Sets with more than n vectors in $\mathbb{R}^n$

Let S be a set of $m \in \mathbb{N}$ vectors in $\mathbb{R}^n$, where $m > n$. Given a vector $\vec{v} \in S$ and the set of all vectors in S except $\vec{v}$ (call this set $\tilde{S}$), there are two possibilities:

- $\tilde{S}$ is a linearly dependent set in $\mathbb{R}^n$. In this case, the addition of $\vec{v}$ doesn't change this fact, i.e. the set S as a whole is linearly dependent.
- The set $\tilde{S}$ is linearly independent, and since it has n vectors it forms a basis set of $\mathbb{R}^n$. Therefore, $\vec{v}$ can be written as a linear combination of the vectors in $\tilde{S}$, and thus the inclusion of $\vec{v}$ in S makes S a linearly dependent set.

QED

Let us now take a vector, for example $\vec{v} = \begin{bmatrix}1\\-3\\7\end{bmatrix}$, and span it by three different basis sets:

$$B_1 = \left\{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\right\} \quad B_2 = \left\{\begin{bmatrix}5\\1\\2\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}4\\-1\\1\end{bmatrix}\right\}, \quad B_3 = \left\{\begin{bmatrix}-1\\0\\2\end{bmatrix}, \begin{bmatrix}0\\2\\-3\end{bmatrix}, \begin{bmatrix}2\\2\\3\end{bmatrix}\right\}.$$

As can be seen in Figure 1.7, for each basis set the coefficients (colored) are different. In this context we call the coefficients the **coordinates** of $\vec{v}$ in that basis set. In the basis set

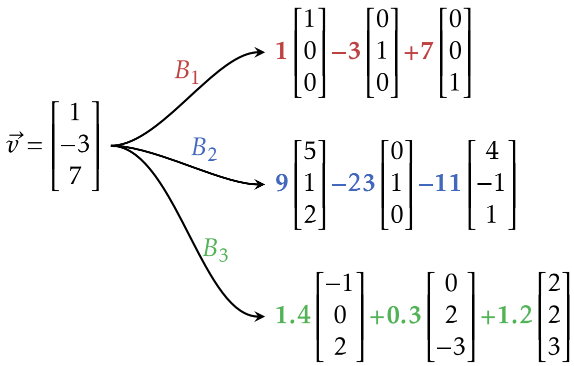


Figure 1.7 The vector $\vec{v} = \begin{bmatrix} 1 \\ -3 \\ 7 \end{bmatrix}$ spanned in three different basis sets.

$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ the coordinates of $\vec{v}$ are $(1, -3, 7)$ (as we will see next, it is not a coincidense that these are equal to its components as a column vector), and in the basis set $\left\{ \begin{bmatrix} 5 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix} \right\}$ its coordinates are $(9, -23, -11)$.

Changing the coordinates of a vector between different basis sets is called **basis transformation**, and is generally done using **matrices**. We will discuss this in more details in the next sections of this chapter. For now, let's look at a graphical representation of a vector being expressed in a different basis set (Figure 1.8): in the figure, we see that the vector $\vec{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ can be written in the basis set $B = \left\{ \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} -4 \\ 2 \end{bmatrix} \right\}$ using the coefficients 2 and $\frac{1}{2}$, i.e.

$$\vec{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 2 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} -4 \\ 2 \end{bmatrix}.$$

Therefore, in the basis set B, the coordinates of $\vec{w}$ are $\left(2, \frac{1}{2}\right)$.

A bsis set B in which all vectors are **orthogonal** (i.e. are at 90°) to each other is called a **orthogonal basis set**. If all vectors are unit vectors as well, i.e. their norms all equal to 1, the basis set is then an **orthonormal basis set**.

Example 1.11 Orthogonal and orthonormal basis sets

The vectors $\vec{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ are linearly independent and thus form a basis set of $\mathbb{R}^2$. We can calculate their respective angles in relation to the x-axis (θ_a and θ_b) to find the angle between them (φ):

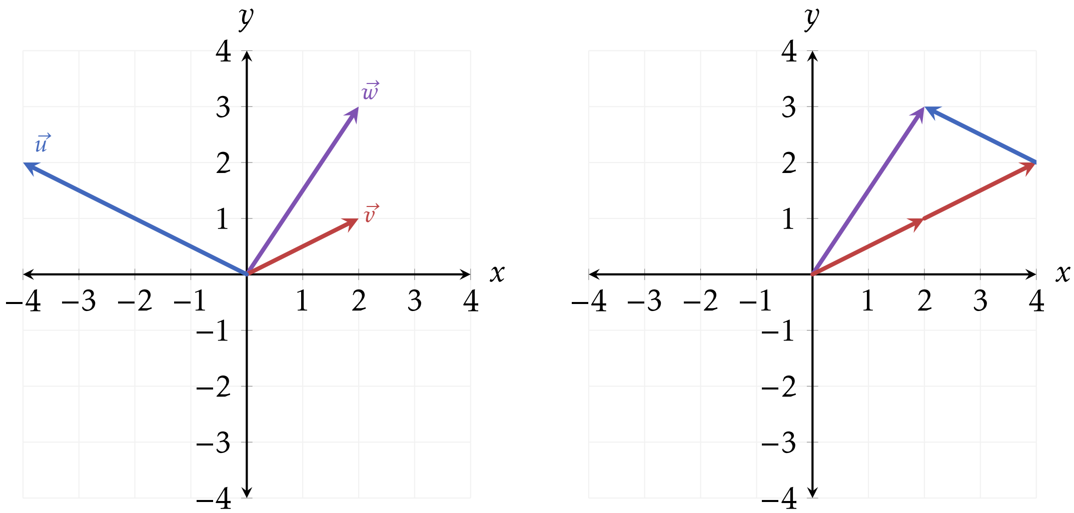
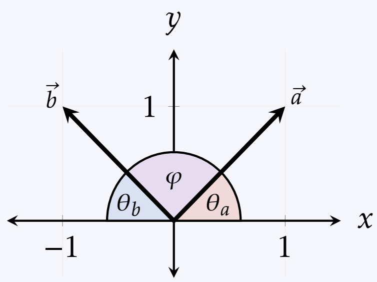


Figure 1.8 The vector $\vec{w} = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$ is spanned using the vectors $\vec{u} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$, yielding the coordinates $\left(2, \frac{1}{2}\right)$ in the basis set B.



The angle of $\vec{a}$ is

$$\theta_a = \arctan\left(\frac{a_y}{a_x}\right) = \arctan(1) = \frac{\pi}{4} \; (= 45°).$$

Similarily, the angle α_b also equals $\frac{\pi}{4}$. Therefore, $\varphi = 2\frac{\pi}{4} = \frac{\pi}{2}$ $(= 90°)$ - i.e. $\vec{a}$ and $\vec{b}$ are orthogonal, and thus form an orthogonal basis set of $\mathbb{R}^2$.
To get a similar *orthonormal* basis set we can simply normalize the two vectors. We start with $\vec{a}$: its norm is

$$\left\|\vec{a}\right\| = \sqrt{1^2 + 1^2} = \sqrt{2}.$$

Thus, the vector $\hat{a} = \frac{1}{\sqrt{2}}\vec{a} = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$ is a unit vector. The same argument is valid for

$\vec{b}$, i.e. $\hat{b} = \frac{1}{2}\vec{b} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$. We therefore get that

$$\left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$$

is an orthonormal basis set of $\mathbb{R}^2$.

Challange 1.2 Orthonormal basis sets of $\mathbb{R}^2$

Show that all orthonormal basis sets of $\mathbb{R}^2$ are rotations of the set

$$\left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} \right\}$$

as a whole (i.e. each rotation angle is applied to both vectors).

?

See example below for such sets in $\mathbb{R}^2$ and $\mathbb{R}^3$.

One common orthonormal basis set in any $\mathbb{R}^n$ is the so-called **standard basis set**. We saw the standard basis set in $\mathbb{R}^3$ in Figure 1.7: it is the set $B_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$. Note how in this set, each vector has a special structure: one of its components is 1 while the rest are 0. In the first basis vector the non-zero component is the first component of the vector, in the second basis vector it is the second component, and in the third basis vector it is the third component. In $\mathbb{R}^2$ the standard basis set is simply $\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$, and generally in $\mathbb{R}^n$ it is

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right\}, \tag{1.1.25}$$

i.e. in the n-th basis vector the n-th component is 1 while the rest are 0. The standard basis vectors are generally labeled as $\hat{e}_1$, $\hat{e}_2$, …, $\hat{e}_n$ - they get the "hat" symbol since they are all unit length.

In $\mathbb{R}^2$ and $\mathbb{R}^3$ we give $\hat{e}_1$, $\hat{e}_2$ and $\hat{e}_3$ special notations: $\hat{x}$, $\hat{y}$ and $\hat{z}$, respectively (obviously $\hat{z}$ doesn't exists in $\mathbb{R}^2$). For historical reasons, these vectors are sometimes denoted in physics textbooks as $\hat{i}$, $\hat{j}$ and $\hat{k}$.

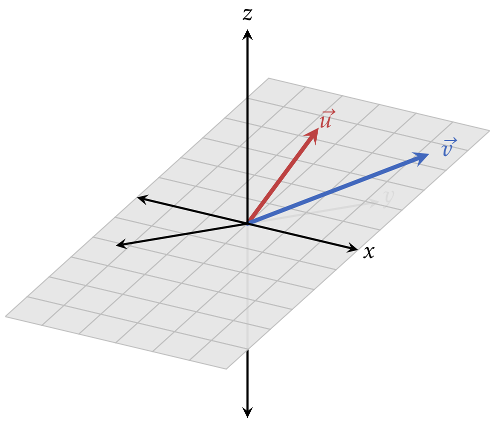


Figure 1.9 The angle between two linearly independent vectors lies on the plane spanned by the vectors.

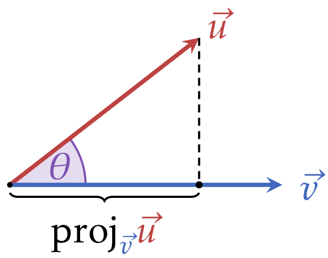


Figure 1.10 The projection of a vector $\vec{u}$ onto another vector $\vec{v}$ in the plane spanned by the two vectors.

1.1.4 The scalar product

When given two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$ it is often useful to know the angle between them: if the two vectors are linearly dependent then the angle is either $\theta = 0$ if they point in the same direction, or $\theta = \pi$ if the point in opposite directions (remember: we measure angles in radians). Otherwise, the angle θ can take any value in $(0, \pi)$. Angles are always measured on a plane, and in the case of two linearly independent vectors that plane is of course the one spanned by the two vectors (Figure 1.9).

If considering only the plane the vectors span, we can rotate it such that one of the vectors, say $\vec{u}$, lies horizotally (see Figure 1.10). We then drop a perpendicular line from the head of the $\vec{u}$ to the horizontal vector $\vec{v}$. We call the length from the origin to the intersection point of $\vec{v}$ and the perpendicular line the **projection** of $\vec{u}$ onto $\vec{v}$, and denote it as $\text{proj}_{\vec{v}}\vec{u}$.

Since the origin, the head of $\vec{u}$ and the intersection point of the perpendicular line with $\vec{v}$ form a right triangle, using basic trigonometry we find that the cosine of the angle θ is

$$\cos(\theta) = \frac{\text{proj}_{\vec{v}}\vec{u}}{\|\vec{u}\|}. \tag{1.1.26}$$

We can now use this construct to define a product between $\vec{u}$ and $\vec{v}$: their **scalar product**. We define it as following:

$$\vec{u}\cdot\vec{v} = \text{proj}_{\vec{v}}\vec{u}\cdot\left\|\vec{v}\right\|. \tag{1.1.27}$$

Subtituting Equation 1.1.26 into Equation 1.1.27 gives a very nice relation between the scalar product of two vectors and the angle between them:

$$\cos(\theta) = \frac{\vec{u}\cdot\vec{v}}{\left\|\vec{u}\right\|\left\|\vec{v}\right\|}. \tag{1.1.28}$$

The angle between the two vectors is then isolated by applying the arccos function on the right-hand side of Equation 1.1.28. A common form of this equation is the following:

$$\vec{u}\cdot\vec{v} = \left\|\vec{u}\right\|\left\|\vec{v}\right\|\cos(\theta). \tag{1.1.29}$$

Note that the scalar product returns a number, i.e. in the terms of linear algebra - a scalar, and hence its name. Since it is commonly denoted with a dot between the two vectors, it is sometimes refered to as the **dot product**. A common notation for the scalar product is the so-called **bracket notation**:

$$\langle\vec{a},\vec{b}\rangle.$$

Sometimes the comma in the notation is replaced by a vertical separator line:

$$\langle\vec{a}\,|\,\vec{b}\rangle.$$

This notation is very common in physics, and especially quantum physics where it is very useful and helps in simplifying many calculations. This will be discussed in more details in chapter/section TBD.

Later in the section we will examine some common properties of the scalar product, and see how we can calculate it directly from the vectors in their column form. Beofre we do that, let's use what we learned about the scalar product so far to solve some easy problems in the examples below.

Example 1.12 Angle between two vectors

Find the scalar product of the vectors

$$\vec{a} = \begin{bmatrix}1\\1\end{bmatrix},\ \vec{b} = \begin{bmatrix}-1\\1\end{bmatrix}.$$

Solution:
As seen in Example 1.11, the angle between $\vec{a}$ and $\vec{b}$ is $\frac{\pi}{2}$. Therefore, their scalar product is

$$\begin{aligned}\vec{a}\cdot\vec{b} &= \left\|\vec{a}\right\|\left\|\vec{b}\right\|\cos(\theta)\\ &= \sqrt{2}\sqrt{2}\cos\left(\frac{\pi}{2}\right)\end{aligned}$$

$$= 2 \cdot 0 = 0.$$

Example 1.13 Scalar product of two vectors

Calculate the scalar product of the two vectors $\vec{u} = \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$, $\vec{v} = \begin{bmatrix} -1 \\ 0 \\ 2 \end{bmatrix}$, given that the angle between them is $\theta \approx 2.069 \approx 118.561°$.

Solution:
The norms of the two vectors are

$$\left\|\vec{u}\right\| = \sqrt{2^2 + 3^2 + (-1)^2} = \sqrt{4 + 9 + 1} = \sqrt{14} \approx 3.742,$$
$$\left\|\vec{v}\right\| = \sqrt{(-1)^2 + 0^2 + 2^2} = \sqrt{1 + 4} = \sqrt{5} \approx 2.236.$$

Therefore, their scalar product is

$$\vec{u} \cdot \vec{v} \approx \sqrt{14}\sqrt{5}\cos(2.069) \approx -4.$$

The scalar product of any two vectors $\vec{u}, \vec{v}$ has two important properties:

- It is commutative, i.e. $\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u}$.
- Scalars can be taken out of the product, i.e. $(\alpha \vec{v}) \cdot \vec{u} = \vec{v} \cdot (\alpha \vec{u}) = \alpha (\vec{u} \cdot \vec{v})$.
- It equals zero in only one of two cases:
  1.1. One of the vectors (or both) is the zero vector, or
  1.2. The angle θ between the vectors is $\frac{\pi}{2}$, since then $\cos(\theta) = \cos\left(\frac{\pi}{2}\right) = 0$.

When the angle between two vectors is $\frac{\pi}{2}$ (remember: this is equivalent to 90°), we say that the two vectors are **orhtogonal** to eacth other. Note that in the special case of 2- and 3-dimensional we say that the vectors are **perpendicular** to each other.

This is such an important fact that we will put effort into framing it nicely, so you (the reader) could memorize it well. How well should you memorize this? Such that if someone wakes you up in the middle of the night and asked you, you could easily repeat it[1].

[1] For a humble fee, I'm willing to do this - just write me an email and we can discuss the terms ;)

$$\vec{u} \cdot \vec{v} = 0$$

$$\Updownarrow$$

$\vec{u}$ and $\vec{v}$ are orthogonal

Calculating the scalar product of two vectors in $\mathbb{R}^n$ using their column form is extremely straight-forward: it is nothing more than the sum of the component-wise product of the two vectors, i.e. given

$$\vec{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}, \ \vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix},$$

the scalar product $\vec{u} \cdot \vec{v}$ is

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{k=1}^{n} u_i v_i. \tag{1.1.30}$$

Example 1.14 Angle between two vectors

Calculate the scalar product of the two vectors $\vec{a} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ using the above formula (Equation 1.1.30).

Solution:
We simply substitute $\vec{a}$ and $\vec{b}$ into the equation:

$$\vec{a} \cdot \vec{b} = 1 \cdot (-1) + 1 \cdot 1 = -1 + 1 = 0,$$

which is exactly the result we got using the previous method.

Example 1.15 Scalar product of two vectors - algebraicly

Calculate the scalar product $\vec{u} \cdot \vec{v}$ from Example 1.13 using Equation 1.1.30.

Solution:

$$\vec{u}\cdot\vec{v} = 2\cdot(-1) + 3\cdot 0 + (-1)\cdot 2 = -2 - 2 = -4,$$

exactly the result we got in Example 1.13.

The norm of a vector can be calculated using the scalar product: given a vector $\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$,

$$\vec{v}\cdot\vec{v} = v_1 v_1 + v_2 v_2 + \cdots + v_n v_n = v_1^2 + v_2^2 + \cdots + v_n^2 = \left\|\vec{v}\right\|^2. \tag{1.1.31}$$

We therefore usually define the norm in terms of the scalar product:

$$\left\|\vec{v}\right\| = \sqrt{\vec{v}\cdot\vec{v}}. \tag{1.1.32}$$

This might seem unconsequential at the moment, but it will become very useful when we generalize linear algebra to more abstract vector spaces (Chapter 3).

Any vector can be **decomposed** into its projections on n orthogonal directions. In fact, this is exactly what we do when we write a vector as a linear combination of the vectors of an orthogonal basis: consider for example the vector

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}.$$

It can be written as the linear combination

$$\vec{v} = v_1\hat{e}_1 + v_2\hat{e}_2 + \cdots + v_n\hat{e}_n = \sum_{i=1}^{n} v_i\hat{e}_i,$$

where in turn any element v_i is the projection of $\vec{v}$ on the basis vector $\hat{e}_i$:

$$v_i = \mathrm{proj}_{\hat{e}_i}\vec{v}, \tag{1.1.33}$$

and thus the component $v_i\hat{e}_i = \left(\mathrm{proj}_{\hat{e}_i}\vec{v}\right)\hat{e}_i$ is itself a vector of norm v_i pointing at the direction $\hat{e}_i$. In general, given an orthogonal basis set $B = \left\{\vec{b}_1,\ \vec{b}_2,\ \cdots,\ \vec{b}_n\right\}$, any vector in $\mathbb{R}^n$ can be decomposed as follows:

$$\vec{v} = \sum_{i=1}^{n}\left(\mathrm{proj}_{\hat{b}_i}\vec{v}\right)\hat{b}_i. \tag{1.1.34}$$

In the case where B is an orthonormal basis set, we know that each of its vector is a unit vector (i.e. $\left\|\vec{b}_i\right\| = 1$), and using Equation 1.1.27 we can re-write Equation 1.1.34 as

$$\vec{v} = \sum_{i=1}^{n}\left(\vec{v}\cdot\hat{b}_i\right)\hat{b}_i. \tag{1.1.35}$$

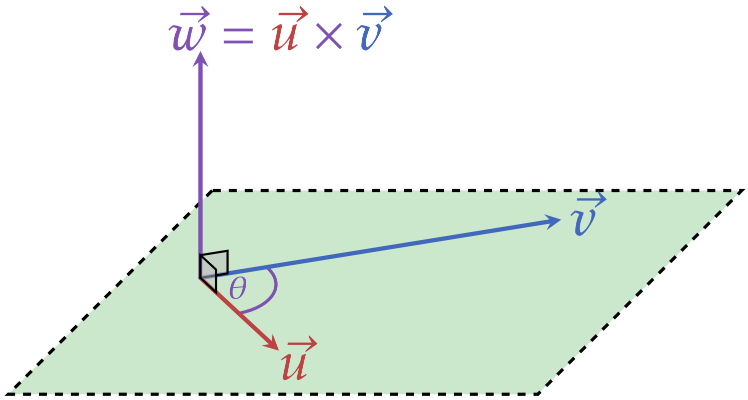


Figure 1.11 The cross product of the vectors $\vec{u}$ and $\vec{v}$ relative to the plane spanned by the two vectors.

Example 1.16 Decomposing a vector

EXAMPLE TBD

1.1.5 The cross product

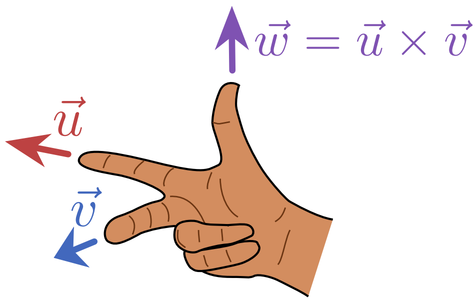
Another commonly used product of two vectors is the so-called **cross product**. Unlike the scalar product, it is only really valid in $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^7$, of which we will focus on $\mathbb{R}^3$ and touch a bit on its uses in $\mathbb{R}^2$. Also in contrast to the scalar product, the cross product in $\mathbb{R}^3$ results in a vector rather than a scalar - therefore the product is sometimes known as the **vector product**. The cross product uses the notation $\vec{a} \times \vec{b}$, from which it derives its name.

Geometrically, the cross product of two vectors $\vec{u}, \vec{v} \in \mathbb{R}^3$ is defined as a vector $\vec{w} \in \mathbb{R}^3$ which is **orthogonal to both** $\vec{u}$ and $\vec{v}$, and has a magnitude

$$\left\|\vec{w}\right\| = \left\|\vec{u}\right\| \left\|\vec{v}\right\| \sin(\theta), \tag{1.1.36}$$

where, like with the scalar product, θ is the angle between $\vec{u}$ and $\vec{v}$.

The direction of $\vec{u} \times \vec{v}$ is determined by the **right-hand rule**: using a person's right hand, when $\vec{u}$ points in the direction of their index finger and $\vec{v}$ points in the direction of their middle finger, then vector $\vec{w} = \vec{u} \times \vec{v}$ points in the direction of their thumb:



The cross product is **anti-commutative**, i.e. changing the order of the vectors results in inverting the product:

$$\vec{u} \times \vec{v} = -(\vec{v} \times \vec{u}).$$

When the vectors are given as column vectors $\vec{u} = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}$, $\vec{v} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$, the resulting cross product is

$$\vec{u} \times \vec{v} = \begin{pmatrix} u_y v_z - u_z v_y \\ u_z v_x - u_x v_z \\ u_x v_y - u_y v_x \end{pmatrix} \tag{1.1.37}$$

Note 1.5 The cross product of the standard basis vectors

The cross product of two of the standard basis vectors in $\mathbb{R}^3$ is the third basis vector. Its sign ($\pm$) is determined by a cyclic rule:

$$\operatorname{sign}\left(\hat{e}_i \times \hat{e}_j\right) = \begin{cases} 1 & \text{if } (i,j) \in \{(1,2),\ (2,3),\ (3,1)\}, \\ -1 & \text{if } (i,j) \in \{(3,2),\ (2,1),\ (1,3)\}, \\ 0 & \text{otherwise.} \end{cases}$$

!

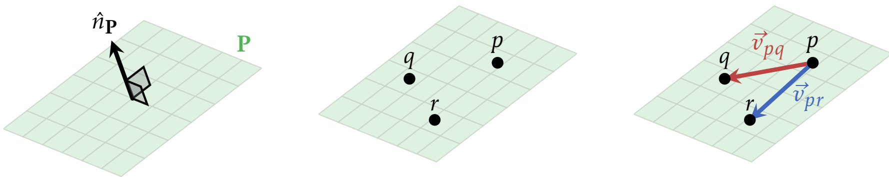
Challange 1.3 Orthogonalily of the cross product

Using component calculation and utilizing the dot product, show that $\vec{a} \times \vec{v}$ is indeed orthogonal to both $\vec{a}$ and $\vec{b}$.

?

1.1.6 Normal vectors

A special kind of vector in $\mathbb{R}^3$ is the so-called **normal vector** to a plane **P**: this vector, usually denoted as $\hat{n}_{\mathbf{P}}$, is pointing at the orthogonal direction to any vector of the plane (see XXX). Given one knows three points on the plane, its normal vector can be calculated: say the following three points in **P** are given (for visualizing the following



1 The normal vector to **P**. 2 Finding three points on the plane. 3 Finding two vectors on the plane.

Figure 1.12 A normal vector $\hat{n}_{\mathbf{P}}$ to the plane **P**.

steps see YYY):

$$
\begin{aligned}
p &= (p_x, p_y, p_z) \\
q &= (q_x, q_y, q_z) \\
r &= (r_x, r_y, r_z),
\end{aligned} \tag{1.1.38}
$$

We can get two vectors lying on the plane by first considering the points as vectors, i.e.

$$
\vec{p} = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix}, \ \vec{q} = \begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}, \ \vec{r} = \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix}. \tag{1.1.39}
$$

Then, we calculate two vectors on the plane by subtraction, e.g.

$$
\begin{aligned}
\vec{v}_{pq} &= \vec{q} - \vec{p} = \begin{bmatrix} q_x - p_x \\ q_y - p_y \\ q_z - p_z \end{bmatrix}, \\
\vec{v}_{pr} &= \vec{r} - \vec{p} = \begin{bmatrix} r_x - p_x \\ r_y - p_y \\ r_z - p_z \end{bmatrix}.
\end{aligned} \tag{1.1.40}
$$

The normal vector $\hat{n}_{\boldsymbol{p}}$ must be orthogonal to both $\vec{v}_{pq}$ and $\vec{v}_{pr}$ - and so we use the cross product to find its direction:

$$
\vec{n}_{\mathbf{P}} = \vec{v}_{pq} \times \vec{v}_{pr} = \begin{bmatrix} (q_y - p_y)(r_z - p_z) - (r_y - p_y)(q_z - p_z) \\ (p_x - q_x)(r_z - p_z) - (r_x - p_x)(q_z - p_z) \\ (q_x - p_x)(r_y - p_y) - (r_x - p_x)(p_y - q_y) \end{bmatrix}. \tag{1.1.41}
$$

Normalizing $\vec{n}_{\mathbf{P}}$ will then yield the normal vector $\hat{n}_{\mathbf{P}}$[2].

[2] I leave this as a challange to the reader, because I'm lazy.

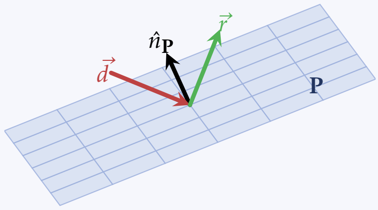
Note 1.6 Sign of normal vectors

The vector $\vec{m} = -\hat{n}_{\mathbf{P}}$ has all the properties of $\hat{n}_{\mathbf{P}}$, and is indeed a normal vector to $\mathbf{P}$. The choice of which of the two vectors to use depends on the application. For now, we do not elaborate on this further.

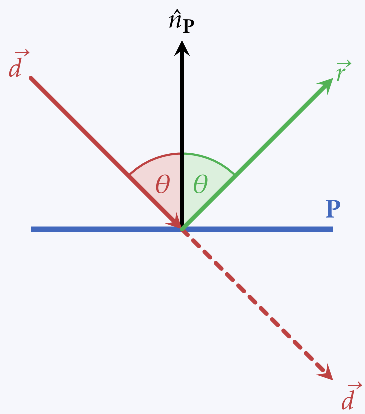
To wrap up the vectors section, we present and solve a single problem in the following example.

Example 1.17 Reflection of light rays

A ray light hits a mirror, modelled by the plane $\mathbf{P}$ which is defined by the normal vector $\hat{n}_{\mathbf{P}}$. The direction of the light ray is given by $\vec{d}$. What is the direction of the reflected light ray $\vec{r}$? Recall that both the incident and reflected rays are at the same angle in respect to the normal vector of $\hat{n}_{\mathbf{P}}$, and that the incident ray lie on the plane defined by $\vec{d}$ and $\hat{n}_{\mathbf{P}}$.



We can rotate our viewpoint of the problem, looking at $\mathbf{P}$ from the side and in such a way that we look head-on at the plane spanned by $\hat{n}_{\mathbf{P}}$ and $\vec{d}$:



(the dashed red vector in the above figure represents the vector incident ray, $\vec{d}$, moved such that its origin lies at the origin of the other vectors)

As with any vector, we can decompose $\vec{d}$ to its projections on the vectors of an orthonormal basis set (Equation 1.1.35). Since we reduced the problem to two dimensions, we need a basis of two orthonormal directions: we choose one to be $\hat{n}_{\mathbf{P}}$, and the other orthogonal to it (in the figure above it is in the horizontal direction) which we call $\hat{p}$. The decomposition of $\vec{d}$ then reads:

$$\vec{d} = \left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}} + \left(\vec{d} \cdot \hat{p}\right) \hat{p}.$$

Since there are only two vectors in the basis set $\{\hat{n}_{\mathbf{P}}, \hat{p}\}$, we can actually write the component $\left(\vec{d} \cdot \hat{p}\right) \hat{p}$ as $\vec{d} - \left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}}$, yielding a rather silly looking expression for $\vec{d}$:

$$\vec{d} = \left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}} + \left[\vec{d} - \left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}}\right].$$

However, in closer inspection the above expression is not at all silly, and is actually very similar to the reflected vector $\vec{r}$: since they are both of same norm and oposing directions with respect to the direction $\hat{n}_{\mathbf{P}}$, we can write $\vec{r}$ as

$$\vec{r} = -\left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}} + \left[\vec{d} - \left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}}\right].$$

From the above expressions for $\vec{d}$ and $\vec{r}$ we can isolate an expression for $\vec{r}$ as a function of $\vec{d}$ and $\hat{n}_{\mathbf{P}}$:

$$\begin{aligned}\vec{r} &= d - \left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}} - \left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}} \\ &= d - 2\left(\vec{d} \cdot \hat{n}_{\mathbf{P}}\right) \hat{n}_{\mathbf{P}}.\end{aligned}$$

1.2 LINEAR TRANSFORMATIONS

In the previous section we introduced real vectors and their most important properties. In this section we explore a special set of operations that can act on vectors, namely **linear transformations**.

Linear transformations are a special set of transformations that are relatively easy to analyze. Their use is extremely widespread all throughout different fields of mathematics and its application, e.g. (just to name a few): in computer graphics, machine learning, quantum physics and engineering.

1.2.1 Definition

As mentioned in Chapter 0, a “transformation“ is simply another name for a function. Thus in our context, linear transformations are some functions that act on vectors: a linear transformation T takes a vector as an input, and outputs another vector, possibly of a different dimension, i.e.

$$T : \mathbb{R}^n \to \mathbb{R}^m. \tag{1.2.1}$$

What makes linear transformations more "special" than other functions is their property of **linearity**, which entails the following two properties:

- Scalability: for any scalar α and vector $\vec{v}$,

$$T(\alpha\vec{v}) = \alpha T(\vec{v}).$$

- Additivity: for any two vectors $\vec{u}, \vec{v}$

$$T(\vec{u}+\vec{v}) = T(\vec{u}) + T(\vec{v}).$$

Example 1.18 A linear transformation

Claim: the following $\mathbb{R}^3 \to \mathbb{R}$ transformation is linear:

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = 2x + 3y - z.$$

Proof: We can show this using the properties of linear transformations.

- Scalability: given a scalar $\alpha \in \mathbb{R}$,

$$T\left(\begin{bmatrix} \alpha x \\ \alpha y \\ \alpha z \end{bmatrix}\right) = 2(\alpha x) + 3(\alpha y) - (\alpha z) = \alpha(2x + 3y - z) = \alpha T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right).$$

- Additivity: given two vectors $\vec{u} = \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$,

$$\begin{aligned} T\left(\begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} + \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}\right) &= T\left(\begin{bmatrix} u_x + v_x \\ u_y + v_y \\ u_z + v_z \end{bmatrix}\right) \\ &= 2(u_x + v_x) + 3\left(u_y + v_y\right) - (u_z + v_z) \\ &= T\left(\begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}\right) + T\left(\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}\right). \end{aligned}$$

Example 1.19 A non-linear transformation

Claim: the following $\mathbb{R}^3 \to \mathbb{R}$ transformation is **not** a linear transformation:

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = 2x^2 + 3y - z.$$

Proof: this time we only need to show a single case where linearity breaks - let's choose *scalability*. Given the vector $\vec{v} = \begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix}$, on one hand

$$T\left(\alpha \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}\right) = 2\left(\alpha u_x\right)^2 + 3\alpha u_y - \alpha u_z = 2\alpha^2 u_x^2 + 3\alpha u_y - \alpha u_z.$$

On the other hand

$$\alpha T\left(\begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix}\right) = \alpha\left(2u_x^2 + 3u_y - u_z\right) = 2\alpha u_x^2 + 3\alpha u_y - \alpha u_z.$$

For any $a \notin \{0, 1\}$ we get that $T\left(\alpha\vec{v}\right) \neq \alpha T\left(\vec{v}\right)$. Therefore, T is not linear.

1.2.2 Developing intuition

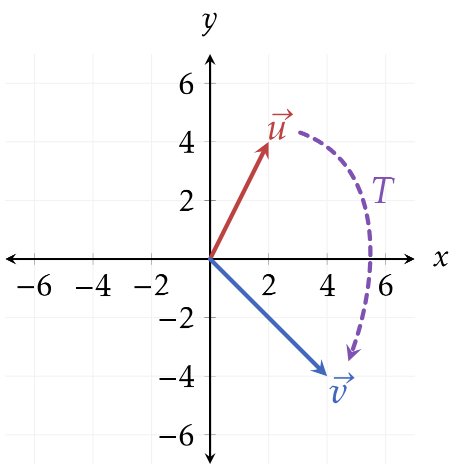
Before moving on to explore the algebraic properties of linear transformations, we first shift our focus to gain some intuition about them. Much like in the last section, we do this using graphical representations of linear transformations in $\mathbb{R}^2$ and $\mathbb{R}^3$. We start with a single vector under transformation: let $\vec{u} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ and $T : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} 2x \\ -y \end{bmatrix}. \tag{1.2.2}$$

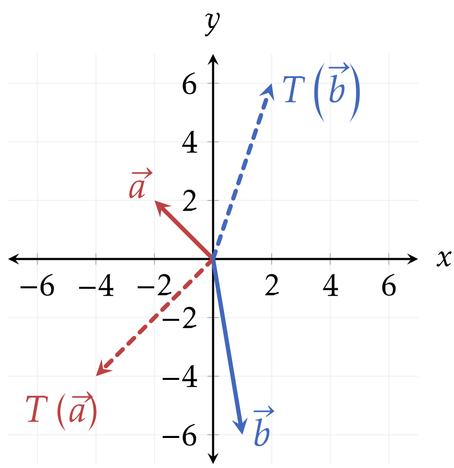
(to the reader: verify that this transformation is indeed linear)

Applying T to $\vec{u}$ yields the vector $\vec{v} = \begin{bmatrix} 4 \\ -4 \end{bmatrix}$ (see Figure 1.131), i.e. it scales the x-component of $\vec{u}$ by 2 and flippes over its y-component.

If we take other vectors, e.g. $\vec{a} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$ and $\vec{b} = \begin{bmatrix} 1 \\ -6 \end{bmatrix}$ we see that T transforms them in the exact same manner: it scales their x-components by 2 and flipps over their y-components (Figure 1.132). This is a fundamental aspect of linear transformations: they always transform all vectors in the exact same manner. We can use this fact to help visualize transformations, by looking at how they transform the entire space. For example, we can draw all grid lines and observe how they are transformed.



1 The vector $\vec{u} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$ is transformed by T yielding the vector $\vec{v} = \begin{bmatrix} 4 \\ -4 \end{bmatrix}$.



2 The vectors $\vec{a} = \begin{bmatrix} -2 \\ 2 \end{bmatrix}$ and $\vec{b}$ are transformed by the same T.

In Figure 1.14 a schematic of $\mathbb{R}^2$ is shown before and after the application of a linear transformation T, by placing a transformed grid (blue) ontop of an untouched grid (gray). In this view, one can see how each point in space is transformed: assuming for example that each two adjacent grid points are 1 unit apart, the gray point at $(-2, 2)$ is transformed to where the blue point is, i.e. $(-1, 1)$ when measured using the original axes.

For comparison, Figure 1.15 shows a non linear transformation applied to $\mathbb{R}^2$.

Figure 1.14 shows some important properties of linear transformation (cf. Figure 1.15):

1.1. The origin stays at the same place after application of the transformation, i.e. $T\left(\vec{0}\right) = \vec{0}$.

1.2. Parallel lines remain parallel after application of the transformation.

1.3. All areas are scaled by the same amount.

It is rather easy to prove the first two properties.

Proof 1.5 Two properties of linear transformations

1.1. Let T be a transformation that does not perserve the origin, i.e.

$$T\left(\vec{0}\right) = \vec{v} \neq \vec{0}.$$

We can scale $\vec{0}$ by a scalar $\alpha \neq 0$, which would yield

$$T\left(\alpha\vec{0}\right) = T\left(\vec{0}\right) = \vec{v}.$$

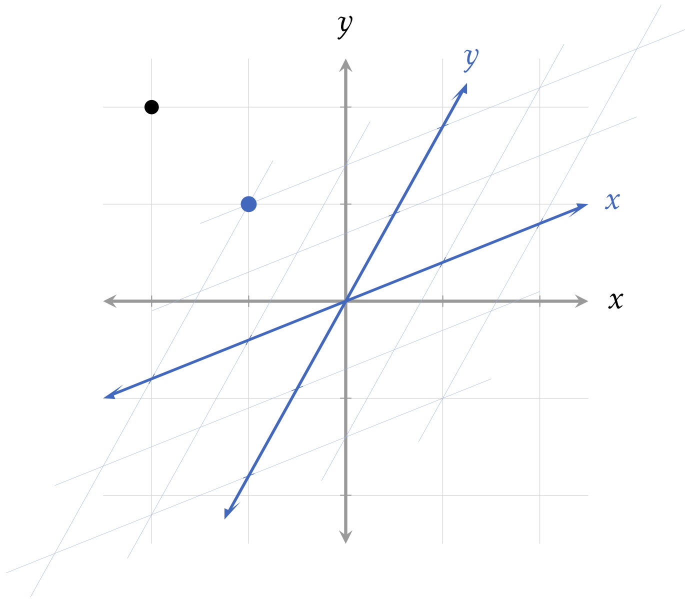


Figure 1.14 $\mathbb{R}^2$ after application of a linear transformation (blue), placed ontop of $\mathbb{R}^2$ before the transformation (gray). Note the black point at the top left at $(-2,2)$ transforming into the blue point at $(-1,1)$.

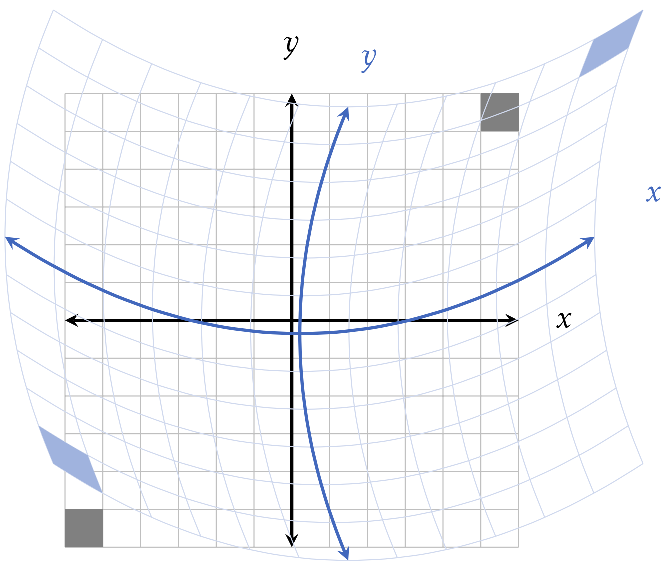


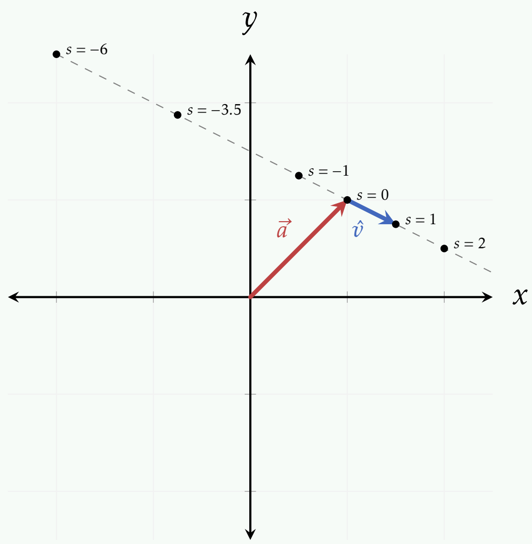
Figure 1.15 A non linear transformation applied to $\mathbb{R}^2$ for comparison.

However, for T to be linear we expect (due to scalability)

$$T\left(\alpha\vec{0}\right) = \alpha\vec{v},$$

but since $\alpha \neq 0$ and $\vec{v} \neq \vec{0}$ this does not happen. Therefore, T can not be linear - and in turn linear transformations must preserve the origin.

1.2. A line is defined using a point $\vec{a}$, and a direction $\hat{v}$ as the set of all the points $\{x = \vec{a} + s\hat{v}, s \in \mathbb{R}\}$:



Parallel lines have the same direction $\hat{s}$, i.e. $x_1 = \vec{a}_1 + s_1\hat{v}$ and $x_2 = \vec{a}_2 + s_2\hat{v}$ are parallel lines. Applying a linear transformation T to these lines yields (using the two defining properties of linear transformations)

$$\begin{aligned} T(x_1) &= T(\vec{a}_1) + s_1 T(\hat{v}), \\ T(x_2) &= T(\vec{a}_2) + s_2 T(\hat{v}). \end{aligned}$$

We can see that the two right-hand side equations represent two new lines with the same direction, i.e. $T(\hat{v})$. Therefore parallel lines remain parallel under a linear transformation.

QED

We will prove the the third property (all areas are scaled by the same amount) later in the chapter.

All linear transformations in $\mathbb{R}^2$ can be created by composing transformations from a set of linear transformation which we will refer to as the **basic linear transformations**[3]. To visualize these basic transformations we apply them on a figure of a tapir[4]:

[3] not an official name.
[4] They are here, they are a trans tapir. Get used to it.

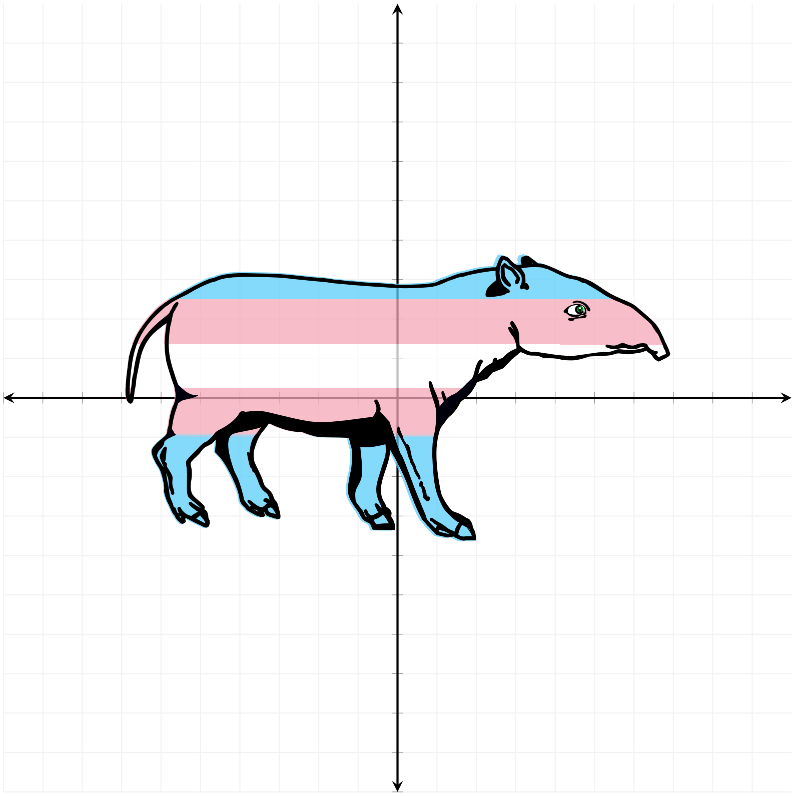
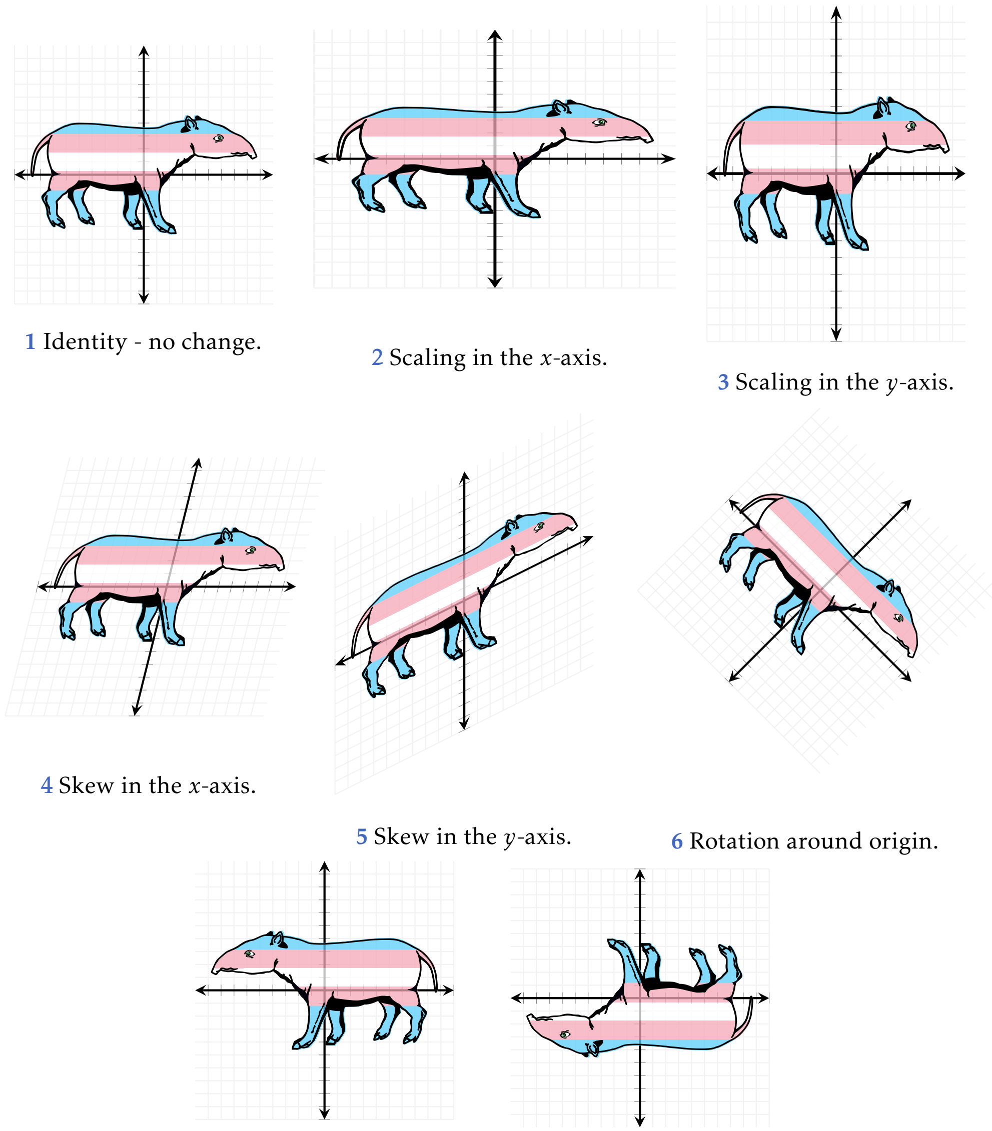


Figure 1.16 shows the basic linear transformations applied to our happy tapir.



1 Identity - no change.

2 Scaling in the x-axis.

3 Scaling in the y-axis.

4 Skew in the x-axis.

5 Skew in the y-axis.

6 Rotation around origin.

7 Reflection about line.

8 Reflection through origin.

Figure 1.16 The basic linear transformations, exemplified using a very happy tapir.

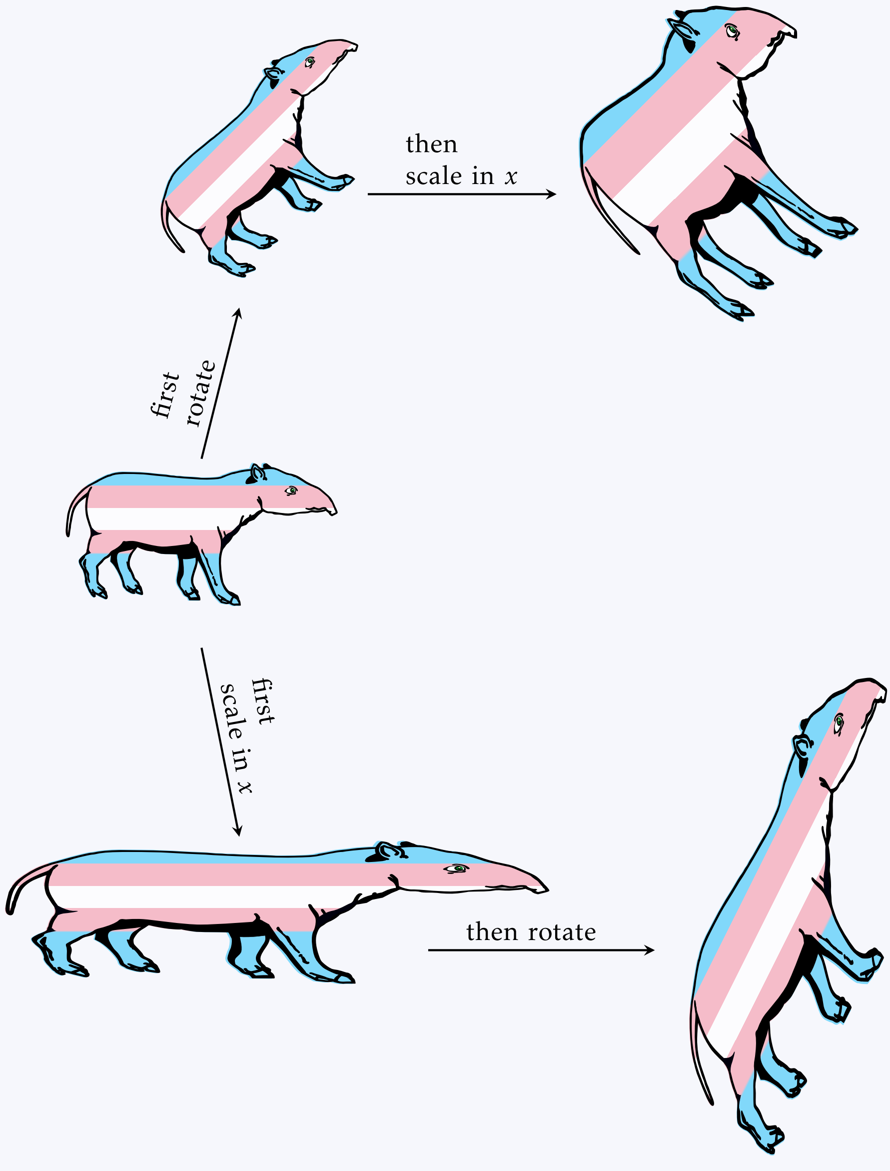
Example 1.20 Composing basic linear transformations

Given the following two linear transformations:

1.1. Scale by 1.5 in the x-direction,

1.2. Rotate by $\frac{\pi}{4}$ anti-clockwise around the origin,

two composite linear transformations can be created: first scale then rotate, and first rotate then scale. As can be seen in the figure bellow, changing the order of composition results in a different linear transformations all together:



This is not a suprising result: in Chapter 0 we learned that function composition is not a commutative operation.

Some of the basic linear transformations can be created as compositions of other basic linear transformations. For example, the composition of reflection across the y-axis followed by reflection across the x-axis results in a 180° rotation around the origin (see

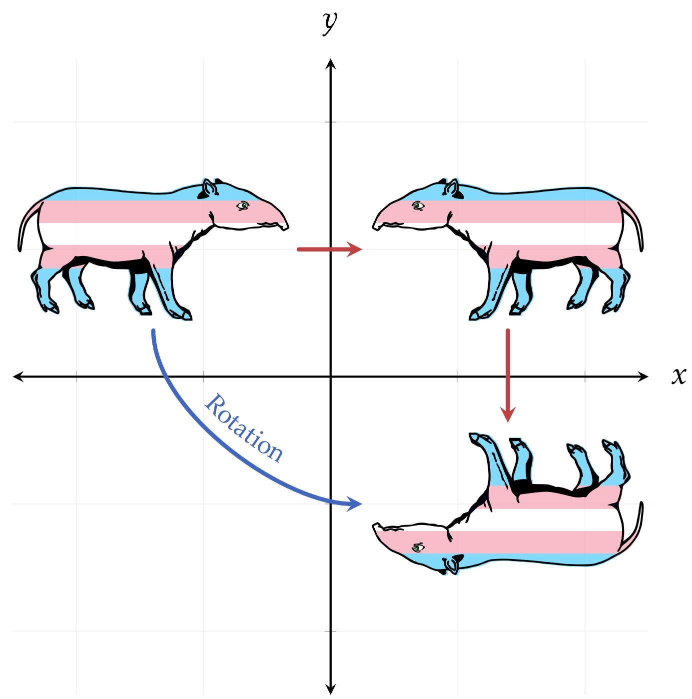


Figure 1.17 Reflection across the y-axis (red arrow) followed by a reflection across the x-axis (red arrow) results in a rotation by 180° around the origin (blue arrow).

Figure 1.17).

1.2.3 3D Linear transformations

The basic linear transformations in $\mathbb{R}^3$ are very much similar to those in $\mathbb{R}^2$ with some small differences worth mentioning. For a start, scaling and skewing can be done in three different directions instead of just two directions (namely, x, y and z instead of just x and y). In addition, there are infinetly many axes of rotation: in $\mathbb{R}^2$ there is just a single axis (actually a point) of rotation - the origin. In $\mathbb{R}^3$ any line that goes through the origin can be an axis of rotation (see Figure 1.18). Lastly, there are three types of reflections: about the origin (a point), across a line going through the origin, and across a plane (see fig?).

1.3 MATRICES

In the previous section we described linear transformations in a rather abstract way: what they are, how they behave qualitatively and how they look like in 2- and 3-dimensions. In this section we introduce a numerical method of representing linear

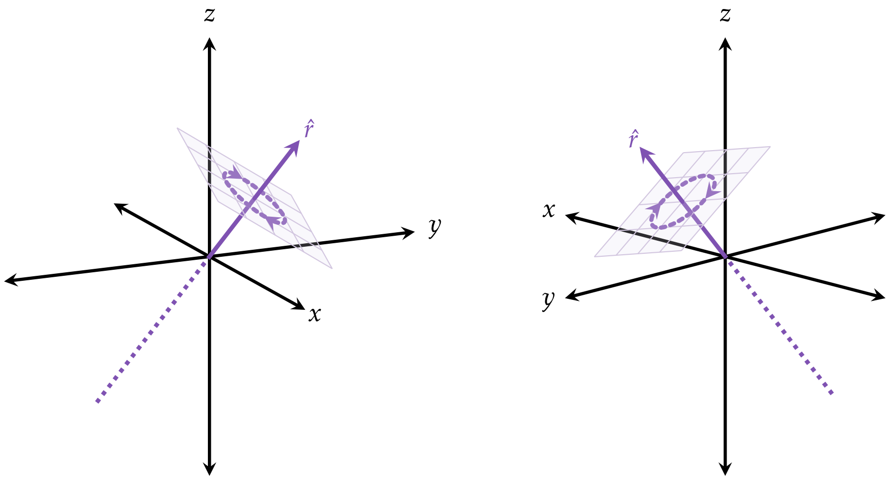


Figure 1.18 Rotation around a direction $\hat{r}$ from two different viewpoints, showing an example of a plane orthogonal to $\hat{r}$ on which the rotation happens.

transformations: matrices.

1.3.1 Linear transformation of basis vectors

Recall that any vector $\vec{v} \in \mathbb{R}^n$ can be written as a linear combination of basis vectors $\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n$:

$$\vec{v} = \sum_{i=1}^{n} \alpha_i \vec{b}_i = \alpha_1 \vec{b}_1 + \alpha_2 \vec{b}_2 + \cdots + \alpha_n \vec{b}_n. \tag{1.3.1}$$

Applying a linear transformation T on $\vec{v}$ yields, using the properties of linear transformations,

$$\begin{aligned} T(\vec{v}) &= T\left(\alpha_1 \vec{b}_1 + \alpha_2 \vec{b}_2 + \cdots + \alpha_n \vec{b}_n\right) \\ \text{additivity} \longrightarrow &= T\left(\alpha_1 \vec{b}_1\right) + T\left(\alpha_2 \vec{b}_2\right) + \cdots + T\left(\alpha_n \vec{b}_n\right) \\ \text{scalability} \longrightarrow &= \alpha_1 T\left(\vec{b}_1\right) + \alpha_2 T\left(\vec{b}_2\right) + \cdots + \alpha_n T\left(\vec{b}_n\right). \end{aligned} \tag{1.3.2}$$

This result is pretty neat: it means that by knowing how a linear transformation T changes the basis vectors, we know exactly how any vector is transformed by T. This true for any basis, and thus specifically to the standard basis, where the coefficients $\alpha_1, \alpha_2, \dots, \alpha_n$ are actually the components of the vector, i.e. $v_1, v_2, \dots, v_n$. Thus in the standard basis:

$$T(\vec{v}) = v_1 T(\hat{e}_1) + v_2 T(\hat{e}_2) + \cdots + v_n T(\hat{e}_n). \tag{1.3.3}$$

Example 1.21 Vector transformation via a basis

Applying the transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$, defined as

$$T\left(\begin{bmatrix} x \\ y \\ z \end{bmatrix}\right) = \begin{bmatrix} x+y-2z \\ 2x+z \\ -x-y-z \end{bmatrix}$$

on the vector $\vec{v} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}$ yields the following vector:

$$T(\vec{v}) = T\left(\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}\right) = \begin{bmatrix} 2+(-1)-2\cdot 3 \\ 2\cdot 2+3 \\ -2-(-1)-3 \end{bmatrix} = \begin{bmatrix} 2-1-6 \\ 4+3 \\ -2+1-3 \end{bmatrix} = \begin{bmatrix} -5 \\ 7 \\ -4 \end{bmatrix}.$$

Now, let us apply T first to the three standard basis vectors $\hat{x}, \hat{y}, \hat{z}$:

$$T(\hat{x}) = T\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 1+0-2\cdot 0 \\ 2\cdot 1+0 \\ -1-0-0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix},$$

$$T(\hat{y}) = T\left(\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right) = \begin{bmatrix} 0+1-2\cdot 0 \\ 2\cdot 0+0 \\ -0-1-0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix},$$

$$T(\hat{z}) = T\left(\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right) = \begin{bmatrix} 0+0-2\cdot 1 \\ 2\cdot 0+1 \\ -0-0-1 \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix}.$$

Taking these results and applying Equation 1.3.2 yields

$$\begin{aligned} T(\vec{v}) &= 2T(\hat{x}) - T(\hat{y}) + 3T(\hat{z}) \\ &= 2\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + 3\begin{bmatrix} -2 \\ 1 \\ -1 \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \begin{bmatrix} -6 \\ 3 \\ -3 \end{bmatrix} \\ &= \begin{bmatrix} 2-1-6 \\ 4-0+3 \\ -2-(-1)+(-3) \end{bmatrix} \\ &= \begin{bmatrix} -5 \\ 7 \\ -4 \end{bmatrix}, \end{aligned}$$

which is indeed what we got when we applied T directly to $\vec{v}$.

1.3.2 From transformations to matrices

The most general linear transformation $T:\mathbb{R}^2 \to \mathbb{R}^2$ has the following form:

$$T\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}, \tag{1.3.4}$$

where $a,b,c,d \in \mathbb{R}$. If we apply this transformation to $\hat{x}$ and $\hat{y}$ we get, respectively,

$$T(\hat{x}) = \begin{bmatrix} a \\ c \end{bmatrix}, \quad T(\hat{y}) = \begin{bmatrix} b \\ d \end{bmatrix}. \tag{1.3.5}$$

We can now collect these two vectors to form a new structure, which we call a **matrix** (in this specific casr a 2×2 matrix):

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}. \tag{1.3.6}$$

We then define the product of M with a vector $\vec{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ to yield $T(\vec{v})$, i.e.

$$M\vec{v} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax+by \\ cx+dy \end{bmatrix}. \tag{1.3.7}$$

This defintion can be re-written as following:

$$M\vec{v} = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} M_1 \cdot \vec{v} \\ M_2 \cdot \vec{v} \end{bmatrix}, \tag{1.3.8}$$

i.e. the i-th component of the resulting vector is the scalar product of the i-th **row** of the matrix with the vector $\vec{v}$.

Example 1.22 Matrix-vector product

Some matrix-vector products:

$$\begin{bmatrix} 1 & -2 \\ 0 & 5 \end{bmatrix}\begin{bmatrix} -3 \\ 2 \end{bmatrix} = \begin{bmatrix} 1\cdot(-3)+(-2)\cdot 2 \\ 0\cdot(-3)+5\cdot 2 \end{bmatrix} = \begin{bmatrix} -7 \\ 10 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 5 \\ -4 \end{bmatrix} = \begin{bmatrix} 1\cdot 5+2\cdot(-4) \\ 1\cdot 5+2\cdot(-4) \end{bmatrix} = \begin{bmatrix} -3 \\ -3 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ -2 \end{bmatrix} = \begin{bmatrix} 2\cdot 2+0\cdot(-2) \\ 0\cdot 2+2\cdot(-2) \end{bmatrix} = \begin{bmatrix} 4 \\ -4 \end{bmatrix}.$$

Challange 1.4 Proof of linearity

Prove that the transformation T in Equation 1.3.4 is indeed linear.

?

The most general form of a linear transformation is $T : \mathbb{R}^n \to \mathbb{R}^m$, i.e. a transformation which takes n-dimensional vectors as input and returns m-dimensional vectors as output:

$$\mathbb{R}^n \ni \qquad T\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}, \qquad \in \mathbb{R}^m \tag{1.3.9}$$

where $a_{ij} \in \mathbb{R}$, $i = 1, 2, 3, \dots, m$ and $j = 1, 2, 3, \dots, n$.

Challange 1.5 Proof of linearity

Prove that the above transformation T is indeed linear.

?

Respectively, we define an $m \times n$ matrix (m rows by n columns) by collecting all the coefficients a_{ij} into a single structure:

$$M = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}. \tag{1.3.10}$$

The product $M\vec{v}$ (where $\vec{v} \in \mathbb{R}^n$) is then defined as

$$M\vec{v} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{bmatrix}. \tag{1.3.11}$$

Again, note that the i-th component of the resulting vector is the scalar product $M_i \cdot \vec{v}$.

Note 1.7 When is a matrix-vector product defined

In order for a matrix-vector product to be defined, the vector must be of the same dimension as the number of **columns** in the matrix - i.e. given an $a \times b$ matrix, a vector must be b-dimensional for the product to be defined.

!

Example 1.23 Some matrix-vector products

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Etiam lobortis facilisis sem. Nullam nec mi et neque pharetra sollicitudin. Praesent imperdiet mi nec ante. Donec ullamcorper, felis non sodales commodo, lectus velit ultrices augue, a dignissim nibh lectus placerat pede. Vivamus nunc nunc, molestie ut, ultricies vel, semper in, velit. Ut porttitor. Praesent in sapien. Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Duis fringilla tristique neque. Sed interdum libero ut metus. Pellentesque placerat. Nam rutrum augue a leo. Morbi sed elit sit amet ante lobortis sollicitudin. Praesent blandit blandit mauris. Praesent lectus tellus, aliquet aliquam, luctus a, egestas a, turpis. Mauris lacinia lorem sit amet ipsum. Nunc quis urna dictum turpis accumsan semper.

The structure of an $m \times n$ matrix M has a nice property: given that the transformation in represented in some basis $B = \left\{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n\right\}$, the i-th column of the matrix always shows how $\vec{b}_i$ is transformed by the product $M\vec{b}_n$. This is easy to see in the case of the standard basis, which we anyway use througout this chapter:

$$M = \begin{array}{c} \begin{array}{cccc} T(\hat{e}_1) & T(\hat{e}_2) & & T(\hat{e}_n) \end{array} \\ \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \end{array}.$$

Example 1.24 Matrices

The product of the following matrix M with each of the vectors $\hat{e}_1, \hat{e}_2, \hat{e}_3$ (i.e. $\hat{x}, \hat{y}$ and $\hat{z}$, respectively) returns the respective column of the matrix:

$$M\hat{e}_1 = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 3 & 4 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \cdot 1 + 2 \cdot 0 + 0 \cdot 0 \\ -1 \cdot 1 + 3 \cdot 0 + 4 \cdot 0 \\ 0 \cdot 1 + 1 \cdot 0 + 3 \cdot 0 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix},$$

$$M\hat{e}_2 = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 3 & 4 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \cdot 0 + 2 \cdot 1 + 0 \cdot 0 \\ -1 \cdot 0 + 3 \cdot 1 + 4 \cdot 0 \\ 0 \cdot 0 + 1 \cdot 1 + 3 \cdot 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix},$$

$$M\hat{e}_3 = \begin{bmatrix} 1 & 2 & 0 \\ -1 & 3 & 4 \\ 0 & 1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1\cdot 0 + 2\cdot 0 + 0\cdot 1 \\ -1\cdot 0 + 3\cdot 0 + 4\cdot 1 \\ 0\cdot 0 + 1\cdot 0 + 3\cdot 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 3 \end{bmatrix}.$$

1.4 SYSTEMS OF LINEAR EQUATIONS

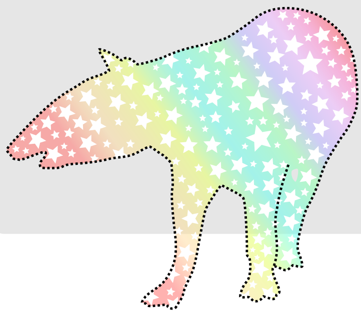
1.5 EIGENVECTORS AND EIGENVALUES

1.6 DECOMPOSITIONS

1.7 SOME REAL LIFE USES OF LINEAR ALGEBRA

1.8 EXERCISES

CHAPTER 2



CALCULUS IN 1D

2.1 SEQUENCES AND SERIES

2.2 LIMITS OF REAL FUNCTIONS

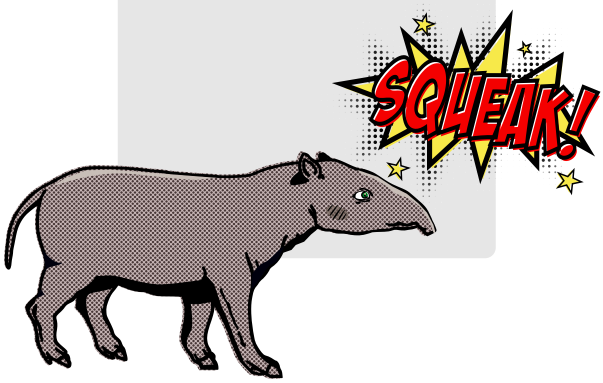
2.3 DERIVATIVES

2.4 INTEGRALS

2.5 ANALYZING REAL FUNCTIONS

CHAPTER

3



LINEAR ALGEBRA

(RIGOROUS APPROACH)

Something about formalism, theorems, proofs, etc.

Note 3.1 Why present rigorous mathematics in this book?

Rigorous mathematics is rarely necessary for those who are interested in the tools mathematics provides us with, rather than the full and deep understanding of the concepts these tools are based on. However, it can be useful to students of scientific fields to experience rigorous mathematics at least once in their course of study. Usually, the choice for the topic to be analyzed rigorously is between linear algebra and calculus - for this book the latter was chosen.

!

3.1 FIELDS

We begin our dive into the rigorous analysis of linear algebra by defining an algebraic construction call a **field**, which we need in order to properly define vector spaces later. In essence, a field has most of the important properties of the real numbers, namely the closure, commutativity, associativity, identity and inverse of addition and multiplication of any two elements in the field (except the product inverse of the field equivalent object for the number 0). In a later section we will use fields to construct the general notion of **vector spaces**.

Definition 3.1 Field

A field $\mathbb{F}$ is a set of objects together with two operations called **addition** and **multiplication** (denoted $+$ and $\cdot$, respectively), for which the following axioms hold:

- **Closure of under addition and multiplication**: for any $a, b \in \mathbb{F}$,
  1. $(a+b) \in \mathbb{F}$,
  2. $(a \cdot b) \in \mathbb{F}$.

- **Commutativity under addition multiplication**: for any $a, b \in \mathbb{F}$,
  1. $a+b=b+a$,
  2. $a \cdot b = b \cdot a$.

- **Associativity under addition and multiplication**: for any $a, b, c \in \mathbb{F}$,
  1. $a+(b+c)=(a+b)+c$,
  2. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

- **Additive and multiplicative identity**: there exist an element in $\mathbb{F}$ called the *additive identity* and denoted by 0, for which $a+0=a$ for any $a \in \mathbb{F}$.

  Similarity, there exists an element in $\mathbb{F}$ called the *multiplicative identity* and denoted by 1, for which $a \cdot 1 = a$ for any $a \in \mathbb{F}$.

- **Additive and multiplicative inverses**: for any element $a \in \mathbb{F}$ (except the additive identity) there exists:
  1. $b \in \mathbb{F}$ such that $a+b=0$, and
  2. $c \in \mathbb{F}$ such that $a \cdot c = 1$.

  (usually b is denoted as $-a$, while c is denoted as a^{-1})

- **Distributivity of multiplication over addition**: for any $a, b, c \in \mathbb{F}$,

$$a \cdot (b+c) = (a \cdot b) + (a \cdot c).$$

3.1.1 Infinite fields

We start with one of the most obvious examples of a field: the real numbers together with the standard addition and product.

Theorem 3.1 $\mathbb{R}$ as a field

The set of real numbers $\mathbb{R}$ forms a field together with the standard addition and product.

We leave the proof of 3.1 to the reader, as it is pretty straight forward using the known properties of the standard addition and product over $\mathbb{R}$ (and rather uninteresting). Instead, we jump forward to using 3.1 for proving the same idea about the complex numbers:

Theorem 3.2 $\mathbb{C}$ as a field and more more more

The set of complex numbers $\mathbb{C}$ forms a field together with the addition and product operations as defined in Section 0.6 (namely Equation 0.6.3, Equation 0.6.4 and Equation 0.6.13).

Proof 3.1 $\mathbb{C}$ as a field

(note: in the following proof, equalities marked with ! use the respective property of the real numbers)

- **Closure under both operations**: for any two complex numbers $z_1 = a + \mathrm{i}b$ and $z_2 = c + \mathrm{i}d$,
  - Addition: since addition in $\mathbb{R}$ is closed, $(a+c) \in \mathbb{R}$ and $(b+d) \in \mathbb{R}$. Therefore
  $$z = z_1 + z_2 = a + c + (b + d)\mathrm{i}$$
  is also a complex number with $\mathfrak{Re}(z) = a + c$ and $\mathfrak{Im}(z) = b + d$.
  - Multiplication: since multiplication in $\mathbb{R}$ is also closed, $(ac - bd) \in \mathbb{R}$ and $(ad + bc) \in \mathbb{R}$. Therefore
  $$z = z_1 \cdot z_2 = ac - bd + (ad + bc)\mathrm{i}$$
  is a complex number with $\mathfrak{Re}(z) = ac - bdc$ and $\mathfrak{Im}(z) = ad + bc$.
- **Commutativity of both operation**: for any two complex numbers $z_1 = a + \mathrm{i}b$ and $z_2 = c + \mathrm{i}d$,
  - Addition: since addition in $\mathbb{R}$ is commutative, $a+c = c+a$ and $b+d = d+b$. Therefore
  $$z_1 + z_2 = a + c + (b + d)\mathrm{i} \stackrel{!}{=} c + a + (d + b)\mathrm{i} = z_2 + z_2.$$

• Multiplication: since multiplication in $\mathbb{R}$ is also commutative, $ac - bd = ca - db$ and $ad + bc = da + cb$. Therefore

$$z_1 \cdot z_2 = ac - bd + (ad + bc)\mathrm{i} \overset{!}{=} ca - db + (da + cb)\mathrm{i} = z_2 \cdot z_1.$$

- **Associativity of both operation**: for any three complex numbers $z_1 = a + \mathrm{i}b$, $z_2 = c + \mathrm{i}d$ and $z_3 = g + \mathrm{i}h$ (where $a, b, c, d, g, h \in \mathbb{R}$[a]),

  • Addition: since addition in $\mathbb{R}$ is associative, $a + (c + g) = (a + c) + g$ and $b + (d + h) = (b + d) + h$. Therefore

$$z_1 + (z_2 + z_3) = a + (c + g) + [b + (d + h)]\mathrm{i} \overset{!}{=} (a + c) + g + [(b + d) + h]\mathrm{i} = (z_1 + z_2) + z_3.$$

  • Multiplication: since multiplication in $\mathbb{R}$ is also associative, the following equalities apply:

$$\begin{aligned}
a \cdot (c \cdot g) &= (a \cdot c) \cdot g,\\
b \cdot (c \cdot h) &= (b \cdot c) \cdot h,\\
a \cdot (d \cdot h) &= (a \cdot d) \cdot h,\\
b \cdot (d \cdot g) &= (b \cdot d) \cdot g,\\
a \cdot (c \cdot h) &= (a \cdot c) \cdot h,\\
a \cdot (d \cdot g) &= (a \cdot d) \cdot g,\\
b \cdot (c \cdot g) &= (b \cdot c) \cdot g,\\
b \cdot (d \cdot h) &= (b \cdot d) \cdot h.
\end{aligned}$$

  Therefore,

$$\begin{aligned}
z_1 \cdot (z_2 \cdot z_3) = {} & a \cdot (c \cdot g) - a \cdot (d \cdot h) - b \cdot (c \cdot h) - b \cdot (d \cdot g)\\
& + [a \cdot (c \cdot h) + a \cdot (d \cdot g) + b \cdot (c \cdot g) - b \cdot (d \cdot h)]\mathrm{i}\\
\overset{!}{=} {} & (a \cdot c) \cdot g - (a \cdot d) \cdot h - (b \cdot c) \cdot h - (b \cdot d) \cdot g\\
& + [(a \cdot c) \cdot h + (a \cdot d) \cdot g + (b \cdot c) \cdot g - (b \cdot d) \cdot h]\mathrm{i}\\
= {} & (z_1 \cdot z_2) \cdot z_3.
\end{aligned}$$

- **Identity for both operations**:

  • Addition: the complex number $0 = 0 + 0i$ is the complex addition identity: for any real number $x \in \mathbb{R}$, $x + 0 = x$. Therefore, for any complex number $z = a + \mathrm{i}b$,

$$z + 0 = a + \mathrm{i}b + 0 + 0i = a + 0 + (b + 0)\mathrm{i} \overset{!}{=} a + \mathrm{i}b.$$

  • Multiplication: the complex number $1 = 1 + 0\mathrm{i}$ is the complex multiplication identity: for any real number $x \in \mathbb{R}$, $x \cdot 1 = x$ and $x \cdot 0 = 0$. Therefore, for any complex number $z = a + \mathrm{i}b$,

$$z \cdot 1 = (a + \mathrm{i}b) \cdot (1 + 0\mathrm{i}) \overset{!}{=} a \cdot 1 - \cancel{b \cdot 0\mathrm{i}^2} + (\cancel{a \cdot 0\mathrm{i}} + b \cdot 1)\mathrm{i} = a + \mathrm{i}b.$$

- **Inverse for both operations**:

  - Addition: for any complex number $z_1 = a + \mathrm{i}b$, the number $z_2 = -a - \mathrm{i}b$ is also a complex number for which

  $$z_1 + z_2 = a + \mathrm{i}b + -a - \mathrm{i}b \overset{!}{=} a - a + (b - b)\mathrm{i} = 0 + 0\mathrm{i} = 0.$$

  - Multiplication: for any complex number $z = re^{\theta i}$ where $r \neq 0$, the number $z^{-1} = \frac{1}{r}\mathrm{e}^{-\mathrm{i}\theta}$ is also a complex number for which

  $$z \cdot z^{-1} = r\mathrm{e}^{\mathrm{i}\theta} \cdot \frac{1}{r}\mathrm{e}^{-\mathrm{i}\theta} \overset{!}{=} \frac{r}{r}\mathrm{e}^{\cancel{\mathrm{i}\theta - \mathrm{i}\theta}} = 1 \cdot 1 = 1.$$

  Note: for $z = a + \mathrm{i}b$,

  $$z^{-1} = \frac{1}{r}\mathrm{e}^{-\mathrm{i}\theta} = \frac{1}{|z|} \cdot \frac{a - \mathrm{i}b}{|z|} = \frac{1}{|z|} \cdot \frac{\bar{z}}{|z|} = \frac{\bar{z}}{|z|^2}.$$

  Therefore, for any $z \neq 0$, $z^{-1} = \frac{\bar{z}}{|z|^2}$.

- **Distributivity of multiplication over addition**: for any $z_1 = a + \mathrm{i}b$, $z_2 = c + \mathrm{i}d$ and $z_3 = g + \mathrm{i}h$,

$$\begin{aligned}
z_1 \cdot (z_2 + z_3) &= (a + \mathrm{i}b) \cdot (c + \mathrm{i}d + g + \mathrm{i}h) = (a + \mathrm{i}b) \cdot (c + g + [d + h]\mathrm{i}) \\
&= ac + ag + (bd)\mathrm{i}^2 + (bh)\mathrm{i}^2 \\
&+ (ad)\mathrm{i} + (ah)\mathrm{i} + (bc)\mathrm{i} + (bg)\mathrm{i} \\
&= ac + ag - bd - bh + (ad + ah + bc + bg)\mathrm{i} \\
&= ac - bd + (ad + bc)\mathrm{i} + ag - bh + (ah + bg)\mathrm{i} \\
&= (z_1 \cdot z_2) + (z_1 \cdot z_3).
\end{aligned}$$

[a] The letters g and h are used instead of e and f to avoid confusion with Eurler's constant and the common notation for real functions, respectively.

QED

The sets $\mathbb{R}$ and $\mathbb{C}$ are examples of **infinite fields**, since they each have infinite number of elements. The set $\mathbb{Q}$ (rational numbers) can be shown to also be an infinite field, however unlike $\mathbb{R}$ and $\mathbb{C}$ it has **countable** number of elements, i.e. each number in $\mathbb{Q}$ can be assigned an index $1, 2, 3, \ldots$[1].

[1] For proof, see ...

Challange 3.1 $\mathbb{Q}$ as a field

Prove that $\mathbb{Q}$ (together with the usual addition and product operation) is indeed a field.

?

3.1.2 Finite fields

While all three examples of fields we encountered so far have each an infinite number of elements, some fields only have a finite number of elements (called their **order**). For example, consider the set $S = \{0, 1, a, b\}$ and the addition and product operations described using the following tables (left table describes addition, right table describes multiplication):

+	0	1	a	b
0	0	1	a	b
1	1	0	b	a
a	a	b	0	1
b	b	a	1	0

·	0	1	a	b
0	0	0	0	0
1	0	1	a	b
a	0	a	b	1
b	0	b	1	a

By examining the tables above, several points become clear:

- all the possible combinations of operands in both addition and multiplication give elements from S itself, meaning that the set is closed under both these operations.
- both tables are symmetric around their main diagonal, meaning that both addition and multiplication are commutative operations.
- in the addition table, the first row and first column both show that $x + 0 = x$ for any $x \in S$, meaning that 0 is the additive identity in S.
- in the product table, the second row and second column both show that $x \cdot 1 = x$ for any $x \in S$, meaning that 1 is the multiplicative identity in S.
- in the addition table, the element 0 appears in each row and each column exactly once. This means that every element x has a single additive inverse $y \in S$.
- in the product table, the element 1 appears in each row and each column exactly once, except for the first row and first column. This means that every element $x \neq 0$ has a single multiplicative inverse $z \in S$.

We therefore only need to prove two points to show that S is a field together with the operations described by the above tables: associativity of both operations and distributivity of multiplication over addition. We leave these proofs as a challenge to the reader. Such a field is sometime denoted as $\mathbb{F}_4$. There are, of course, infinitely many finite fields.

3.1.3 Modulo fields

Another example of finite fields are sets of integers of the form $\{0,1,2,3,\ldots,n\}$ where n is a prime, together with **modular addition** and **modular product**. To understand modular arithmetics, we recall the fact that on a circle, an angle can have a negative value but also greater than 360° values are possible (see Figure 0.3): 390° is equivalent to 30°, −30° is equivalent to 330°, etc. The set of integer values $0°, 1°, 2°, \ldots, 359°$ on a circle is an example of a modular set: if for example we add together two angles of values $\deg 100$ and 300° we get the equivalent angle $\deg 60$. If we subtract $\deg 300$ from $\deg 100$ the result is an angle of $\deg 160$.

We say that on a circle, the values $360°, 720°, -360°$ etc. are all **congruent** to 0 modulo 360. In mathematical notation we represent this fact as e.g.

$$720 \equiv 0 \pmod{360}. \tag{3.1.1}$$

Note that from this point forward we drop the degrees unit, and deal with pure integers. The notation for the set $\{0,1,2,\ldots,359\}$ is $\mathbb{Z}_{360}$. Generally speaking, the set $\{0,1,2,\ldots,n\}$ is denoted as $\mathbb{Z}_n$.

Note 3.2 About modulo set notation

It is not common to use the notation $\mathbb{Z}_n$ for the modulo-n set, since it is also used for a different algebraic construct, namely the n-adic ring. However, due to the simplicity of the notation, and the fact that we don't discuss rings in this chapter we are using it in this book. Common notations for the set are $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{Z}/n$.

!

Addition and multiplication on $\mathbb{Z}_n$ is done by the following rather straight forward definition:

Definition 3.2 Operations in $\mathbb{Z}_n$

In the set $\mathbb{Z}_n$ addition and multiplication are defined as the following:

- **Addition**: for any two elements $a, b \in \mathbb{Z}_n,\ a+b := (a+b) \pmod{n}$.

- **Multiplication**: for any two elements $a, b \in \mathbb{Z}_n,\ a \cdot b := (a \cdot b) \pmod{n}$.

π

Example 3.1 Operations in $\mathbb{Z}_n$

The tables below show addition and multiplication results of numbers in different modulo sets $\mathbb{Z}_n$ for some values of n:

n	$2+3$	$2\cdot 3$
4	1	2
5	0	1
6	5	0
7	5	6
8	5	6
9	5	6
10	5	6
11	5	6

n	$4+7$	$4\cdot 7$
8	3	4
9	2	1
10	1	8
15	11	13
20	11	8
27	11	1
28	11	0
30	11	28

Figure 3.1 shows the equivalency between integers and the elements of $\mathbb{Z}_5$.

$\mathbb{Z}$	$\cdots$	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	$\cdots$
$\downarrow$		$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	
$\mathbb{Z}_5$	$\cdots$	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	$\cdots$

Figure 3.1 An example of the periodicity of $\mathbb{Z}_5$: the top numbers are the ordinary integers, each showing their respective congruent modulo 5 below (blue dashed arrow).

Only the sets $\mathbb{Z}_n$ for which n is a prime number are also fields. Let's define this property precisely:

Theorem 3.3 $\mathbb{Z}_p$ is a field

Any modulo set $\mathbb{Z}_p$ where p is a prime number greater than 1 is also a field together with the operations as defined in 3.2.

In order to prove 3.3 we use two lemmas: the first is known as **Bézout's lemma**:

Lemma 3.1 Bézout's lemma

For any two positive integers a, b there exist two integers x, y such that

$$\gcd(a, b) = xa + yb.$$

Note 3.3 $\gcd(a, b)$

$\gcd(a, b)$ is the **greatest common deviser** of the two integers a and b. For example, $\gcd(36, 24) = 12$ since the divisors of 36 are $1,\ 2,\ 3,\ 4,\ 6,\ 9,\ 12,\ 18,\ 36$, and the divisors of 24 are $1,\ 2,\ 3,\ 4,\ 6,\ 8,\ 12,\ 24$.

An example of Bézout's lemma is the following:

Example 3.2 Bézout's lemma in action

For the two positive integers $a = 60,\ y = 114$

$$\gcd(60, 114) = 6.$$

Therefore, Bézout's lemma says that there exist two integers x, y such that

$$6 = 60x + 114y.$$

Indeed, two such integers exist: $x = 2$ and $y = -1$.

(SHOULD WE PROVE THE LEMMA?..)

The second lemma we use is the following:

Lemma 3.2 $\gcd(n, p) = 1$

Given a positive prime number p, then for any positive integer $n < p$,

$$\gcd(p, n) = 1.$$

Proving the lemma:

Proof 3.2 $\gcd(n, p) = 1$

We assume that $\gcd(p, n) \neq 1$. Then there exist an integer $a \leq n < p$ which divides both n and p, meaning that p has a divider, contrary to the assumption that p is a prime number. Therefore $\gcd(n, p)$ must equal 1.

QED

Now we can proceed to the proof of 3.3:

Proof 3.3 $\mathbb{Z}_p$ is a field

- **Closure under both operations**: the definition of the modulo operator limit any $M \pmod{p}$ (where $M \in \mathbb{Z}$) to be in $[0, p-1]$. Therefore the result of using the operators given in 3.2 must be within the same range, and thus in $\mathbb{Z}_p$.

- **Commutativity and associativity of both operations**: for any two numbers $a, b \in \mathbb{Z}_p$ the result $a + b$ and $a \cdot b$ under $\mathbb{Z}$ is both commutative and associative. Therefore the result modulo n is the same no matter the order of operations.

- **Additive identity**: the number $0 \in \mathbb{Z}_p$ is the additive identity, since for each $a \in \mathbb{Z}_p,\ a + 0 = a$.

- **Multiplicative identity**: the number $1 \in \mathbb{Z}_p$ is the additive identity, since for each $a \in \mathbb{Z}_p$, $a \cdot 1 = a$.

- **Additive inverse**: for each $a \in \mathbb{Z}_p$ the element $n = p - a$ is in $\mathbb{Z}_p$ since $p > a$. Adding n to a results in 0:

$$a + n = a + (p - a) = p \equiv 0 \pmod{p}.$$

- **Multiplicative inverse**: let $a \in \mathbb{Z}_p$ and $a \neq 0$. Since p is a prime, $\gcd(a, p) = 1$ and from Bézout's theorem we know that there exist two integers x, y such that

$$xa + yp = 1.$$

Rearrangement gives $p = \frac{1-xa}{y}$ meaning that p divides $1 - xa$, and thus

$$xa \equiv 1 \pmod{p}.$$

Therefore x is the multiplicative inverse of a.

- **Distributivity of multiplication over addition**: ...

QED

The only part of the proof that uses the fact that p is a prime number is the multiplicative inverse. When n is not a prime, $\mathbb{Z}_n$ is not a field.

Challange 3.2 $\mathbb{Z}_n$ is not a field when n is not a prime number

Prove that the modulo set $\mathbb{Z}_n$ where n is **not** a prime number, is not a field. (**hint**: what property of prime numbers is used in the above proof to show that there is always a multiplicative inverse in $\mathbb{Z}_p$ where p is prime?)

?

3.2 VECTOR SPACES

As we've seen in Chapter 1 vectors are found at the heart of linear algebra. We first defined them in a geometric way as objects with magnitude and direction, and later as lists of real numbers, analyzing the connections between these two mostly parallel definitions. We also spoke about vector spaces of the type $\mathbb{R}^n$ as the structures vectors exist in. However, we haven't defined vectors nor vector spaces formally - which is exactly what we do in this section, by defining the concept of **vector spaces**.

Note 3.4 $\mathbb{R}^n$ as a guide to general vector spaces

While reading the definition below, it is worthwhile to reflect on each of the given axioms as it relates to the familiar vector space $\mathbb{R}^n$.

!

Definition 3.3 Vector space

A vector space over a field $\mathbb{F}$ is a set V which, together with two operations described below, fulfils a list of axioms. The two operations are

• Vector addition: an operation which takes two elements of V and returns a single element of V, i.e. $+: V \times V \to V$.

• Scalar multiplication: an operation which takes a single element of $\mathbb{F}$ and a single element of V and returns a single element of V, i.e. $\cdot : \mathbb{F}, V \to V$.

The axioms to be fulfilled are:

• **Commutativity of vector addition**: for any $u, v \in V$,

$$u + v = v + u.$$

• **Associativity of vector addition**: for any $u, v, w \in V$,

$$u + (v + w) = (u + v) + w.$$

• **Additive identity**: there exist an element $0 \in V$ for which, for any $v \in V$,

$$v + 0 = v.$$

• **Scalar multiplicative identity**: for any $v \in V$

$$1 \cdot v = v,$$

where 1 is the multiplicative identity in $\mathbb{F}$.

• **Additive inverse**: for any $v \in V$ there exist an element $u \in V$ for which

$$v + u = 0.$$

• **Associativity of scalar multiplication**: for any $\alpha, \beta \in \mathbb{F}$ and $v \in V$

$$\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v,$$

where $\alpha\beta$ is the multiplication defined for $\mathbb{F}$.

• **Distributivity of vector addition**: for any $\alpha \in \mathbb{F}$ and $u, v \in V$,

$$\alpha \cdot (u + v) = (\alpha \cdot u) + (\alpha \cdot v).$$

- **Distributivity of scalar addition**: for any $\alpha, \beta \in \mathbb{F}$ and $v \in V$,

$$(\alpha + \beta) \cdot v = (\alpha \cdot v) + (\beta \cdot v).$$

The elements of V are then called **vectors**, and the elements of $\mathbb{F}$ are called **scalars**.

Since we discussed $\mathbb{R}^n$ thoroughly in Chapter 1, let's prove that it is indeed a vector space under the above definition. First, the claim:

Theorem 3.4 $\mathbb{R}^n$ is a vector space

The set of elements of the form

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}$$

where $v_i \in \mathbb{R}$, forms a vector space over $\mathbb{R}$ together with the following two operations:

- **Vector addition**:

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}.$$

- **Scalar multiplication**:

$$\alpha \cdot \vec{v} = \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix}.$$

The proof itself is pretty easy, based on the fact that $\mathbb{R}$ is a field:

Proof 3.4 $\mathbb{R}^n$ is a vector space

Since the results of both operations defined for $\mathbb{R}^n$ only depend on the respective components of a vector $v \in \mathbb{R}^n$, all the axioms of a vector space apply, since they derive directly from the fact that $\mathbb{R}$ is a field. As an example, we will elaborate on two of the axioms:

- **Additive inverse**: Given a vector $\vec{v} \in \mathbb{R}^n$, each of its components v_i has an inverse under $\mathbb{R}$, namely $-v_i$. Therefore,

$$\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} + \begin{bmatrix} -v_1 \\ -v_2 \\ \vdots \\ -v_n \end{bmatrix} = \begin{bmatrix} v_1 - v_1 \\ v_2 - v_2 \\ \vdots \\ v_n - v_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \vec{0},$$

which is the additive identity in $\mathbb{R}^n$.

- **Distributivity of vector addition**: for each component of two vectors $\vec{u}, \vec{v} \in \mathbb{R}^n$, given the rules for vector addition and scalar multiplication, together with the distributivity of numbers in $\mathbb{R}$:

$$\begin{aligned} \alpha(\vec{u}+\vec{v}) = \alpha\left(\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}\right) &= \alpha \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix} \\ = \begin{bmatrix} \alpha u_1 + \alpha v_1 \\ \alpha u_2 + \alpha v_2 \\ \vdots \\ \alpha u_n + \alpha v_n \end{bmatrix} &= \begin{bmatrix} \alpha u_1 \\ \alpha u_2 \\ \vdots \\ \alpha u_n \end{bmatrix} + \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix} = \alpha\vec{u} + \alpha\vec{v}. \end{aligned}$$

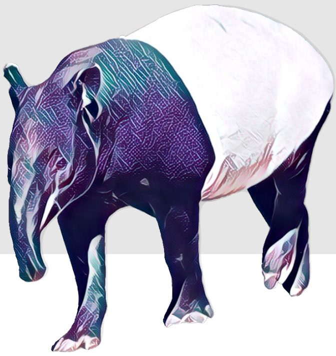
QED

(it is adviceable for the reader to go over the rest of the axioms and prove them for $\mathbb{R}^n$)

3.3 EXERCISES

CHAPTER

4

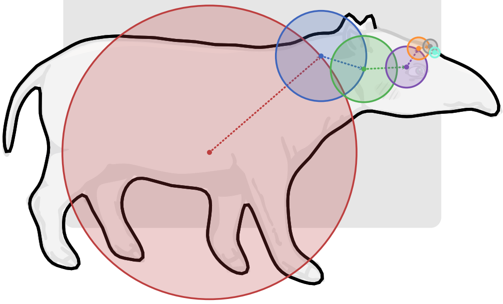


DIFFERENTIAL EQUATIONS

4.1 EXERCISES

CHAPTER

5

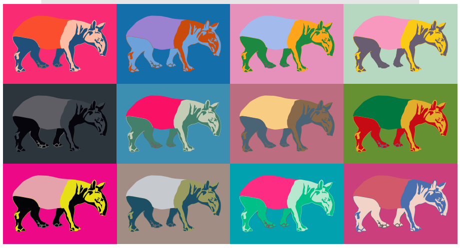


THE FOURIER TRANSFORM

5.1 EXERCISES

CHAPTER

6



SYMMETRY GOURPS

6.1 EXERCISES