MATHEMATICS FOR SCIENCE STUDENTS

An open-source book

Written, illustrated and typeset (mostly) by

PELEG BAR SAPIR

with contributions from others

$$a^b = e^{b \log(a)}$$

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

$$T\left(\alpha\vec{u} + \beta\vec{v}\right) = \alpha T\left(\vec{u}\right) + \beta T\left(\vec{v}\right)$$

$$A = Q \Lambda Q^{-1}$$

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\,\mathrm{d}t$$

$$R(\theta) = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}$$

$$\frac{\mathrm{d}f}{\mathrm{d}x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

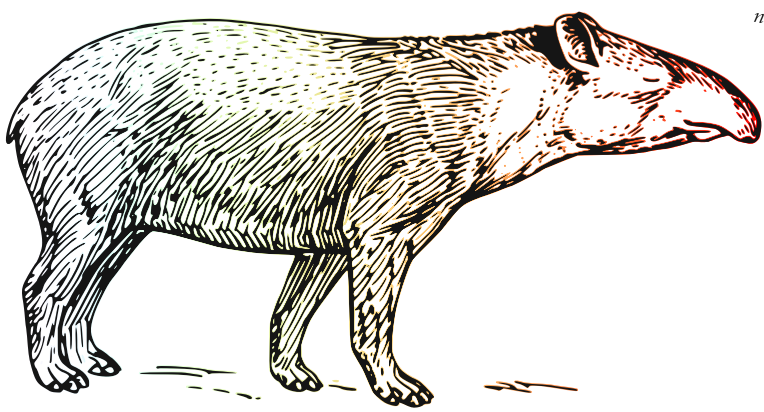
$$\langle \hat{e}_i, \hat{e}_j \rangle = \delta_{ij}$$

$$\vec{v} = \sum_{i=1}^{n} \alpha_i \hat{e}_i$$

$$e^{\pi i} + 1 = 0$$

$$\int_a^b f(x)\,\mathrm{d}x = F(b) - F(a)$$

$$\cos(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n)!} x^{2n}$$

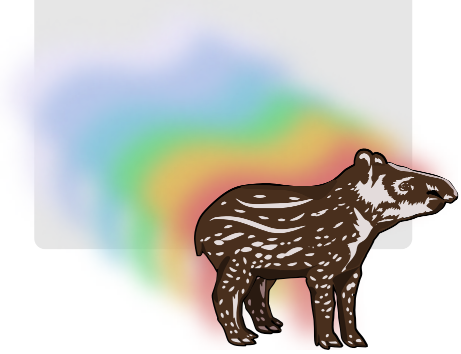


PUBLISHED IN THE WILD

! **To be written**: Rights, lefts, etc. will be written here in the future !

HERE BE TABLE

CHAPTER 0



INTRODUCTION

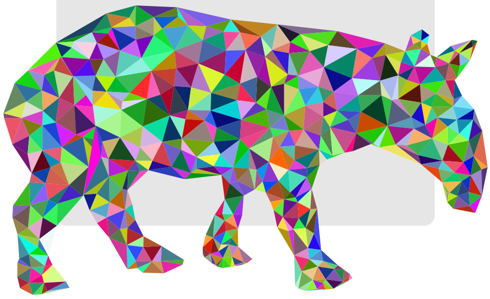
In this chapter we introduce key concepts that will be used in later chapters. For this reason, unlike other chapters it contains many statements, sometimes given without thorough explanations or reasoning. While all of these statements are grounded in deep ideas and can be formulated in a rigorous manner, 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.

!

CHAPTER 1



REAL CALCULUS IN 1D

1.1 SEQUENCES AND SERIES

1.1.1 Basics

A **sequence** is an indexed collection of **elements**. By *indexed* we mean that the order of the elements in a sequence matters (unlike with sets): changing the order of any element changes the sequence as a whole. The following are some examples of sequences composed of real numbers:

- $1, -3, 0, -7, 2, 1.5, 4, 0, 1, -0.35, \sqrt{2}$.
- $0, 1, 2, 1, 1, -1, 0$.
- $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots$

- $0,1,0,1,0,1,0,1,0,1,0,1,0,1,\ldots$

The examples above present two more properties of sequences:

- Elements may repeat (unlike in the case of sets), and
- sequences can be either **finite** (as in the first two examples), or **infinite** (as in the latter two examples).

The number of elements in a sequence is called its **length**. In the case of infinite sequences we say that their length equals ∞ (inifinity). The elements of a sequence a are usually indexed using a subscript, such that a_1 is the first element in the sequence, a_2 is the second element in the sequence, etc. - and generally a_i is the i-th element in the sequence, where $i \in \mathbb{N}$.

We can therefore define a sequence somewhat more formally as a function from a subset of the natural numbers to the real numbers:

$$a : N \to \mathbb{R}, \tag{1.1.1}$$

where $N \subseteq \mathbb{N}$.

Example 1.1 Sequences as functions

The following 9-element sequence a

$$\begin{array}{ccccccccc}
3, & 4, & \frac{1}{2}, & 0, & 2, & 6, & -\frac{2}{3}, & 0, & -1. \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\
a(1) & a(2) & a(3) & a(4) & a(5) & a(6) & a(7) & a(8) & a(9)
\end{array}$$

can be viewed as a function

$$a : \{1,2,\ldots,9\} \to \mathbb{R},$$

or more precisely as a function

$$a : \{1,2,\ldots,9\} \to \left\{-1,-\frac{2}{3},0,\frac{1}{2},2,3,4,6\right\}.$$

The follow infinite sequence b

$$\begin{array}{cccccccc}
1, & \frac{1}{2}, & \frac{1}{3}, & \frac{1}{4}, & \frac{1}{5}, & \frac{1}{6}, & \frac{1}{7}, & \cdots \\
\uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \\
b(1) & b(2) & b(3) & b(4) & b(5) & b(6) & b(7) &
\end{array}$$

can be viewed as a function

$$b : \mathbb{N} \to (0,1].$$

Since sequences can be viewed as functions, they can be defined using formulas: for example, the sequence

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots$$

can be defined using the simple formula

$$a_n = \frac{1}{n}.$$

Example 1.2 Some sequences defined using formulas

$$\begin{aligned}
(-1)^n &\quad\Rightarrow\quad -1, 1, -1, 1, -1, 1, -1, 1, -1, \dots \\
3n+4 &\quad\Rightarrow\quad 7, 10, 13, 16, 19, 22, \dots \\
(n+1)^2 &\quad\Rightarrow\quad 4, 9, 16, 25, 36, 49, \dots \\
\begin{cases} 2n+1 & \text{if } n \text{ is odd,} \\ n-1 & \text{if } n \text{ is even.} \end{cases} &\quad\Rightarrow\quad 3, 1, 7, 3, 11, 5, 15, 7, \dots
\end{aligned}$$

Sequences can also be defined using **recursion**, where the value of an element is defined using previous values and a **starting value**. For example:

$$a_n = a_{n-1}^2 - 2,$$

with the starting value $a_1 = 3$. We the get that

$$a_2 = a_1^2 - 2 = 3^2 - 2 = 7,$$

and thus

$$a_3 = a_2^2 - 2 = 7^2 - 2 = 47,$$

etc.

Example 1.3 The Fibonacci sequence

The **Fibonacci sequences** is a well-known sequence defined using the following recursive rule:

$$F_n = F_{n-1} + F_{n-2},$$

with $F_1 = F_2 = 1$. The first few elements of the sequence are therefore

$$1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, \dots$$

See Figure 1.1 for a graphical representation of the Fibonacci sequence.

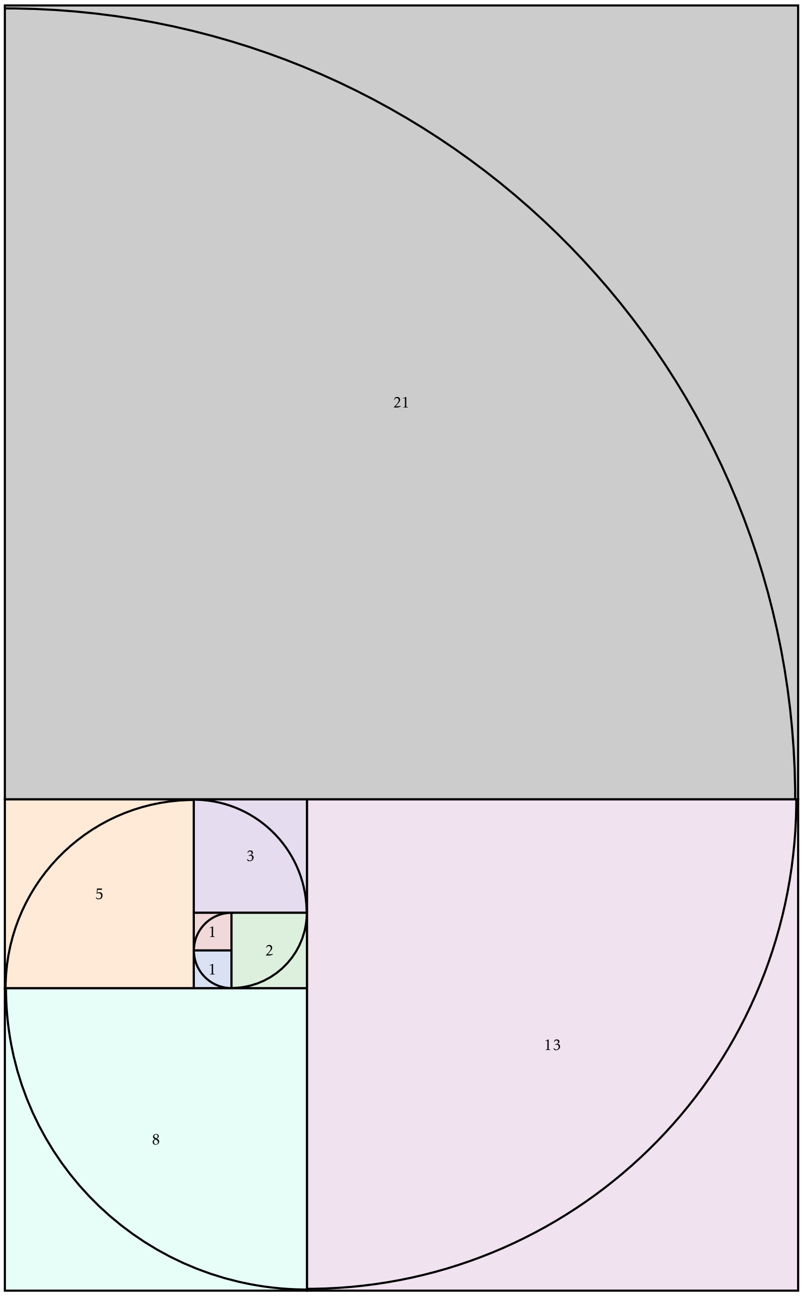


Figure 1.1 A graphical representation of the Fibonacci sequence: two squares of side 1 are placed adjacent to each other on the plane. In each subsequent step a new square is place such that its side is equal to the combined sides of the previous two squares. This way, the side of each square in the sequence follows the Fibonacci sequence. In each square we draw a quarter circle centered on one of the vertices, such that we get the famous **golden ratio** helix.

Note 1.1 Focus of section

From now on in the section we will focus on infinite sequences only.

1.1.2 Types of sequences

Consider the sequence $a_n = n^2$. Since $n \in \mathbb{N}$, for any n, $a_{n+1} > a_n$, since $(n+1)^2 > n^2$ (see Figure 1.2). We say that such a sequence is **increasing**. In fact, for a sequence to be increasing some sequential elements can be equal: for example, the sequence $c_n = 1,1,1,2,2,2,3,3,3,4,4,4,5,5,5,\dots$ is also a increasing sequence. Thus, the definition of a increasing sequence is the following:

Definition 1.1 increasing sequence

sequence a_n is said to be *increasing* if for any $n \in \mathbb{N}$, $a_{n+1} \geq a_n$.

If we change the condition to $a_{n+1} > a_n$ we say that such a sequence is **strictly increasing**. In the above examples a_n is a strictly increasing sequence, while c_n is just increasing (since for some indeces n, $c_{n+1} = c_n$).

Similarly, a **decreasing** sequence is a sequence b_n for which for any $n \in \mathbb{N}$, $b_{n+1} \leq b_n$. An example of such sequence is $b_n = \frac{1}{n}$ (see Figure 1.3). ANd of course, if we change the condition to $b_{n+1} < b_n$ then the sequence is **strictly decreasing**.

Generally, a sequence that is either increasin or decreasing is said to be **monotone**. If a sequence is monotone starting only from a certain n, we say that the sequence is **eventually monotone** (i.e. *eventually increasing* or *eventually decreasing*). An example of such sequence is $d_n = (n-5)^2$ (Figure 1.4): for $N \in 1,2,3,4,5$ it is decreasing, but starting from $n = 5$ it is increasing for any n.

As an example of a sequence which isn't monote, consider the sequence $e_n = \sin(n)$: for some values of n, $e_{n+1} > e_n$ and for some other values $e_{n+1} < e_n$ (see Figure 1.5).

The following are two ways to determine whether a sequence a_n is monotone:

- **Difference test**: if $a_{n+1} - a_n \geq 0$ for all $n \in \mathbb{N}$, then the sequence is increasing. If $a_{n+1} - a_n \leq 0$ for all $n \in \mathbb{N}$ then the sequence is decreasing.

- **Ratio test**: if $\frac{a_n+1}{a_n} \geq 1$ for all $n \in \mathbb{N}$ then the sequence is increasing, and if $\frac{a_{n+1}}{a_n} < 1$ for all $n \in \mathbb{N}$ then the sequence is decreasing.

Example 1.4 Difference test

Given the sequence $a_n = \frac{n}{n+1}$, we look at the difference $a_{n+1} - a_n$:

$$a_{n+1} - a_n = \frac{n+1}{n+2} - \frac{n}{n+1} = \frac{(n+1)(n+1) - (n+2)n}{(n+1)(n+2)}$$

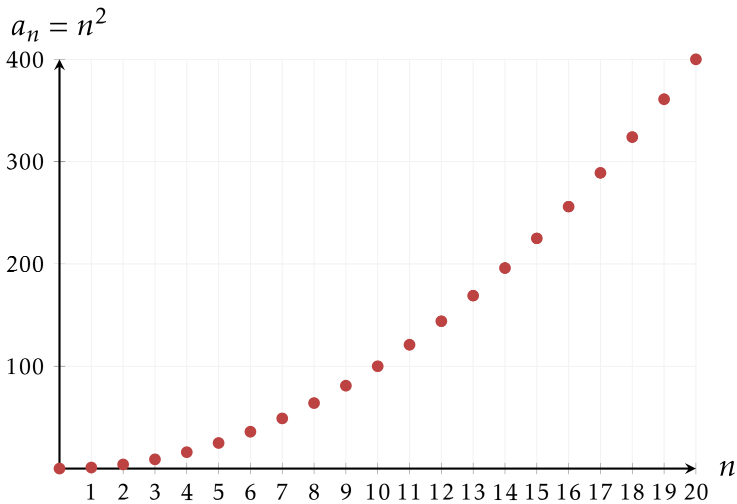


Figure 1.2 The sequence $a_n = n^2$ is increasing, and is in fact *strictly* increasing.

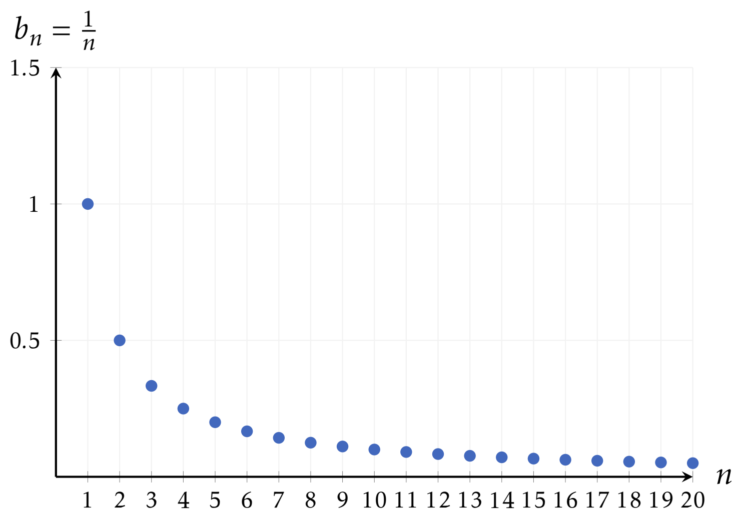


Figure 1.3 The sequence $b_n = \frac{1}{n}$ is decreasing, and is in fact *strictly* decreasing.

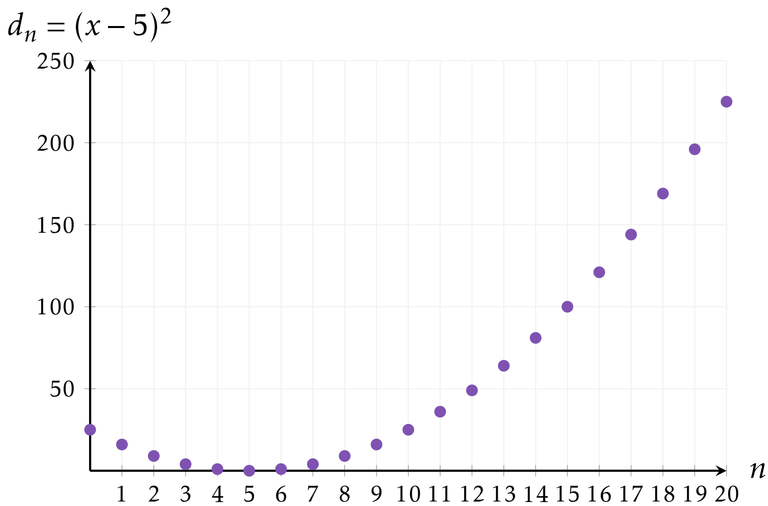


Figure 1.4 The sequence $d_n = (n-5)^2$ starts as a decreasing sequence, but starting from $n = 5$ it is increasing, making it an *eventually increasing sequence*.

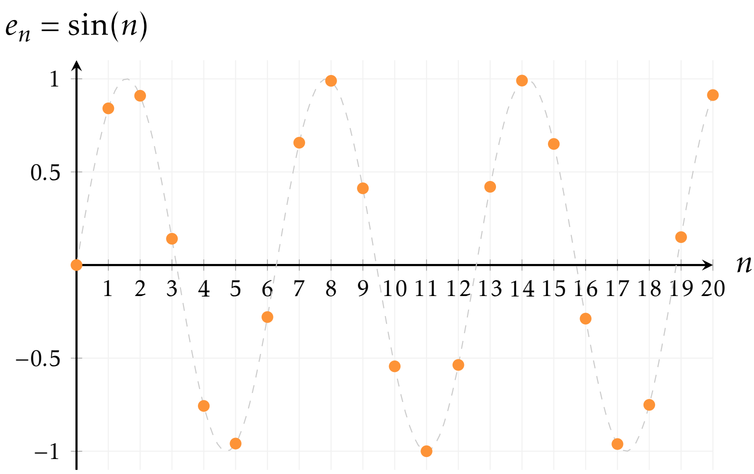


Figure 1.5 The sequence $e_n = \sin(n)$ is neither increasing nor decreasing. For reference, the function $\sin(x)$ is plotted as a dashed line behind e_n.

$$= \frac{n^2 + 2n + 1 - n^2 - 2n}{(n+1)(n+2)} = \frac{1}{(n+1)(n+2)} < 1 \quad \forall n \in \mathbb{N}.$$

The last (in)equality stems from the fact that no matter what n we subtitute into $(n+1)(n+2)$, the result will be greater than 1, and thus $\frac{1}{(n+1)(n+2)}$ is always smaller than 1. Therefore, a_n is a decreasing sequence.

Example 1.5 Ratio test

Given the sequence $b_n = \frac{2^n}{n^2}$, the ratio of a_{n+1} to a_n is

$$\frac{a_{n+1}}{a_n} = \frac{\frac{2^{n+1}}{(n+1)^2}}{\frac{2^n}{n^2}} = 2\frac{n^2}{(n+1)^2}.$$

Let's look at the first few approximated values of the ratio $\frac{n^2}{(n+1)^2}$:

n	$\frac{n^2}{(n+1)^2}$
0	0
1	0.25
2	0.44444...
3	0.5625
4	0.64
5	0.69444...
6	0.7346938775510204
7	0.765625
8	0.7901234567901234
9	0.81
10	0.8264462809917356
11	0.840277777...
12	0.8520710059171598
13	0.8622448979591837

We see that for any $n \geq 3$, $\frac{n^2}{(n+1)^2} > \frac{1}{2}$, and therefore $2\frac{n^2}{(n+1)^2} > 1$. Thus, the sequence is eventually increasing.

Some sequences are **bounded** from below: this means that their elements never get smaller than some constant $\underline{M} \in \mathbb{R}$. For example, consider the simple sequence $a_n = n$, where $n = \{1, 2, 3, 4, \ldots\}$: there is no element in the sequence that is smaller than 1. Therefore, a_n is bounded from below by 1. Of course, one may argue that b_n is also bounded from below by 0, or -6, or in fact any negative number. This is true, however we are usually interested in the *maximal* number $\underline{M}$ that bounds the sequence from below, which in this case is $\underline{M} = 1$. We call that number the **infimum** of the sequence, and denote it as $\inf a_n$.

Similarly, a sequence a_n can be bounded from above by some number $\overline{M} \in \mathbb{R}$, i.e. there exist no n for which $a_n > \overline{M}$. We call the *minimal* such number the **supremum** of the sequence a_n, denoted $\sup a_n$. For example, the sequence $b_n = \frac{1}{n}$ is bounded from above by any real number $x \geq 1$, and therefore $\sup b_n = 1$. In fact, b_n is also bounded from below by $\underline{M} = 0$, and therefore we say that it is **bounded**. Another example of a sequence that is bounded is $e_n = \sin(n)$, which is bounded from below by $\underline{M} = -1$ and from above by $\overline{M} = 1$.

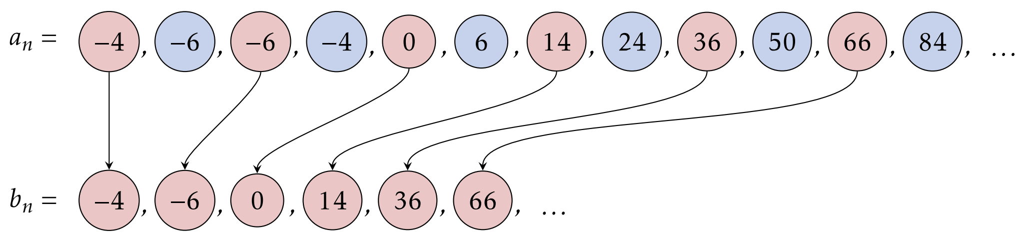
Example 1.6 Bounded and unbounded sequences

The following table shows some examples of sequences that are bounded from below, from above, or neither:

a_n	**First 5 elements**	$\inf a_n$	$\sup a_n$
$n^2 - n$	$0, 2, 6, 12, 20, \ldots$	0	-
$\frac{n}{n+1}$	$\frac{1}{2}, \frac{2}{3}, \frac{3}{4}, \frac{4}{5}, \frac{5}{6}, \ldots$	$\frac{1}{2}$	1
e^{-n}	$e^{-1}, e^{-2}, e^{-3}, e^{-4}, e^{-5}, \ldots$	0	e^{-1}
$\log(n)$	$0, \log(2), \log(3), \log(4), \log(5), \ldots$	0	-
$(-1)^n$	$-1, 1, -1, 1, -1, \ldots$	-1	1
$(-1)^n n$	$-1, 2, -3, 4, -5, \ldots$	-	-
$(-2)^n$	$-2, 4, -8, 16, -32, \ldots$	-	-

1.1.3 Subsequences

Given any sequence a_n, we can remove from it any number of its elements (including 0 elements) and get a new sequence b_n which is a **subsequence** of a_n. For example, let $a_n = n^2 - 5n$. We can remove each 2nd element from a_n (i.e. those with indices $2, 4, 6, 8, \ldots$) and get the following sequence b_n:



Note 1.2 Order of elements in a subsequence

A subsequence must preserve the order of the original sequence, since all we do in practice is removing elements from the original sequence, without changing the order of the remaining elements.

!

Let's look at a more formal defintion of a subsequence, which uses the choice of indices instead of removing elements:

Definition 1.2 Subsequence

A subsequence of a sequence a_n is a sequence a_{n_k}, where n_k is a **strictly increasing** sequence of natural numbers.

π

To create a subsequence using the above defintion, one can first create a sequence of indices n_k, and then subtitute only those indices into n in a_n. For example, given the sequence $a_n = n^2 - 5n$ from before, we can define a sequence of indices $n_k = 1, 3, 5, 7, 9, 11, \dots$ which would then yield the subsequence b_n shown before.

The reason we define n_k to be strictly increasing is to avoid changing the order of the elements from the original sequeunce a_n: for example, if we allowed n_k to be "just" increasing, we might end up with a case where there are two subsequent equal indices, e.g. $n_k = 1, 3, 5, 8, 9, 9, 10, \dots$. That would mean that we repeat an element from a_n **twice or more** in the subsequence (in the example this would be a_9), rendering it invalid as a subsequence, since as mentioned before - a subsequence must perserve the order of the original sequence.

Subsequences share all of the above-mentioned properties of the original sequence: if the original sequence is increasing or decreasing - so do all of its subsequences, and if it is bounded from above or below - so do all of its subsequences. Let's prove two of these properties:

Proof 1.1 Rising sequences and their subsequences

Claim: given an **increasing** sequence a_n, all of its subsequences are also increasing sequences themselves.

Proof: using contradiction. Let a_n be an increasing sequence, and b_n a subsequence of a_n which isn't increasing. From the fact that b_n is not an increasing sequence we know that there exist at least two indices k, m such that $k < m$ but $b_k > b_m$. Since any b_n is an element of a_n without change of order, we can substitute $b_k = a_i$ for some index i and $b_m = a_j$ for some index j, such that $i < j$ (since $k < m$ - this is exactly the idea of preserving the order of a_n). We therefore get that $a_i = b_k > b_m = a_j$, or simply $a_i > a_j$ even though $i < j$ - in contradiction to a_n being an increasing sequence. Therefore there can be no subsequence of a_n that isn't an increasing sequence.

QED

Proof 1.2 Bounded sequences and their subsequences

Claim: given a sequence a_n which is **bounded from below**, all of its subsequences are also bounded from below.

Proof: also using contradiction. Let a_n be a sequence bounded from below by $\inf a_n = \underline{M}$. Let b_n be a subsequence of a_n that isn't bounded from below - i.e. there exist an element b_i such that $b_i < \underline{M}$. Since b_n is a subsequence of a_n, $b_i = a_j$ for some index j. Therefore $a_j = b_i < \underline{M}$ in contradiction to $\underline{M}$ being the infimum of a_n. Therefore, a subsequence of a sequence bounded from below can not be unbounded from below.

QED

Challenge 1.1 Further proofs

1. Prove that all subsequences of a decreasing sequence are themselves decreasing.
2. Prove that all subsequences of a sequence bounded from above are themselves bounded from above.
3. Prove that all subsequences of a **strictly** increasing/decreasing sequence are themselves increasing or decreasing, respectively.
4. Given a bounded sequence a_n with $\inf a_n = \underline{M}$, can it have a subseqeunce b_n with $\inf b_n \neq \underline{M}$? If yes - give an example. If no - prove your claim.

?

1.1.4 Limits

As you probably noticed by now, some infinite sequences seem to approach a certain value as we increase n. That is to say, the bigger n is, the closer such a sequence a_n gets to a certain value $L \in \mathbb{R}$. For example, the sequence $b_n = \frac{1}{n}$ approaches to $L = 0$ as we increase n (see again Figure 1.3). The sequence $a_n = \frac{1}{n^2+1}$ approaches the value $L = 1$ as we increase n (see Figure 1.6). On the contrary, the sequence $d_n = (n-5)^2$ eventually increase in such a way that it "approaches" $L = \infty$, while $e_n = \sin(n)$ doesn't approach any value and instead endlessly "jumps" around in a repeated manner.

The formal term for the behaviour of a_n and b_n is called **convergance**, and it is one of the most important properties of infinite sequences. In this subsection we will define, analyze, and explain it in detail. To begin, we can divide all infinite sequences into two separate categories:

1. Sequences which converge to a finite number $L \in \mathbb{R}$.
2. Sequences which do not converge to any finite number.

Sequences in the second category are said to be **diverging**, and they can be further split into two separate categories:

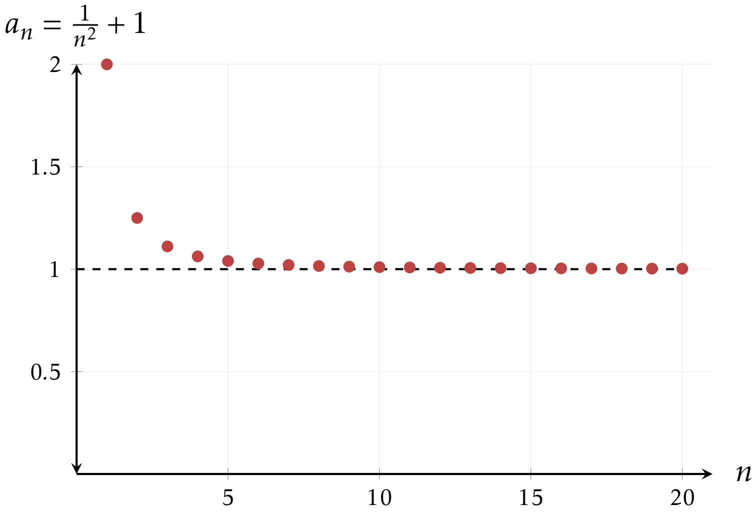


Figure 1.6 The sequence $a_n = \frac{1}{n^2} + 1$ approaches the value $L = 1$ as n increases.

i. Sequences which diverge to either positive or negative infinity.

ii. Sequences which neither converge nor diverge to $\pm\infty$.

Let us start with a more precise analysis of sequences that diverge to $\pm\infty$. In essence, a sequence which diverges to positive infinity is a sequence that is bounded from below but not from above, that is - given any real number R the sequence eventually passes it. In other words, all the elements in the suquence *after a certain value of* n are greater than R, for any R that we choose.

Take for example the sequence $a_n = n^2$. It most certainly has a lower boundry, namely $\inf a_n = 1$. On the other hand, given any $R \in \mathbb{R}$ eventually the values of a_n pass it. For example, given $R = 100$, the elements of a_n pass it after just 10 elements (since $a_{11} = 11^2 = 121 > 100$). The number $R = 1,000,000$ is passed after 1000 elements, etc. No matter how big R is, **eventually** a_n will pass it. Therefore, we say that a_n *goes to infinity*, and denote it by writing

$$\lim_{n\to\infty} a_n = \infty. \tag{1.1.2}$$

The notation lim is short for **limit**. While it can be argued that a divergent sequence has no limit, sometimes the term is used in the case of divergence to $\pm\infty$.

Note 1.3 Another limit notation

Another common notation used to denote that a sequence a_n is going to infinity as n increases is the following:

$$a_n \xrightarrow{n\to\infty} \infty.$$

!

A more formal definition of this behaviour is as follows:

Definition 1.3 Sequence going to inifnity

Let a_n be an infinite sequence. If for any $R \in \mathbb{R}$ there exist $n_R \in \mathbb{N}$ such that for any $n > n_R$, $a_n > R$, the sequence is said to be going to positive infinity. We denote this fact by writing

$$\lim_{n\to\infty} a_n = \infty.$$

π

Note 1.4 Chain of inequality

Note the following: say we have 5 real numbers $k_1, \ldots, k_5$. If

$$k_1 > k_2 > k_3 = k_4 \geq k_5.$$

then we know that

$$k_1 > k_5.$$

This might seem obvious - but is worth noting before we move on, as we will be using similar chains of inequalities in up-coming proofs.

!

Of course, for negative infinity the behaviour is very similar: a sequence a_n with an upper boundry $\overline{M}$ and no lower bound $\underline{M}$ is said to be **going to negative infinity**, since for any $R \in \mathbb{R}$ there exist an $n_R \in \mathbb{N}$ for which if $n > n_R$ then $a_n < R$.

Generally speaking, proving that a sequence goes to either positive or negative infinity follows a certain pattern, which we will examplify using the sequence $a_n = \frac{n}{2}$ (Figure 1.7). It should be clear that the sequence goes to positive infinity as n increases, since we can make the values of a_n as large as we want by substituting a respective n into $\frac{n}{2}$: for example, given $R = 1000$ we can subtitute $n = 2000$, yielding $a_{2000} = \frac{2000}{2} = 1000$, and thus any a_n where $n > 2000$ will be bigger than $R = 1000$. For $R = 1,000,000$ we can substitute $n = 2,000,000$ and so forth.

To show that this is true for any $R \in \mathbb{R}$ we should do the following: given a real number R find an index n_0 such that $a_{n_0} \geq R$. Since a_n is a strictly increasing sequence, that would mean that for any $n > n_0$, $a_n > a_{n_0} \geq R$, or simply $a_n > R$. In the case of $a_n = \frac{n}{2}$ we can simply choose the closest integer to $2R$ that is also bigger than $2R$ (i.e. if $R = 2.3$ we choose $n_0 = 3$, if $R = 100.7$ we choose $n_0 = 101$, etc.).

To always get an integer *equal to or bigger than* R we can use the **ceiling** operator. For any given $x \in \mathbb{R}$, its ceiling (denoted $\lceil x \rceil$) is the closest integer which is bigger than or equal to x, or more formally:

Definition 1.4 Ceiling and floor operators

Let $x \in \mathbb{R}$. Then

$$\lceil x \rceil = \min\left(\{n \in \mathbb{N} | n \geq x\}\right). \tag{1.1.3}$$

$$\lfloor x \rfloor = \max\left(\{n \in \mathbb{N} | n \leq x\}\right). \tag{1.1.4}$$

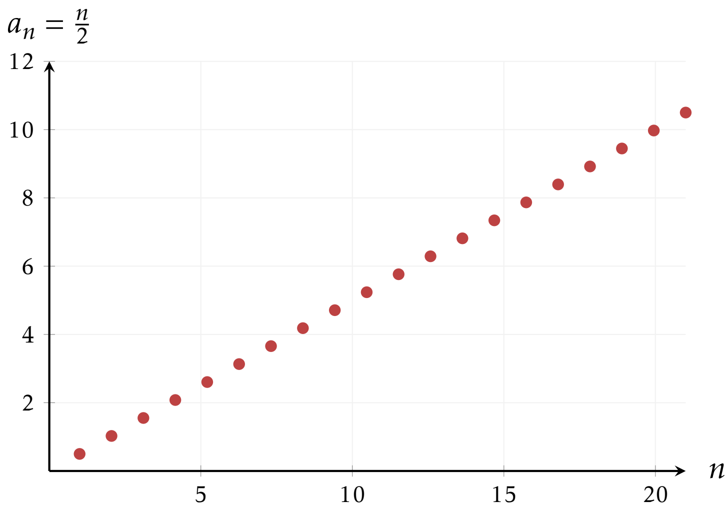


Figure 1.7 The sequence $a_n = \frac{n}{2}$ goes to positive infinity as n increases.

i.e. - $\lceil x \rceil$ is the **minimal** integer n that is **bigger than or equal** to x, and $\lfloor x \rfloor$ is the **maximal** integer n that is **smaller than or equal** to x.

π

Using the value $n_0 = \lceil 2R \rceil$ we can show, step by step, that indeed $\lim\limits_{n\to\infty} = \infty$ by using several subtitutions: for any $n > n_0$ we get that

$$a_n = \frac{n}{2} \overset{\text{since } n > n_0}{>} \frac{n_0}{2} = \frac{\lceil 2R \rceil}{2} \underset{\text{since } \lceil 2R \rceil \geq 2R}{\geq} \frac{2R}{2} = R,$$

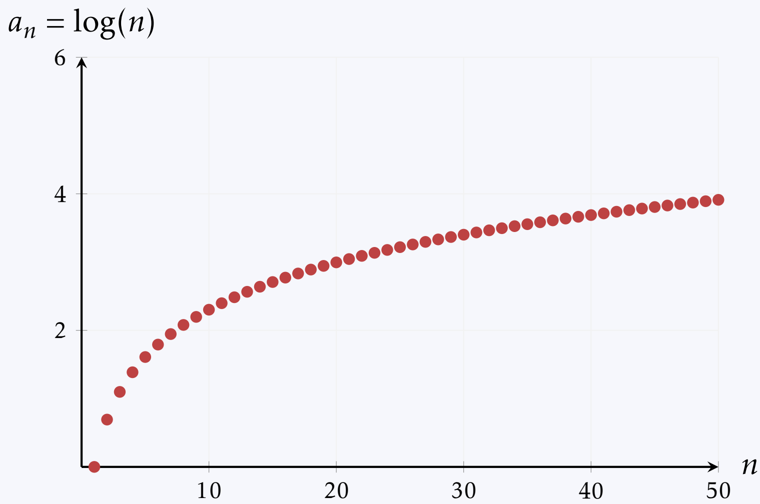
or simply $a_n > R$.

Example 1.7 The sequence $\log(n)$

Let's show that the sequence $a_n = \log(n)$ (see graph below) goes to infinity as n increases: first, we note that $\log(n)$ is a strictly increasing sequence bounded from below by $\underline{M} = 0$. Now, given some positive $R \in \mathbb{R}$ we chose $n_0 = \lceil e^R \rceil$ and thus get that for any $n > n_0$

$$a_n > a_{n_0} = \log(n_0) = \log\left(\lceil e^R \rceil\right) \geq \log\left(e^R\right) = R.$$

altogther we get $a_n > R$, and therefore $\lim\limits_{n\to\infty} \log(n) = \infty$.



Note: this proof works because $\log(n)$ is a *strictly* increasing sequence. If it were only an increasing sequence we would not be guaranteed that $n > n_0$ means that $a_n > a_{n_0}$, and the entire process would not yield that $\log(n)$ always passes any given real number R. Indeed, by naively looking at the graph above we may be mistaken to think that $\log(n)$ actually approaches some finite number, say 4.5 or so, and doesn't go to infinity. This is of course not the case.

Example 1.8 A sequence which goes to negative infinity

Let us now show that the sequence $a_n = -\sqrt{n}$ goes to negative infinity. We first note that $-\sqrt{n}$ is always negative, is decreasing and bouned from above by $\overline{M} = 0$. For any given negative $R \in \mathbb{R}$ we choose $n_0 = \lfloor R^2 \rfloor$, and thus for any $n > n_0$ we get

$$a_n < a_{n_0} = -\sqrt{n_0} = -\sqrt{\lfloor R^2 \rfloor} \leq -\sqrt{R^2} = -|R| = R.$$

! **To be written**: is this example actually necessery? It seems reduntant unless we show somethin
!

The next type of sequences we ananlyze are those sequences that converge to a real number L as n increases - i.e. as n increases, the terms of the sequence get closer and closer to L. A classical example for such a sequence is $a_n = \frac{1}{n}$ (Figure 1.3): as n increases, the terms $\frac{1}{n}$ become smaller and smaller (while always being positive) and the sequence as a whole approaches $L = 0$.

Sequences don't have to approach a limit from above: some sequences approach a limit from below (Figure 1.8), while others may **oscillate** around the limit (Figure 1.9).

We define convergence in a similar way to how we defined that a sequence goes to ∞: in that case we had to show that for any $R \in \mathbb{R}$ the sequence eventually surpasses R and

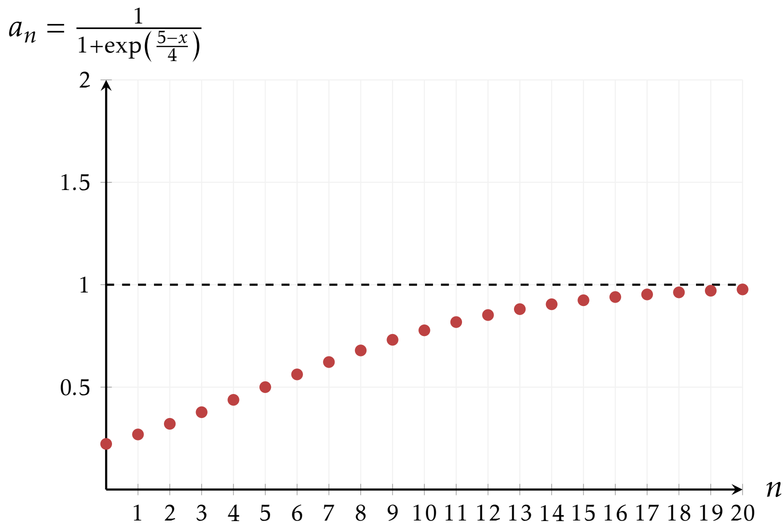


Figure 1.8 The logistic sequence $a_n = \frac{1}{1+\exp\left(\frac{5-x}{4}\right)}$ approaches the value $L = 1$ from below as $n \to \infty$.

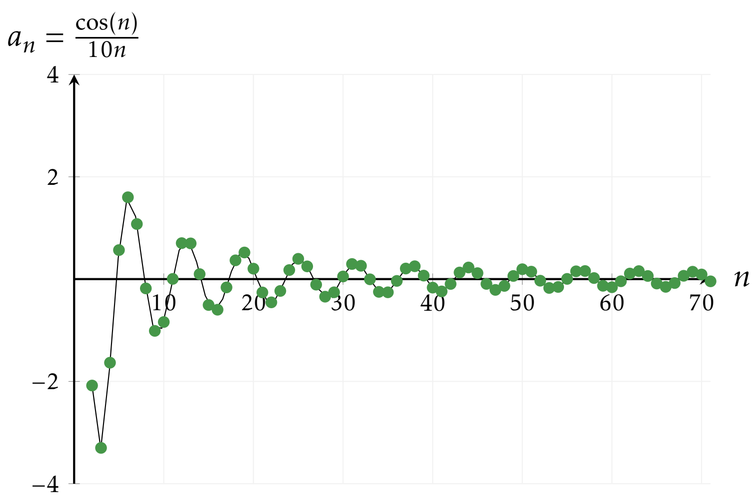


Figure 1.9 The sequence $a_n = \frac{\cos(n)}{10n}$ approaches, with oscillations, the limit $L = 0$. A line connecting the elements is drawn to help see the progression of the sequence.

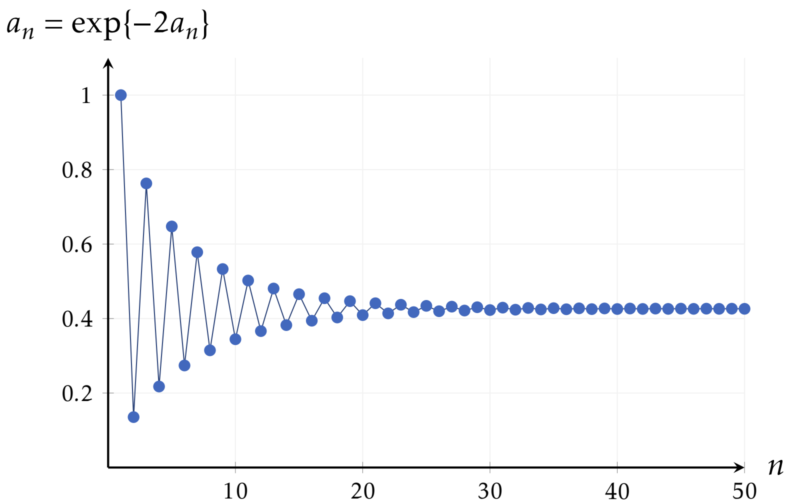


Figure 1.10 The sequence defined by the recursive formula $a_{n+1} = \exp(-2a_n)$ with $a_1 = 1$ converges to the limit $L \approx 0.4263027510068627$, or more precisely $\frac{1}{2}W(2)$, where W is the Lambert W function.

no element is ever again equal to or smaller than R. In the case of convergence to some finite number $L \in \mathbb{R}$ we will show that for any distance ε from L - no matter how small! - the sequence eventually stays within ε of L.

Let us use the simplest convergence example to explain this idea: $a_n = \frac{1}{n}$, which converges to $L = 0$. Given a small number ε, say $\varepsilon = \frac{1}{10}$, eventually the sequence stays at most within $\frac{1}{10}$ of $L = 0$. This happens starting from $n_0 = 10$: any element thereafter is smaller than $\frac{1}{10}$, which means that it is withing $\varepsilon = \frac{1}{10}$ of $L = 0$ (see Figure 1.11).

We can repeat this for different value of ε: given $\varepsilon = \frac{1}{100}$, for any $n > 100$ the elements a_n are guaranteed to be within $\pm\frac{1}{100}$ of $L = 0$. Given $\varepsilon = \frac{1}{5000}$, for any $n > 5000$ the elements a_n are within $\pm\frac{1}{5000}$ of $L = 0$, etc. In general, given any real ε, no matter how small, we can set $n_0 = \lceil \frac{1}{\varepsilon} \rceil$, and then for any $n > n_0$ we get that

$$a_n \overset{\text{since } a_n \text{ is strictly decreasing}}{<} a_{n_0} = \frac{1}{n_0} = \frac{1}{\lceil \frac{1}{\varepsilon} \rceil} \underset{\text{since } \lceil \frac{1}{x} \rceil \geq \frac{1}{x}}{\leq} \frac{1}{\frac{1}{\varepsilon}} = \varepsilon, \tag{1.1.5}$$

i.e. altogther

$$a_n < \varepsilon, \tag{1.1.6}$$

and therefore a_n is within $\pm\varepsilon$ of $L = 0$.

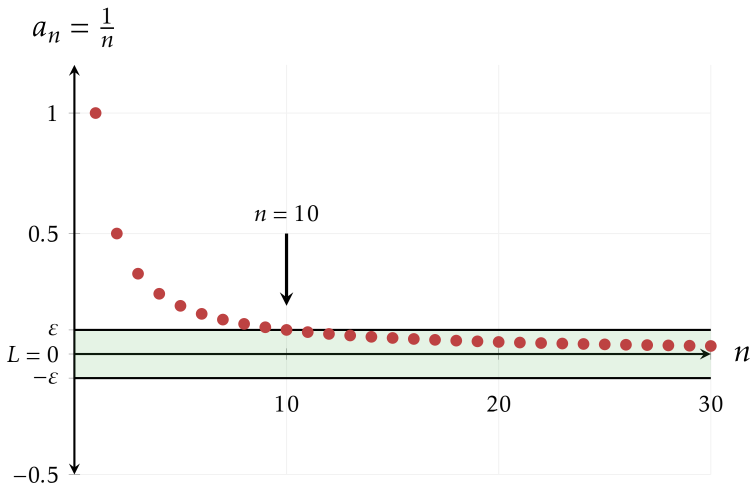


Figure 1.11 The sequence $a_n = \frac{1}{n}$. For any $n > 10$, the element a_n is withing $\varepsilon = \frac{1}{10}$ of the limit $L = 0$. The interval $(-\varepsilon, \varepsilon) = \left(-\frac{1}{10}, \frac{1}{10}\right)$ on the y-axis is highlighted in green.

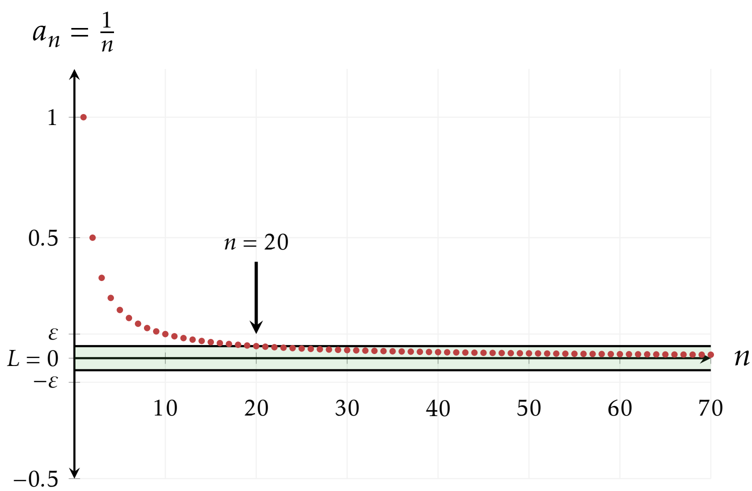


Figure 1.12 Same as Figure 1.11 except here $\varepsilon = 0.05$, and thus $n_0 = \lceil \frac{1}{0.05} \rceil = 20$. Therefore, starting fom $n = 20$ all the elements of the sequence are within the interval $(-0.05, 0.05)$. Note: the values of the sequence are drawn with smaller filled circles to prevent overlap between subsequent points.

Example 1.9 Convergence

Let's prove that the sequence $a_n = \frac{n+1}{n^2}$ converges to $L = 0$ as $n \to \infty$. For any $\varepsilon \in \mathbb{R}$ we chose $n_0 = \lceil \frac{\varepsilon}{2} \rceil$. Since a_n is a strictly decreasing sequence[a], we get that for any $n > n_0$:

$$a_n < a_{n_0} = \frac{n_0 + 1}{n_0^2}.$$

Note the following: for any $x > 1$, $x^3 > x > 1$. Therefore if we replace both n_0 and 1 in the numerator of the expression $\frac{n_0+1}{n_0^2}$ by n_0^3 we are guaranteed to get a number bigger than $\frac{n_0+1}{n_0^2}$. We can therefore continue with the substitution:

$$\frac{n_0 + 1}{n_0^2} < \frac{n_0^3 + n_0^3}{n_0^2} = \frac{2n_0^3}{n_0^2} = 2n_0 = \cancel{2}\frac{\varepsilon}{\cancel{2}} = \varepsilon,$$

and therefore altogther we get that

$$a_n < \varepsilon,$$

and thus a_n converges to $L = 0$ as n increases to ∞.

[a] Show that!

1.2 VECTORS

1.2.1 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 1.5 Real vectors

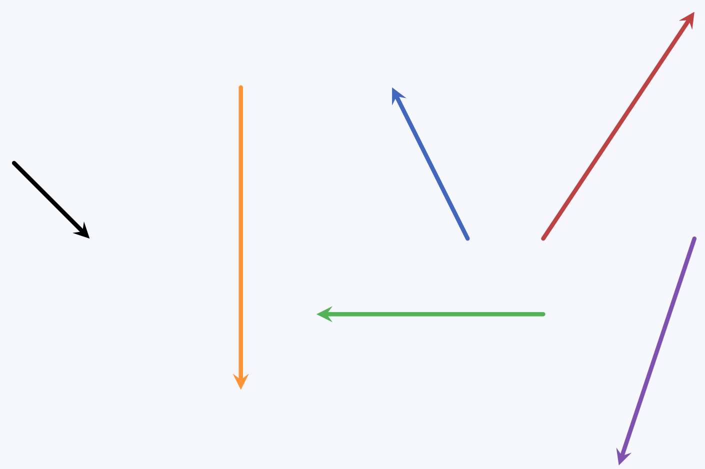
A *real vector* is an object with a **magnitude** (also called **norm**) and a **direction**.

π

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

Example 1.10 Real vectors

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



Vectors are usually denoted in one of the following ways:

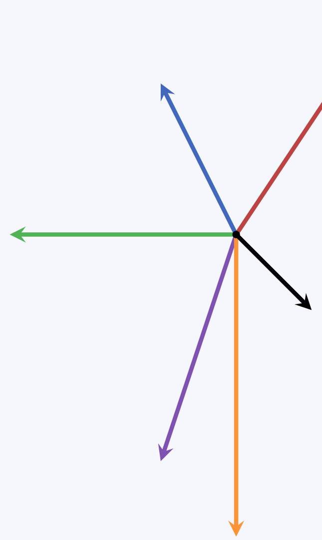
- **Arrow above letter**: $\vec{u}, \vec{v}, \vec{x}, \vec{a}, \ldots$
- **Bold letter**: $\boldsymbol{u}, \boldsymbol{v}, \boldsymbol{x}, \boldsymbol{a}, \ldots$
- **Bar below letter**: $\underline{u}, \underline{v}, \underline{x}, \underline{a}, \ldots$

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 1.11 Real vectors

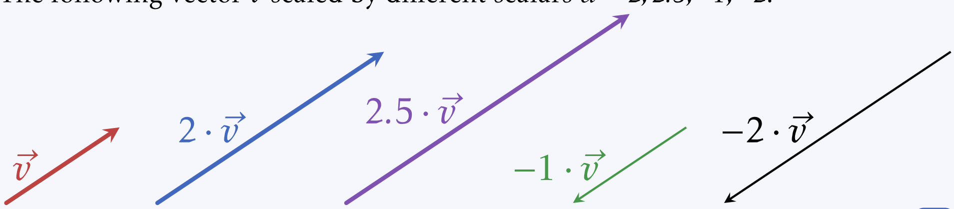
The vectors from the previous translated (moved) such that their origins all lie on the same point:



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 1.12 Scaling vectors

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



Note 1.5 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 chapter we use the following notation for the norm of a vector $\vec{v}$: $\left\|\vec{v}\right\|$.

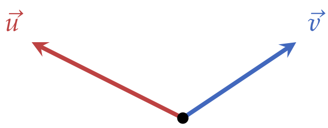
A vector $\vec{v}$ with norm $\left\|\vec{v}\right\| = 1$ is called a **unit vector**, and is usually denoted by replacing the arrow symbol by a hat symbol: $\hat{v}$. Any vector (except $\vec{0}$) can be scaled into a unit vector by scaling the vector by 1 over its own norm, i.e.

$$\hat{v} = \frac{1}{\left\|\vec{v}\right\|}\vec{v}. \tag{1.2.1}$$

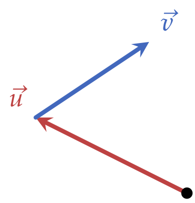
The result of normalization is a vector of unit norm which points in the same direction of the original vector.

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 1.13):

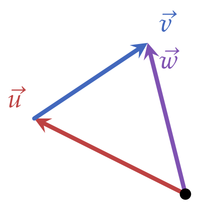
1. Move (translate) $\vec{v}$ such that its origin lies on the head of $\vec{u}$.
2. The vector $\vec{w}$ is the vector drawn from the origin of $\vec{u}$ to the head of $\vec{v}$.



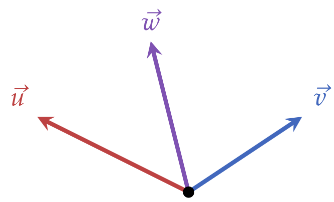
1 The vectors $\vec{u}$ and $\vec{v}$.



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 1.13 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 1.14: 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.

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 $\vec{v}$,

$$\vec{v}+\vec{0}=\vec{v}. \tag{1.2.2}$$

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

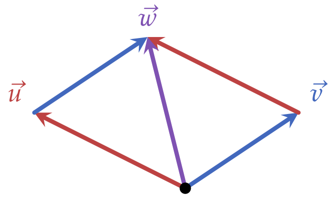


Figure 1.14 The parallogram law of vector addition.

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.

$$\vec{v}+\left(-\vec{v}\right)=\vec{0}. \tag{1.2.3}$$

1.2.2 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 1.15), 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:

$$\vec{v}=\begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}. \tag{1.2.4}$$

Note 1.6 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.

!

Example 1.13 Vector components in two dimensions

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

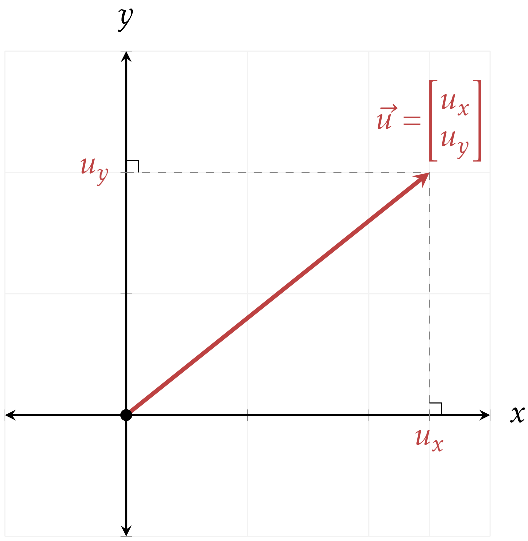
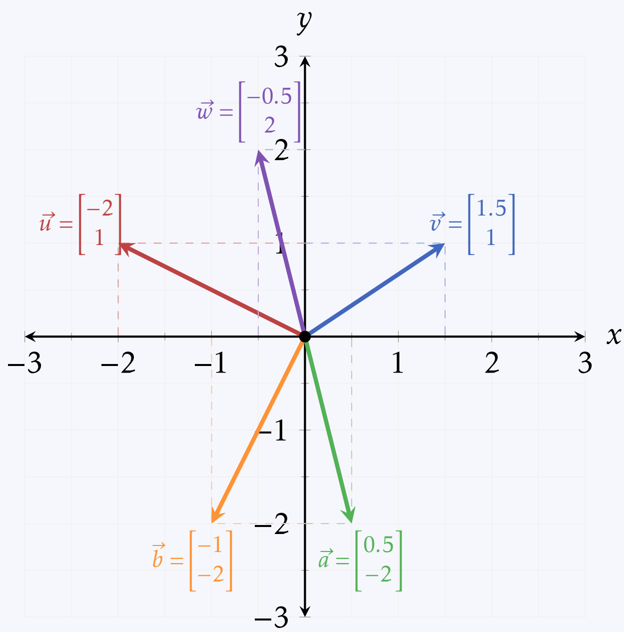
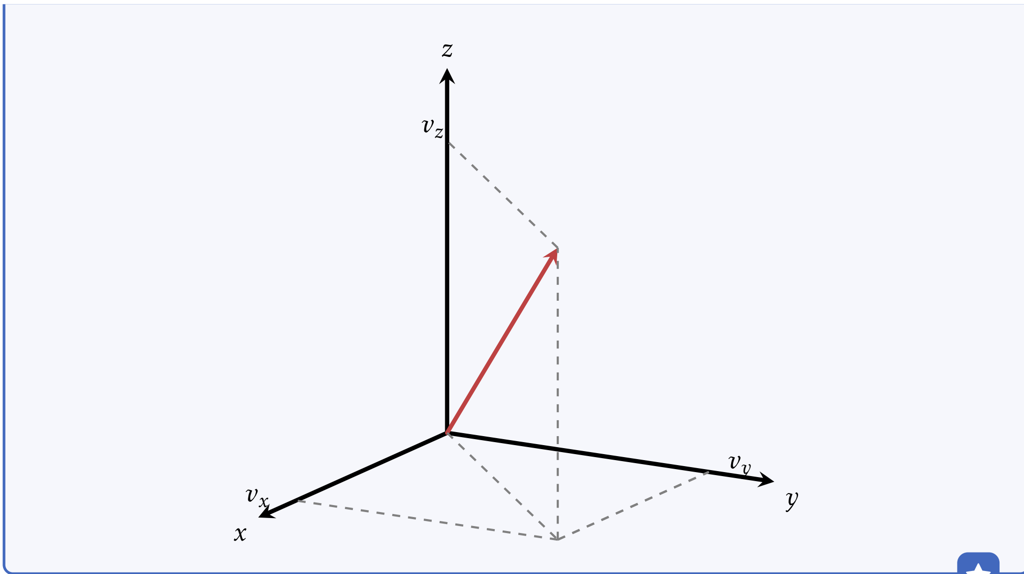


Figure 1.15 Placing a 2-dimensional vector $\vec{u}$ on the 2-dimensional Cartesian coordinate system, showing its x- and y-components.



Example 1.14 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. $(v_1, v_2, \dots, v_n)$. 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 clearer in **??** and will be used to its fullest extent in future chapters when discussing **covariant vectors**, **contravarient vectors**, and in general **tensors**. For now, we stick with the column form of vectors to stay consistent with common notation.

However, the row form of vectors hints at 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.

$$\mathbb{R}^n = \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n}. \tag{1.2.5}$$

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.

1.2.3 Norm, polar coordinates and spherical coordinates

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 1.16), we can use the Pythagorean theorem to calculate the norm of the vector, which is equal to the hypotenous of said triangle:

$$\left\|\vec{v}\right\| = \sqrt{v_x^2 + v_y^2}. \tag{1.2.6}$$

Much like complex numbers, vectors in $\mathbb{R}^2$ can be expressed using **polar coordinates**, i.e. using the norm of the vector and its angle θ relative to the x-axis (cf. **??** and **??**). The relation between the cartesian and polar coordinates is

$$\begin{aligned} v_x &= \left\|\vec{v}\right\| \cos(\theta), \\ v_y &= \left\|\vec{v}\right\| \sin(\theta). \end{aligned} \tag{1.2.7}$$

To calculate θ from v_x and v_y we use the definition of $\tan(\theta)$ (see **??**), and get that

$$\tan(\theta) = \frac{v_y}{v_x}, \tag{1.2.8}$$

i.e.

$$\theta = \arctan\left(\frac{v_y}{v_x}\right). \tag{1.2.9}$$

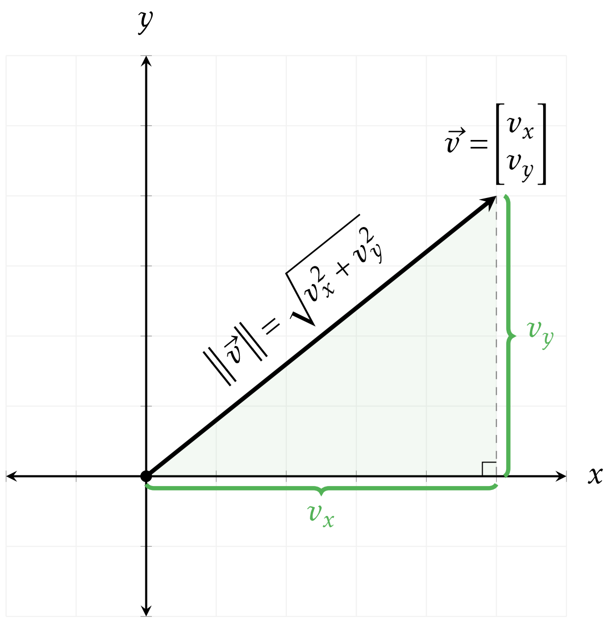


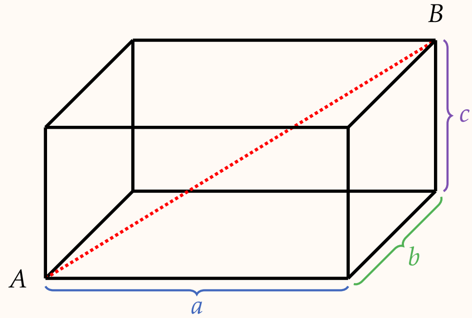
Figure 1.16 Calculating the norm of a 2-dimensional column vector.

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

$$\left\|\vec{v}\right\| = \sqrt{v_x^2 + v_y^2 + v_z^2}. \tag{1.2.10}$$

Challenge 1.2 Norm of a 3D vector

Show why Equation 1.2.10 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:

$$\left\|\vec{v}\right\| = \sqrt{v_1^2 + v_2^2 + v_3^2 + \cdots + v_n^2} = \sqrt{\sum_{i=1}^{n} v_i^2}. \tag{1.2.11}$$

Note 1.7 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.

!

$\mathbb{R}^3$ has its own version of polar coordinates, sometimes refered to as **cylindrical coordinates**. These coordinates are similar to the polar coordinates in $\mathbb{R}^2$, with an additional "height" component: the three coordinates are ρ, φ and z, where

- ρ is the norm of the projection of $\vec{v}$ onto the xy-plane[1],
- φ is the angle between the projection of $\vec{v}$ and the x-axis.
- z is the distance between the head of $\vec{v}$ to the xy-plane.

The conversion between cylindrical and cartesian coordinates is given by

$$\begin{aligned} x &= \rho\cos\left(\varphi\right), \\ y &= \rho\sin\left(\varphi\right), \\ z &= z. \end{aligned} \tag{1.2.12}$$

[1] ρ is used instead of r to prevent confusion with the polar coordinates in $\mathbb{R}^2$

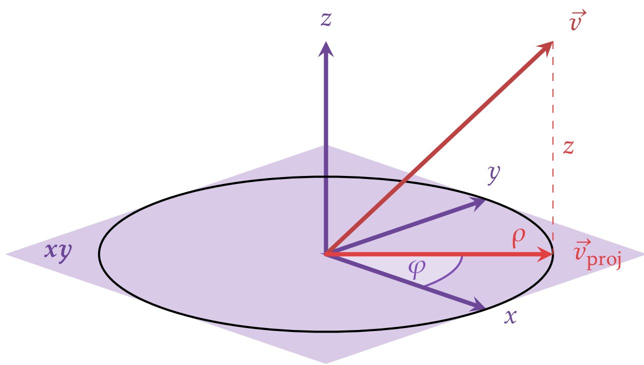


Figure 1.17 The cylindrical coordinates ρ, φ, z.

Yet another useful set of coordinates in $\mathbb{R}^3$ are the **spherical coordinates**. Given a vector $\vec{v}$, instead of using two length coordinates, the spherical coordinate system uses two angles φ and θ: φ is the angle between the projection of $\vec{v}$ onto the xy-plane, and θ the angle between $\vec{v}$ and the z-axis. The third coordinate is then the norm of $\vec{v}$, denoted r. See **??** for a graphical representation.

! **To be written**: a nice figure for spherical coordinates !

1.2.4 Operations

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.

$$\alpha\vec{v} = \begin{bmatrix} \alpha v_1 \\ \alpha v_2 \\ \vdots \\ \alpha v_n \end{bmatrix}. \tag{1.2.13}$$

We can prove Equation 1.2.13 by directly calculating the norm of a scaled vector $\vec{w} = \alpha\vec{v}$:

Proof 1.3 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 1.2.1), by showing that dividing each component of a vector by its norm gives a vector of unit norm:

Proof 1.4 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 $\frac{1}{\left\| \vec{v} \right\|}$ 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 1.15 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.

$$\vec{u} + \vec{v} = \begin{bmatrix} u_1 + v_1 \\ u_2 + v_2 \\ \vdots \\ u_n + v_n \end{bmatrix}. \tag{1.2.14}$$

! **To be written**: how this addition is the same as the one shown in Figure 1.13. !

Note 1.8 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.

1.2.5 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 1.6 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

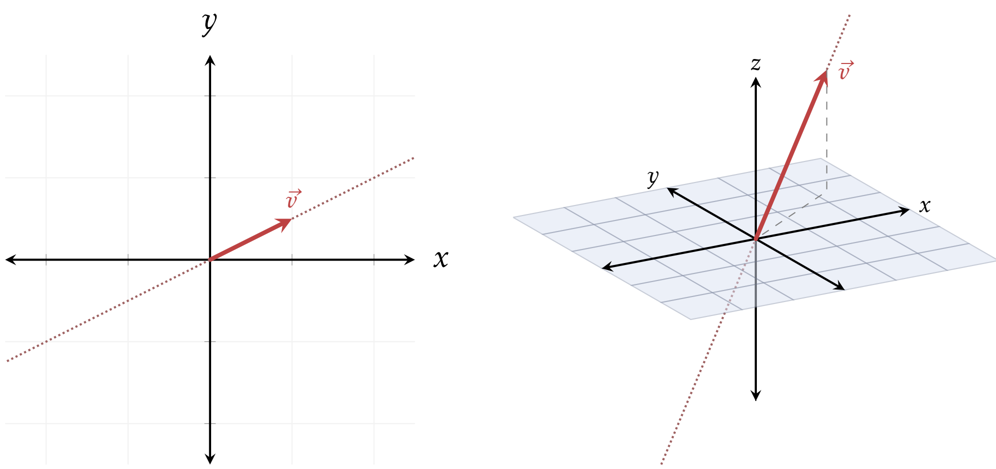
$$\vec{w} = \alpha_1 \vec{v}_1 + \alpha_2 \vec{v}_2 + \cdots + \alpha_n \vec{v}_n = \sum_{i=1}^{n} \alpha_i \vec{v}_i. \tag{1.2.15}$$

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.

$$V = \left\{ \alpha \vec{v} \,|\, \alpha \in \mathbb{R} \right\}. \tag{1.2.16}$$

The set V represents a line in the direction of $\vec{v}$ going through the origin (see Figure 1.18). 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 1.18 The span of a single vector $\vec{v}$, shown as a dashed line: in $\mathbb{R}^2$ (left) and $\mathbb{R}^3$ (right).

Similarily, 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 $\vec{v} = \alpha \vec{u}$),

$$V = \left\{ \alpha \vec{u} + \beta \vec{v} \,|\, \alpha, \beta \in \mathbb{R} \right\}, \tag{1.2.17}$$

is a plane that goes through the origin (see Figure 1.19). Such vectors are also said to be **non-collinear**.

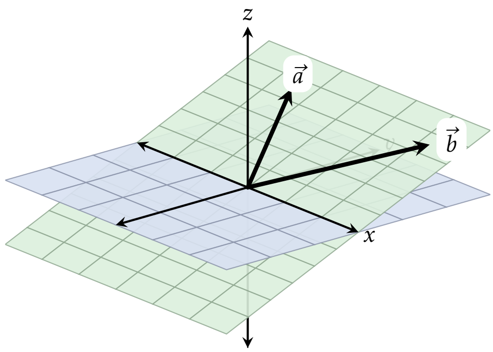


Figure 1.19 Two vectors $\vec{a}$ and $\vec{b}$ span a plane (colored green) in $\mathbb{R}^3$. The xy-plane (i.e. $z = 0$) is shown in blue for emphasis.

Example 1.16 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

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

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 1.5 $\mathbb{R}^2$ 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

$$\vec{u} = \alpha\vec{v} \tag{1.2.18}$$

has no solution, i.e. the system

$$\begin{cases} u_x = \alpha v_x \\ u_y = \alpha v_y \end{cases} \tag{1.2.19}$$

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

$$u_x v_y \neq u_y v_x. \tag{1.2.20}$$

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

$$\begin{cases} w_x = \alpha u_x + \beta v_x \\ w_y = \alpha u_y + \beta v_y \end{cases} \tag{1.2.21}$$

Isolating α using the first equation yields

$$\alpha = \frac{w_x - \beta v_x}{u_x}, \tag{1.2.22}$$

and subtituting it into the second equation yields

$$\beta = \frac{w_y - \alpha u_y}{v_y} = \frac{w_y - \frac{w_x - \beta v_x}{u_x}}{v_y}, \tag{1.2.23}$$

which rearranges into

$$\beta = \frac{u_x w_y - u_y w_x}{u_x v_y - u_y v_x}, \tag{1.2.24}$$

and thus

$$\alpha = \frac{-v_x w_y + v_y w_x}{u_x v_y - u_y v_x}. \tag{1.2.25}$$

We can see that α and β exist iff $u_x v_y \neq u_y v_x$, which is guaranteed by Equation 1.2.20. 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 1.7 Linear dependent set of vectors

A set of n vectors
$$S = \left\{\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n\right\} \tag{1.2.26}$$
is said to be linearly dependent if there exist a linear combination
$$\alpha_1\vec{v}_1 + \alpha_2\vec{v}_2 + \cdots + \alpha_n\vec{v}_n = \vec{0}, \tag{1.2.27}$$
and **at least** one the coefficients $\alpha_i \neq 0$.

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

Example 1.17 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 1.18 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 1.19 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:

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

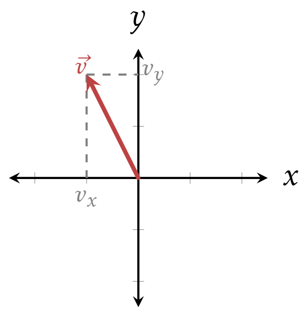
i.e.

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

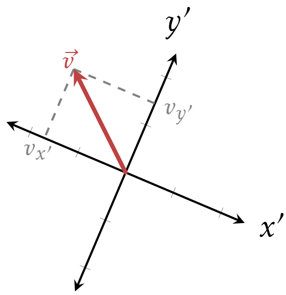
(you, the reader, should verify this!)

The way we described what a basis set is, while being accurate and general, does not give us any particular intuition about what basis sets actually *do*, and why do we even bother with them. To understand this, consider some vector $\vec{v} \in \mathbb{R}^2$. Without defining some frame of reference, as far as we're considered $\vec{v}$ is merely some arrow floating in space:

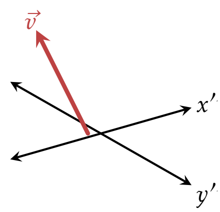
Note that $\vec{v}$ still has all the properties any other general vector has: it a norm and a direction. However, we can't say anything meaningful about this direction, except maybe that it is roughly pointing up and to left. In order to make any sense of $\vec{v}$ we have to choose some frame of reference, i.e. two axes. We can of course use the usual horizontal and vertical directions (which we usually call x and y):



Having this frame of reference, i.e. the x- and y-axes, we can calculate the components of $\vec{v}$ **in relation to these axes** by droping two perpendicular lines, one for each axis. But there's nothing really special about these axes, they are just convinent to draw on a flat paper. We could use any other two non-colinear directions, for example the following x' and y':



Notice how $\vec{v}$ stays the same, the only difference is how we will describe its components using the x', y' axes system. We are of course not restricted to having two perpendicular axes, e.g. the following x'', y'' axes:



! **To be written**: improve the above figure !

The axis system we use as a reference is the basis set we use to describe vectors, except one detail: a basis set also tells us what is the unit of measurement in the direction of each basis vector. This is of course the norm of that basis vector.

! **To be written**: show a vector drawn from integer amounts of 2 basis vectors !

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

Definition 1.8 Basis sets

Let B be a **linearly independent set** of vectors in $\mathbb{R}^n$. If any vector $\vec{v} \in \mathbb{R}^{\propto}$ 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.

π

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

$$\dim(V) = |B|. \tag{1.2.28}$$

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 1.6 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 1.20, 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 $\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \right\}$ the coordinates of $\vec{v}$ are $(1, -3, 7)$ (as we will see next, it is not a

$$\vec{v} = \begin{bmatrix}1\\-3\\7\end{bmatrix} \xrightarrow{B_1} 1\begin{bmatrix}1\\0\\0\end{bmatrix} - 3\begin{bmatrix}0\\1\\0\end{bmatrix} + 7\begin{bmatrix}0\\0\\1\end{bmatrix}$$

$$\vec{v} = \begin{bmatrix}1\\-3\\7\end{bmatrix} \xrightarrow{B_2} 9\begin{bmatrix}5\\1\\2\end{bmatrix} - 23\begin{bmatrix}0\\1\\0\end{bmatrix} - 11\begin{bmatrix}4\\-1\\1\end{bmatrix}$$

$$\vec{v} = \begin{bmatrix}1\\-3\\7\end{bmatrix} \xrightarrow{B_3} 1.4\begin{bmatrix}-1\\0\\2\end{bmatrix} + 0.3\begin{bmatrix}0\\2\\-3\end{bmatrix} + 1.2\begin{bmatrix}2\\2\\3\end{bmatrix}$$

Figure 1.20 The vector $\vec{v} = \begin{bmatrix}1\\-3\\7\end{bmatrix}$ spanned in three different basis sets.

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 1.21): 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 basis 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 1.20 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):

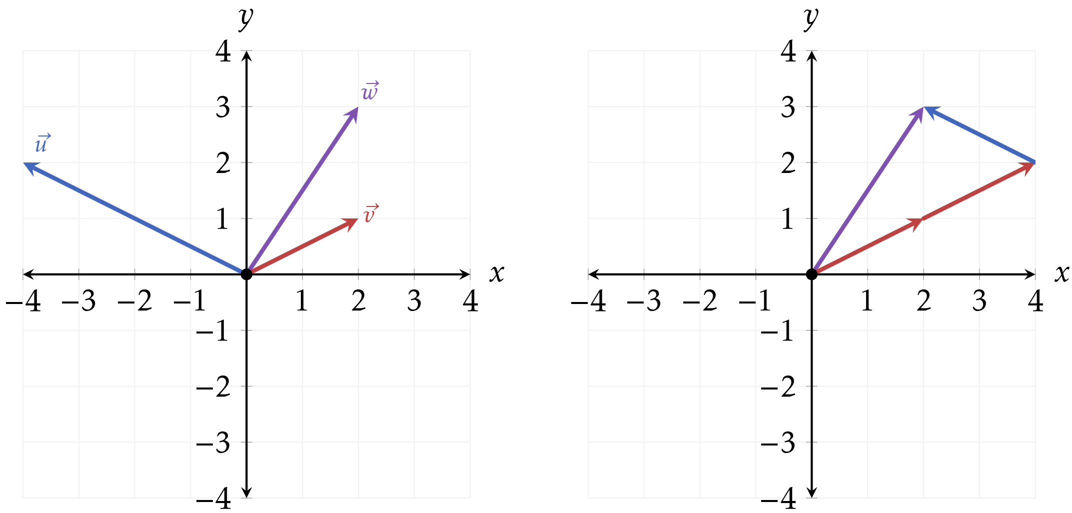
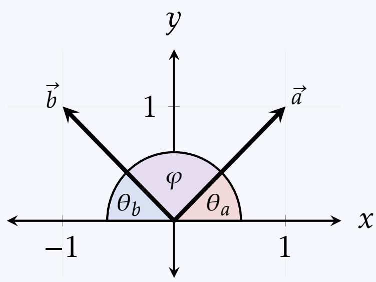


Figure 1.21 The vector $\vec{w}=\begin{bmatrix}2\\3\end{bmatrix}$ is spanned using the vectors $\vec{u}=\begin{bmatrix}2\\1\end{bmatrix}$ and $\vec{v}=\begin{bmatrix}-4\\2\end{bmatrix}$, yielding the coordinates $\left(2,\frac{1}{2}\right)$ in the basis set B.



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 $\mathbb{R}^2$.

Challenge 1.3 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 1.20: 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

$$B = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \right\}, \tag{1.2.29}$$

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 $\hat{k}$.

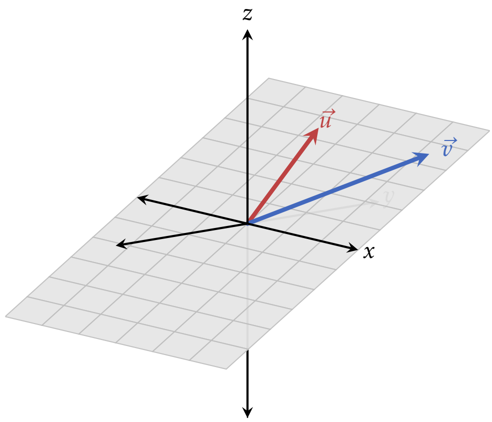


Figure 1.22 The angle between two linearly independent vectors lies on the plane spanned by the vectors.

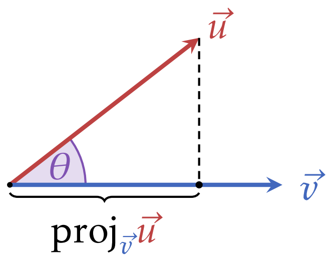


Figure 1.23 The projection of a vector $\vec{u}$ onto another vector $\vec{v}$ in the plane spanned by the two vectors.

1.2.6 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 1.22).

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 1.23). 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

$$\cos(\theta) = \frac{\text{proj}_{\vec{v}}\vec{u}}{\left\|\vec{u}\right\|}. \tag{1.2.30}$$

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

$$\vec{u} \cdot \vec{v} = \text{proj}_{\vec{v}} \vec{u} \cdot \left\| \vec{v} \right\|. \tag{1.2.31}$$

Subtituting Equation 1.2.30 into Equation 1.2.31 gives a very nice relation between the scalar product of two vectors and the angle between them:

$$\cos(\theta) = \frac{\vec{u} \cdot \vec{v}}{\left\| \vec{u} \right\| \left\| \vec{v} \right\|}. \tag{1.2.32}$$

The angle between the two vectors is then isolated by applying the arccos function on the right-hand side of Equation 1.2.32. A common form of this equation is the following:

$$\vec{u} \cdot \vec{v} = \left\| \vec{u} \right\| \left\| \vec{v} \right\| \cos(\theta). \tag{1.2.33}$$

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 1.21 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 1.20, the angle between $\vec{a}$ and $\vec{b}$ is $\frac{\pi}{2}$. Therefore, their scalar product is

$$\begin{aligned} \vec{a} \cdot \vec{b} &= \left\| \vec{a} \right\| \left\| \vec{b} \right\| \cos(\theta) \\ &= \sqrt{2}\sqrt{2} \cos\left(\frac{\pi}{2}\right) \\ &= 2 \cdot 0 = 0. \end{aligned}$$

Example 1.22 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

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

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. $\left(\alpha\vec{v}\right) \cdot \vec{u} = \vec{v} \cdot \left(\alpha\vec{u}\right) = \alpha\left(\vec{u} \cdot \vec{v}\right)$.
- It equals zero in only one of two cases:
  1. One of the vectors (or both) is the zero vector, or
  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 **orthogonal** 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[2].

[2] 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

$$\vec{u} \cdot \vec{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n = \sum_{k=1}^{n} u_i v_i. \tag{1.2.34}$$

Example 1.23 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 1.2.34).

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 1.24 Scalar product of two vectors - algebraicly

Calculate the scalar product $\vec{u} \cdot \vec{v}$ from Example 1.22 using Equation 1.2.34.

Solution:

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

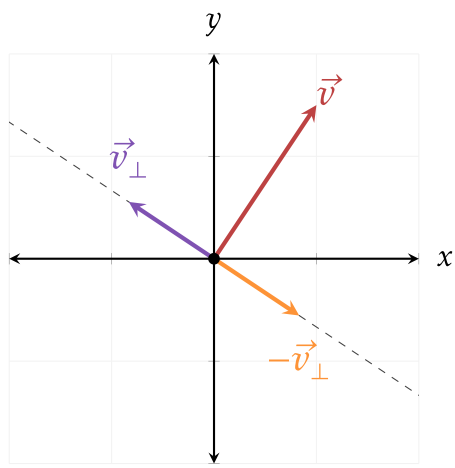


Figure 1.24 A vector $\vec{v}$ and its orthogonal direction, signified by a dashed line. Two vectors $\vec{v}^\perp$ and $\vec{v}^\perp$ are drawn on the orthogonal direction.

exactly the result we got in Example 1.22.

For any given a 2-dimensional vector $\vec{v} = \begin{bmatrix} x \\ y \end{bmatrix}$ there is only a single orthogonal direction (Figure 1.24). We can use Equation 1.2.34 to find a general formula for a vector $\vec{v^\perp}$ representing this direction:

$$0 = \vec{v} \cdot \vec{v}^\perp = \begin{bmatrix} x \\ y \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = xa + yb.$$

The solution for the above equation is the vector

$$\vec{v}^\perp = \begin{bmatrix} -y \\ x \end{bmatrix}. \tag{1.2.35}$$

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

$$\vec{v} \cdot \vec{v} = v_1 v_1 + v_2 v_2 + \cdots + v_n v_n = v_1^2 + v_2^2 + \cdots + v_n^2 = \left\| \vec{v} \right\|^2. \tag{1.2.36}$$

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

$$\left\| \vec{v} \right\| = \sqrt{\vec{v} \cdot \vec{v}}. \tag{1.2.37}$$

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

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 $\hat{e}_i$:

$$v_i = \operatorname{proj}_{\hat{e}_i}\vec{v}, \tag{1.2.38}$$

and thus the component $v_i\hat{e}_i = \left(\operatorname{proj}_{\hat{e}_i}\vec{v}\right)\hat{e}_i$ 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:

$$\vec{v} = \sum_{i=1}^{n}\left(\operatorname{proj}_{\hat{b}_i}\vec{v}\right)\hat{b}_i. \tag{1.2.39}$$

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 1.2.31 we can re-write Equation 1.2.39 as

$$\vec{v} = \sum_{i=1}^{n}\left(\vec{v}\cdot\hat{b}_i\right)\hat{b}_i. \tag{1.2.40}$$

Example 1.25 Decomposing a vector

EXAMPLE TBD

1.2.7 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.

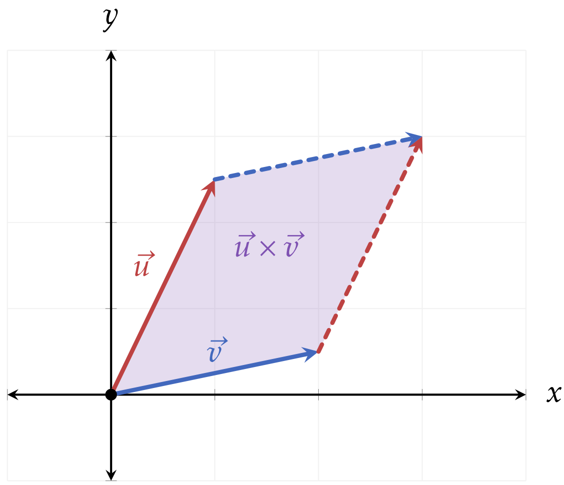


Figure 1.25 The cross product in $\mathbb{R}^2$ of two vectors $\vec{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} c \\ d \end{bmatrix}$ as the signed area of the parallogram defined by the vectors.

We start with the definition of the cross product in $\mathbb{R}^2$: the cross product of two vectors $\vec{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} c \\ d \end{bmatrix}$ is the (signed) area of the parallelogram defined by the two vectors (see Figure 1.25).

The value of the parallelogram defined by $\vec{u}$ and $\vec{v}$ is

$$\vec{u} \times \vec{v} = \|\vec{u}\| \|\vec{v}\| \sin(\theta), \tag{1.2.41}$$

where θ is the angle between the vectors. This is extremely similar to the scalar product, and we can use this fact to find how to calculate the cross product from vectors in column form: if we replace $\vec{u}$ by a vector orthogonal to it, denoted by $\vec{u}^{\perp}$, the cross product is then

$$\vec{u} \times \vec{v} = \|\vec{u}^{\perp}\| \|\vec{v}\| \sin\left(\theta + \frac{\pi}{2}\right), \tag{1.2.42}$$

since the angle between $\vec{u}^{\perp}$ and $\vec{v}$ is $\frac{\pi}{2}$ more than that between $\vec{u}$ and $\vec{v}$. Using the fact that $\sin\left(\theta + \frac{\pi}{2}\right) = \cos(\theta)$, we get the equality

$$\begin{aligned} \vec{u} \times \vec{v} &= \|\vec{u}^{\perp}\| \|\vec{v}\| \sin\left(\theta + \frac{\pi}{2}\right) \\ &= \|\vec{u}^{\perp}\| \|\vec{v}\| \cos(\theta) \\ &= \vec{u}^{\perp} \cdot \vec{v}. \end{aligned} \tag{1.2.43}$$

In $\mathbb{R}^2$, any vector $\vec{u} = \begin{bmatrix} a \\ b \end{bmatrix}$ has two vectors orthogonal to it: $\begin{bmatrix} -b \\ a \end{bmatrix}$ and $\begin{bmatrix} b \\ -a \end{bmatrix}$. Choosing the former gives

$$\vec{u} \times \vec{v} = \begin{bmatrix} -b \\ a \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = -bc + ad, \tag{1.2.44}$$

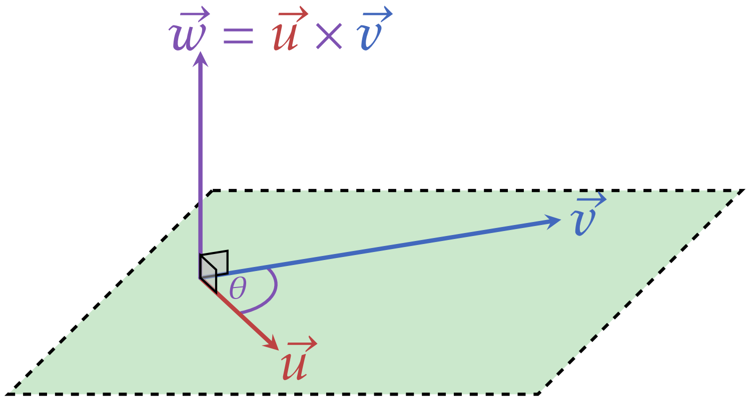


Figure 1.26 The cross product of the vectors $\vec{u}$ and $\vec{v}$ relative to the plane spanned by the two vectors.

while choosing the latter gives

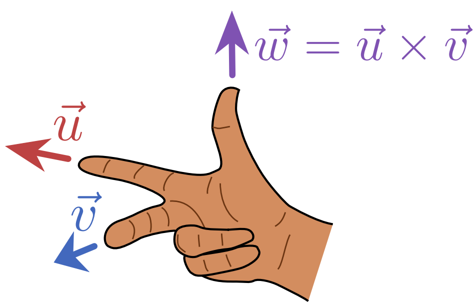
$$\vec{u} \times \vec{v} = \begin{bmatrix} b \\ -a \end{bmatrix} \cdot \begin{bmatrix} c \\ d \end{bmatrix} = bc - ad. \tag{1.2.45}$$

These two forms are the opposite of each other - i.e. if one yields the value 4, the other yields the value -4. We will see which one is used in a moment.

On to $\mathbb{R}^3$: 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 with norm of the same magnitude as the product would have in $\mathbb{R}^2$, i.e.

$$\left\|\vec{w}\right\| = \left\|\vec{u}\right\|\left\|\vec{v}\right\|\sin(\theta). \tag{1.2.46}$$

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:

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

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

$$\vec{u} \times \vec{v} = \begin{pmatrix} u_y v_z - u_z v_y \\ u_z v_x - u_x v_z \\ u_x v_y - u_y v_x \end{pmatrix} \tag{1.2.47}$$

Note 1.9 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:

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

!

Challenge 1.4 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 $\vec{b}$.

?

1.2.8 The Gram–Schmidt process

While all basis sets of a given space are equaly good at spanning that space[3], as humans we sometimes prefer using orthonormal basis sets due to their nice properties. One such property of orthonormal basis sets, which we will use in a later section, is that the scalar product of any two vectors of the set is the Kronecker's delta - i.e. given an orthonormal basis set

$$B = \left\{\vec{b}_1, \vec{b}_2, \dots, \vec{b}_n\right\},$$

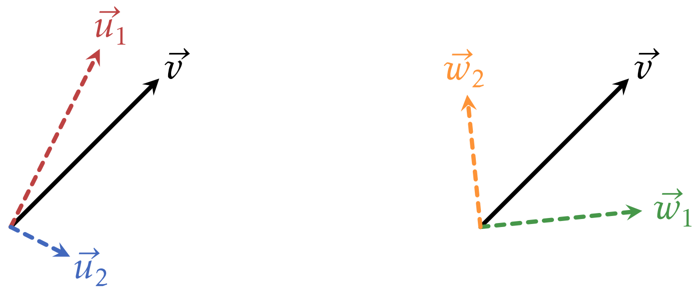
for any two basis vectors $\vec{b}_i$ and $\vec{b}_j$,

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

However, most basis sets are not orthonormal[4]. How can we construct an orthonormal basis set from a given basis set?

[3] they're good basis sets Brent

[4] since there are countlessly infinitely many basis sets for any space, the meaning of "most" in this context is that the probability that a random basis set is not orthonormal is greater than the probability that it is orthonormal.



1 2

Figure 1.27 The same vector $\vec{v}$ decomposed into two sets of orthogonal components: (a) u_1 and u_2; (b) w_1 and w_2. There are infinitely many such orthogonal sets on any plane containing $\vec{v}$.

In the unlikely case that the given basis set is orthogonal, the answer is simple: normalize each of the basis vectors. When the given basis set in not orthogonal we can use the **Gram-Schmidt process** (GSP), which takes a basis set and transforms it into an orthonormal basis set.

In order to understand the GSP, one must first understand the following fact: given a vector $\vec{v}$ and a plane P which contains the vector, we can always write $\vec{v}$ as the sum of any two orthogonal vectors $\vec{a}$ and $\vec{b}$ in P, i.e.

$$\vec{v} = \vec{a} + \vec{b}, \tag{1.2.48}$$

where all the above vectors lie on the same plane (see Figure 1.27).

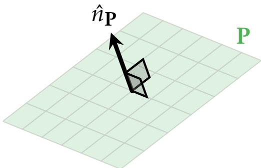
Rearranging Equation 1.2.48, we get that

$$\begin{aligned} \vec{a} &= \vec{v} - \vec{b}, \text{ and} \\ \vec{b} &= \vec{v} - \vec{a}. \end{aligned} \tag{1.2.49}$$

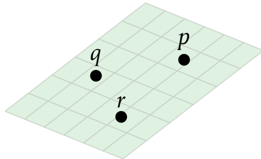
! **To be written**: Finish this subsection !

1.2.9 Normal vectors

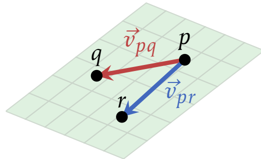
A special kind of vector in $\mathbb{R}^3$ is the so-called **normal vector** to a plane **P**: this vector, usually denoted as $\hat{n}_{\mathbf{P}}$, is pointing at the orthogonal direction to any vector of the plane (see XXX). Given one knows three points on the plane, its normal vector can be calculated: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



1 The normal vector to **P**.



2 Finding three points on the plane.



3 Finding two vectors on the plane.

Figure 1.28 A normal vector $\hat{n}_\mathbf{P}$ to the plane **P**.

see YYY): ! **To be written**: Change XXX and YYY to the right refs !

$$p = (p_x, p_y, p_z)$$
$$q = (q_x, q_y, q_z)$$
$$r = (r_x, r_y, r_z), \tag{1.2.50}$$

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

$$\vec{p} = \begin{bmatrix} p_x \\ p_y \\ p_z \end{bmatrix}, \ \vec{q} = \begin{bmatrix} q_x \\ q_y \\ q_z \end{bmatrix}, \ \vec{r} = \begin{bmatrix} r_x \\ r_y \\ r_z \end{bmatrix}. \tag{1.2.51}$$

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

$$\vec{v}_{pq} = \vec{q} - \vec{p} = \begin{bmatrix} q_x - p_x \\ q_y - p_y \\ q_z - p_z \end{bmatrix},$$
$$\vec{v}_{pr} = \vec{r} - \vec{p} = \begin{bmatrix} r_x - p_x \\ r_y - p_y \\ r_z - p_z \end{bmatrix}. \tag{1.2.52}$$

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

$$\vec{n}_\mathbf{P} = \vec{v}_{pq} \times \vec{v}_{pr} = \begin{bmatrix} (q_y - p_y)(r_z - p_z) - (r_y - p_y)(q_z - p_z) \\ (p_x - q_x)(r_z - p_z) - (r_x - p_x)(q_z - p_z) \\ (q_x - p_x)(r_y - p_y) - (r_x - p_x)(p_y - q_y) \end{bmatrix}. \tag{1.2.53}$$

Normalizing $\vec{n}_\mathbf{P}$ will then yield the normal vector $\hat{n}_\mathbf{P}$[5].

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

Note 1.10 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.

1.2.10 Examples

To wrap up the vectors section, we solve some problems which cover the material presented in the section:

Example 1.26 Vector form of gravitational force

According to Newton's law of gravity, given two objects O_1 and O_2 with masses m_1 and m_2 respectively, each of them would feel a gravitational force of attraction **in the direction of the other object** with the following magnitude:

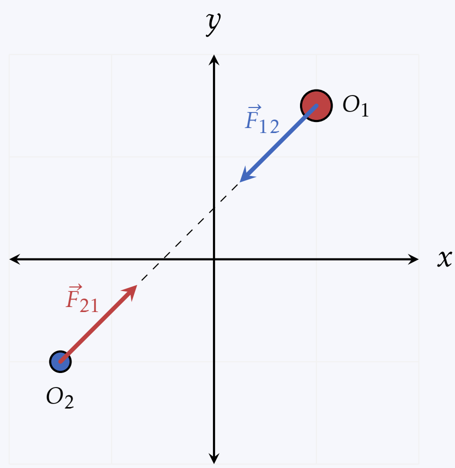
$$F = G\frac{m_1 m_2}{r^2},$$

where G is a universal constant and r is the distance between the two objects.

Say we put O_1 and O_2 on an axis system such that their positions are $\vec{r}_1 = \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}$ and $\vec{r}_2 = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix}$, respectively (see below figure for a 2-dimensions representation)...

The system therefore has two forces:

- $\vec{F}_{12}$: the gravitational force acting on object O_1 as a result of O_2. It points from O_1 towards O_2.
- $\vec{F}_{21}$: the gravitational force acting on O_2 as a result of O_1. It points from O_2 towards O_1.



Find the explicit form of each of the two force vectors.

Solution
The two vectors $\vec{F}_{12}$ and $\vec{F}_{21}$ both lie on the line connecting O_1 and O_2. Therefore, their orientations are exactly opposite, and since their magnitudes have to be equal (see the force definition above), the two vectors are simply the opposite of eachother, i.e.

$$\vec{F}_{12} = -\vec{F}_{21}.$$

We therefore need to calculate only a single force vector $\vec{F}$ and we automatically get the other force vector as $-\vec{F}$. We will thus first find the explicit form of the vector $\vec{F} = \vec{F}_{12}$, and using this form easily find $\vec{F}_{21}$ as $-\vec{F}$.
To get the explicit form of $\vec{F}$ we should find its magnitude and direction. The magnitude is given by

$$\left\|\vec{F}\right\| = G\frac{m_1 m_2}{r^2},$$

and since we are given G, m_1 and m_2 as parameters, we only need to express r^2 as a function of the positions $\vec{r}_1$ and $\vec{r}_2$ of the two objects. Given any two vectors, we can find their relative distance by simply *subtracting* one vector from the other. The result is a vector connecting the two given vectors, and its norm would be the distance between the vectors. Therefore, we find that

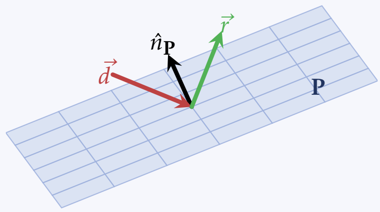
$$r^2 = \left\|\vec{r}_1 - \vec{r}_2\right\|^2 = \left\|\begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{bmatrix}\right\|^2 = (x_2 - x_1)^2 + \left(y_2 - y_1\right)^2 + (z_2 - z_1)^2.$$

(note that the order of subtraction forces the resulting vector to point from O_1 towards O_2. This will become handy soon)
As with any norm, the expression $(x_2 - x_1)^2 + \left(y_2 - y_1\right)^2 + (z_2 - z_1)^2$ can be written as the dot product a vector with itself, namely $\vec{\Delta r} = \vec{r}_2 - \vec{r}_1 = \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{bmatrix}$, and we get the following simple relationship:

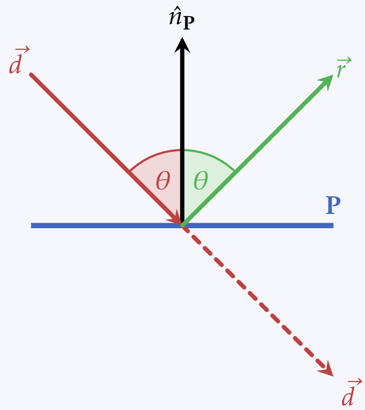
$$r^2 = \langle r_2 - r_1 | r_2 - r_1 \rangle.$$

Example 1.27 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 $\hat{n}_{\mathbf{P}}$.



Solution
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 1.2.40). 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 $\left\{\hat{n}_{\mathbf{P}}, \hat{p}\right\}$, 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 $\hat{n}_{\mathbf{P}}$:

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

! **To be written**: discussion about right- and left-handed spaces/orientations !