# Introductory Mathematics for Computer Science (COMP0011)

Raphael Li

Year 1 Term 2, 2024–25

---

## Contents

# 1 Complex numbers

The foundation of the *complex numbers* is given by the imaginary unit $i$, defined either as $i = \sqrt{-1}$ or as $i^2 = -1$.

A complex number $z$ can be written as $a + bi$, where $a, b \in \mathbb{R}$. The real numbers $a$ and $b$ are known as the *real part* and the *complex part* of $z$ respectively.

The set of all complex numbers is denoted as $\mathbb{C}$. Note that the set of real numbers $\mathbb{R}$ is a subset of $\mathbb{C}$. See figure 1.

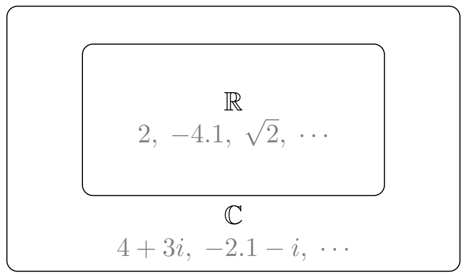



Figure 1: The set of real numbers $\mathbb{R}$ is a subset of the set of complex numbers $\mathbb{C}$. All real numbers are complex numbers.

## 1.1 Basic arithmetic with complex numbers, and complex conjugates

To add or subtract two complex numbers, we deal with the real and imaginary parts separately.

$$(2+3i)+(5-8i) = (2+5)+(3+(-8))i = 7-5i \qquad \text{(Addition)}$$
$$(2+3i)-(5-8i) = (2-5)+(3-(-8))i = -3+11i \qquad \text{(Subtraction)}$$

The multiplication of complex numbers is also straightforward as long as we bear in mind that $i^2 = -1$.

$$\begin{aligned}(3+4i)(-2+3i) &= -6+9i-8i+12i^2 \\ &= (-6-12)+(9-8)i \\ &= -18+i\end{aligned}$$

To divide a complex number by another, e.g.

$$\frac{a+bi}{c+di}$$

we multiply both the numerator and denominator by $c - di$, which is obtained by flipping the sign of the imaginary part of the denominator. For example, if we want to compute

$$\frac{2+3i}{5-4i},$$

we flip the sign of the imaginary part of $5 - 4i$ to get $5 + 4i$. We then multiply both the numerator and denominator of the fraction by this $5 + 4i$ to get

$$\begin{aligned}\frac{2+3i}{5-4i} &= \frac{(2+3i)(5+4i)}{(5-4i)(5+4i)} \\ &= \frac{10+8i+15i-12}{25+20i-20i+16} \\ &= \frac{-2+23i}{41} \\ &= \frac{-2}{41}+\frac{23}{41}i.\end{aligned}$$

Notice how multiplying $5 - 4i$ with $5 + 4i$ produces the real number $41$. By flipping the sign of the imaginary part of a complex number, we obtain what's called its *complex conjugate*. The complex conjugate of $z$ is denoted as $\overline{z}$. By writing $z$ as $a + bi$, we can easily prove that the product of any complex number with its conjugate must equal a real number:

$$z \times \overline{z} = (a + bi)(a - bi) = a^2 + b^2 \in \mathbb{R}.$$

## 1.2 Visualising complex numbers

Given some complex number $z = x + yi$, we can treat its real and imaginary parts as Cartesian coordinates, thus mapping it to a point on the 2D plane.

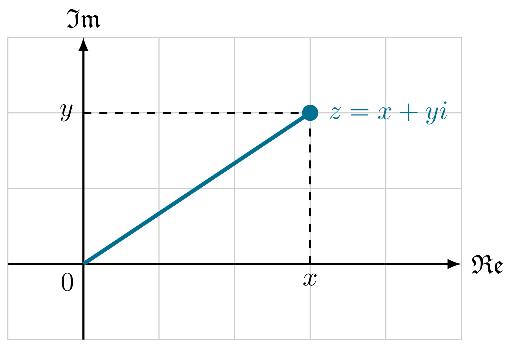



Figure 2: The complex number $z = x + yi$ as a point on the 2D plane

## 1.3 Exponential form

Recall that it is possible to express a point on a 2D plane using polar coordinates $(R, \theta)$ as well. Indeed, given any complex number $z = x + yi$, we can find its corresponding pair of values $R$ and $\theta$.

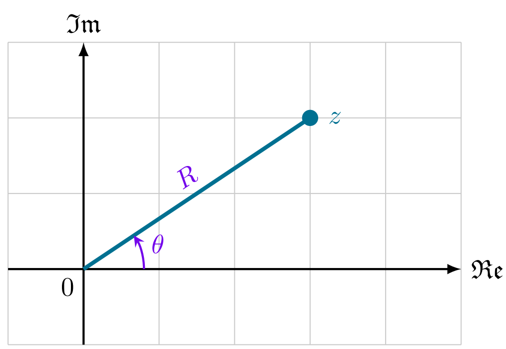



Figure 3: The position of a complex number on the 2D plane can be represented using polar coordinates.

Based on this idea, we introduce a new notation as follows.

> If the position of a complex number $z$ on the 2D plane can be represented by the polar coordinates $(R, \theta)$, then we have
> $$z = R \times e^{i\theta}$$
> where $R, \theta \in \mathbb{R}$ and $R \geq 0$.
>
> $R$ is called the *absolute value* or *modulus* of $z$ and is denoted as $|z|$. This represents the point's position from the origin.

$\theta$ is called the *argument* of $z$ and is denoted as $\arg(z)$. This represents the angle from horizontal.

This way of representing complex numbers is known as the *exponential form*. (This is a natural result of Euler's formula $e^{i\theta} = \cos\theta + i\sin\theta$.)

Now consider two complex numbers expressed in exponential form.

$$z_1 = R_1 \times e^{i\theta_1}$$
$$z_2 = R_2 \times e^{i\theta_2}$$

These two numbers are considered equal if both of the following conditions hold.

$$R_1 = R_2$$
$$\theta_1 = \theta_2 \textcolor{red}{+2k\pi} \qquad \text{(for some } k \in \mathbb{Z}\text{)}$$

Note that the red part is necessary because a rotation of $2\pi$ radians has no effect on a point's position.

The exponential form makes the multiplication and division of complex numbers a lot easier.

**Multiplication**

$$(1 \times e^{\frac{\pi}{6}i}) \times (2 \times e^{-\frac{\pi}{4}i}) = 2 \times e^{\frac{\pi}{6}i - \frac{\pi}{4}i}$$
$$= 2 \times e^{-\frac{\pi}{12}i}$$

**Division**

$$\frac{1 \times e^{\frac{\pi}{6}i}}{2 \times e^{-\frac{\pi}{4}i}} = \frac{1}{2} \times \frac{e^{\frac{\pi}{6}i}}{e^{-\frac{\pi}{4}i}}$$
$$= 2 \times e^{\frac{5\pi}{12}i}$$

## 1.4 Converting between Cartesian and exponential forms

The methods used to convert between the Cartesian form $x + yi$ and the exponential form $R \times e^{i\theta}$ are outlined below.

- **Given the Cartesian form of a complex number, find its exponential form.**

  Given the Cartesian form $z = x + yi$, we can find the modulus using Pythagoras' theorem.

  $$|z| = \sqrt{x^2 + y^2}$$

  The argument can be found using the arctangent.

  $$\arg(z) = \arctan\left(\frac{y}{x}\right)$$

- **Given the exponential form of a complex number, find its Cartesian form.**

  Given the exponential form $z = R \times e^{i\theta}$, we can find the Cartesian coordinates using simple trigonometry.

  $$x = R\cos\theta$$
  $$y = R\sin\theta$$

To speed up conversion processes, it is often useful to memorize the Cartesian coordinates of some special points on the unit circle. See figure 4 and table 1.

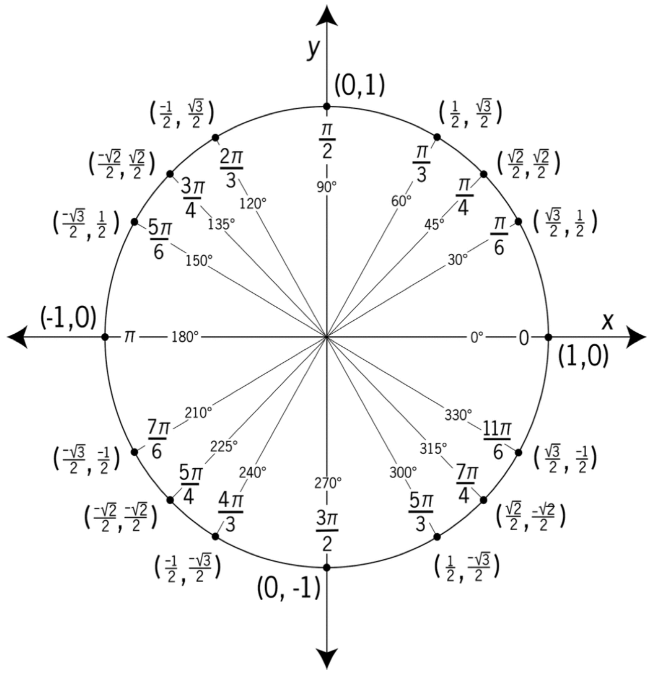



Figure 4: It is important to know the coordinates of points on the circle corresponding to classic angles.

| $\theta$ (radians) | $\pi/6$ | $\pi/4$ | $\pi/3$ |
|---|---|---|---|
| $\theta$ (degrees) | $30°$ | $45°$ | $60°$ |
| $\sin\theta$ | $\frac{1}{2}$ | $\frac{\sqrt{2}}{2}$ | $\frac{\sqrt{3}}{2}$ |
| $\cos\theta$ | $\frac{\sqrt{3}}{2}$ | $\frac{\sqrt{2}}{2}$ | $\frac{1}{2}$ |

Table 1: The values of $\sin\theta$ and $\cos\theta$ for some classic angles $\theta$.

## 1.5 Visualising arithmetic on complex numbers

When visualised on the 2D plane, the addition of complex numbers is similar to that of vectors. We join the arrows in a tip-to-tail manner in order to determine the sum, as shown in figure 5.

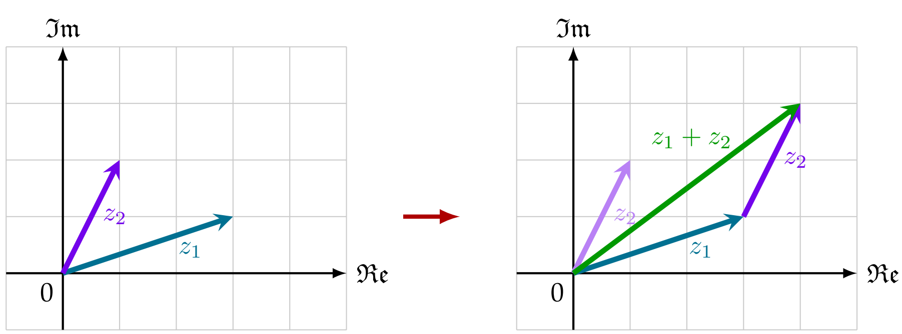



Figure 5: Addition of complex numbers.

The above figure also illustrates another key idea. Notice how in the figure on the right, the vectors of

$z_1$, $z_2$ and $z_1 + z_2$ form a triangle. This means their absolute values must fulfil the triangle inequality.

$$|z_1| + |z_2| \geq |z_1 + z_2|$$

To visualise multiplication we consider the exponential form. As shown in figure 6, when two complex numbers are multiplied, their arguments are added together to produce a rotation, while their moduli are multiplied.

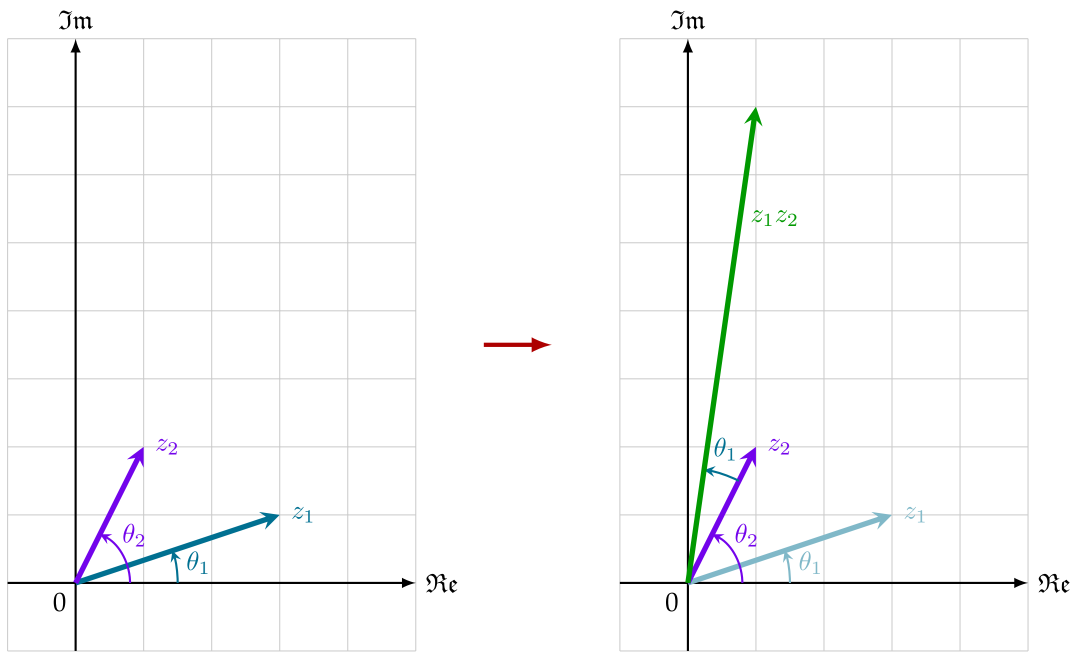



Figure 6: Multiplication of complex numbers.

## 1.6 Roots of unity

The *roots of unity* are the solutions to the equation

$$z^n = 1, \tag{1}$$

where $n$ is a positive integer.

Solving this equation for values of $n$ such as 2 and 4 is straightforward:

$$\begin{aligned} n = 2 &\implies z^2 = 1 \implies z = 1 \text{ or } -1 \\ n = 4 &\implies z^4 = 1 \implies z = 1,\ i,\ -1 \text{ or } -i \end{aligned}$$

but solving it for other values of $n$ requires us to express $z$ in its exponential form, i.e.

$$z = R \times e^{i\theta}. \qquad (R, \theta \in \mathbb{R} \text{ and } R \geq 0)$$

This allows us to rewrite the equation as

$$\begin{aligned} (R \times e^{i\theta})^n &= 1 \times e^{0i} \\ R^n \times e^{in\theta} &= 1 \times e^{0i} \end{aligned}$$

which yields the following.

$$\begin{cases} R^n = 1 \\ n\theta = 0 + 2k\pi = 2k\pi \end{cases} \qquad (\text{for some } k \in \mathbb{Z})$$

Since $R \geq 0$ and $R \in \mathbb{R}$, we must have $R = 1$. Furthermore, the second equation gives us

$$\theta = \frac{2k\pi}{n}$$

i.e.

$$\theta \in \left\{ \cdots,\ -3 \cdot \frac{2\pi}{n},\ -2 \cdot \frac{2\pi}{n},\ -\frac{2\pi}{n},\ 0,\ \frac{2\pi}{n},\ 2 \cdot \frac{2\pi}{n},\ 3 \cdot \frac{2\pi}{n},\ \cdots \right\}$$

which seemingly means that there are infinitely many roots of unity. However, this is impossible because by the fundamental theorem of algebra, equation (1) (which is a polynomial equation of degree $n$) can only have $n$ solutions.

To resolve this apparent paradox, let us visualise the problem on a 2D plane. For the sake of simplicity let us assume $n = 3$. We know that all solutions to (1) must have a modulus of $R = 1$, so they must lie on the unit circle.

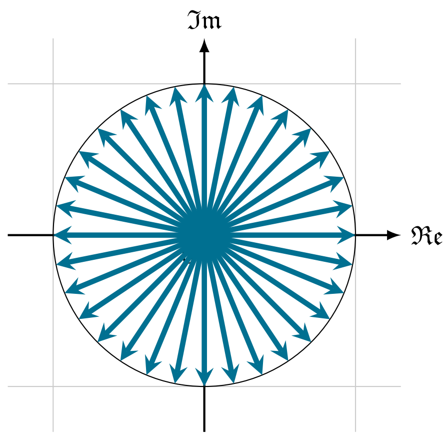



Figure 7: The roots of unity must lie somewhere on the unit circle.

We want to find the angles $\theta$ such that if we start at the point $1$ and then rotate anticlockwise by $\theta$ radians $n = 3$ times, we end up back at $1$.

- Obviously we can have $\theta = 0$.
- Another obvious solution is $\theta = 2\pi/3$. If we rotate by this angle $3$ times, we will have completed a full $2\pi$ radians, bringing us back to the initial point.
- Moreover, we can also have $\theta = 4\pi/3$. Rotating by this angle $3$ times creates a total rotation of $4\pi$ radians (i.e. 2 full cycles), bringing us once again back to the starting point.
- Continuing this pattern, it appears that $\theta = 6\pi/3$ is also a solution. However, this is in fact the same as $\theta = 0$, since angles differing by $2\pi$ are considered equivalent. The same applies for $\theta = 8\pi/3$ (equivalent to $2\pi/3$), $\theta = 10\pi/3$ (equivalent to $4\pi/3$), and so on.

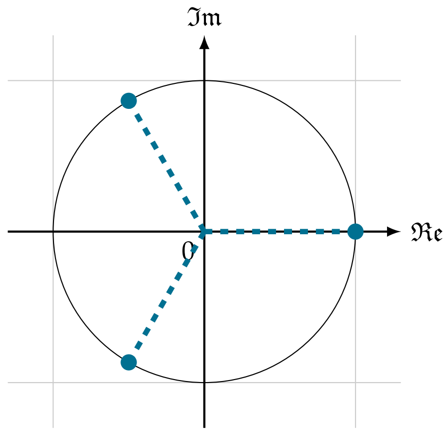



Figure 8: The third roots of unity.

This resolves the above paradox — we were right in thinking that the possible values of $\theta$ are given by

$$\theta = \frac{2k\pi}{n}$$

or

$$\theta \in \left\{ \cdots, \ -3 \cdot \frac{2\pi}{n}, \ -2 \cdot \frac{2\pi}{n}, \ -\frac{2\pi}{n}, \ 0, \ \frac{2\pi}{n}, \ 2 \cdot \frac{2\pi}{n}, \ 3 \cdot \frac{2\pi}{n}, \ \cdots \right\}$$

but these solutions are not all distinct. To make sure we only count distinct solutions, we impose the range $0 \leq k < n$, giving us

$$\theta = \frac{2k\pi}{n} \qquad (k \in \mathbb{N} \text{ and } k < n)$$

or

$$\theta \in \left\{ 0, \ \frac{2\pi}{n}, \ 2 \cdot \frac{2\pi}{n}, \ 3 \cdot \frac{2\pi}{n}, \ \cdots, \ (n-1) \cdot \frac{2\pi}{n} \right\}.$$

This yields the solutions

$$z = 1 \times e^{\frac{2k\pi}{n} i} = e^{\frac{2k\pi}{n} i} \qquad (k \in \mathbb{N} \text{ and } k < n)$$

or

$$z \in \left\{ 0, \ e^{\frac{2\pi}{n} i}, \ e^{\frac{4\pi}{n} i}, \ \cdots, \ e^{\frac{2(n-1)\pi}{n} i} \right\}.$$

## 2 Continuous functions

A function $f$ maps elements of a set $A$ to elements of another set $B$. We denote this as $f : A \to B$. In practice, most functions we consider will have type $\mathbb{R} \to \mathbb{R}$.

If a function maps a number $x$ to its square $x^2$, we can denote this by $x \mapsto x^2$. (Note the difference in the arrow symbol used — the symbol $\mapsto$ is read as "maps to".)

A function $y = f(x)$ can be represented graphically as the set of points $(x, y)$. See figure 9.

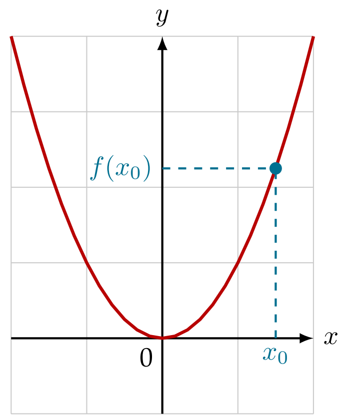



Figure 9: The graph of the function $y = x^2$.

In the next few subsections, we will be looking at some classic mathematical functions.

### 2.1 Trigonometric functions

Consider a point $P$ on the unit circle. If we let $\theta$ be the angle between $OP$ and the horizontal axis, then the coordinates of $P$ can be expressed as $(\cos\theta,\ \sin\theta)$. This is illustrated in figure 10.

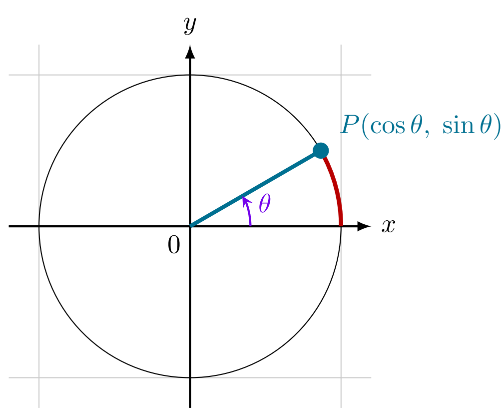



Figure 10: The trigonometric functions $\cos\theta$ and $\sin\theta$ can be defined using the unit circle. Note that if we are measuring $\theta$ in radians, then the length of the arc highlighted in red must be equal to $\theta$.

We've previously seen the values of $\sin\theta$ and $\cos\theta$ for some classic angles $\theta$ in table 1. Plotting these functions on a graph results in figure 11.

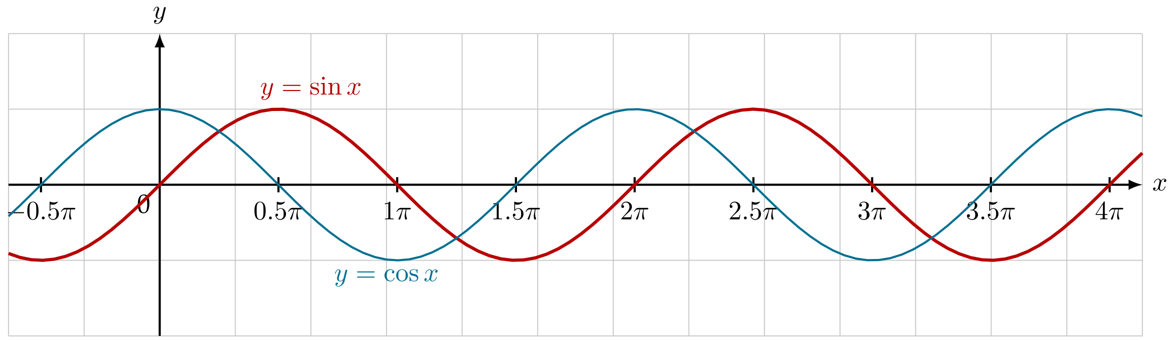



Figure 11: The graph of the functions $\sin x$ and $\cos x$.

## 2.2 Exponential and logarithm

One way to define the exponential function $\exp$ is as follows.

$$\begin{aligned}\exp(x+y) &= \exp(x)\cdot\exp(y)\\ \exp(0) &= 1\\ \frac{d}{dx}\exp(x) &= \exp(x)\end{aligned}$$

Note that the first two relationships can be satisfied by any function of the form $f(x) = a^x$ where $a \in \mathbb{R}$. However, if we take all three conditions into account, the only function satisfying them is $\exp(x) = e^x$, where $e = 2.71828\cdots$ is Euler's number.

The exponential function $\exp(x) = e^x$ is plotted in figure 12. Note that:

- For all values of $x$, we have $\exp(x) > 0$.
- When $x$ is negative, $\exp(x)$ is very small.
- The value of $\exp(x)$ grows very fast as $x$ increases.

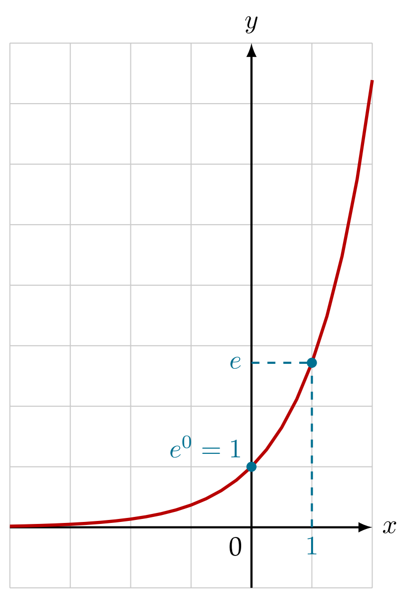



Figure 12: The graph of the function $y = \exp(x) = e^x$.

The natural logarithm $\ln x$ is the inverse of the exponential, meaning that $\ln(e^x) = x$. This results in the following properties.

$$\begin{aligned}
\ln(ab) &= \ln a + \ln b \\
\ln\left(\frac{a}{b}\right) &= \ln a - \ln b \\
\ln 1 &= 0 \\
\ln e &= 1 \\
a^x &= e^{x \ln a} \\
\ln(a^x) &= x \ln a
\end{aligned}$$

Since $e^x > 0$ for all $x$, the natural logarithm $\ln x$ is only defined for positive values of $x$.

The plot of $y = \ln x$ is given in figure 13. Note that:

- For $x < 1$, we have $\ln x < 0$.
- The curve intersects the $x$-axis at $(1, 0)$.
- For $x > 1$, the value of $\ln x$ grows very slowly as $x$ increases.

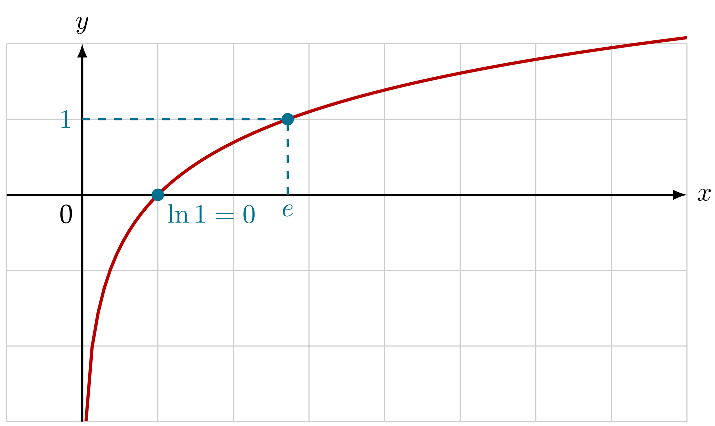



Figure 13: The graph of the function $y = \exp(x) = e^x$.

## 2.3 Introduction to limits

The idea of limits is simple.

> As the input $x$ approaches a value $p$, the output $f(x)$ also approaches a value $L$. (Both $p$ and $L$ possibly infinite.) To denote this we write $f(x) \to L$ as $x \to p$, or $\lim_{x \to p} f(x) = L$.

How do you formally define something like that?

Let us consider the simplest case, where both $p$ and $L$ are finite. We give the following definition.

> **Definition of a limit, with both $p$ and $L$ finite.**
>
> The main idea is that $f(x)$ can get *arbitrarily close* to $L$ as long as $x$ is close enough to $p$.
>
> In other words, no matter how close we want our output to be to $L$, we can always find a range of inputs around $p$ such that the output is within that range. This is written as
>
> $$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x \in \mathbb{R},\ 0 < |x - p| < \delta \Rightarrow |f(x) - L| < \epsilon.$$

This is called the epsilon-delta or $(\epsilon, \delta)$ definition of a limit. See figure 14 for an illustration of this.

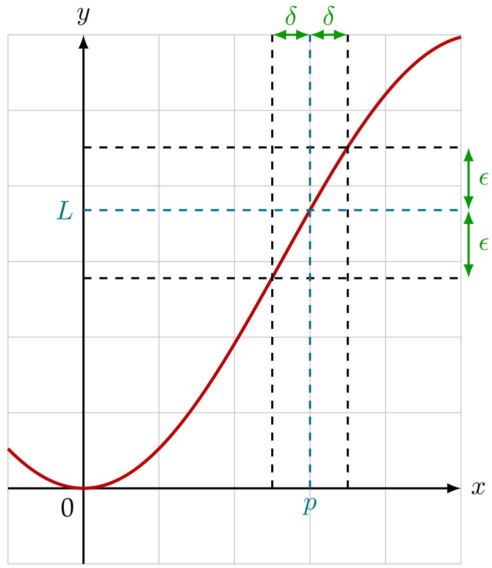



Figure 14: As $x$ approaches $p$, $f(x)$ approaches $L$.

An example problem utilising the definition is shown below.

**Problem.** Using the epsilon-delta definition of a limit, show that $\lim_{x\to 2} 2x + 3 = 7$.

**Intuition.** We want to show that

$$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x \in \mathbb{R},\ 0 < |x - 2| < \delta \Rightarrow |2x + 3 - 7| < \epsilon$$

i.e.

$$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x \in \mathbb{R},\ 0 < |x - 2| < \delta \Rightarrow |2x - 4| < \epsilon$$

Given some $\epsilon > 0$, we need to work out how to choose a $\delta$ value such that $0 < |x - 2| < \delta$ implies $|2x - 4| < \epsilon$. Simple observation shows that we can choose $\delta = \epsilon/2$, which allows us to construct the proof below.

**Proof.** We want to show that

$$\forall \epsilon > 0,\ \exists \delta > 0,\ \forall x \in \mathbb{R},\ 0 < |x - 2| < \delta \Rightarrow |2x - 4| < \epsilon.$$

For any $\epsilon > 0$, let $\delta = \epsilon/2 > 0$. Then if $0 < |x - 2| < \delta$ for some real $x$, we have

$$\begin{aligned} |x - 2| &< \delta \\ |x - 2| &< \frac{\epsilon}{2} \\ 2|x - 2| &< \epsilon \\ |2x - 4| &< \epsilon \end{aligned}$$

which concludes the proof.

Now, what happens if $L$ is infinite? We can define this as follows.

**Definition of a limit, with $p$ finite and $L$ infinite.**

The main idea is that $f(x)$ can become arbitrarily large as $x$ gets close enough to $p$.

In other words, for any value $d$, we can always find a range of inputs around $p$ such that the output is greater than $d$. This is written as

$$\forall d > 0,\ \exists \delta > 0,\ \forall x \in \mathbb{R},\ 0 < |x - p| < \delta \Rightarrow f(x) > d.$$

See figure 15.

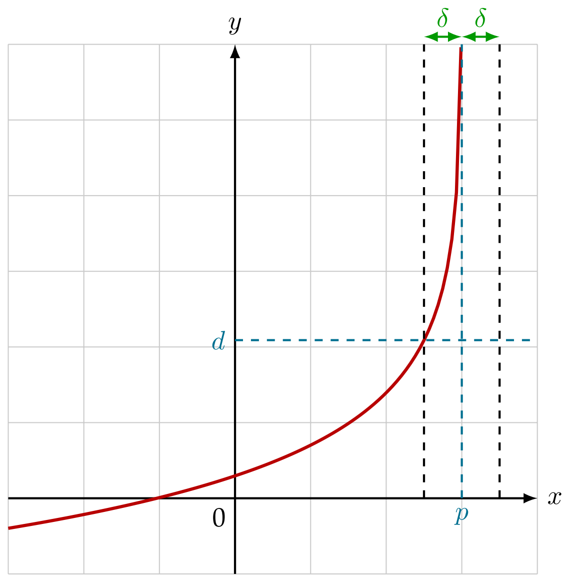



Figure 15: As $x$ approaches $p$, $f(x)$ approaches infinity.

Now let us consider the opposite scenario where $L$ is finite but $p$ is infinite. We modify our definition like so.

**Definition of a limit, with $p$ infinite and $L$ finite.**

The main idea is that $f(x)$ can become arbitrarily close to $L$ as long as $x$ is large enough.

In other words, no matter how close we want our output to be to $L$, we can always find a value $c$ such that as long as $x$ is greater than $c$, the output is within that range. This is written as

$$\forall \epsilon > 0,\ \exists c > 0,\ \forall x \in \mathbb{R},\ x > c \Rightarrow |f(x) - L| < \epsilon.$$

See figure 16.

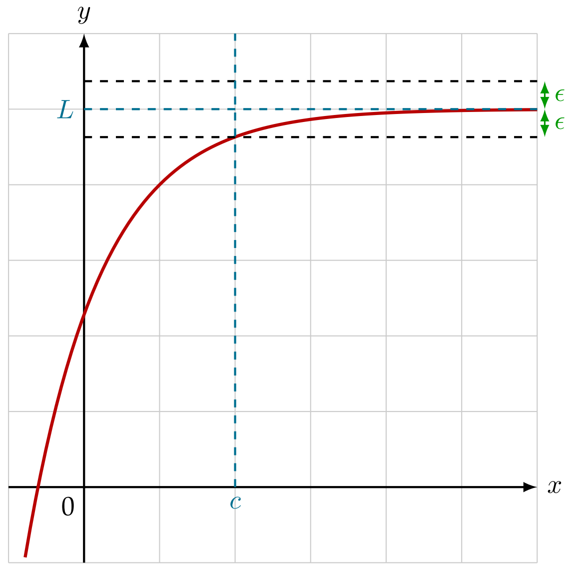



Figure 16: As $x$ approaches infinity, $f(x)$ approaches $L$.

The final case is where both $p$ and $L$ are infinite. The definition for this is as follows.

**Definition of a limit, with both $p$ and $L$ infinite.**

The main idea is that $f(x)$ can become arbitrarily large as $x$ gets large enough.

In other words, for any value $d$, we can always find a value $c$ such that as long as $x$ is greater than $c$, the output is greater than $d$. This is written as

$$\forall d > 0,\ \exists c > 0,\ \forall x \in \mathbb{R},\ x > c \Rightarrow f(x) > d.$$

See figure 17.

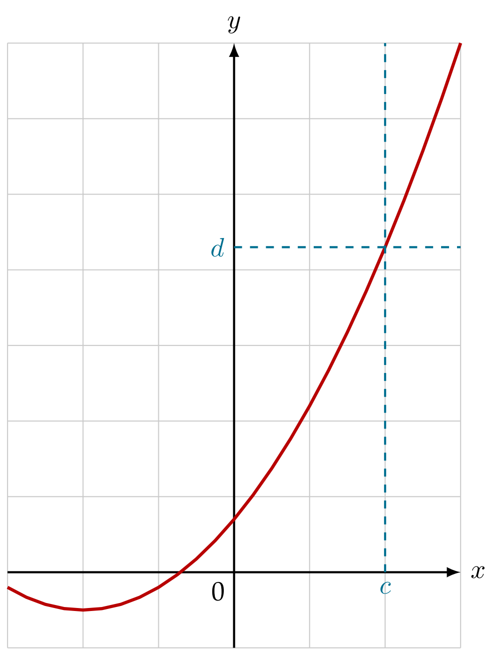



Figure 17: As $x$ approaches infinity, so does $f(x)$.

## 2.4 Handling infinities and indeterminate forms

A limit that exists is known as a *finite* limit. Finite limits can be combined in a natural way.

$$\lim_{x\to a}(f(x)+g(x)) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)$$
$$\lim_{x\to a}(f(x)\cdot g(x)) = \lim_{x\to a} f(x) \cdot \lim_{x\to a} g(x)$$
$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\lim_{x\to a} f(x)}{\lim_{x\to a} g(x)}$$

We use the following rules to handle infinities.

$$a \times \infty = \infty$$
$$\frac{a}{\infty} = 0$$

If a limit involves $x$ approaching zero, we may sometimes have to specify the direction in which $x$ is approaching it, i.e. whether it is approaching zero as a positive number (from the right) or as a negative number (from the left).

$$\lim_{x\to 0^+} \frac{1}{x} = \infty$$
$$\lim_{x\to 0^-} \frac{1}{x} = -\infty$$

There are certain cases where we *cannot* combine limits. These are called *indeterminate forms*, and there is no general rule for figuring out what these indeterminate forms evaluate to. Examples of indeterminate forms are given below.

$$\frac{0}{0},\ \frac{\infty}{\infty},\ 0\times\infty,\ \infty-\infty,\ 0^0,\ 1^\infty,\ \infty^0$$

## 2.5 Little o and big O notation

It is often useful to talk about the rate at which some function changes as its input increases (or decreases), without worrying to much about the detailed form. To do this, we introduce two types of notation: little $o$ and big $O$.

To compare the order of growth of two functions $f(x)$ and $g(x)$, we can look at the ratio of the two functions as their input approaches infinity, i.e.

$$\lim_{x\to\infty}\left|\frac{f(x)}{g(x)}\right|.$$

Let's look at how this limit behaves for different functions.

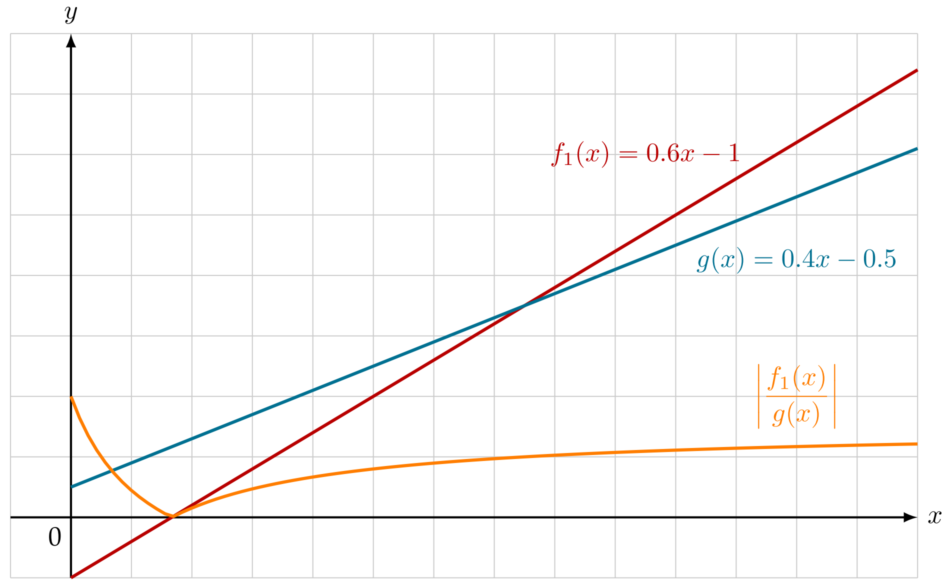



Figure 18: Graphs of a linear function $f_1(x)$, another linear function $g(x)$ and the absolute value of their quotient $|f(x)/g(x)|$.

Figure 18 shows the graphs of two functions $f_1(x) = 0.6x - 1$ and $g(x) = 0.4x + 0.5$. Both of these functions are linear, so they should have the same order of growth. When this happens, the limit of the ratio of the two functions, as $x$ approaches infinity, should be a finite constant. As we see in the graph, this is indeed the case, with the orange curve converging to a value of $1.5$.

Now consider the logarithmic function $f_2(x) = \ln x$ and its relationship with $g(x)$. As shown in figure 19, the limit of their ratio once again converges to a finite constant: zero. This is because the logarithmic function grows much slower than the linear function.

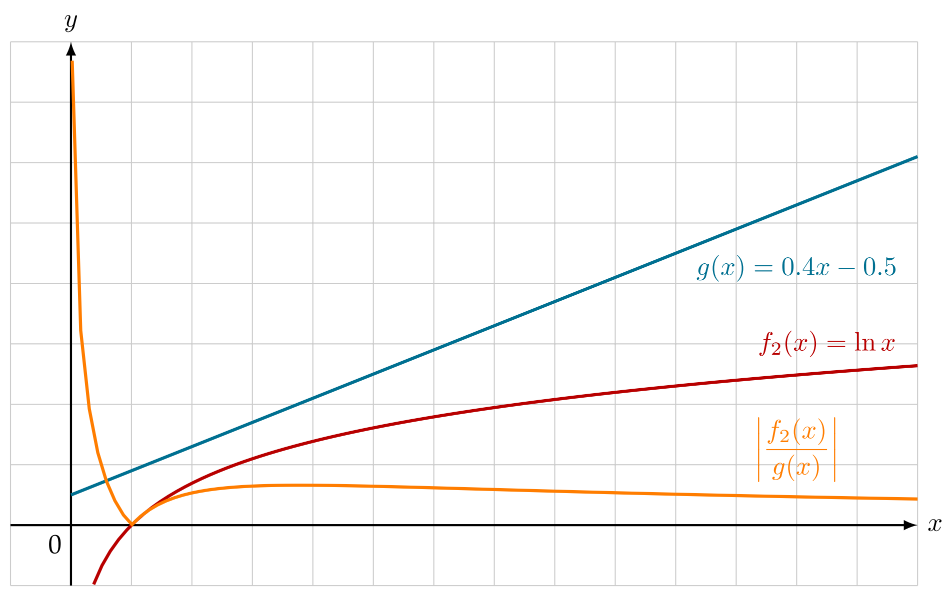



Figure 19: Graphs of a logarithmic function $f_2(x)$, a linear function $g(x)$ and the absolute value of their quotient $|f(x)/g(x)|$.

From this we can gather that the limit of the ratio of two functions $f(x)$ and $g(x)$ is bounded (i.e. a finite constant) when $f(x)$ grows *as fast as* or *slower* than $g(x)$. We can denote this with big $O$ notation, as $f(x) = O(g(x))$. The definition of big $O$ notation is given below.

**Big O notation.**

We write $f = O(g)$ near infinity if $\lim_{x\to\infty} \left|\frac{f(x)}{g(x)}\right|$ is bounded, i.e.

$$\exists M \in \mathbb{R},\ \lim_{x\to\infty} \left|\frac{f(x)}{g(x)}\right| < M.$$

Sometimes we want to consider the growth rate of functions not just near infinity but near a specific point $x = b$. In this case, we write $f = O(g)$ near $b$ if

$$\exists M \in \mathbb{R},\ \lim_{x\to b} \left|\frac{f(x)}{g(x)}\right| < M.$$

Remember that big $O$ notation covers two cases:

- $f(x)$ grows as fast as $g(x)$ (meaning that the two functions have the same order of growth), or
- $f(x)$ grows slower than $g(x)$ (meaning that the former has a lower order of growth than the latter).

If we want to be more specific and consider only the second case where $f(x)$ grows strictly slower than $g(x)$, we can use little $o$ notation to write $f(x) = o(g(x))$. The definition of little $o$ notation is given below.

**Little o notation.**

We write $f = o(g)$ near infinity if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$.

Again, we sometimes want to consider the growth rate of functions near a specific point $x = b$. For this we say that $f = o(g)$ near $b$ if $\lim_{x\to b} \frac{f(x)}{g(x)} = 0$.

Note that since little $o$ is a special case of big $O$, we have $f = o(g) \ \Rightarrow\ f = O(g)$.

## 2.6 Continuity

A function $f$ is continuous if for all $a$ where $f(a)$ is defined, we have $\lim_{x\to a} f(x) = f(a)$.

In practice, this means that the graph of $y = f(x)$ is a single unbroken curve. The exponential and logarithm functions, for example, are both continuous.

An important result of this is the *intermediate value theorem*.

**Intermediate value theorem.**

Assume for a continuous function $f$ that $a < b$ and $f(a) < f(b)$. For any value $y$ such that $f(a) < y < f(b)$, there exists a (not necessarily unique) value $x$ such that $a < x < b$ and $f(x) = y$.

See figure 20 and 21.

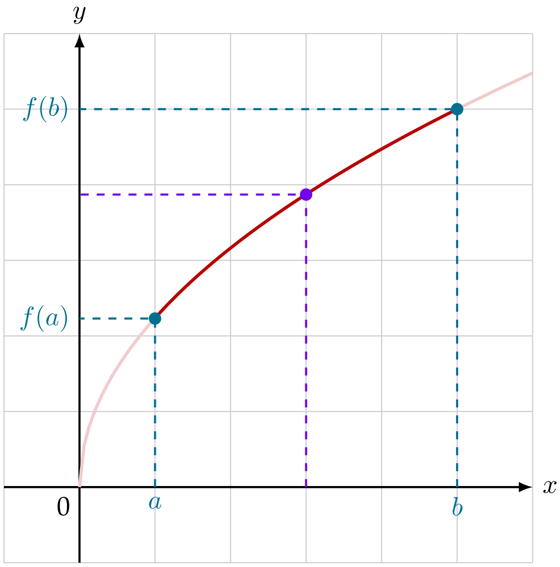



Figure 20: The intermediate value theorem.

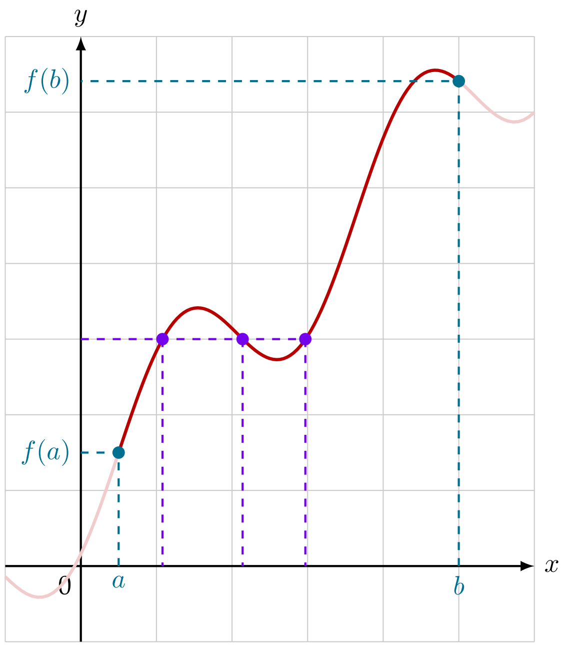


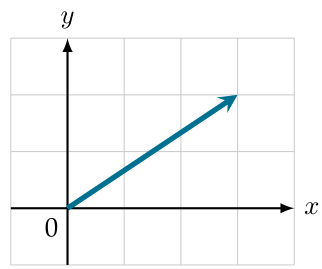

Figure 21: In the intermediate value theorem, for a given value $y$, the value of $x$ does not necessarily have to be unique.

# 3 Vector spaces

We're used to thinking of vectors as something like

$$\begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

or




but this is only a fraction of what the term "vector" encompasses. As we shall see in this section, given the correct prerequisites, even something like

$$2x^2 + 3x + 5$$

can be a vector!

## 3.1 What is a vector space?

To start, we define a *vector space* as follows.

> **Definition of a vector space.**
>
> For a field $K$, a non-empty set $V$ is a $K$-vector space (or a vector space over $K$) if, for any $\vec{u}, \vec{v} \in V$ and any scalar $a \in K$, we have
>
> $$\vec{u} + \vec{v} \in V \qquad \text{(closure under vector addition)}$$
> $$a\vec{u} \in V \qquad \text{(closure under scalar multiplication)}$$
>
> The elements of $V$ are called *vectors*.

We note the following:

- Here, $K$ is a *field*, typically either $\mathbb{R}$ or $\mathbb{C}$. This field tells us what counts as a "scalar" in this vector space.
- The pair of properties listed in the definition give rise to what's called *linearity* — it's what makes linear algebra linear.
- Vector addition and scalar multiplication are governed by the following rules.

$$\begin{aligned}
(a+b)\vec{v} &= a\vec{v} + b\vec{v} \\
(a(b\vec{v})) &= (ab)\vec{v} \\
\vec{u} + \vec{v} &= \vec{v} + \vec{u} \\
\vec{u} + (\vec{v} + \vec{w}) &= (\vec{u} + \vec{v}) + \vec{w} \\
a(\vec{u} + \vec{v}) &= a\vec{u} + a\vec{v} \\
\vec{v} &= \vec{v} + \vec{0} \\
0\vec{v} &= \vec{0} \\
1\vec{v} &= \vec{v}
\end{aligned}$$

  where $\vec{u}, \vec{v}, \vec{w}$ are vectors, $a, b$ are scalars in $K$, and $\vec{0}$ represents the zero vector.

## 3.2 Examples of vector spaces

Based on the definition above, what counts as a vector space? (For now, let us set $K = \mathbb{R}$ and consider only $\mathbb{R}$-vector spaces.)

Unsurprisingly, the set of pairs of real numbers, denoted as $\mathbb{R}^2$, is a vector space:

$$\mathbb{R}^2 = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} \middle| \, x, y \in \mathbb{R} \right\}$$

and each of the pairs $\begin{bmatrix} x \\ y \end{bmatrix}$ is a vector. This is because of the fact that the sum of any two vectors in $\mathbb{R}^2$ must also be in $\mathbb{R}^2$, and that the scaled version of any vector in $\mathbb{R}^2$ must also be in $\mathbb{R}^2$. These vectors can be visualised as arrows in 2D space, as shown in figures 22 and 23.

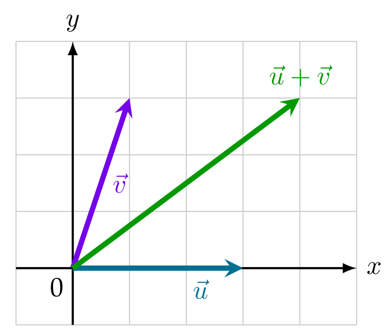



Figure 22: Vectors in $\mathbb{R}^2$ are closed under addition. For any given pair of vectors $\vec{u}$ and $\vec{v}$ both in $V$, their sum $\vec{u} + \vec{v}$ must also be a vector.

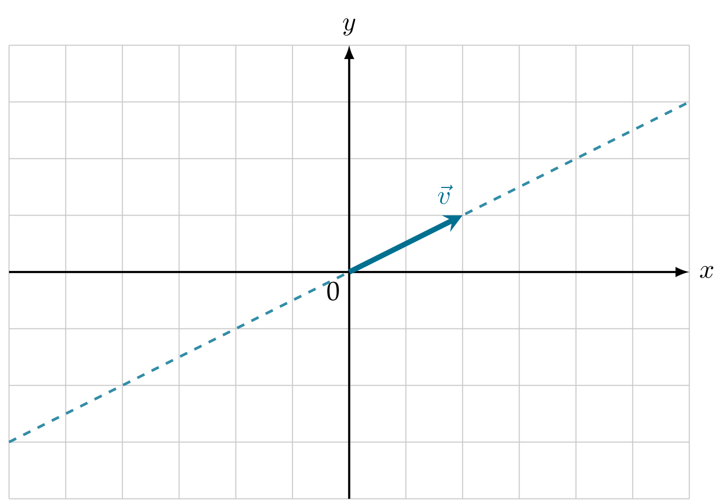



Figure 23: Vectors in $\mathbb{R}^2$ are closed under scalar multiplication. For any given vector $v \in \mathbb{R}^2$, the product between $\vec{v}$ and any scalar $a$ (i.e. any vector lying on the dashed line) is also a vector.

The same goes with the set of triplets of real numbers $\begin{bmatrix} x \\ y \\ z \end{bmatrix}$, which is denoted as $\mathbb{R}^3$ and whose vectors can be visualised as arrows in 3D space. In fact, for any natural number $n$, the set of $n$-tuples of real numbers (denoted as $\mathbb{R}^n$) is a vector space.

So far this is not very exciting as it is equivalent to our usual notion of what a "vector" is. To step things up a notch, let us consider the set of polynomials of degree at most 2. Is this set a vector space?

$$\{ax^2 + bx + c \mid a, b, c, \in \mathbb{R}\}$$

To answer this, we notice that:

- Given any two polynomials in this set,

$$\alpha = a_1x^2 + b_1x + c_1$$
$$\beta = a_2x^2 + b_2x + c_2$$

  their sum $\alpha + \beta$ must also be an element of this set.

- If we multiply a polynomial of degree at most 2 with a scalar $a$, the resultant product must also be a polynomial of degree at most 2.

This means that this set is indeed a vector space!

Table 2 lists some examples of sets that are vector spaces, and some that aren't.

| **Vector spaces** | **Not vector spaces** |
|---|---|
| The set of real numbers $\mathbb{R}$ | The set of natural numbers $\mathbb{N}$ |
| The set of complex numbers $\mathbb{C}$ | The set of polynomials of degree exactly 2 |
| The set of continuous functions that act on $\mathbb{R}$ | The set of irrational numbers $\mathbb{R} \setminus \mathbb{Q}$ |

Table 2: Some sets are vector spaces while some are not.

Our previous definition of a vector space consisted of two properties. We can group those properties together to produce the following alternative but equivalent definition.

> **Alternative definition of a vector space.**
>
> For a field $K$ (typically $\mathbb{R}$ or $\mathbb{C}$), a non-empty set $V$ is a $K$-vector space if, for any $\vec{u}, \vec{v} \in V$ and any scalar $a \in K$, we have $a\vec{u} + \vec{v} \in V$.

A vector space always contains a zero vector. This can be shown by setting $\vec{u} = \vec{v}$ and $a = -1$ in the definition above.

## 3.3 Collinearity, span and subspaces

Vectors $\vec{u}$ and $\vec{v}$ are said to be collinear if $\vec{u} = a\vec{v}$ for some scalar $a$. See figure 24.

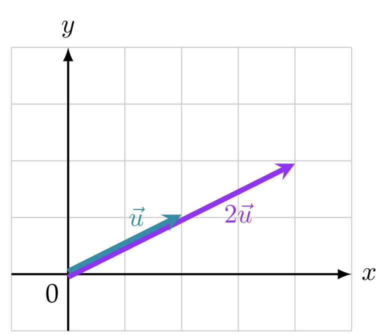



Figure 24: Two collinear vectors.

Given some vectors $\vec{u_1}, \vec{u_2}, \cdots, \vec{u_n}$ and some scalars $a_1, a_2, \cdots, a_n$, the sum $a_1\vec{u_1} + a_2\vec{v_2} + \cdots + a_n\vec{v_n}$ is called a *linear combination* of those vectors. This is illustrated in figure 25.

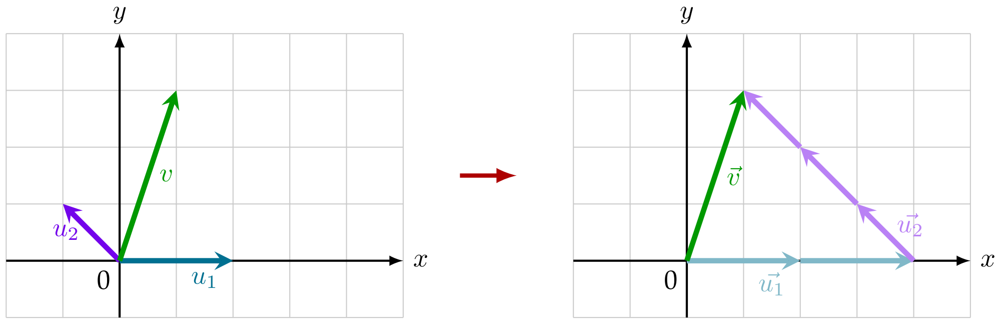



Figure 25: The green vector $v$ can be expressed as a linear combination of the vectors $u_1$ and $u_2$, i.e. $v = 2u_1 + 3u_2$.

Given a family of vectors $S = (\vec{u_1}, \vec{u_2}, \cdots, \vec{u_n})$, the *span* of $S$ is defined as the set of all linear combinations of the vectors in $S$. For instance, consider the vectors shown in figure 26.

- The vectors $\vec{u}$ and $\vec{v}$ span the $xy$-plane.
- The vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ together span the entire three-dimensional space $\mathbb{R}^3$.
- Since the vectors $\vec{w}$ and $2\vec{w}$ are collinear, their span is along a single straight line.

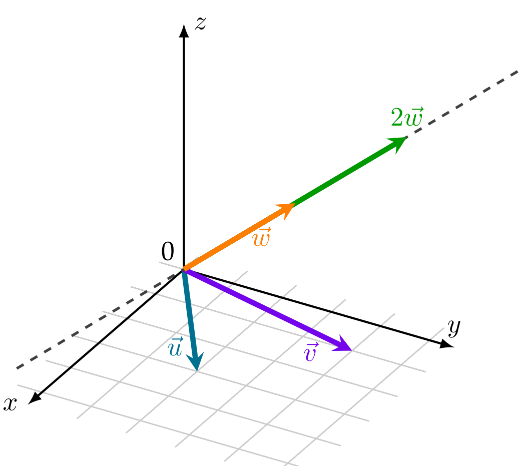



Figure 26: Three vectors in $\mathbb{R}^3$. Both $\vec{u}$ and $\vec{v}$ lie flat on the $xy$-plane. The dashed line represents the span of $\vec{w}$ and $2\vec{w}$.

Given some vector space $V$, a subset $U$ of $V$ is called a *subspace* if it is stable under linearity (i.e. also a vector space). For example, the $xy$-plane is a subspace of the three-dimensional vector space $\mathbb{R}^3$.

## 3.4 Linear independence

The idea of *linear independence* can be defined in several ways.

**Definition of linear independence.**

The vectors $\vec{u_1},\ \vec{u_2},\ \cdots,\ \vec{u_n}$ are said to be *linearly independent* if:

- There exist no scalars $a_1,\ a_2,\ \cdots,\ a_n$ (not all equal to zero) such that

$$a_1\vec{u_1} + a_2\vec{u_2} + \cdots + a_n\vec{u_n} = 0. \qquad (*)$$

- None of the vectors can be expressed as a linear combination of the others.
- None of the vectors belong to the span of the others.

These three statements are equivalent.

To prove linear independence, we can make use of the first statement and show that for equation (*) to hold, the scalars $a_1,\ a_2,\ \cdots,\ a_n$ must all be equal to zero.

To disprove linear independence, we can use the second statement and show that one of the vectors can be expressed as a linear combination of the others. (Alternatively, we can provide a solution to equation (*) where not all of $a_1,\ a_2,\ \cdots,\ a_n$ are zero.)

For example, in figure 26, the vectors $\vec{u}$, $\vec{v}$ and $\vec{w}$ are linearly independent as none of the vectors belong to the span of the others.

The example problems below demonstrate how linear independence can be proved or disproved.

**Problem.** Is this family of polynomials linearly independent? Explain.

$$\{3x,\ 4x^2-6x,\ 2x^2\}$$

**Solution 1.** No. This is because one of the polynomials can be written as a linear combination of the others:

$$4x^2-6x=2(2x^2)+(-2)(3x)$$

**Solution 2.** No, because $2(3x)+(4x^2-6x)-2(2x^2)=0$.

**Problem.** Is this family of vectors linearly independent? Explain.

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

**Solution.** Yes. Assume there exists a linear combination such that

$$a_1\begin{bmatrix}2\\2\end{bmatrix}+a_2\begin{bmatrix}1\\-1\end{bmatrix}=0.$$

This produces the following system of simultaneous equations:

$$\begin{cases}2a_1+a_2=0\\2a_1-a_2=0\end{cases}$$

for which the only solution is $a_1=a_2=0$. Hence, the family of vectors are indeed linearly independent.

### 3.5 Basis

Let $V$ be a vector space. A basis[1] of $V$ is a set $S$ of linearly independent vectors that spans the entirety of $V$. For instance, in figure 26, the set of vectors $\{\vec{u},\vec{v}\}$ is a basis of the $xy$-plane.

If the basis $S$ is finite, then the size (cardinality) of $S$ is referred to as the *dimension* of $V$. This means that the $xy$-plane in figure 26 has a dimension of 2.

[1]Plural: bases.

One of the most obvious bases for $\mathbb{R}^n$ is

$$\left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \cdots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}$$

which is known as the *canonical basis*. See figures 27 and 28.

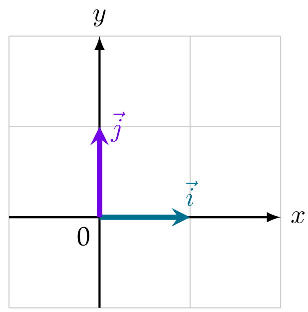



Figure 27: The canonical basis $\{\vec{i}, \vec{j}\}$ of $\mathbb{R}^2$.

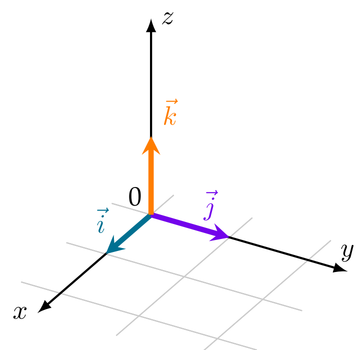



Figure 28: The canonical basis $\{\vec{i}, \vec{j}, \vec{k}\}$ of $\mathbb{R}^3$.

We note that any set of $n$ linear independent vectors in $\mathbb{R}^n$ must span $\mathbb{R}^n$, and must thus be a basis of $\mathbb{R}^n$.

## 3.6 Linear maps

We define a *linear map* or *linear mapping* as follows.

**Definition of a linear map.**

For two $K$-vector spaces $V$ and $W$, consider a function $f : V \to W$ that maps each vector in $V$ to a vector in $W$.

This function is said to be a *linear map* if, for any two vectors $\vec{u}, \vec{v} \in V$ and any scalar $a \in K$, we have

$$f(\vec{u} + \vec{v}) = f(\vec{u}) + f(\vec{v})$$
$$f(a\vec{u}) = af(\vec{u}).$$

Once again we can combine the two equations above to get an alternative but equivalent definition.

**Alternative definition of a linear map.**

> The function $f : V \to W$ is said to be a *linear map* if, for any two vectors $\vec{u}, \vec{v} \in V$ and any scalar $a \in K$, we have $f(a\vec{u} + \vec{v}) = af(\vec{u}) + f(\vec{v})$.

For any linear map $f$ we must have $f(\vec{0}) = \vec{0}$. This can be shown by setting $a = 0$ and $\vec{v} = \vec{0}$ in the definition above.

An example of a linear map in $\mathbb{R}^2$ is a function $f$ that rotates vectors by $90°$ anticlockwise, i.e.

$$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} -y \\ x \end{bmatrix}$$

To prove this, consider any two vectors $\vec{u} = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $\vec{v} = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$. For any scalar $a \in \mathbb{R}$, we have

$$\begin{aligned}
f(a\vec{u} + \vec{v}) &= f\left(a\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) \\
&= f\left(\begin{bmatrix} ax_1 \\ ay_1 \end{bmatrix} + \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) \\
&= f\left(\begin{bmatrix} ax_1 + x_2 \\ ay_1 + y_2 \end{bmatrix}\right) \\
&= \begin{bmatrix} -ay_1 - y_2 \\ ax_1 + x_2 \end{bmatrix} \\
&= \begin{bmatrix} -ay_1 \\ ax_1 \end{bmatrix} + \begin{bmatrix} -y_2 \\ x_2 \end{bmatrix} \\
&= a\begin{bmatrix} -y_1 \\ x_1 \end{bmatrix} + \begin{bmatrix} -y_2 \\ x_2 \end{bmatrix} \\
&= af\left(\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}\right) + f\left(\begin{bmatrix} x_2 \\ y_2 \end{bmatrix}\right) \\
&= af(\vec{u}) + f(\vec{v})
\end{aligned}$$

which shows that $f$ is a linear map. In fact, rotations, scalings and projections are all examples of linear maps in $\mathbb{R}^2$.

If a linear map maps vectors from a vector space to the same vector space, i.e.

$$f : V \to V$$

then this mapping is known as an *endomorphism*[2]. For instance, the $90°$ rotation function we saw before maps vectors in $\mathbb{R}^2$ to vectors in $\mathbb{R}^2$, so it is an endomorphism.

Let $U$, $V$ and $W$ be vector spaces. Given two linear maps:

$$\begin{aligned}
f &: U \to V \\
g &: V \to W
\end{aligned}$$

we can *compose* them into a new linear map

$$g \circ f : U \to W$$

where $f$ is applied first, then $g$.

[2]From "endo-" (meaning "internal" in Greek) and "morphism" (a mathematical term for structure-preserving mappings).

### 3.7 Kernel and image

Consider the linear map $f : V \to W$, where $V$ and $W$ are vector spaces. We define the following:

- The *kernel* of $f$, denoted as $\mathrm{Ker}(f)$, is the set of vectors $v \in V$ for which $f(v) = \vec{0}$.
- The *image* of $f$, denoted as $\mathrm{Im}(f)$, is the set of $f(v)$ for every $v \in V$.

Both $\mathrm{Ker}(f)$ and $\mathrm{Im}(f)$ are vector spaces. Note that $\mathrm{Ker}(f)$ always contains $\vec{0}$ since $f(\vec{0}) = \vec{0}$.

As an example, consider the following linear map.

$$g\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \begin{bmatrix} x+y \\ x+y \end{bmatrix}$$

To find its kernel $\mathrm{Ker}(g)$, we solve

$$\begin{aligned} g\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) &= \vec{0} \\ \begin{bmatrix} x+y \\ x+y \end{bmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ x+y &= 0 \\ y &= -x \end{aligned}$$

which gives us

$$\mathrm{Ker}(g) = \left\{ \begin{bmatrix} x \\ -x \end{bmatrix} \middle| \, x \in \mathbb{R} \right\} = \mathrm{Span}\left(\begin{bmatrix} 1 \\ -1 \end{bmatrix}\right).$$

To find the image $\mathrm{Im}(g)$, we want to look for vectors $\vec{v}$ that verify $g(\vec{u}) = \vec{v}$ for some $\vec{u}$. To do this, it might be useful to look at some examples and search for patterns.

$$\begin{aligned} g\left(\begin{bmatrix} 1 \\ 2 \end{bmatrix}\right) &= \begin{bmatrix} 3 \\ 3 \end{bmatrix} \\ g\left(\begin{bmatrix} 3 \\ 4 \end{bmatrix}\right) &= \begin{bmatrix} 7 \\ 7 \end{bmatrix} \\ g\left(\begin{bmatrix} -5 \\ 3 \end{bmatrix}\right) &= \begin{bmatrix} -2 \\ -2 \end{bmatrix} \end{aligned}$$

We notice that the entries in the output vectors are always identical. We formalise this as follows — for $\vec{v} = g(\vec{u})$ to be true, we must have

$$\begin{aligned} \exists x, y,\ \vec{v} &= \begin{bmatrix} x+y \\ x+y \end{bmatrix} \\ \exists z,\ \vec{v} &= \begin{bmatrix} z \\ z \end{bmatrix} \end{aligned}$$

which yields $\mathrm{Im}(g) = \mathrm{Span}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix}\right)$.

# 4 Matrices

A vector can be multiplied by a matrix as follows.

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

Figure 29 shows a diagrammatic representation of this procedure.

$$\begin{bmatrix} x \\ y \end{bmatrix}$$

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$$

Figure 29: Multiplying a matrix by a vector.

Notice how a matrix maps a vector, $\begin{bmatrix} x \\ y \end{bmatrix}$, to a new one, $\begin{bmatrix} ax + by \\ cx + dy \end{bmatrix}$, just like a linear map does. This is not a coincidence — it is in fact possible to express any linear map as a matrix, as we shall see below.

## 4.1 Expressing linear maps as matrices

Consider two vector spaces:

- A vector space $U$ of $m$ dimensions, with a specified set of basis vectors.
- A vector space $V$ of $n$ dimensions, with a specified set of basis vectors.

Every linear map $f : U \to V$ can then be represented by a matrix $M$ with $m$ rows and $n$ columns. Each column represents the image of a basis vector of $U$, expressed in the basis of $V$.

It's important to take a moment to digest what this means. Several examples are included below.

### 4.1.1 Example 1: An endomorphism with unchanged bases

**Setup.** We set both $U$ and $V$ to $\mathbb{R}^2$. We will use the canonical basis for both vector spaces:

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}.$$

We define $f$ to be a linear map that scales the length of a vector by a factor of $1.5$ and then rotates it by $30°$. See figure 30.

We want to find the matrix $M$ that represents this endomorphism.

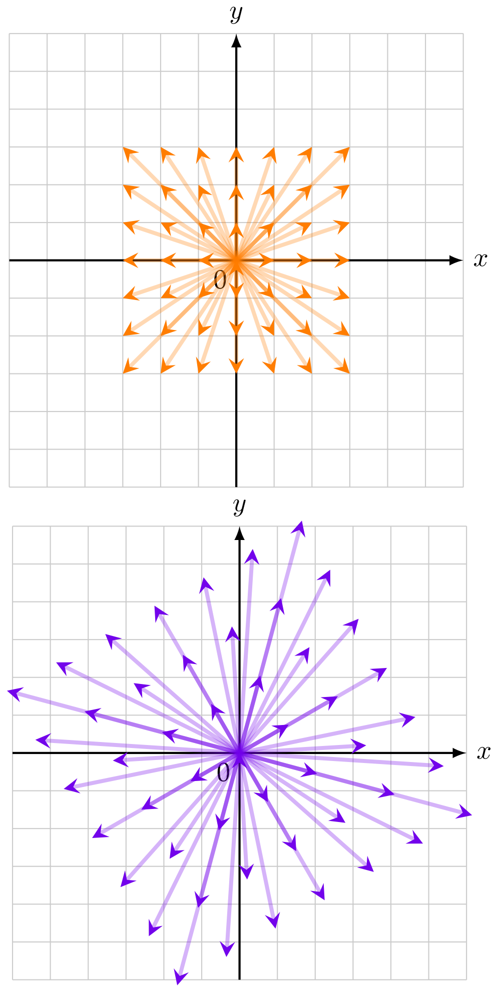



Figure 30: The linear map $f$ takes a vector (in orange) and scales it by a factor of 1.5 before rotating it by $30°$ anticlockwise. This results in a new vector (in purple).

Let's start with the size of the matrix, also known as its *order*. We know both $U$ and $V$ are 2-dimensional vector spaces, so the matrix is going to have an order of $2 \times 2$, with two rows and two columns.

$$M = \begin{bmatrix} \square & \square \\ \square & \square \end{bmatrix}$$

Remember that each column of the matrix represents the image of a basis vector of $U$, expressed in the basis of $V$. This means that the first column (in red) will represent where the first basis vector of $U$ will get mapped to, as expressed using the basis vectors of $V$.

- The first basis vector of $U$ is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Following the rule of our transformation, the linear map will take us to $\begin{bmatrix} 2\cos 30° \\ 2\sin 30° \end{bmatrix}$. Since the bases are the same for both vector spaces, this will be the first column of our matrix $M$.
- Similarly, the second basis vector of $U$ is $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. This maps to $\begin{bmatrix} 2\cos 120° \\ 2\sin 120° \end{bmatrix}$, or $\begin{bmatrix} -2\sin 30° \\ 2\cos 30° \end{bmatrix}$. Again, the bases are the same for both vector spaces, so this will be the second column of our matrix $M$.

Hence, the linear map $f$ is represented by the matrix

$$M = \begin{bmatrix} 2\cos 30° & -2\sin 30° \\ 2\sin 30° & 2\cos 30° \end{bmatrix} = \begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$$

and this matrix can be applied to any vector in $U$. For instance, consider the input vector $\vec{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$. To find out what this vector maps to, we simply work out the following expression:

$$\begin{aligned} M\vec{v} &= \begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix} \begin{bmatrix} 3 \\ 2 \end{bmatrix} \\ &= \begin{bmatrix} 3\sqrt{3} - 2 \\ 2\sqrt{3} + 3 \end{bmatrix} \end{aligned}$$

to get our answer. This is verified in figure 31.

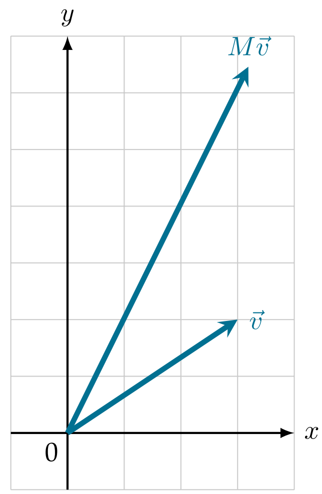



Figure 31: Verifying the equivalence between the linear map $f$ and the matrix $M = \begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$ by using an example vector $\vec{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

#### 4.1.2 Example 2: An endomorphism with changed bases

**Setup.** Like the previous example, we set both $U$ and $V$ to $\mathbb{R}^2$, and we want to find a matrix $M$ that corresponds to a linear map $f$ where vectors are scaled 1.5 times and rotated by 30° anticlockwise.

Only this time, we will use the canonical basis for the vector space $U$:

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}$$

and the following basis for $V$. (Note that this is a basis since the two vectors are linearly independent.)

$$\left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$$

The process is really the same. We know from the previous example that the canonical basis vectors will be mapped to $\begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ \sqrt{3} \end{bmatrix}$ respectively. All we have to do now is to express these vectors using the basis of $V$.

Let's take the first vector $\begin{bmatrix}\sqrt{3}\\1\end{bmatrix}$ as an example. We want to find scalars $x_1$ and $y_1$ such that

$$x_1\begin{bmatrix}2\\0\end{bmatrix}+y_1\begin{bmatrix}-1\\1\end{bmatrix}=\begin{bmatrix}\sqrt{3}\\1\end{bmatrix}.$$

This has the solution of $x_1=(\sqrt{3}+1)/2$ and $y_1=1$. In other words, the image of our first basis vector can be expressed using the basis of $V$ as $\begin{bmatrix}(\sqrt{3}+1)/2\\1\end{bmatrix}$. This will be the first column of our matrix $M$.

Similarly, we want to find scalars $x_2$ and $y_2$ such that

$$x_2\begin{bmatrix}2\\0\end{bmatrix}+y_2\begin{bmatrix}-1\\1\end{bmatrix}=\begin{bmatrix}-1\\\sqrt{3}\end{bmatrix}.$$

This has the solution of $x_2=(\sqrt{3}-1)/2$ and $y_2=\sqrt{3}$. This gives us the second column of our matrix: $\begin{bmatrix}(\sqrt{3}-1)/2\\\sqrt{3}\end{bmatrix}$.

Hence, the matrix that is equivalent to $f$ is given by

$$M=\begin{bmatrix}(\sqrt{3}+1)/2 & (\sqrt{3}-1)/2\\1 & \sqrt{3}\end{bmatrix}.$$

Comparing this to the previous example, we see that the same linear map can be represented as different matrices depending on the chosen bases.

#### 4.1.3 Example 3: Non-endomorphism

**Setup.** Now for something a bit different. Let us set $U=\mathbb{R}^3$ and $V=\mathbb{R}^2$.

Define $f$ as the linear map that projects a vector onto the $xy$-plane. In other words, the linear map takes some vector $\vec{v}$ in three-dimensional space, casts its shadow onto the $xy$-plane, and returns a vector that represents that shadow. See figure 32.

We will use the canonical bases for both $U$ and $V$, i.e.

$$\left\{\begin{bmatrix}1\\0\\0\end{bmatrix},\begin{bmatrix}0\\1\\0\end{bmatrix},\begin{bmatrix}0\\0\\1\end{bmatrix}\right\}$$

and

$$\left\{\begin{bmatrix}1\\0\end{bmatrix},\begin{bmatrix}0\\1\end{bmatrix}\right\}$$

respectively.

We want to find a matrix $M$ that represents this linear map.

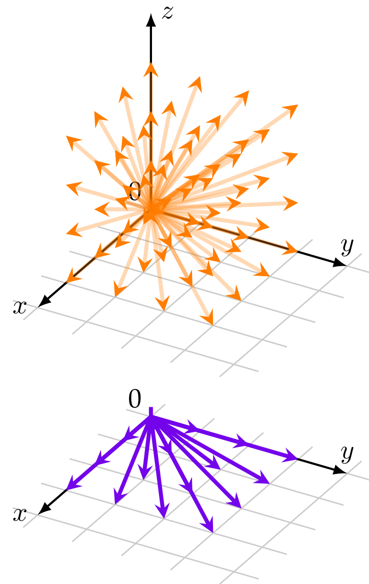



Figure 32: The linear map $f$ projects a three-dimensional vector (in orange) onto the $xy$-plane. This results in a new two-dimensional vector (in purple).

Firstly, we note that since we are mapping from a three-dimensional vector space to a two-dimensional one, our matrix must have the order $3 \times 2$ with three rows and two columns.

$$M = \begin{bmatrix} \square & \square & \square \\ \square & \square & \square \end{bmatrix}$$

Again, each column of the matrix represents the image of a basis vector of $U$, expressed in the basis of $V$. Following this principle:

- The first basis vector of $U$ is $\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$. Its projection is $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$.
- The second basis vector is $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$, which has the projection $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.
- The third basis vector is $\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$, with the projection $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$.

We're basically just losing the $z$-coordinate. Each of these projections is a column in our matrix $M$. This yields

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

which is equivalent to the linear map $f$.

### 4.1.4 Kernels and images

Given a linear map $f$ and its equivalent matrix $M$, we have the following.

$$\mathrm{Ker}(M) = \mathrm{Ker}(f)$$
$$\mathrm{Im}(M) = \mathrm{Im}(f)$$

## 4.2 The elements of a matrix

The numbers inside a matrix are called *elements* or *coefficients*. A matrix of order $m \times n$ can be indexed in the following manner.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}$$

Here, the element in the $i$-th row and $j$-th column is denoted as $a_{i,j}$.

## 4.3 Matrix arithmetic

Let $\mathcal{M}_{m,n}(\mathbb{R})$ denote the set of matrices with $m$ rows and $n$ columns containing real coefficients.

These matrices can be added or subtracted element-by-element.

$$M + Q = \begin{bmatrix} \quad m_{i,j} + q_{i,j} \quad \end{bmatrix}$$

$$M - Q = \begin{bmatrix} \quad m_{i,j} - q_{i,j} \quad \end{bmatrix}$$

They can also be multiplied by a scalar.

$$aM = \begin{bmatrix} \quad a \cdot m_{i,j} \quad \end{bmatrix}$$

Notice that the order of matrices do not change when they are added or multiplied. This means that $\mathcal{M}_{m,n}(\mathbb{R})$ is itself a vector space.

Moreover, matrices can also be multiplied, but only if their sizes match. The product of two matrices $MQ$ is only valid when

$$M \in \mathcal{M}_{m,n}(\mathbb{R})$$
$$Q \in \mathcal{M}_{n,k}(\mathbb{R})$$

for some integers $m$, $n$ and $k$. See figure 33.

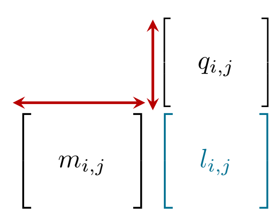



Figure 33: The product of two matrices $L = MQ$ is only valid when the number of columns in $M$ matches the number of rows in $Q$.

Note that in particular, all $n \times n$ matrices can be multiplied.

The multiplication process follows the following rule:

$$l_{i,j} = \sum_{r=1}^{n} m_{i,r} \cdot q_{r,j}$$

and has the following properties:

- If the matrices $M$ and $Q$ represent the linear maps $f$ and $g$ respectively, then the product $MQ$ represents the composition $f \circ g$.
- Square matrices (i.e. matrices of order $n \times n$ for some integer $n$) are associative.
- Square matrices have the following matrix as the unit (i.e. multiplicative identity).

$$I = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

- The multiplication between a matrix and a vector is a special case of matrix multiplication.
- Matrix multiplication is not commutative. For two matrices $M$ and $Q$, the equivalence $MQ = QM$ does not generally hold.

## 4.4 The inverse of a matrix

A square matrix describes a linear map from a vector space to another vector space of the same number of dimensions. Sometimes, the map described by the matrix just so happens to be bijective, meaning that it is possible to *invert* it.

A square matrix $M$ is said to be *invertible* when there is some *inverse* matrix $M^{-1}$ such that

$$M^{-1}M = MM^{-1} = I.$$

For any square matrix $M$, the following statements are equivalent:

- $M$ has an inverse.
- The columns of $M$ are linearly independent.
- The rows of $M$ are linearly independent.
- $\mathrm{Ker}(M) = \{0\}$.

Note that $I$ is its own inverse. Also, a matrix in which one row or column consists entirely of zeroes must be non-invertible.

## 4.5 How to invert a matrix

There are many ways to invert a matrix. Here we introduce one method that inverts a matrix by solving a system of linear equations via Gaussian elimination.

The best way to illustrate this method is with an example. Suppose we want to find the inverse of the matrix below.

$$M = \begin{bmatrix} 1 & 1 & -2 \\ 2 & 3 & 0 \\ 1 & 2 & 3 \end{bmatrix}$$

To do this, we place it inside an equation like so.

$$
\begin{aligned}
M\begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\
\begin{bmatrix} 1 & 1 & -2 \\ 2 & 3 & 0 \\ 1 & 2 & 3 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} a \\ b \\ c \end{bmatrix} \qquad (1)
\end{aligned}
$$

This is equivalent to a system of linear equations of three unknowns.

$$
\begin{cases} 1x + 1y - 2z = a \\ 2x + 3y + 0z = b \\ 1x + 2y + 3z = c \end{cases}
$$

We can solve this system via Gaussian elimination. We convert the augmented matrix into an equivalent upper triangular matrix, where all entries below the main diagonal are zero.

$$
\begin{aligned}
\left(\begin{array}{ccc|c} 1 & 1 & -2 & a \\ 2 & 3 & 0 & b \\ 1 & 2 & 3 & c \end{array}\right) &\sim \left(\begin{array}{ccc|c} 1 & 1 & -2 & a \\ 0 & 1 & 4 & -2a+b \\ 0 & 1 & 5 & -a+c \end{array}\right) \\
&\sim \left(\begin{array}{ccc|c} 1 & 1 & -2 & a \\ 0 & 1 & 4 & -2a+b \\ 0 & 0 & 1 & a-b+c \end{array}\right)
\end{aligned}
$$

Therefore we have

$$
z = a - b + c
$$

$$
\begin{aligned}
y + 4z &= -2a + b \\
y &= -2a + b - 4(a - b + c) \\
y &= -6a + 5b - 4c
\end{aligned}
$$

$$
\begin{aligned}
x + y - 2z &= a \\
x &= a - y + 2z \\
x &= a - (-6a + 5b - 4c) + 2(a - b + c) \\
x &= 9a - 7b + 6c
\end{aligned}
$$

which gives us the following set of solutions.

$$
\begin{aligned}
x &= 9a - 7b + 6c \\
y &= -6a + 5b - 4c \\
z &= a - b + c
\end{aligned}
$$

This can be expressed more concisely as follows.

$$
\begin{aligned}
\begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} 9a - 7b + 6c \\ -6a + 5b - 4c \\ a - b + c \end{bmatrix} \\
\begin{bmatrix} x \\ y \\ z \end{bmatrix} &= \begin{bmatrix} 9 & -7 & 6 \\ -6 & 5 & -4 \\ 1 & -1 & 1 \end{bmatrix}\begin{bmatrix} a \\ b \\ c \end{bmatrix} \qquad (2)
\end{aligned}
$$

Now compare our final equation (2) with our original system (1). We see that by solving the system, we have computed the inverse of our matrix $M$. Hence we conclude that

$$M^{-1} = \begin{bmatrix} 9 & -7 & 6 \\ -6 & 5 & -4 \\ 1 & -1 & 1 \end{bmatrix}.$$

## 4.6 Determinants

Each matrix is associated with a scalar quantity known as the *determinant*. For each matrix $A$, we denote its determinant as $\det(A)$, $\det A$ or $|A|$.

The determinant is helpful when we want to know whether a matrix is invertible, without having to compute the actual inverse. This is because this quantity has the following property.

$$\det(A) \neq 0 \;\Leftrightarrow\; A \text{ is invertible}$$

The determinant of a $2 \times 2$ matrix can be evaluated using the simple formula below.

$$\det \begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad - bc$$

Computing the determinant for larger matrices is slightly more complicated. To begin with, consider a square matrix $M$. Let $m_{i,j}$ be the element in the $i$-th row and $j$-th column of this matrix. We then define the following terminology.

- The *minor* of this element, denoted as $M_{i,j}$, refers to the determinant of the matrix obtained by removing the $i$-th row and $j$-th column from $M$. For example, if

  $$M = \begin{bmatrix} 1 & 1 & -2 \\ 2 & 3 & 0 \\ 1 & 2 & 3 \end{bmatrix}$$

  then the minor $M_{2,3}$ is given by

  $$M_{2,3} = \det \begin{bmatrix} 1 & 1 & \square \\ \square & \square & \square \\ 1 & 2 & \square \end{bmatrix} = \det \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} = 1 \times 2 - 1 \times 1 = 1$$

- The *cofactor* of this element, denoted as $C_{i,j}$, is exactly the same as its *minor*, except the sign (positive or negative) may have to be flipped depending on its position.

  $$C_{i,j} = (-1)^{i+j} M_{ij}$$

  Whether the sign have to flipped for the cofactor of an element can be visualised as a checkerboard pattern across the matrix.

  $$\begin{bmatrix} \color{blue}\square & \color{red}\square & \color{blue}\square & \color{red}\square & \cdots \\ \color{red}\square & \color{blue}\square & \color{red}\square & \color{blue}\square & \cdots \\ \color{blue}\square & \color{red}\square & \color{blue}\square & \color{red}\square & \cdots \\ \color{red}\square & \color{blue}\square & \color{red}\square & \color{blue}\square & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

  Here, the elements highlighted in blue are where the sign of the cofactor is identical to that of the minor. The ones highlighted in red are where the sign has to be flipped.

To calculate the determinant of a matrix, we pick any one of its rows or columns. Then, for each element in that row/column, multiply the element by its cofactor. The sum of the products calculated across the row/column is the determinant of the matrix.

An example might help. Consider yet again the following matrix.

$$M = \begin{bmatrix} 1 & 1 & -2 \\ 2 & 3 & 0 \\ 1 & 2 & 3 \end{bmatrix}$$

Let us choose the top row, with entries $1$, $1$ and $-2$.

- The minor of the first element is given by

$$\det \begin{bmatrix} \square & \square & \square \\ \square & 3 & 0 \\ \square & 2 & 3 \end{bmatrix} = \det \begin{bmatrix} 3 & 0 \\ 2 & 3 \end{bmatrix} = 3 \times 3 - 0 \times 2 = 9.$$

  For this element, the cofactor is exactly the same as the minor, with no sign flipping needed. The cofactor is $9$.

- The minor of the second element is given by:

$$\det \begin{bmatrix} \square & \square & \square \\ 2 & \square & 0 \\ 1 & \square & 3 \end{bmatrix} = \det \begin{bmatrix} 2 & 0 \\ 1 & 3 \end{bmatrix} = 2 \times 3 - 0 \times 1 = 6.$$

  For this element, we flip the sign of the minor to get the cofactor, which is $-6$.

- The minor of the last element is given by

$$\det \begin{bmatrix} \square & \square & \square \\ 2 & 3 & \square \\ 1 & 2 & \square \end{bmatrix} = \det \begin{bmatrix} 2 & 3 \\ 1 & 2 \end{bmatrix} = 2 \times 2 - 3 \times 1 = 1.$$

  For this element, the cofactor is exactly the same as the minor. The cofactor is $1$.

We now multiply each element by its cofactor to get its determinant.

$$\begin{aligned} \det M &= 1 \times 9 + 1 \times (-6) + (-2) \times 1 \\ &= 9 - 6 - 2 \\ &= 1 \end{aligned}$$

Since the determinant is nonzero, this matrix is invertible.

When computing determinants on paper using this method, we can speed up the process by picking a row or column that contains a zero. For instance, for the matrix $M$, if we had picked the second row, we would only have to compute two cofactors instead of three.

$$\det M = 2 \times (\text{cofactor of } 2) + 3 \times (\text{cofactor of } 3) + 0 \times (\text{cofactor of } 0)$$

Finally, we note the following:

- If a matrix has a row or column that consists entirely of zeroes, its determinant is zero.
- The determinant of an upper triangular matrix is equal to the product of the entries on its main diagonal.

# 5 Sequences and series

## 5.1 What is a sequence?

A *sequence* is essentially an ordered list of numbers. Examples include the harmonic sequence and the Fibonacci sequence, both of which are shown below.

$$1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \cdots \qquad \text{(Harmonic sequence)}$$

$$0, 1, 1, 2, 3, 5, 8, 13, \cdots \qquad \text{(Fibonacci sequence)}$$

We denote the sequence $u_0, u_1, u_2, u_3, \cdots$ as $(u_n)$, with $n \in \mathbb{N}$. Note that zero-based indexing is used here.

For instance, we can define the harmonic sequence using

$$u_n = \frac{1}{n+1}$$

while the Fibonacci sequence can be described by the recurrence relation

$$u_{n+2} = u_{n+1} + u_n.$$

## 5.2 Arithmetic, geometric, monotone, bounded and recursive sequences

An *arithmetic sequence* is one where each term is obtained by adding a constant increment $d$ to the previous term. The general term of an arithmetic sequence $(u_n)$ is given by

$$u_n = a + nd$$

where $a = u_0$ is the first term of the sequence.

A *geometric sequence* is one where each term is obtained by multiplying the previous term by a constant ratio $r$. The general term of a geometric sequence $(u_n)$ is given by

$$u_n = ar^n$$

where $a = u_0$ is the first term of the sequence.

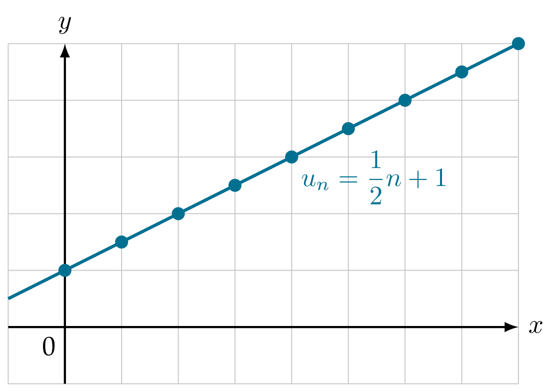



Figure 34: The graph of an arithmetic sequence.

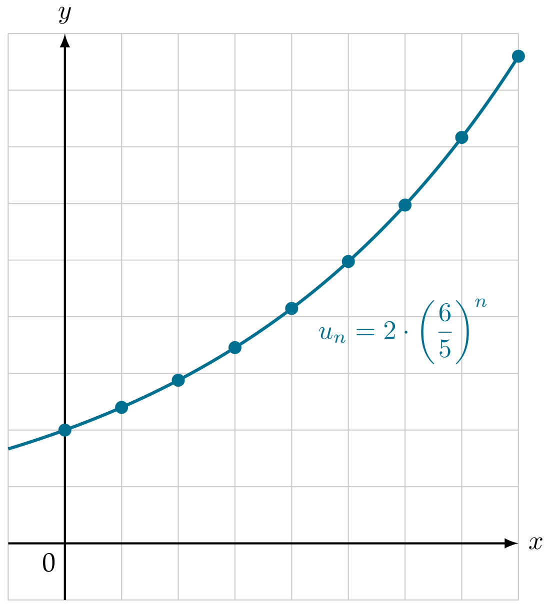



Figure 35: The graph of an geometric sequence.

A sequence $(u_n)$ is said to be...

- ...*increasing* if for all $n \in \mathbb{N}$, we have $u_{n+1} \geq u_n$.
- ...*decreasing* if for all $n \in \mathbb{N}$, we have $u_{n+1} \leq u_n$.
- ...*monotone* if it is either increasing or decreasing.
- ...*bounded* if there exists some real number $M$ such that for all $n \in \mathbb{N}$, we have $|u_n| \leq M$.

Notice that in non-strict inequalities (as opposed to strict inequalities) are used in the first two definitions. This means that a constant sequence is both increasing and decreasing.

In the last definition, both strict or non-strict inequalities can be used — they are equivalent in this context.

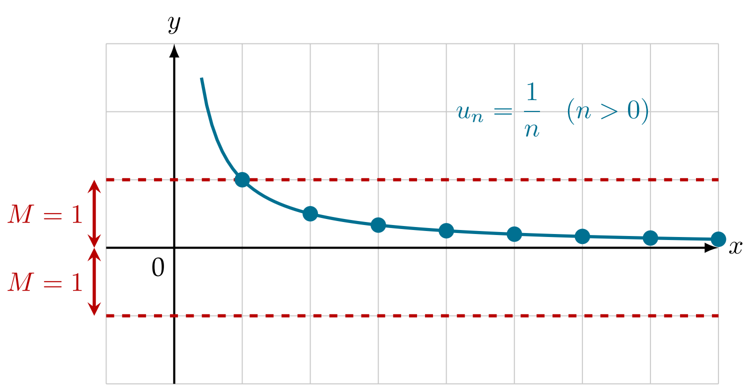



Figure 36: The graph of a decreasing and bounded sequence. Note that none of the terms have an absolute value that exceeds $M = 1$, as shown by the red dashed line.

A *recursive sequence* is defined by a recurrence relation along with some initial values. In other words, each term of the sequence is expressed as a function of the previous terms. For instance, the Fibonacci sequence can be defined as follows.

$$\begin{cases} u_0 = 0 \\ u_1 = 1 \\ u_{n+2} = u_{n+1} + u_n \end{cases}$$

## 5.3 Convergence of a sequence

Some sequences $(u_n)$ have a limit as $n$ approaches infinity[3]. For instance, as shown in figure 36, the harmonic sequence has a limit of zero. In other words, it *converges* to zero.

We say that a sequence $(u_n)$ converges to a limit $L$ if the terms $u_n$ can get arbitrarily close to $L$ for sufficiently large values of $n$, i.e.

$$\forall \epsilon > 0,\ \exists n_0 \in \mathbb{N},\ n > n_0 \implies |u_n - L| < \epsilon$$

where $n_0$ is called a *rank*.

For example, we can prove that the harmonic sequence converges to zero as follows.

**Problem.** Prove that the harmonic sequence converges to zero, i.e.

$$\lim_{n \to \infty} \frac{1}{n} = 0$$

**Intuition.** Recall the definition of a limit of a seqeence.

$$\forall \epsilon > 0,\ \exists n_0 \in \mathbb{N},\ n > n_0 \implies |u_n - L| < \epsilon$$

Substituting $u_n = 1/n$ and $L = 0$, we can simplify the statement above to

$$\forall \epsilon > 0,\ \exists n_0 \in \mathbb{N},\ n > n_0 \implies \frac{1}{n} < \epsilon.$$

Notice that we can rearrange the inequality at the end as follows.

$$\begin{aligned} \frac{1}{n} &< \epsilon \\ n &> \frac{1}{\epsilon} \end{aligned}$$

This means that if we choose $n_0 = 1/\epsilon$, then the inequality holds for all $n > n_0$.

**Solution.** Let $\epsilon > 0$. If we take $n_0 = 1/\epsilon$, then for all $n > n_0$, we have

$$\begin{aligned} n &> n_0 \\ n &> \frac{1}{\epsilon} \\ \epsilon &> \frac{1}{n} \\ \frac{1}{n} &< \epsilon \\ u_n &< \epsilon \\ |u_n - 0| &< \epsilon \end{aligned}$$

[3]Note that as sequences as discrete, it doesn't make sense to consider limits as $n$ approaches any value other than infinity. Consequently, whenever we mention a limit of a sequence, it is implicitly assumed that we mean the limit as $n$ goes to infinity.

which concludes the proof.

Note that

- Arithmetic sequences do not converge unless the increment is zero.
- Geometric sequences converge if the ratio $r$ satisfies either of the following conditions.

$$r = 1$$
$$-1 < r < 1$$

  When $r = 1$, the sequence is a constant sequence, so it converges to that constant term. When $-1 < r < 1$, the sequence converges to zero.

- The convergence of a recursive sequence depends on several factors, including the properties of the function used to define the recurrence relation.

## 5.4 Asymptotic behaviour of sequences

Like with functions, we can compare the behaviour of sequences near infinity using little $o$ and big $O$ notation. For example, we have:

$$\begin{aligned} 2n + 5 &= o(n^2) \\ 2n + 5 &= O(n) \\ \frac{1}{n} &= o(1). \end{aligned}$$

## 5.5 Important thereoms about sequences and their limits

Here we introduce a couple of theorems regarding the limit of sequences.

- **Theorem of monotone convergence.** A monotone sequence converges if and only if it is bounded.

  For example, consider $u_n = 1/n!$. Since

$$u_{n+1} = \frac{1}{(n+1)!} = \frac{1}{n+1} \cdot u_n \leq u_n$$

  the sequence is decreasing. Furthermore, we have

$$\left| \frac{1}{n!} \right| < 1$$

  so the sequence is bounded. Therefore, $(u_n)$ must converge.

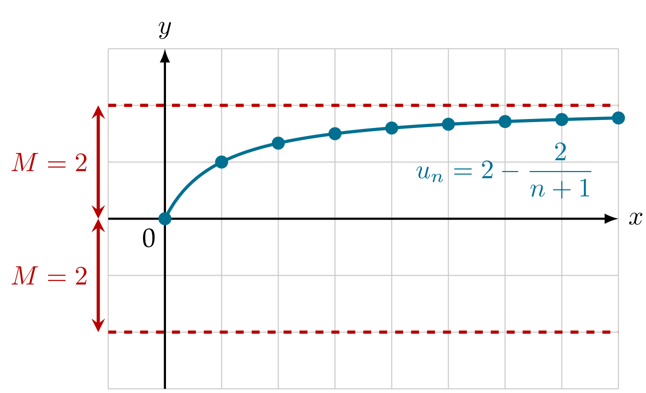



Figure 37: The graph of a increasing and bounded sequence that converges to 2.

- **Squeeze theorem.** Let $(u_n)$, $(v_n)$, and $(w_n)$ be sequences such that for all $n \in \mathbb{N}$, we have

  $$u_n \leq v_n \leq w_n.$$

  If both $(u_n)$ and $(w_n)$ converge to the same limit $L$, then $(v_n)$ also converges to $L$.

  For example, suppose we want to compute the limit of the following sequence. (Assume $n > 0$.)

  $$v_n = \frac{\sin n}{n}$$

  Doing this using the classic definition of a limit can be quite difficult. Now consider the following sequences.

  $$u_n = -\frac{1}{n}$$
  $$w_n = \frac{1}{n}$$

  Notice that that both of these sequences converge to zero. Also, for all $n \in \mathbb{N}$, we have

  $$-1 \leq \sin n \leq 1$$
  $$-\frac{1}{n} \leq \frac{\sin n}{n} \leq \frac{1}{n}$$
  $$u_n \leq v_n \leq w_n$$

  By the squeeze theorem, the sequence $(v_n)$ also converges to zero. See figure 38.

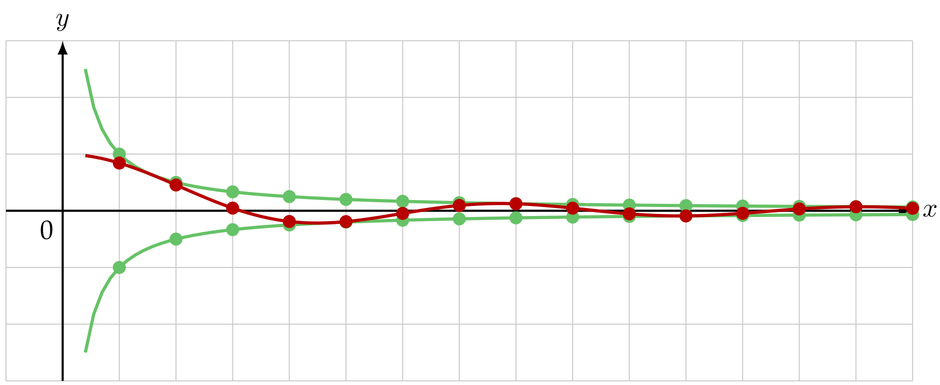



Figure 38: A demonstration of the squeeze theorem. The red curve represents the sequence $v_n = (\sin x)/x$. The green curves represent the sequences $u_n = -1/n$ and $w_n = 1/n$.

## 5.6 What is a series?

Consider the following sequence.

$$1, 2, 3, 4, 5, 6, \cdots$$

What happens when we add up the first $k$ terms of this sequence?

$$
\begin{aligned}
k &= 1 \qquad && \underbrace{1}_{1} + 2 + 3 + 4 + 5 + 6 + \cdots \\
k &= 2 \qquad && \underbrace{1 + 2}_{3} + 3 + 4 + 5 + 6 + \cdots \\
k &= 3 \qquad && \underbrace{1 + 2 + 3}_{6} + 4 + 5 + 6 + \cdots \\
k &= 4 \qquad && \underbrace{1 + 2 + 3 + 4}_{10} + 5 + 6 + \cdots \\
k &= 5 \qquad && \underbrace{1 + 2 + 3 + 4 + 5}_{15} + 6 + \cdots
\end{aligned}
$$

Each of these sums is called a *partial sum*, and together they form a new sequence, which in this case is the triangular numbers.

$$1, 3, 6, 10, 15, 21, \cdots$$

Whenever we take an existing sequence $(u_n)$ and compute its partial sums $\left(\sum_{n=0}^{k=n} u_k\right)$ to form a new sequence, this new sequence is called a *series*, denoted as $\sum u_n$. The series of triangular numbers in particular can be described by the general term $S_n = n(n+1)/2$. (This is a special case where we opt for one-based indexing!)[4]

Now consider an arithmetic sequence

$$u_n = a + nd$$

where $a = u_0$ is the first term. The series of this sequence is given by

$$\begin{aligned} u_0 + u_1 + u_2 + \cdots + u_n &= a + (a+d) + (a+2d) + \cdots + (a+nd) \\ &= (n+1)a + (1+2+\cdots+n)d \\ &= (n+1)a + \frac{n(n+1)}{2}d \qquad \text{(using the formula for triangular numbers)} \end{aligned}$$

What about the partial sums of a geometric sequence instead? Consider the sequence

$$u_n = ar^n$$

where again $a = u_0$ is the first term. Denote the sum $u_0 + u_1 + u_2 + \cdots + u_n$ as $S_n$. We can write

$$\begin{aligned} S_n &= a + ar + ar^2 + \cdots + ar^n \\ rS_n &= ar + ar^2 + ar^3 + \cdots + ar^{n+1} \end{aligned}$$

Subtracting the bottom equation from the top one gives

$$(1-r)S_n = a - ar^{n+1}$$

which gives us the following formula for the series of a geometric sequence.

$$S_n = a + ar + ar^2 + \cdots + ar^n = \frac{a}{1-r}(1-r^{n+1})$$

## 5.7 Convergence of a series

Some series do not converge.

$$1 + 2 + 3 + 4 + \cdots \to \infty$$

However, some series do converge — for instance, consider the following series. As the number of terms approach infinity, the sum approaches 2.

$$1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 2$$

To see why, we first note that this is a geometric series with partial sums

$$\begin{aligned} S_n &= \frac{1}{1-\frac{1}{2}}\left(1 - \left(\frac{1}{2}\right)^{n+1}\right) \\ &= 2\left(1 - \left(\frac{1}{2}\right)^{n+1}\right) \end{aligned}$$

As $n$ approaches infinity, the term in red approaches zero, so the sum converges to 2. We can generalise this result as follows.

[4] The formula can be proved as follows. Suppose we want to find $S_n = 1 + 2 + 3 + \cdots + (n-2) + (n-1) + n$. This can be rewritten as $S_n = n + (n-1) + (n-2) + \cdots + 3 + 2 + 1$. We can add these two equations together to get $2S_n = (n+1) + (n+1) + \cdots + (n+1) = n(n+1)$. Dividing both sides by 2 yields $S_n = n(n+1)/2$.

**Convergence of a geometric series.**

A geometric series $\sum ar^n$ converges to $a/(1-r)$ if $-1 < r < 1$.

In other words, we have

$$\sum_{n=0}^{\infty} ar^n = \frac{a}{1-r}$$

when $-1 < r < 1$.

This can be further generalised into what is called the *ratio test* (also known as d'Alembert's criterion). But first, let us define the term *absolute convergence*. A series $\sum u_n$ is said to be *absolutely convergent* if the series $\sum |u_n|$ converges.

And now, the ratio test.

**Ratio test for convergence of a series.**

To determine whether a series $\sum a_n$ converges (assuming $a_n \neq 0$ for all $n$), we consider the sequence of ratios $\dfrac{|a_{n+1}|}{|a_n|}$.

$$\begin{aligned}
\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} > 1 &\implies \text{the series diverges} \\
\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} < 1 &\implies \text{the series converges absolutely} \\
\lim_{n\to\infty} \frac{|a_{n+1}|}{|a_n|} = 1 &\implies \text{no conclusion reached}
\end{aligned}$$

For example, consider the series $\sum (n+1)/2^n$. Applying the ratio test, we consider the limit

$$\begin{aligned}
\lim_{n\to\infty} \frac{(n+2)/2^{n+1}}{(n+1)/2^n} &= \lim_{n\to\infty} \frac{n+2}{2(n+1)} \\
&= \lim_{n\to\infty} \frac{(n+1)+1}{2(n+1)} \\
&= \lim_{n\to\infty} \frac{(n+1)+1}{2(n+1)} \\
&= \lim_{n\to\infty} \left(\frac{1}{2} + \frac{1}{2(n+1)}\right) \\
&= \frac{1}{2} \\
&< 1
\end{aligned}$$

which means the series is absolutely convergent. Since all terms are positive, the series is also convergent.

We also have a comparison test for series convergence. Consider two series $\sum a_n$ and $\sum b_n$, where $a_n, b_n \geq 0$. Then,

- If we have $a_n \leq b_n$ and $\sum b_n$ is convergent, then $\sum a_n$ is also convergent (smaller than a convergent series).
- If we have $a_n \geq b_n$ and $\sum b_n$ is divergent, then $\sum a_n$ is also divergent (larger than a divergent series).

As an example, consider the series $\sum 1/(2^n+1)$. We can easily show that $1/(2^n+1) \leq \frac{1}{2^n}$. Since the series $\sum 1/2^n$ is convergent, the series $\sum 1/(2^n+1)$ is also convergent.

Similarly, consider the series $\sum_{n\geq 3}(\ln n)/n$. When $n \geq 3$, we have $\ln n \geq 1$, so $(\ln n)/n \geq 1/n$. Since the series $\sum 1/n$ is divergent, the series $\sum(\ln n)/n$ is also divergent.

# 6 Differential calculus

## 6.1 What is a derivative?

Differential calculus is the study of infinitesimal variations. Consider for instance a function $y = f(x)$. Given some interval from $a$ and $a + h$, we can study the variation of $f$ in this interval by considering the ratio

$$\frac{f(a+h) - f(a)}{h}.$$

This is visualised in figure 39.

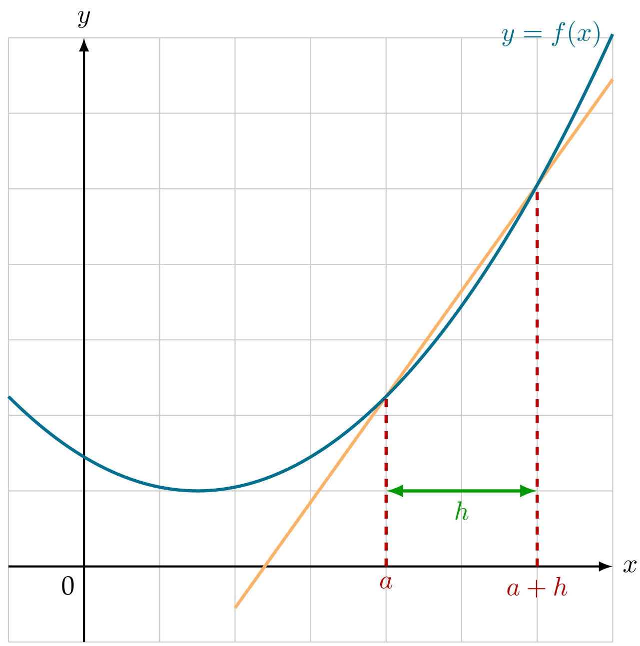



Figure 39: Studying the variation of a function $f$ between $x = a$ and $x = a + h$. The ratio $\frac{f(a+h)-f(a)}{h}$ can be visualised as the slope of the orange line which passes through the points $(a, f(a))$ and $(a + h, f(a + h))$.

To study the variations close to $a$, we consider the limit of this ratio as $h$ tends to zero. See figure 40.

$$\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

This limit is referred to as the *derivative* of $f$ in $a$. If this limit exists, then the function $f$ is said to be *differentiable* in $a$.

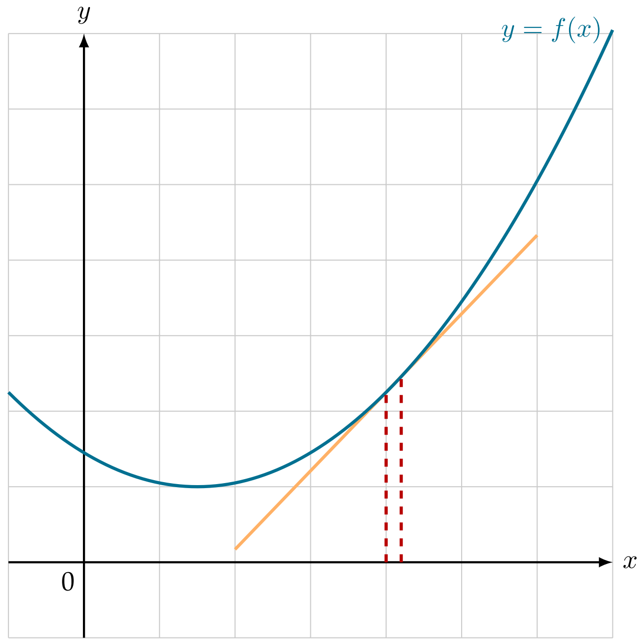



Figure 40: To find the derivative of a function $f$ in $a$, we consider the slope of the orange line as the length of the interval $h$ tends to zero.

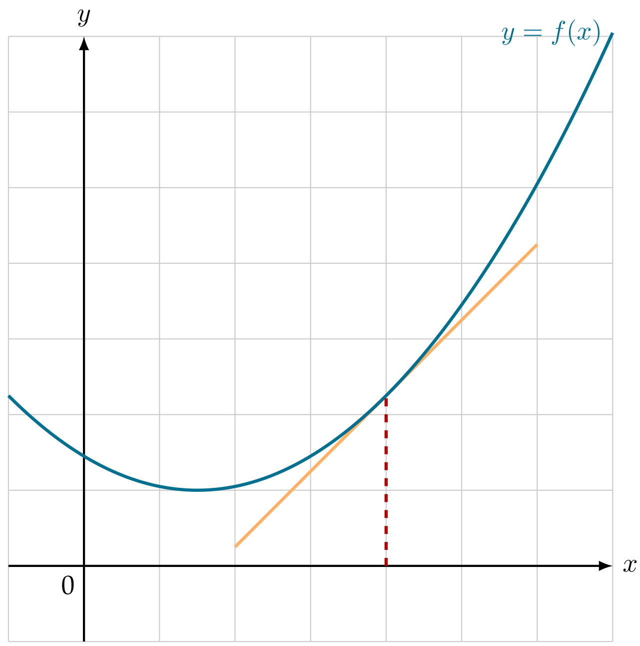



Figure 41: The derivative of a function $f$ in a point $a$ can be thought of as the slope of the tangent to the graph of $f$ at the point where $x = a$.

The derivative of a function $f$ in some value $a$ is denoted as $f'(a)$.

Note that the derivative of $f$ is itself a function, which is denoted here as $f'$. We can also write this as

$$\frac{df}{dx} \quad \text{or} \quad f^{(1)}.$$

Note that $\frac{df}{dx}$ is not a quotient or fraction. It can be read as applying a differentiation operator $\frac{d}{dx}$ to a function $f$. The differentiation operator specifies the variable of differentiation, which in this case is $x$.

Note that

- If the derivative of a function $f$ is $g$, i.e.
$$f' = g$$
then $f$ is called the *primitive* or *antiderivative* of $g$.

- Applying the differentiation operator to a function $f$ twice, i.e.
$$\frac{d}{dx}\left(\frac{df}{dx}\right)$$
can be written more compactly as
$$\frac{d^2 f}{dx^2}.$$
This produces a *second-order derivative*, which is denoted as $f''(x)$ or $f^{(2)}(x)$.

## 6.2 Finding the derivative of a function

To find the derivative of a function, we can use the definition of the derivative as a limit. Consider for instance the function $f(x) = x^2$. We can find the derivative of $f$ by computing the limit

$$\begin{aligned}
\lim_{h\to 0} \frac{f(x+h) - f(x)}{h} &= \lim_{h\to 0} \frac{(x+h)^2 - x^2}{h} \\
&= \lim_{h\to 0} \frac{x^2 + 2hx + h^2 - x^2}{h} \\
&= \lim_{h\to 0} \frac{2hx + h^2}{h} \\
&= \lim_{h\to 0} (2x + h) \\
&= \lim_{h\to 0} 2x \\
&= 2x
\end{aligned}$$

which means that the derivative of $f(x) = x^2$ at any point $a$ is given by $f'(a) = 2a$. In other words, the tangent to the graph of $f(x)$ when $x = a$ must have a slope of $2a$.

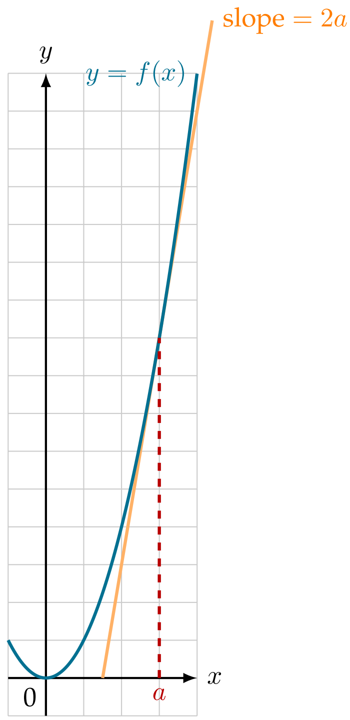



Figure 42: The derivative of $x^2$ is $2x$. This means the derivative of $x^2$ (or the slope of the tangent to the graph of $y = x^2$) at any point $a$ is $2a$.

Finding derivatives as a limit can be cumbersome and tedious. To speed this up, it might be helpful to memorise a few differentiation rules, as listed below. Here, $f$ and $g$ are differentiable functions, $c$ is a constant, and $n$ is any real number.

$$c \xrightarrow{\frac{d}{dx}} 0 \tag{1}$$

$$x \xrightarrow{\frac{d}{dx}} 1 \tag{2}$$

$$\sin x \xrightarrow{\frac{d}{dx}} \cos x \tag{3}$$

$$\cos x \xrightarrow{\frac{d}{dx}} -\sin x \tag{4}$$

$$e^x \xrightarrow{\frac{d}{dx}} e^x \tag{5}$$

$$\ln x \xrightarrow{\frac{d}{dx}} \frac{1}{x} \tag{6}$$

$$f(x) \pm g(x) \xrightarrow{\frac{d}{dx}} f'(x) \pm g'(x) \tag{7}$$

$$c \cdot f(x) \xrightarrow{\frac{d}{dx}} c \cdot f'(x) \tag{8}$$

$$x^n \xrightarrow{\frac{d}{dx}} nx^{n-1} \tag{9}$$

$$f(x) \cdot g(x) \xrightarrow{\frac{d}{dx}} f(x) \cdot g'(x) + g(x) \cdot f'(x) \tag{10}$$

$$\frac{f(x)}{g(x)} \xrightarrow{\frac{d}{dx}} \frac{g(x) \cdot f'(x) - f(x) \cdot g'(x)}{(g(x))^2} \tag{11}$$

$$f(g(x)) \xrightarrow{\frac{d}{dx}} f'(g(x)) \cdot g'(x) \tag{12}$$

Several notes:

- From equation (7) we have

$$f(x)+g(x) \xrightarrow{\frac{d}{dx}} f'(x)+g'(x).$$

  Combining this with equation (8) we see that differentiation is a linear operation.

- The power rule (equation (9)) can be applied with any real number $n$. A few examples are shown below.

$$\begin{aligned} \frac{1}{x} = x^{-1} &\xrightarrow{\frac{d}{dx}} -x^{-2} = -\frac{1}{x^2} \\ \sqrt{x} = x^{\frac{1}{2}} &\xrightarrow{\frac{d}{dx}} \frac{1}{2}x^{-\frac{1}{2}} = \frac{1}{2\sqrt{x}} \end{aligned}$$

- The quotient rule (equation (11)) can be derived from the power rule (equation (9)), the product rule (equation (10)), and the chain rule (equation (12)).

An example problem is shown below.

**Problem.** Calculate $\frac{d}{dx}\, 3e^{x^2+2}$.

**Solution 1.** Let $f(x) = 3e^x$ and $g(x) = x^2+2$. We want to find $\frac{d}{dx} f(g(x))$.

Note that $f'(x) = 3e^x$ and $g'(x) = 2x$. Hence, we have

$$\begin{aligned} \frac{d}{dx} f(g(x)) &= f'(g(x)) \cdot g'(x) && \text{(chain rule)} \\ &= f'(x^2+2) \cdot 2x \\ &= 3e^{x^2+2} \cdot 2x \\ &= 6xe^{x^2+2}. \end{aligned}$$

**Solution 2.** Those who are more fluent with the rules of differentiation can simply apply the rules to find:

$$\begin{aligned} \frac{d}{dx} 3e^{x^2+2} &= 3 \cdot \left(e^{x^2+2}\right) \cdot (2x) \\ &= 6xe^{x^2+2}. \end{aligned}$$

Generalising from this example problem, we can show that a function of the form

$$f(x) = c \cdot e^{g(x)} \qquad \text{(where } c \text{ is a constant)}$$

must have the derivative

$$\begin{aligned} f'(x) &= c \cdot e^{g(x)} \cdot g'(x) \\ &= f(x) \cdot g'(x). \end{aligned}$$

This will be useful for solving differential equations later.

## 6.3 More on differentiability

Recall that a function $f(x)$ is said to be differentiable if the derivative of $f'(x)$ exists when $x = a$.

We note the following theorem.

**Differentiability is stronger than continuity.**

If a function $f$ is differentiable in $a$, then $f$ is continuous in $a$.

We can prove this as follows.

**Proof that differentiability implies continuity.**

We assume differentiability, meaning that the limit

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

exists. We want to prove that

$$\lim_{x \to a} f(x) = f(a).$$

Consider the following limit.

$$\begin{aligned}
\lim_{x \to a} (f(x) - f(a)) &= \lim_{x \to a} \left( (f(x) - f(a)) \cdot \frac{x-a}{x-a} \right) && (\because x - a \neq 0) \\
&= \lim_{x \to a} \left( \frac{f(x) - f(a)}{x-a} \cdot (x-a) \right) \\
&= \left( \lim_{x \to a} \frac{f(x) - f(a)}{x-a} \right) \cdot \left( \lim_{x \to a} (x-a) \right) \\
&= f'(a) \cdot 0 \\
&= 0
\end{aligned}$$

Rearranging, we get

$$\lim_{x \to a} f(x) = f(a)$$

which concludes the proof.

Intuitively, we can think of differentiability as whether the tangent to the graph of a function $f$ at a point $a$ exists. For instance, the absolute value function $|x|$ is not differentiable at $x = 0$ because at $x = 0$, the graph of $y = |x|$ consists of a sharp angular point where no tangent can be drawn. See figure 43.

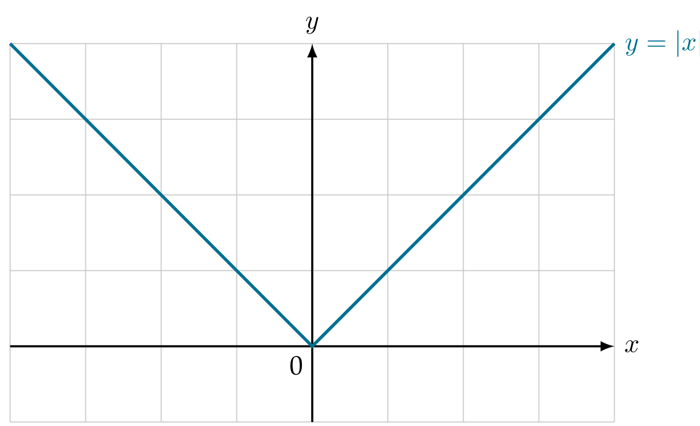



Figure 43: The absolute value function is non-differentiable at $x = 0$.

## 6.4 Stationary points

If we have $\frac{df}{dx} > 0$ at some point $a$, we say that $f$ is *increasing* at that point. Conversely, if we have $\frac{df}{dx} < 0$ at some point $a$, we say that $f$ is *decreasing* at that point. But what if $\frac{df}{dx} = 0$?

When this happens, we say that $f$ has a *stationary point* at $a$. This could mean a number of things:

- A local minimum;
- A local maximum; or
- An inflection point[5].

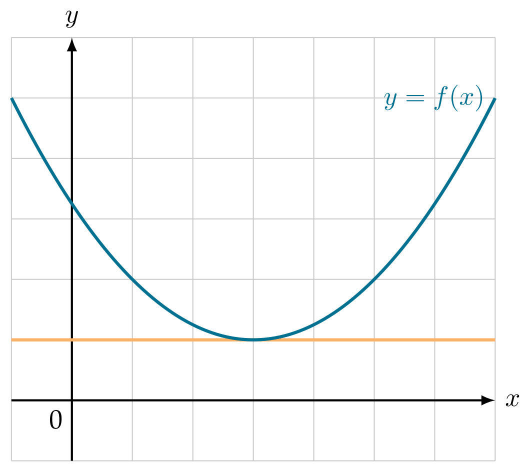



Figure 44: At a local minimum, the derivative of a function is zero.

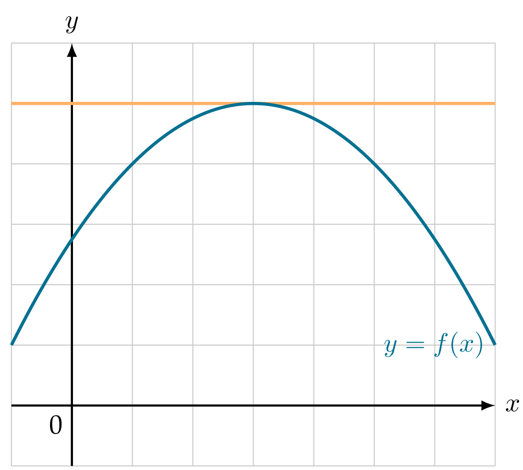



Figure 45: At a local maximum, the derivative of a function is zero.

[5] Also called a "point of inflection".

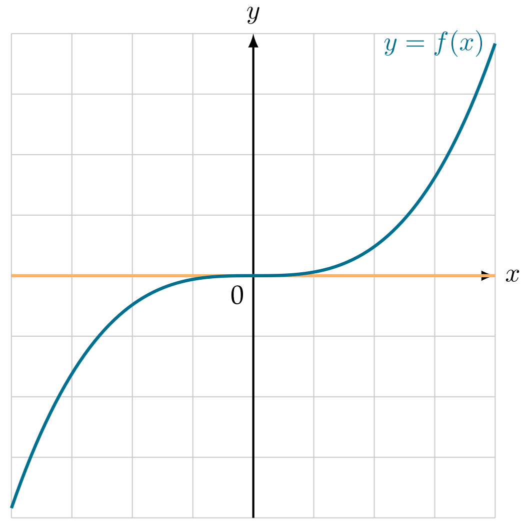



Figure 46: At an inflection point, the derivative of a function is zero.

There are several ways to distinguishing between these cases, as will be demonstrated using the following function.

$$f(x) = 2x^3 - 9x^2 + 12x - 4$$

To find the stationary points of $f$, we first compute its derivative and set it equal to zero.

$$\begin{aligned} f'(x) = 6x^2 - 18x + 12 &= 0 \\ x^2 - 3x + 2 &= 0 \\ (x-1)(x-2) &= 0 \\ x &= 1 \text{ or } 2 \end{aligned}$$

This means that this function has two stationary points: one at $x = 1$ and one at $x = 2$. But are these local maxima, local minima, or inflection points?

One straightforward way for working this out is the *first derivative test*, where we consider the sign of the derivative around the stationary point.

| $x$ | 0 | 1 | 1.5 | 2 | 3 |
|---|---|---|---|---|---|
| $f'(x)$ | $+$ | 0 | $-$ | 0 | $+$ |

We can infer from the table above that $f$ has a local maximum at $x = 1$ and a local minimum at $x = 2$, which can be verified using a graph.

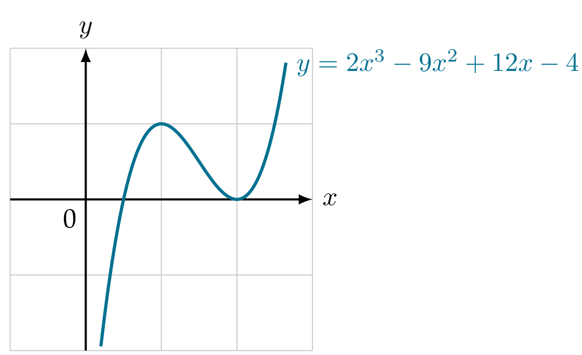



Figure 47: The function $f(x) = 2x^3 - 9x^2 + 12x - 4$ has a local maximum at $x = 1$ and a local minimum at $x = 2$.

Another way to identify the nature of a stationary point is the *second derivative test*. This involves computing the second derivative of the function at each stationary point.

- If the second derivative is positive, then the function has a local minimum at that point.
- If the second derivative is negative, then the function has a local maximum at that point.
- If the second derivative is zero, then the test is inconclusive. The point can be a local maximum, a local minimum, or an inflection point.

Applying this test to the function above, we have

$$f''(x) = 12x - 18$$

which gives us

$$f''(1) = -6 < 0$$
$$f''(2) = 6 > 0$$

This leads to the same result as before: $f$ has a local maximum at $x = 1$ and a local minimum at $x = 2$.

## 6.5 Differential equations

Differential equations are ones that involve both functions and their derivatives. Recall from earlier that a function of the form

$$f(x) = c \cdot e^{g(x)}$$

must have the derivative

$$f'(x) = c \cdot e^{g(x)} \cdot g'(x) = f(x) \cdot g'(x)$$

where $c \in \mathbb{R}$. We will make use of this relationship a lot when solving differential equations, as we will see in the following example problems.

**Problem.** Solve $f(x) = f'(x)$.

**Solution.** $f(x) = C \cdot e^x$ where $C \in \mathbb{R}$.

**Problem.** Solve $f'(x) = -3f(x)$.

**Solution.** $f(x) = C \cdot e^{-3x}$ where $C \in \mathbb{R}$.

Now consider the slightly more complicated case of having to solve

$$f'(x) = a(x) \cdot f(x).$$

To do this, we find the primitive (antiderivative) of $a(x)$, which we will denote $A(x)$. In other words, we have $A'(x) = a(x)$. The solutions to this differential equation are then given by $f(x) = C \cdot e^{A(x)}$, where $C \in \mathbb{R}$. An example is given below.

**Problem.** Solve $f'(x) = (2x + 1) \cdot f(x)$.

**Solution.** The antiderivative of $2x + 1$ is $x^2 + x$, so the solution to the equation is given by $f(x) = C \cdot e^{x^2+x}$ where $C \in \mathbb{R}$.

These solutions, which contain unspecified multiplicative constants, are called *general solutions* of the differential equation. Sometimes we might be asked to find specific values for these constants by using initial conditions.

**Problem.** What is the solution of $f'(x) = (2x + 1) \cdot f(x)$ that verifies $f(0) = 3$?

**Solution.** Continuing from the previous example problem, we know that the solution to the differential equation is given by $f(x) = C \cdot e^{x^2+x}$, where $C \in \mathbb{R}$. We can find the value of $C$ by using the initial condition $f(0) = 3$.

$$\begin{aligned} C \cdot e^{0^2+0} &= 3 \\ C &= 3 \end{aligned}$$

This gives us the specific solution $f(x) = 3 \cdot e^{x^2+x}$.

# 7 Integral calculus

Calculating discrete sums is rather easy arithmetic, but how about infinite sums? Is it possible to, say, compute the area under a given curve?

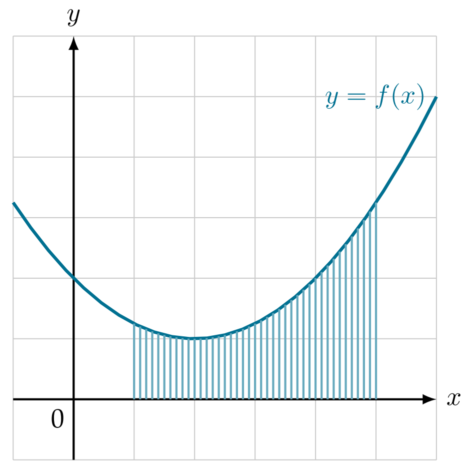



Figure 48: The area under a quadratic curve.

Let us consider a simpler example, where the curve in question is merely a straight line. Suppose we have a linear function $f(x) = 2x$. How can we calculate the area under its graph, evaluated between $x = 0$ and some $x = a$?

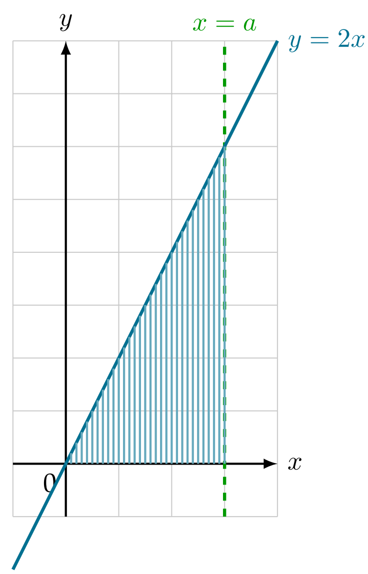



Figure 49: The area under the straight line given by the function $y = 2x$.

Geometry tells us that the area of this triangle can be calculated as

$$\frac{1}{2}(a)(2a) = a^2.$$

If we rename our variable $a$ as $x$, we see that the area under the straight line $y = 2x$ evaluated between $0$ and some real number $x$ can be expressed as $x^2$. Notice that differentiating $x^2$ gives us $2x$, which is our original linear function!

This is by no means a coincidence — in fact, the process of finding the area under a curve is the inverse operation of differentiation. This process of finding a primitive or antiderivative of a function is known as *integration*, which will be the focus of this section.

## 7.1 What is an indefinite integral?

Recall from the previous section that if $f$ is the derivative of $F$, then $F$ is called a *primitive* or *antiderivative* of $f$. For example, $x^2$ is a primitive of $2x$.

$$x^2 \rightleftarrows 2x$$

However, that this is not the only primitive of $2x$. For instance, $x^2 + 5$ and $x^2 - 3$ are both valid antiderivatives. In fact, as we shall prove below, all primitives of $f$ differ by a constant.

> **Theorem.** for any function $f$, all primitives of $f$ differ by a constant.
>
> **Proof.** Let $F_1$ and $F_2$ be two primitives of $f$. We want to prove that $F_1 - F_2$ is a constant.
>
> To do this, we consider the derivative of the expression $F_1 - F_2$.
>
> $$\begin{aligned}(F_1 - F_2)' &= F_1' - F_2' \\ &= f - f \\ &= 0.\end{aligned}$$
>
> Since $(F_1 - F_2)' = 0$, the difference $F_1 - F_2$ must be a constant.

From this theorem, we conclude that $x^2 + C$ is a primitive of $2x$ for any constant $C$. We can express this fact using an *indefinite integral*, as shown below.

$$\int 2x\,dx = x^2 + C$$

Here, the constant $C$ is called the *constant of integration*[6].

## 7.2 Rules of integration

We can modify our rules of differentiation from the last section into rules of integration, some of which are shown below.

$$c \xrightarrow{\int} cx + C \tag{1}$$

$$x \xrightarrow{\int} \frac{1}{2}x^2 + C \tag{2}$$

$$x^p \xrightarrow{\int} \frac{1}{p+1}x^{p+1} + C \tag{3}$$

$$e^x \xrightarrow{\int} \frac{1}{p+1}e^x + C \tag{4}$$

$$\frac{1}{x} \xrightarrow{\int} \ln|x| + C \tag{5}$$

$$\sin x \xrightarrow{\int} -\cos x + C \tag{6}$$

$$\cos x \xrightarrow{\int} \sin x + C \tag{7}$$

Here, $c$ is a constant, $p$ is a real number, and $C$ is the constant of integration[7]. We've left out the product rule and chain rule for now, but we'll see how they can be applied to integration later on.

[6] Or: the *integration constant*.

[7] The rules for integration are essentially the reverse of the rules for differentiation, with a few exceptions. For example, the integral of $1/x$ is $\ln|x|$ rather than $\ln x$, as the natural logarithm is only defined for positive numbers.

## 7.3 Integration by substitution

*Integration by substitution* is a powerful technique for integrating functions that are not immediately obvious. Consider, for example, the integral below.

$$\int 2x\sqrt{x^2+1}\,dx$$

This integral is seems a little tricky at first glance, but we can make it easier by substituting $u = x^2 + 1$. This substitution gives us

$$\frac{du}{dx} = 2x.$$

Not so rigorously, we can rearrange this equation to obtain

$$du = 2x\,dx$$

the right-hand-side of which is part of our original integral. We can now rewrite our integral in terms of $u$ like so:

$$\begin{aligned}\int 2x\sqrt{x^2+1}\,dx &= \int \sqrt{u}\,du \\ &= \int u^{\frac{1}{2}}\,du \\ &= \frac{2}{3}u^{\frac{3}{2}} + C \\ &= \frac{2}{3}(x^2+1)^{\frac{3}{2}} + C\end{aligned}$$

which gives us the answer.

Upon closer examination, one can find that integration by substitution is but the chain rule in disguise, as shown below.

$$\begin{aligned}\frac{d}{dx}f(g(x)) &= f'(g(x)) \cdot g'(x) && \text{(chain rule)} \\ \int \frac{d}{dx}f(g(x))\,dx &= \int f'(g(x)) \cdot g'(x)\,dx && \text{(integrate both sides)} \\ f(g(x)) + C &= \int f'(g(x)) \cdot g'(x)\,dx && \text{(integral \& derivative cancel out)} \\ \int f'(g(x)) \cdot g'(x)\,dx &= f(g(x)) + C && \text{(integration by substitution)}\end{aligned}$$

## 7.4 Integration by parts

Another useful technique for integration is *integration by parts*. As an example, consider the following integral.

$$\int x\cos x\,dx$$

We know that

$$\frac{d}{dx}\sin x = \cos x$$

which again after some not-so-rigorous rearrangement gives us

$$d(\sin x) = \cos x\,dx.$$

This allows us to transform our original integral into

$$\int x\cos x\,dx = \int x\,d(\sin x)$$

but now we're stuck.

Fortunately, integration by parts tells us that whenever we have an integral of the form

$$\int f(x)\,d(g(x))$$

we can rewrite it as

$$\begin{aligned} & f(x)\cdot g(x) - \int g(x)\,d(f(x)) \\ &= f(x)\cdot g(x) - \int g(x)f'(x)\,dx. \end{aligned}$$

Hence, we can continue our calculation as follows.

$$\begin{aligned} \int x\,d(\sin x) &= x\cdot\sin x - \int \sin x\,d(x) \\ &= x\cdot\sin x - \int \sin x\,dx \\ &= x\sin x + \cos x + C \end{aligned}$$

Comparing this technique to our rules of differentiation, we see that integration by parts is essentially the product rule in reverse. We demonstrate this below.

$$\begin{aligned} \frac{d}{dx}(f(x)\cdot g(x)) &= f(x)\cdot g'(x) + g(x)\cdot f'(x) && \text{(product rule)} \\ \int \frac{d}{dx}(f(x)\cdot g(x))\,dx &= \int (f(x)\cdot g'(x) + g(x)\cdot f'(x))\,dx && \text{(integrate both sides)} \\ f(x)\cdot g(x) + C &= \int (f(x)\cdot g'(x) + g(x)\cdot f'(x))\,dx && \text{(integral \& derivative cancel out)} \\ f(x)\cdot g(x) + C &= \int f(x)\cdot g'(x)\,dx + \int g(x)\cdot f'(x)\,dx && \text{(integral of sum)} \\ \int f(x)\cdot g'(x)\,dx &= f(x)\cdot g(x) - \int g(x)\cdot f'(x)\,dx && \text{(rearranging to eliminate } C\text{)} \\ \int f(x)\cdot\, d(g(x)) &= f(x)\cdot g(x) - \int g(x)\cdot\, d(f(x)) && \text{(integration by parts)} \end{aligned}$$

## 7.5 What is a definite integral?

Consider a function $f$ that is continuous and defined on the interval $[a, b]$. The area under the graph of $f$ between $a$ and $b$ can be calculated using a *definite integral*, denoted by

$$\int_a^b f(x)\,dx = F(b) - F(a)$$

where $F$ is a primitive[8] of $f$. For convenience, we also write

$$F(b) - F(a) = [F(x)]_a^b.$$

Note that where the graph is under the horizontal axis, the area is measured in negative values. Therefore, we have the following.

$$\begin{aligned} \int_0^6 (4-x)\,dx &= 6 \\ \int_0^{2\pi} \sin x\,dx &= 0 \end{aligned}$$

[8] Any primitive will work here as the constant of integration will ultimately cancel out.

See figures 50 and 51.

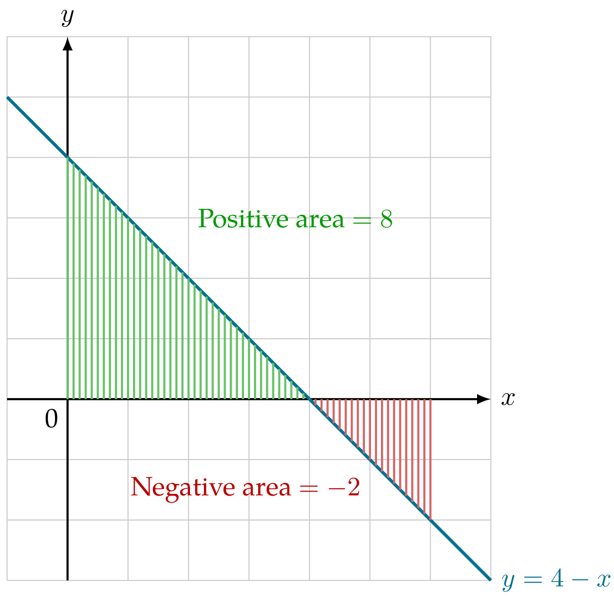



Figure 50: In the context of definite integrals, the area under a curve is always signed.

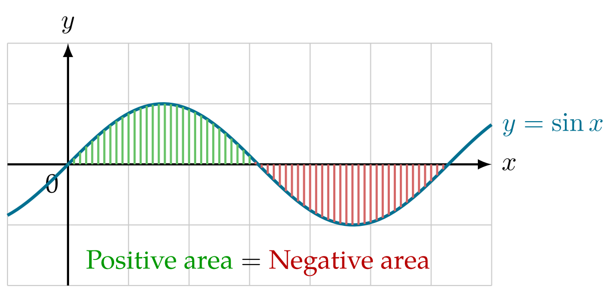



Figure 51: Positive and negative areas may cancel out.

Mathematically, we can compute the integrals above as follows.

$$\begin{aligned}\int_0^6 (4-x)\,dx &= \left[4x - \frac{1}{2}x^2\right]_0^6 \\ &= (24-18) - 0 \\ &= 6\end{aligned}$$

$$\begin{aligned}\int_0^{2\pi} \sin x\,dx &= [-\cos x]_0^{2\pi} \\ &= -1 - (-1) \\ &= 0\end{aligned}$$

Both integration by substitution and integration by parts can be applied to definite integration. For example, consider the integral below.

$$\int_0^2 \frac{x}{\sqrt{x^2+1}}\,dx$$

We proceed with integration by substitution. Let $u = x^2 + 1$, so $\frac{du}{dx} = 2x$, or $du = 2xdx$. When $x = 0$, we have $u = 1$; and when $x = 2$, we have $u = 5$. This allows us to rewrite our integral as

$$\begin{aligned}
\int_0^2 \frac{x}{\sqrt{x^2+1}}\, dx &= \int_1^5 \frac{1}{2\sqrt{u}}\, du \\
&= \frac{1}{2}\int_1^5 u^{-\frac{1}{2}}\, du \\
&= \frac{1}{2}\left[2u^{\frac{1}{2}}\right]_1^5 \\
&= \frac{1}{2}(2\sqrt{5} - 2\sqrt{1}) \\
&= \sqrt{5} - 1
\end{aligned}$$

which gives us the answer.

Now suppose we want to evaluate the integral below.

$$\int_0^e x \ln x\, dx$$

We proceed with integration by parts.

$$\begin{aligned}
\int_0^e x \ln x\, dx &= \frac{1}{2}\int_0^e 2x \ln x\, dx \\
&= \frac{1}{2}\int_0^e \ln x\, d(x^2) \\
&= \frac{1}{2}\left([x^2 \ln x]_0^e - \int_0^e x^2\, d(\ln x)\right) \\
&= \frac{1}{2}\left(e^2 - \int_0^e x^2 \cdot \frac{1}{x}\, d(x)\right) \\
&= \frac{1}{2}\left(e^2 - \int_0^e x\, dx\right) \\
&= \frac{1}{2}\left(e^2 - \left[\frac{1}{2}x^2\right]_0^e\right) \\
&= \frac{1}{2}\left(e^2 - \frac{1}{2}e^2\right) \\
&= \frac{1}{4}e^2
\end{aligned}$$

## 7.6 Numerical methods for computing definite integrals

In practice, how can we approximate the area under a given curve when the corresponding function is unknown (or is horrendous)? One way to do this is to estimate that area using rectangular bars. See figure 52.

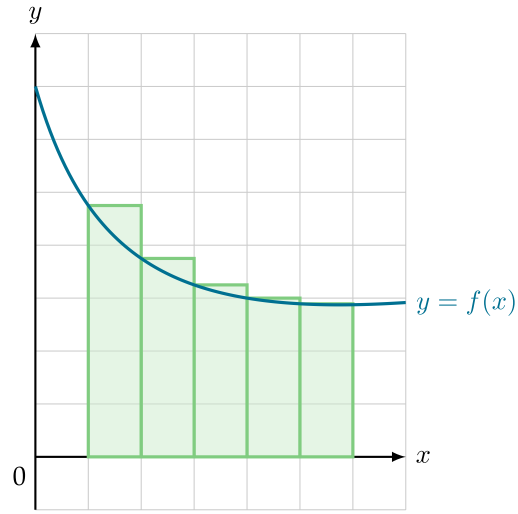



Figure 52: Using rectangles to approximate the area under the curve $y = f(x)$ between $x = 1$ and $x = 5$.

For a total of $n$ rectangles, the area under the curve $y = f(x)$ between $x = a$ and $x = b$ can be approximated as

$$\sum_{i=0}^{n-1} \underbrace{\frac{b-a}{n}}_{\text{Width}} \cdot \underbrace{f\left(a + \frac{b-a}{n} \cdot i\right)}_{\text{Height of } i\text{-th rectangle}}.$$

As $n$ approaches infinity, the subdivisions become more and more precise, and this sum gradually approaches the true area under the curve.

$$\lim_{n\to\infty} \sum_{i=0}^{n-1} \frac{b-a}{n} \cdot f\left(a + \frac{b-a}{n} \cdot i\right) = \int_a^b f(x)\, dx$$

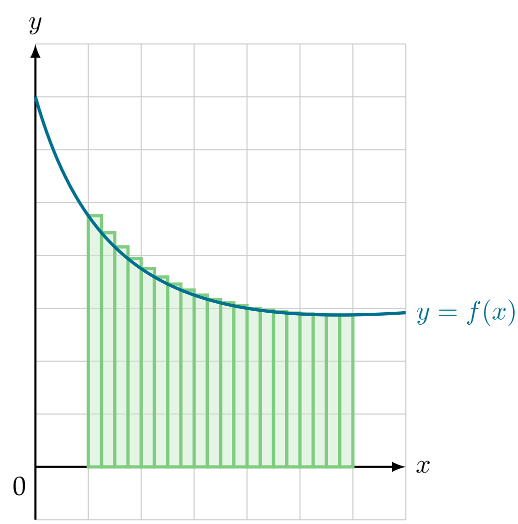



Figure 53: As $n$ approaches infinity, the subdivisions become more and more precise, and the sum of the rectangular areas gradually approaches the true area under the curve.

If we want to obtain an even more accurate result, we can use trapeziums instead of rectangles. See figure 54.

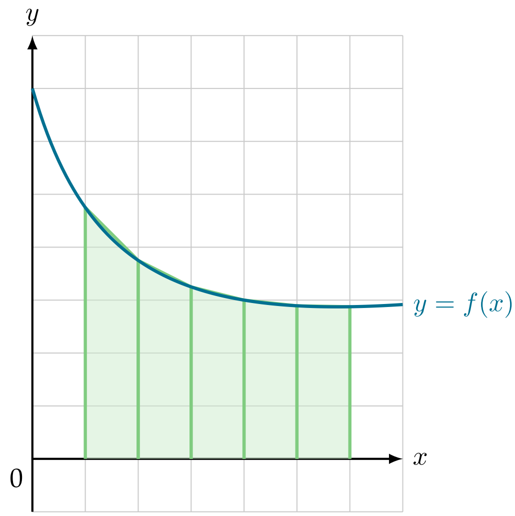



Figure 54: Using trapeziums to approximate the area under the curve.

## 7.7 What is an improper integral?

An *improper integral* is a definite integral that violates the usual assumptions we make when performing integration. An improper integral can take any one of the following forms.

$$\int_a^{\infty} f(x)\,dx \qquad \text{(see figure 55)}$$

$$\int_{-\infty}^{b} f(x)\,dx \qquad \text{(see figure 56)}$$

$$\int_{-\infty}^{\infty} f(x)\,dx \qquad \text{(see figure 57)}$$

$$\int_a^{b} f(x)\,dx \text{ where } f(x) \text{ is undefined somewhere on } [a,b] \qquad \text{(see figure 58)}$$

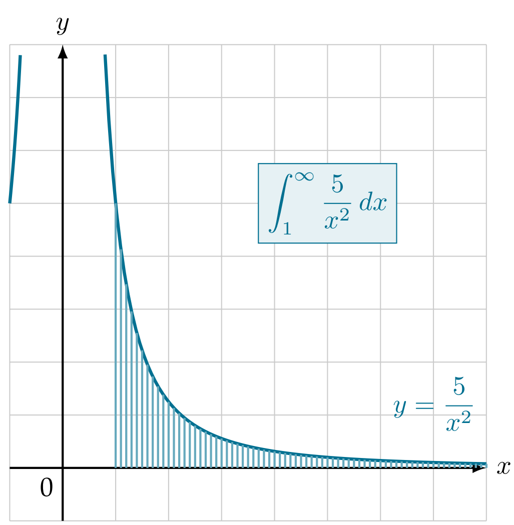



Figure 55: An improper integral evaluated on the interval $[1, \infty)$.

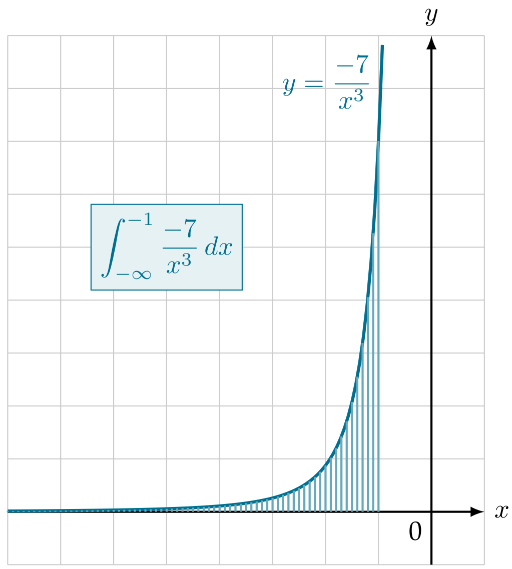



Figure 56: An improper integral evaluated on the interval $(-\infty, -1]$.

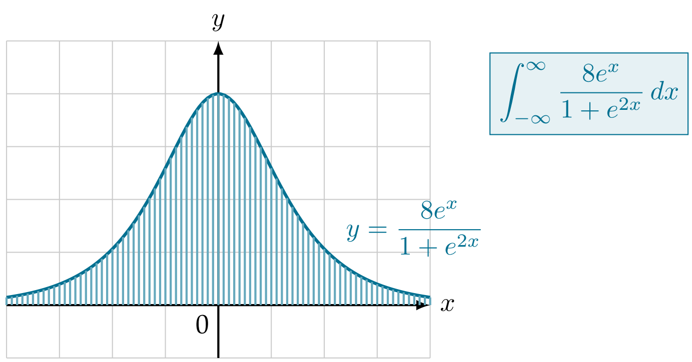



Figure 57: An improper integral evaluated on the interval $(-\infty, \infty)$.

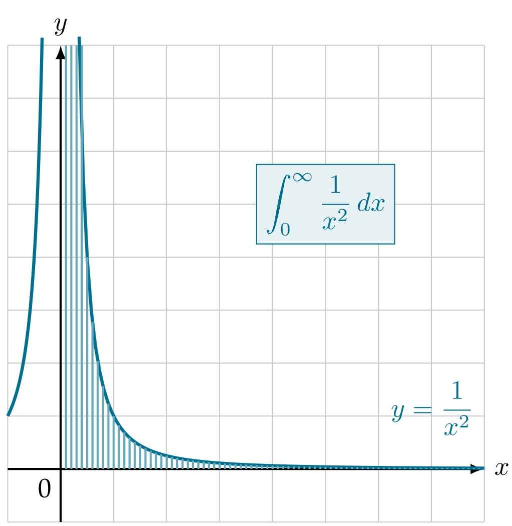



Figure 58: An improper integral involving an asymptote at $x = 0$, where the function $y = 1/x^2$ is undefined.

All of the following can be formally defined as follows.

- If $\int_a^t f(x)\,dx$ exists for all $t > a$, then we define

$$\int_a^\infty f(x)\,dx = \lim_{t\to\infty} \int_a^t f(x)\,dx$$

  assuming that this limit exists and is finite.

- Similarly, if $\int_t^b f(x)\,dx$ exists for all $t < b$, then we define

$$\int_{-\infty}^b f(x)\,dx = \lim_{t\to-\infty} \int_t^b f(x)\,dx$$

  assuming that this limit exists and is finite.

- If for some $c \in \mathbb{R}$, the improper integrals $\int_{-\infty}^c f(x)\,dx$ and $\int_c^\infty f(x)\,dx$ both exist, then we define the following.

$$\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^c f(x)\,dx + \int_c^\infty f(x)\,dx$$

- Improper integrals involving asymptotes are defined in a similar manner using limits.

Note that these integrals are well-defined only when the corresponding limits exist. For instance, the integral

$$\int_0^\infty 2\,dx$$

is undefined. See figure 59.

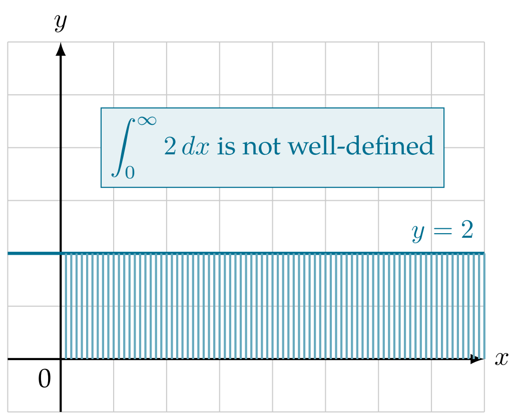



Figure 59: An improper integral involving an asymptote at $x = 0$, where the function $y = 1/x^2$ is undefined.

Sometimes, it might be difficult to directly tell whether an improper integral is well-defined. Take for example the following integral.

$$\int_1^\infty \frac{1}{x}\,dx$$

To determine whether this is a valid integral, we calculate

$$\begin{aligned}
\int_1^\infty \frac{1}{x}\,dx &= \lim_{t\to\infty} \int_1^t \frac{1}{x}\,dx \\
&= \lim_{t\to\infty} [\ln x]_1^t \\
&= \lim_{t\to\infty} (\ln t - \ln 1) \\
&= \lim_{t\to\infty} \ln t
\end{aligned}$$

which does not exist. Therefore, this integral is undefined.

On the contrary, consider the following integral.

$$\int_1^\infty \frac{1}{x^2}\,dx$$

This time, we calculate

$$\begin{aligned}
\int_1^\infty \frac{1}{x^2}\,dx &= \lim_{t\to\infty} \int_1^t \frac{1}{x^2}\,dx \\
&= \lim_{t\to\infty} [-x^{-1}]_1^t \\
&= \lim_{t\to\infty} \left(-\frac{1}{t} - (-1)\right) \\
&= \lim_{t\to\infty} \left(1 - \frac{1}{t}\right) \\
&= 1
\end{aligned}$$

which means the integral is defined and equals to $1$.

In general we have the following theorem.

**Theorem.** Let $a > 0$. The integral

$$\int_a^\infty \frac{1}{x^p}\,dx$$

is defined if $p > 1$ and undefined if $p \leq 1$. On the other hand, the integral

$$\int_0^a \frac{1}{x^p}\,dx$$

is defined if $p < 1$ and undefined if $p \geq 1$.

This is summarised in table 3 and visualised in figure 60. As we approach infinity, the functions with higher exponents become "flat" enough to be integrable. The opposite occurs near zero, where lower exponents give flatter and integrable curves.

| $p$ | $< 1$ | $= 1$ | $> 1$ |
|---|---|---|---|
| $\int_a^\infty \frac{1}{x^p}\,dx$ defined? | ✗ | ✗ | ✓ |
| $\int_0^a \frac{1}{x^p}\,dx$ defined? | ✓ | ✗ | ✗ |

Table 3: A table showing when improper integrals of the function $1/x^p$ are defined. Assume $a > 0$.

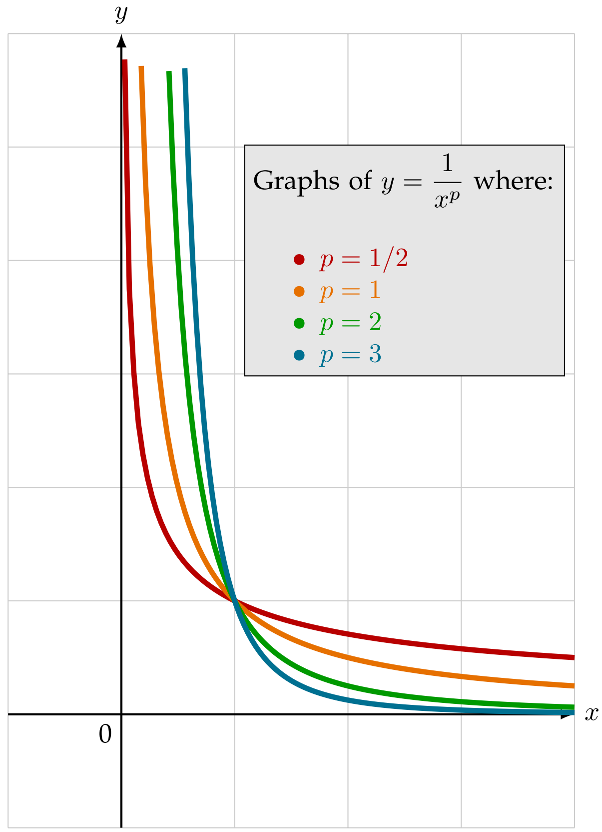



Figure 60: Graphs of $y = 1/x^p$ with varying values of $p$.

# 8 Polynomials

Given a field $\mathbb{K}$ and a variable $x$, a *polynomial* $P$ of $\mathbb{K}[x]$ is defined as a linear combination of powers of $x$.

$$P = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_2 x^2 + a_1 x + a_0 \qquad \text{(assuming } a_n \neq 0\text{)}$$

Note that

- The numbers $a_i$ should all belong to the field $\mathbb{K}$. They are called *coefficients*.
- The products $a_i x^i$ are called *terms*.
- Here, $n$ is the highest power in $P$. This is known as the *degree* of the polynomial.

Examples of polynomials include

$$x^2 \qquad \text{(degree } 2\text{)}$$
$$y - 1.5y^3 + 2 \qquad \text{(degree } 3\text{)}$$
$$z + 3z^4 \qquad \text{(degree } 4\text{)}$$

Although any symbol can be used to denote the variable of a polynomial, we will most use the letter $x$ in this section.

A polynomial of degree $0$ is simply a constant. On the other hand, a polynomial of degree $1$, such as $2x + 1$, is said to be *linear*.

## 8.1 Polynomials are not functions

One important distinction to make is that technically speaking, polynomials are not functions:

- A *polynomial* is a strictly algebraic object. It is sometimes represented as a vector in a vector space. (We will further explore this idea below.)
- A *function* is a mapping — a rule that maps elements from a set to elements of another set.
- A *polynomial function* is a specific type of function. While there is a one-to-one correspondence between polynomials and polynomial functions, they do not refer to the same idea.

Despite this, polynomials can be evaluated by substituting the variable with a value. For example, given the polynomial $P = x^2 - 5x$, we can substitute $x = 7$ to get

$$P(7) = 7^2 - 5 \times 7 = 14.$$

## 8.2 Addition, multiplication and division of polynomials

Polynomials can be added by summing up their like terms.

$$\begin{aligned}(7x^3 + 8x^2 + 3) + (2x^2 + 9x - 4) &= 7x^3 + (8x^2 + 2x^2) + 9x + (3 - 4) \\ &= 7x^3 + 10x^2 + 9x - 1\end{aligned}$$

Polynomials can also be multiplied by an element of $\mathbb{K}$.

$$2(7x^3 + 8x^2 + 3) = 14x^3 + 16x^2 + 6$$

Let $\mathbb{K}_n[x]$ be the set of all polynomials with coefficients in $K$ and of degree at most $n$. Since this set is closed under addition and scaling, it is a vector space.

We can multiply two polynomials by using expansion via distributivity.

$$\begin{aligned}(8x+3)(9x-4) &= (8x)(9x)+(8x)(-4)+(3)(9x)+(3)(-4)\\ &= 72x^2-32x+27x-12\\ &= 72x^2-5x-12\end{aligned}$$

If we denote the degree of a polynomial $P$ as $\deg(P)$, then:

$$\begin{aligned}&\deg(P+Q) \le \max(\deg(P)+\deg(Q)) && \text{(highest-degree terms may cancel out)}\\ &\deg(P\times Q) \le \deg(P)+\deg(Q)\end{aligned}$$

The fact that we can multiply polynomials implies that polynomials can have *factors* — for instance, we say that the polynomial $72x^2-5x-12$ has factors $8x+3$ and $9x-4$.

A polynomial with no non-constant factors is said to be *irreducable*[9]. For example, $7x+4$ is irreducable, but the following polynomials are not.

$$\begin{aligned}x^2+7x+12 &= (x+4)(x+3)\\ x^3+x^2+2x+2 &= (x^2+2)(x+1)\end{aligned}$$

A polynomial is said to *split* if it has only linear factors. Therefore, of the two polynomials listed above, the first one splits but the second one doesn't.

Lastly, we can perform division on polynomials, as explained below.

**Euclidean division of polynomials.**

Given two polynomials $A$ and $B$, there exists polynomials $Q$ and $R$ such that

$$A = QB + R$$

where $R$ has a lower degree than $B$. Here,

- $A$ is the dividend and $B$ is the divisor.
- $Q$ is the quotient and $R$ is the remainder.

Given a dividend and a divisor, we can use long division to identify the corresponding quotient and remainder. An example of this is given in 61, with the division

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

Note how the remainder ($R=5$) has a lower degree than the divisor $B=x-3$.

$$\begin{array}{r|l} & \quad x^2 + \quad x + 3 \\ \cline{2-2} x-3 & x^3 - 2x^2 + 0x - 4 \\ & \underline{x^3 - 3x^2} \qquad\qquad \\ & \qquad +x^2 + 0x \quad \\ & \qquad \underline{+x^2 - 3x} \quad \\ & \qquad\qquad +3x - 4 \\ & \qquad\qquad \underline{+3x - 9} \\ & \qquad\qquad\qquad +5 \end{array}$$

Figure 61: An example of performing long division on polynomials.

[9]This is analogous to how prime numbers work.

## 8.3 Roots of polynomials

A number $a$ in $\mathbb{K}$ is said to be a root of a polynomial $P$ if $P(a) = 0$.

For example,

- The linear polynomial $P = 5x + 2$ has the root $x = -2/5$ because $P(-2/5) = 5(-2/5) + 2 = 0$.
- The polynomial $Q = x^2 + 3x + 2$ has a root $x = -2$ because $Q(-2) = (-2)^2 + 3(-2) + 2 = 0$.

We now introduce the *factor theorem*, which is illustrated below.

**Factor theorem.** A number $a$ is a root of a polynomial $P$ if and only if $(x - a)$ is a factor of $P$.

**Proof.** We prove this statement in two directions.

$(\Leftarrow)$:

$$\begin{aligned} (x-a) \text{ is a factor of } P &\implies P = (x-a)Q \qquad \text{for some polynomial } Q \\ &\implies P(a) = (a-a)Q \\ &\implies P(a) = 0 \\ &\implies a \text{ is a root of } P \end{aligned}$$

$(\Rightarrow)$: Assume $a$ is a root of $P$, so $P(a) = 0$.

We divide $P$ by $x - a$. By Euclidean division, there exists a polynomial $Q$ and a constant $r$ such that

$$P = Q \cdot (x - a) + r.$$

(Recall that the remainder must have a lower degree than the divisor $x - a$. Since the divisor $(x - a)$ has degree $1$, the remainder must be a constant with degree $0$.)

We evaluate both sides of the equation with $x = a$.

$$\begin{aligned} P(a) &= Q(a) \cdot (a - a) + r \\ P(a) &= r \\ r &= 0 \end{aligned}$$

Hence $P = Q \cdot (x - a)$, so $x - a$ is a factor of $P$.

This theorem is extremely useful for finding the roots of a polynomial. For instance, we can factor the polynomial

$$\begin{aligned} 2x^3 - x^2 - 8x + 4 &= x^2(2x-1) - 4(2x-1) \\ &= (x^2-4)(2x-1) \\ &= (x+2)(x-2)(2x-1) \\ &= 2(x+2)(x-2)\left(x - \frac{1}{2}\right) \end{aligned}$$

to show that it has the roots $-2$, $2$ and $1/2$. But are these the only roots?

Yes, they are. We can prove this using the following theorem.

**Theorem.** The number of roots[10] of a polynomial in $\mathbb{K}$ cannot exceed its degree.

[10] In this case, identical roots are counted separately. For example, the polynomial $x^2 - 2x + 1 = (x - 1)^2$ is treated as having two roots, both of value 2. We will dive deeper into this technicality later in this section when we talk about the multiplicity of a root.

**Proof.** We label the roots of a polynomial $P$ as $r_1,\ r_2,\ r_3,\ \cdots,\ r_n$. Hence, by the factor theorem, we have

$$\begin{aligned} P &= (x-r_1)(x-r_2)(x-r_3)\cdots(x-r_n)Q && \text{(for some polynomial } Q\text{)} \\ \deg(P) &= \deg((x-r_1)(x-r_2)(x-r_3)\cdots(x-r_n)Q) \\ &= n + \deg(Q) \\ &> n \end{aligned}$$

Therefore $n < \deg(P)$.

Note that the theorem above implies that the number of roots of a polynomial in $\mathbb{K}$ may not necessarily equal its degree. To see why this is, we will have to take a closer look at polynomials of degree 2.

## 8.4 On the real roots of polynomials of degree 2

The simplest polynomial of degree 2 takes the form $P = x^2 - a$. Finding its roots is equivalent to solving the equation

$$x^2 = a$$

which has the solutions $\sqrt{a}$ and $-\sqrt{a}$.

In general, however, a polynomial of degree 2 takes the form $ax^2 + bx + c$. The corresponding function $f(x) = ax^2 + bx + c$ is quadratic — its graph is a parabola. The roots of the polynomial correspond to the points where the graph crosses the $x$-axis. See figures 62, 63 and 64.

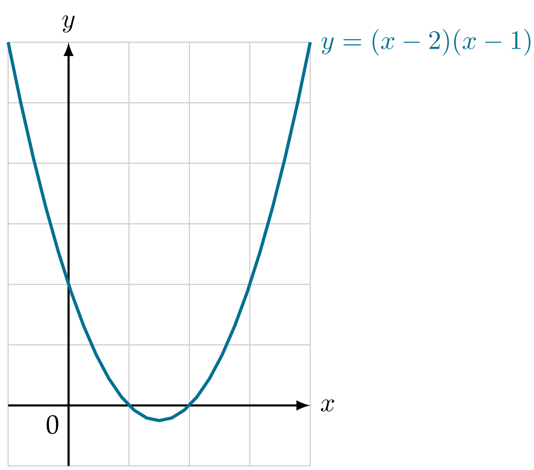



Figure 62: An example of a polynomial of degree 2 with 2 distinct roots. The corresponding function has a graph has two $x$-intercepts.

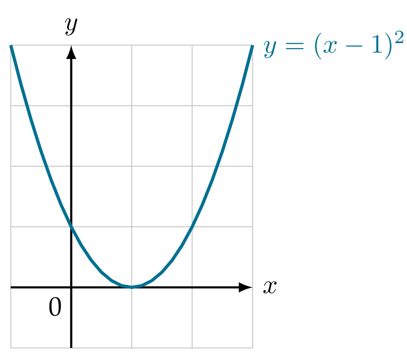



Figure 63: An example of a polynomial of degree 2 with 1 root. The corresponding function has a graph has one $x$-intercept.

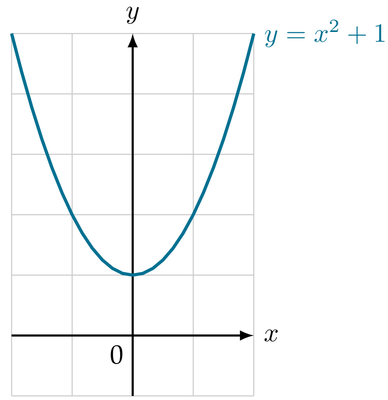



Figure 64: An example of a polynomial of degree 2 with no roots in $\mathbb{R}$. The corresponding function has a graph has no $x$-intercepts.

Usually, we want to work out the number of roots of a polynomial of degree 2 without plotting its graph. This can be done by analysing its *discriminant*. For a polynomial $P = ax^2 + bx + c$, its determinant is defined as $\Delta = b^2 - 4ac$.

- If $\Delta > 0$, then $P$ has two distinct roots in $\mathbb{R}$ and can be factored into the form $P = a(x-r_1)(x-r_2)$.
- If $\Delta = 0$, then $P$ has one root[11] in $\mathbb{R}$ and can be factored into the form $P = a(x-r)^2$.
- If $\Delta < 0$, then $P$ has no roots in $\mathbb{R}$ and is irreducable.

In the first two cases, the root(s) of $P$ are given by the quadratic formula, as shown below.

$$x = \frac{-b \pm \sqrt{\Delta}}{2a} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

## 8.5 On the complex roots of polynomials of degree 2

If we allow complex roots, then a polynomial of degree 2 always have two (not necessarily distinct) roots $r_1$ and $r_2$ in $\mathbb{C}$, as given by the quadratic formula. In other words, every polynomial of degree 2 splits in $\mathbb{C}$ and can be factored into the form $a(x-r_1)(x-r_2)$.

The two roots of a degree 2 polynomial

$$\begin{aligned} r_1 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \\ r_2 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b - \sqrt{b^2 - 4ac}}{2a} \end{aligned}$$

are complex conjugates. This is proved below.

**Theorem.** The two roots of a degree 2 polynomial must be complex conjugates.

**Proof.** If $\Delta \geq 0$, then both roots are real and therefore must be conjugates in $\mathbb{C}$.

If $\Delta < 0$, then $\sqrt{\Delta}$ is purely imaginary and can be expressed as $di$ for some $d \in \mathbb{R}$. Hence,

$$\begin{aligned} r_1 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b + di}{2a} = \frac{-b}{2a} + \frac{d}{2a}i \\ r_2 &= \frac{-b + \sqrt{\Delta}}{2a} = \frac{-b - di}{2a} = \frac{-b}{2a} - \frac{d}{2a}i \end{aligned}$$

[11] This is technically two non-distinct roots.

are complex conjugates.

For example, the polynomial $x^2+1$ has no real roots but can be factored in $\mathbb{C}$ as $(x-i)(x+i)$. Its roots, $i$ and $-i$, are complex conjugates.

Below shows three useful identities for factoring polynomials of degree 2.

$$\begin{aligned}(a+b)^2 &= a^2+b^2+2ab\\(a-b)^2 &= a^2+b^2-2ab\\(a+b)(a-b) &= a^2-b^2\end{aligned}$$

## 8.6 On the roots of polynomials of arbitrary degree

We introduce the following theorems for polynomials of abritrary degree.

- **Theorem on factorisations in $\mathbb{R}$.**

  In $\mathbb{R}$, only polynomials of degree 2 are irreducable[12]. Therefore, every polynomial $P \in \mathbb{R}[x]$ of degree $n > 0$ has a unique factorisation in $\mathbb{R}$ of the form

  $$P = c\underbrace{(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)}_{\text{linear factors}}\underbrace{(x^2+a_1x+b_1)(x^2+a_2x+b_2)\cdots(x^2+a_kx+b_k)}_{\text{quadratic factors}}$$

  where

  – the constants $c,\ \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_m,\ a_1,\ a_2,\ \cdots a_k,\ b_1,\ b_2,\ \cdots b_k$ are real numbers.

  – Each quadratic factor is irreducable with a negative determinant, i.e. $\Delta_i = a_i^2 - 4b_i < 0$ for $1 \leq i \leq k$.

- **Theorem on factorisations in $\mathbb{C}$.**

  Each polynomial $P \in \mathbb{R}[x]$ of degree $n > 0$ has a unique factorisation in $\mathbb{C}$ of the form

  $$P = c(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$$

  where $c, \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n \in \mathbb{C}$.

  In other words, every real polynomial splits in $\mathbb{C}$ with $n$ (not necessarily distinct) complex roots $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$.

When roots are not distinct, a polynomial can be written as

$$P = c(x-\lambda_1)^{k_1}(x-\lambda_2)^{k_2}(x-\lambda_3)^{k_3}\cdots(x-\lambda_j)^{k_j}$$

where $k_1,\ k_2,\ k_3,\ \cdots,\ k_j \geq 1$. We say that $k_i$ is the *multiplicity* of $\lambda_i$.

## 8.7 A theorem on real polynomials of odd degrees

Finally, we introduce an interesting theorem regarding real polynomials of odd degrees.

**Theorem.**

Any polynomial $P$ of odd degree and with real coefficients has at least one real root.

To prove this, we can either use algebra or calculus.

[12] Note that it is still possible for polynomials of degree greater than 2 to have no real roots. This occurs when they are made entirely of irreducable quadratic factors. For example, the polynomial $x^4+1$ can be factored into $x^4+2x^2+1-2x^2 = (x^2+1)^2-(\sqrt{2}x)^2 = (x^2+2\sqrt{x}+1)(x^2-2\sqrt{x}+1)$.

- **Proof 1: An algebraic proof.**

  We first prove the following lemma: If $z$ is a complex and non-real root of $P$, then so is its conjugate $\bar{z}$.

  We write $P$ as follows:

  $$P = c\underbrace{(x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_m)}_{\text{linear factors}}\underbrace{(x^2+a_1x+b_1)(x^2+a_2x+b_2)\cdots(x^2+a_kx+b_k)}_{\text{quadratic factors}}$$

  where the constants $c,\ \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_m,\ a_1,\ a_2,\ \cdots a_k,\ b_1,\ b_2,\ \cdots b_k$ are real numbers; and each quadratic factor is irreducable with a negative determinant.

  We assume $z$ is a root of $P$, which means that $(x-z)$ is a factor. Since $z \notin \mathbb{R}$, we have $z \neq \lambda_1,\ \lambda_2,\ \cdots,\ \lambda_m$. Hence, $(x-z)$ must be a factor of one of the quadratic factors, i.e. $(x^2+a_jx+b_j)$ where $1 \leq j \leq k$.

  We know that the two roots of a quadratic polynomial are always complex conjugates (This was proved earlier using the quadratic formula.) Therefore, $(x-\bar{z})$ is also a factor of $(x^2+a_jx+b_j)$.

  This means that $(x-\bar{z})$ is a factor of $P$, so $\bar{z}$ is a root of $P$. This concludes the proof for the lemma.

  We now prove the required theorem. Since $P$ has an odd degree, it must also have an odd number of roots. By the previously proved lemma, for any complex and non-real root $z$ of $P$, its conjugate $\bar{z} \neq z$ is also a root. Hence, the number of non-real roots of $P$ must be even. This means that the number of real roots of $P$ must be odd, i.e. at least one. Hence proved.

- **Proof 2: A calculus proof.**

  We write $P$ as follows:

  $$P = a_nx^n + a_{n-1}x^{n-1} + \cdots + a_1x + a_0 \qquad (a_n \neq 0)$$

  where $n$ is odd. WLOG assume that $a_n > 0$. This means that as $x$ approaches positive infinity, so does $P$. Hence

  $$\forall d > 0, \exists c > 0, x > c \implies P(x) > d.$$

  Substituting $d = 1$ (or any positive value) tells us that there exists some positive constant $c$ such that $x > c \implies P(x) > 1$. This means that there exists some $x_{\text{pos}}$ for which $P(x_{\text{pos}})$ is positive.

  Similarly, notice that as $x$ approaches negative infinity, so does $P$ (since $n$ is odd). This means that

  $$\forall d < 0, \exists c < 0, x > c \implies P(x) < d.$$

  Substituting $d = -1$ (or any negative value) tells us that there exists some negative constant $c$ such that $x < c \implies P(x) < -1$. This means that there exists some $x_{\text{neg}}$ for which $P(x_{\text{neg}})$ is negative.

  Combining the two results above with the intermediate value theorem, we see that there must exist some value $x_0$ such that $x_{\text{neg}} < x_0 < x_{\text{pos}}$ and $P(x_0) = 0$. This $x_0$ is a real root of $P$. Hence proved.