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.

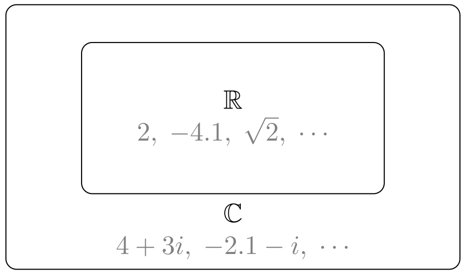


Figure 1: The set of real numbers $\mathbb{R}$ is a subset of the set of complex numbers $\mathbb{C}$. All real numbers are complex numbers.

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.

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

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.

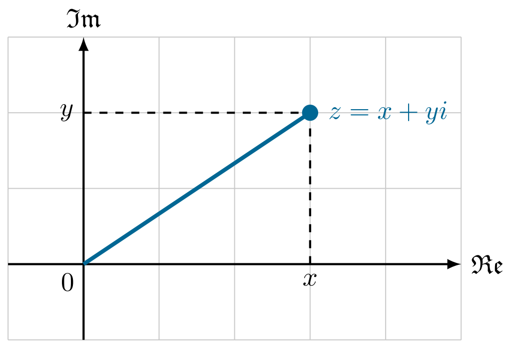


Figure 2: The complex number $z = x + yi$ as a point on the 2D plane

1.3 Exponential form

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

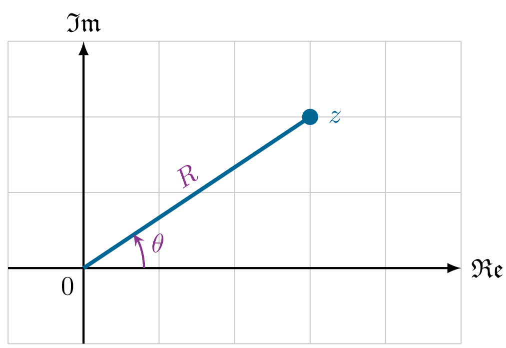


Figure 3: The position of a complex number on the 2D plane can be represented using polar coordinates.

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, θ), 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.

θ 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.

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

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

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

Multiplication

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

Division

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

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.

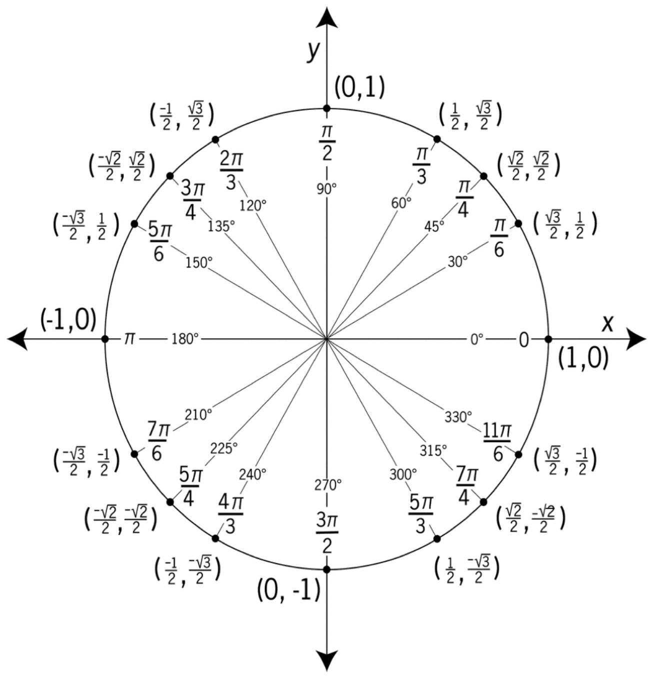


Figure 4: It is important to know the coordinates of points on the circle corresponding to classic angles.

θ (radians)	$\pi/6$	$\pi/4$	$\pi/3$
θ (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 θ.

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.

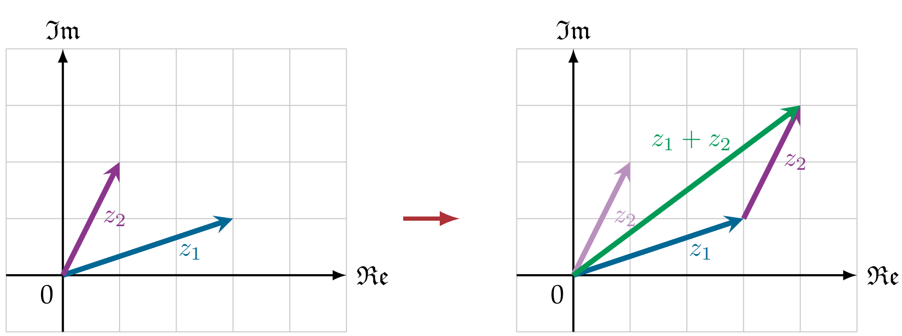


Figure 5: Addition of complex numbers.

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.

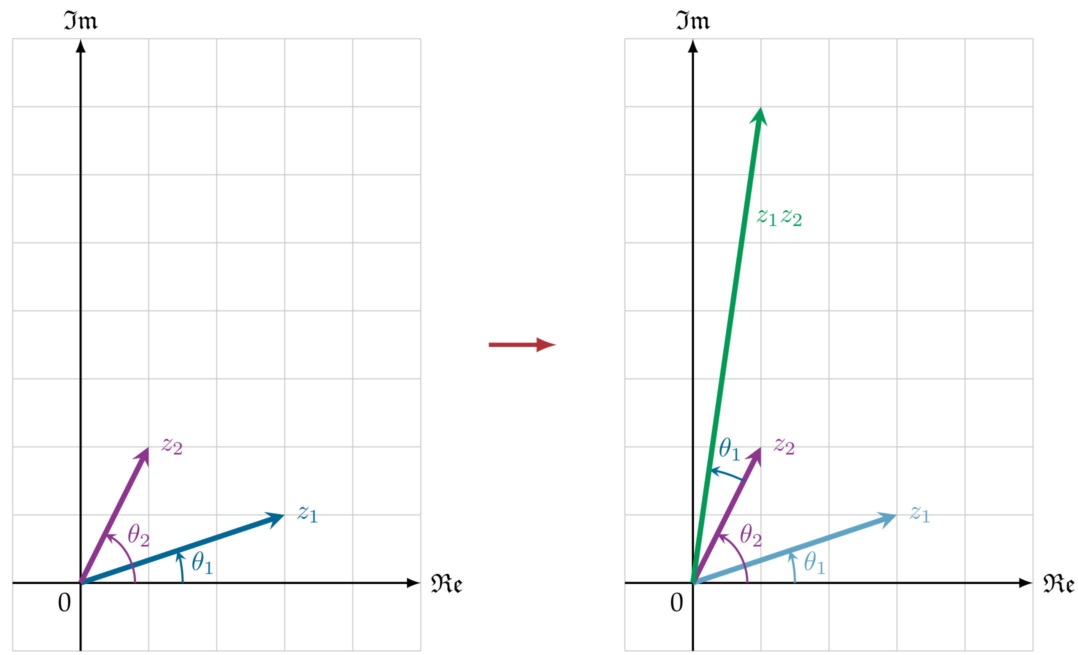


Figure 6: Multiplication of complex numbers.

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.

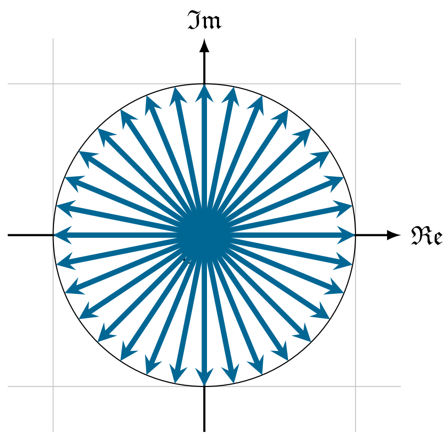


Figure 7: The roots of unity must lie somewhere on the unit circle.

We want to find the angles θ such that if we start at the point 1 and then rotate anticlockwise by θ 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 2π 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 4π 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 2π 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.

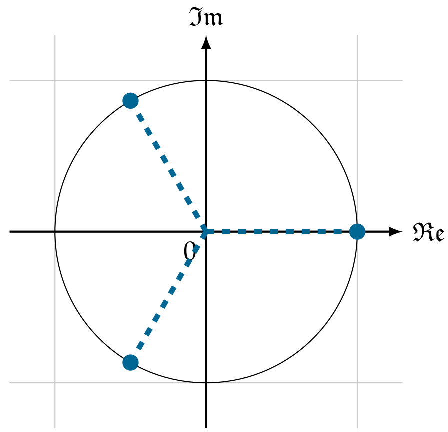


Figure 8: The third roots of unity.

This resolves the above paradox — we were right in thinking that the possible values of θ 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.

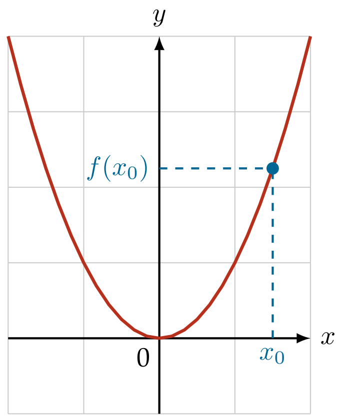


Figure 9: The graph of the function $y = x^2$.

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 θ 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.

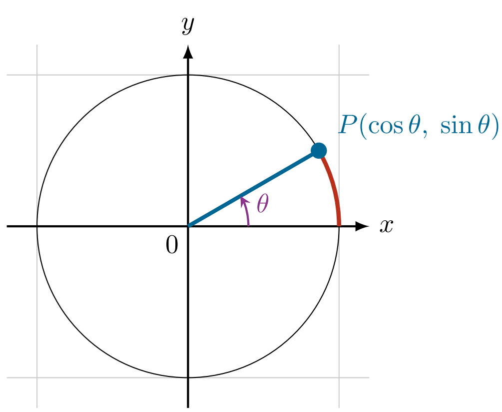


Figure 10: The trigonometric functions $\cos\theta$ and $\sin\theta$ can be defined using the unit circle. Note that if we are measuring θ in radians, then the length of the arc highlighted in red must be equal to θ.

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

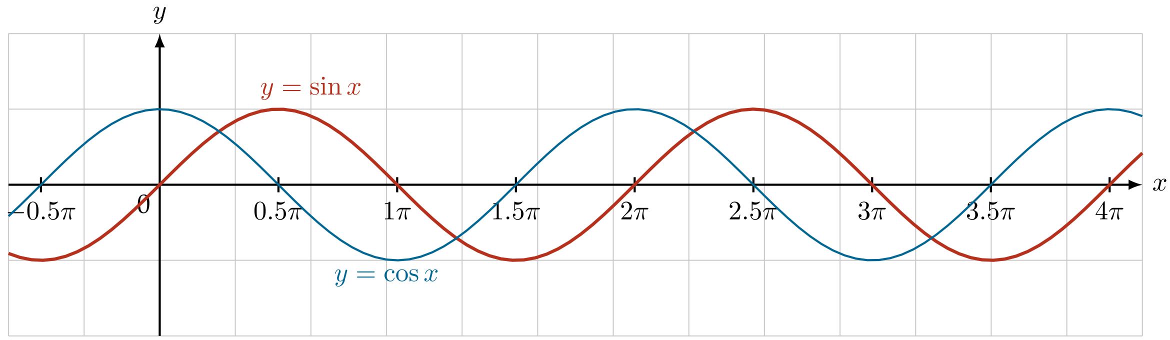


Figure 11: The graph of the functions $\sin x$ and $\cos x$.

2.2 Exponential and logarithm

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

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

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.

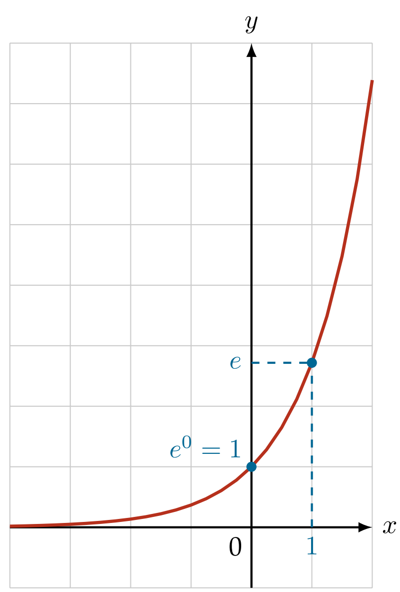


Figure 12: The graph of the function $y = \exp(x) = e^x$.

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.

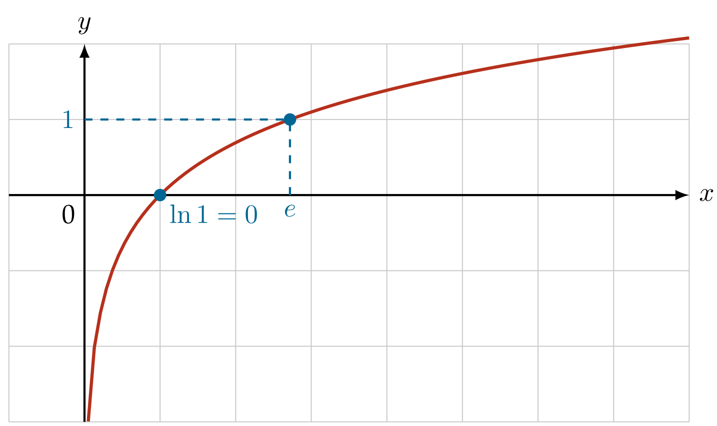


Figure 13: The graph of the function $y = \exp(x) = e^x$.

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 (ϵ, δ) definition of a limit. See figure 14 for an illustration of this.

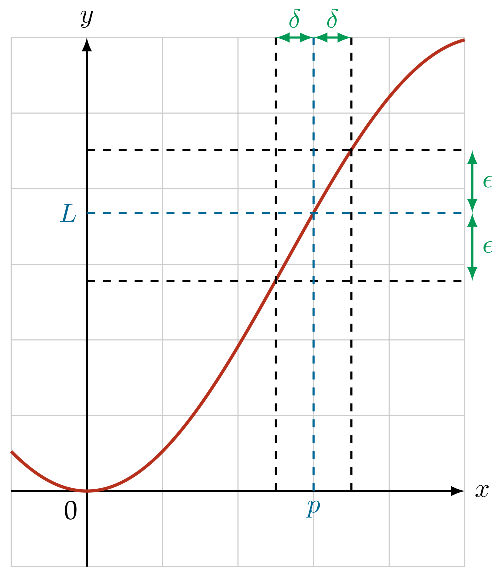


Figure 14: As x approaches p, $f(x)$ approaches L.

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 δ 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.

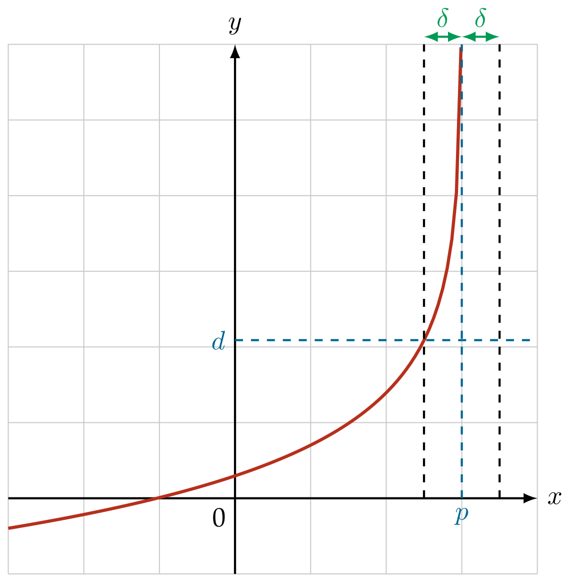


Figure 15: As x approaches p, $f(x)$ approaches infinity.

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.

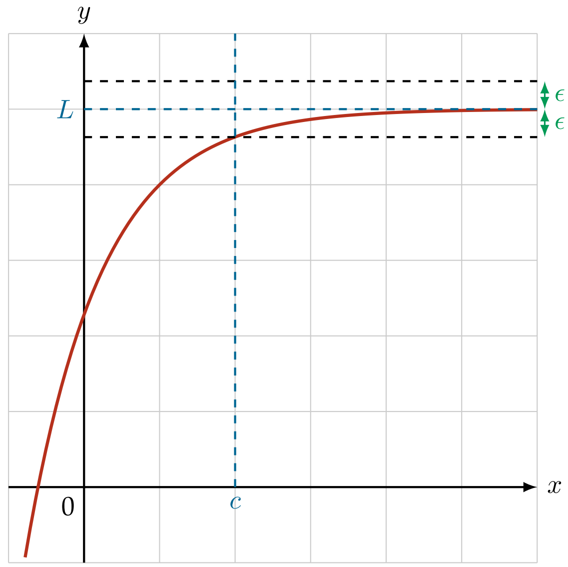


Figure 16: As x approaches infinity, $f(x)$ approaches L.

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.

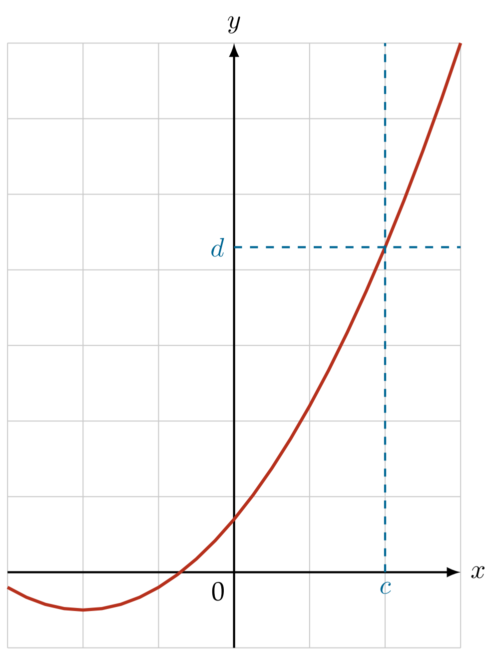


Figure 17: As x approaches infinity, so does $f(x)$.

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.

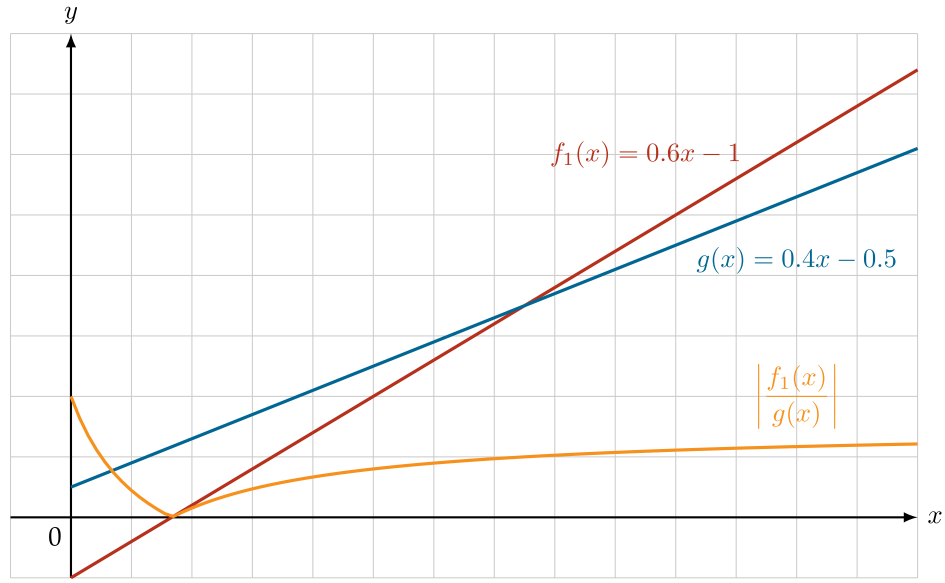


Figure 18: Graphs of a linear function $f_1(x)$, another linear function $g(x)$ and the absolute value of their quotient $|f(x)/g(x)|$.

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.

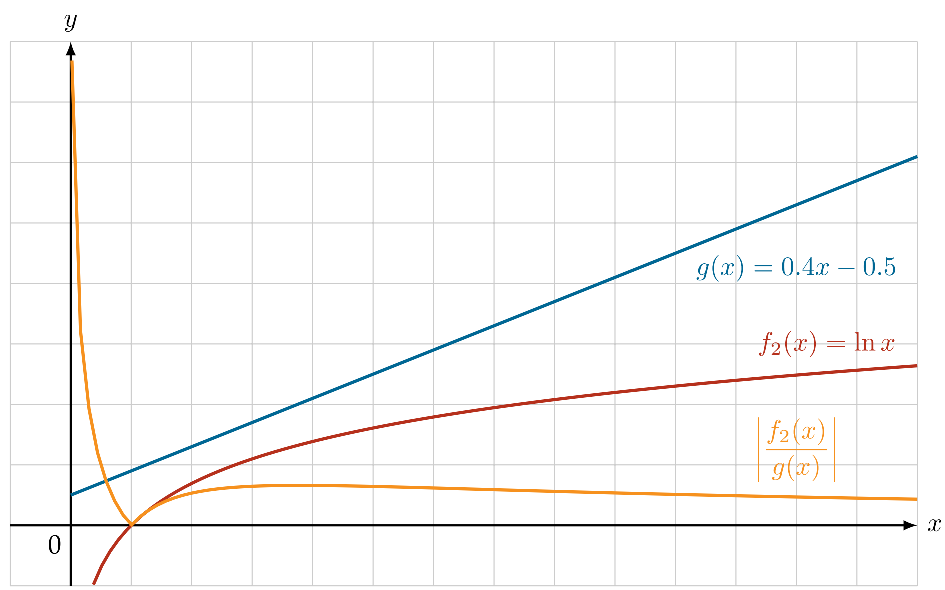


Figure 19: Graphs of a logarithmic function $f_2(x)$, a linear function $g(x)$ and the absolute value of their quotient $|f(x)/g(x)|$.

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.

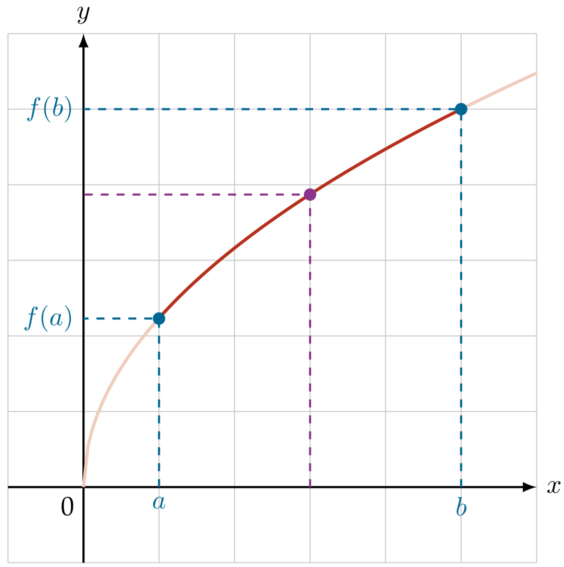


Figure 20: The intermediate value theorem.

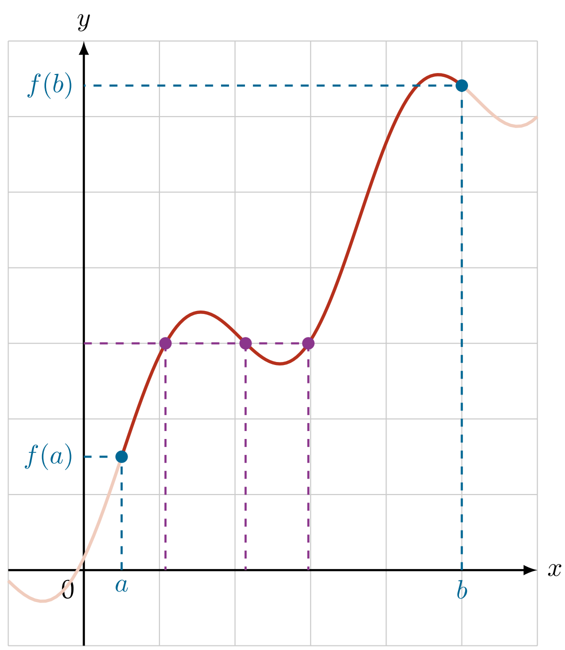


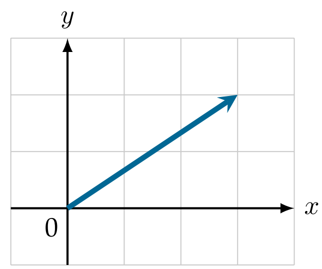
Figure 21: In the intermediate value theorem, for a given value y, the value of x does not necessarily have to be unique.

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.

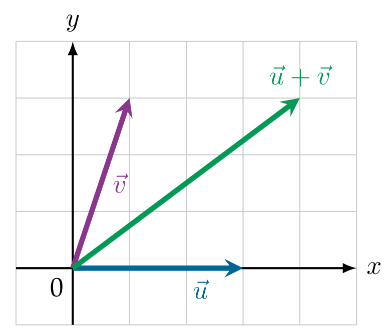


Figure 22: Vectors in $\mathbb{R}^2$ are closed under addition. For any given pair of vectors $\vec{u}$ and $\vec{v}$ both in V, their sum $\vec{u} + \vec{v}$ must also be a vector.

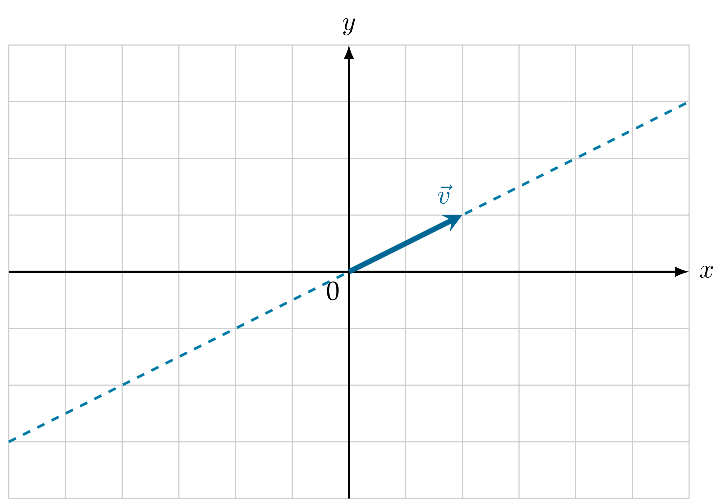


Figure 23: Vectors in $\mathbb{R}^2$ are closed under scalar multiplication. For any given vector $v \in \mathbb{R}^2$, the product between $\vec{v}$ and any scalar a (i.e. any vector lying on the dashed line) is also a vector.

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.

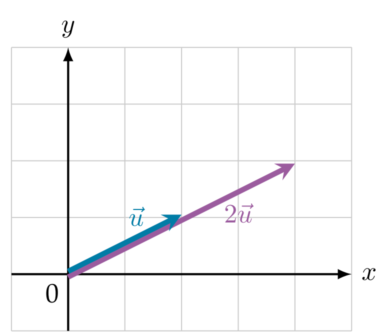


Figure 24: Two collinear vectors.

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.

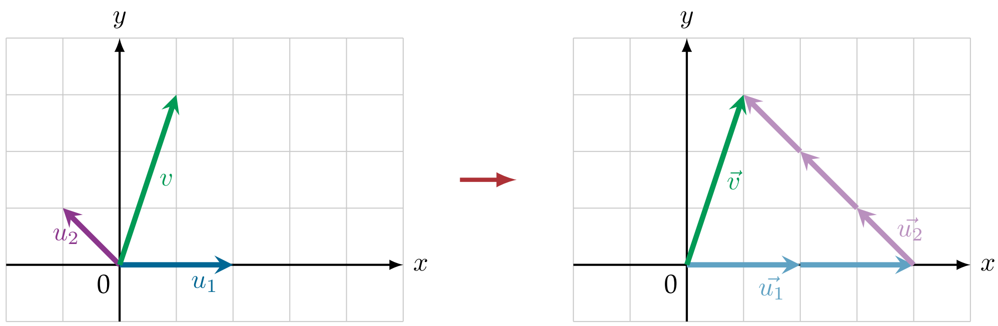


Figure 25: The green vector v can be expressed as a linear combination of the vectors u_1 and u_2, i.e. $v = 2u_1 + 3u_2$.

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.

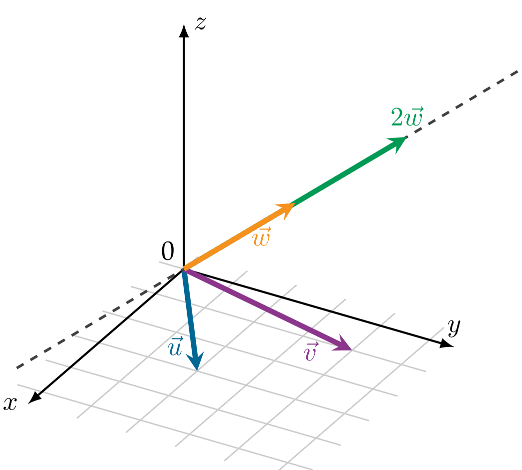


Figure 26: Three vectors in $\mathbb{R}^3$. Both $\vec{u}$ and $\vec{v}$ lie flat on the xy-plane. The dashed line represents the span of $\vec{w}$ and $2\vec{w}$.

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

[1] Plural: bases.

that the xy-plane in figure 26 has a dimension of 2.

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

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

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

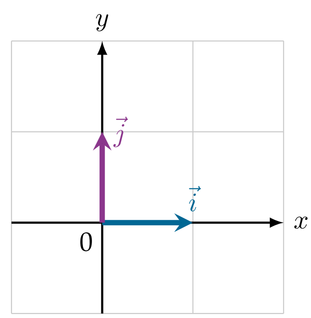


Figure 27: The canonical basis $\{\vec{i}, \vec{j}\}$ of $\mathbb{R}^2$.

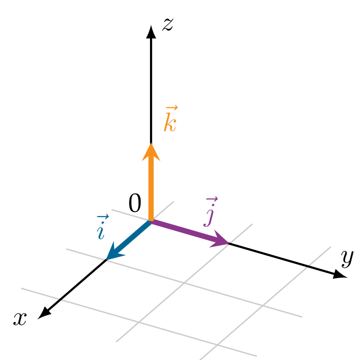


Figure 28: The canonical basis $\{\vec{i}, \vec{j}, \vec{k}\}$ of $\mathbb{R}^3$.

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