

Introductory Mathematics for Computer Science

(COMP0011)

Raphael Li

Year 1 Term 2, 2024–25

Contents

1 Continuous functions	2
1.1 Trigonometric functions	2
1.2 Exponential and logarithm	3
1.3 Introduction to limits	4
1.4 Handling infinities and indeterminate forms	8
1.5 Little o and big O notation	8
1.6 Continuity	10
2 Vector spaces	12
2.1 What is a vector space?	12
2.2 Examples of vector spaces	13
2.3 Collinearity, span and subspaces	14
2.4 Linear independence	15
2.5 Basis	16
2.6 Linear maps	17
2.7 Kernel and image	19
3 Matrices	20
3.1 Expressing linear maps as matrices	20
3.1.1 Example 1: An endomorphism with unchanged bases	20
3.1.2 Example 2: An endomorphism with changed bases	22
3.1.3 Example 3: Non-endomorphism	23
3.1.4 Kernels and images	24
3.2 The elements of a matrix	25
3.3 Matrix arithmetic	25
3.4 The inverse of a matrix	26
3.5 How to invert a matrix	26
3.6 Determinants	28
4 Sequences and series	30
4.1 What is a sequence?	30
4.2 Arithmetic, geometric, monotone, bounded and recursive sequences	30
4.3 Convergence of a sequence	32
4.4 Asymptotic behaviour of sequences	33
4.5 Important theorems about sequences and their limits	33
4.6 What is a series?	34
4.7 Convergence of a series	35

1 Continuous functions

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

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 1.

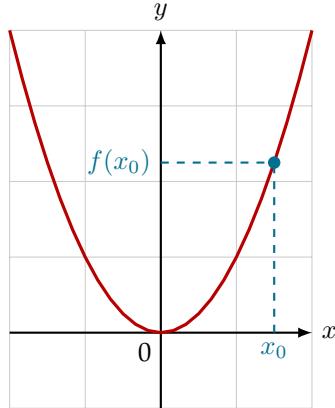


Figure 1: The graph of the function $y = x^2$.

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

1.1 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 2.

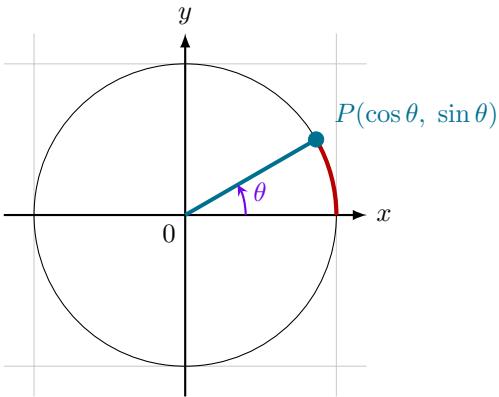


Figure 2: The trigonometric functions $\cos \theta$ and $\sin \theta$ can be defined using the unit circle. Note that if we are measuring θ in radians, then the length of the arc highlighted in red must be equal to θ .

We've previously seen the values of $\sin \theta$ and $\cos \theta$ for some classic angles θ in table ???. Plotting these functions on a graph results in figure 3.

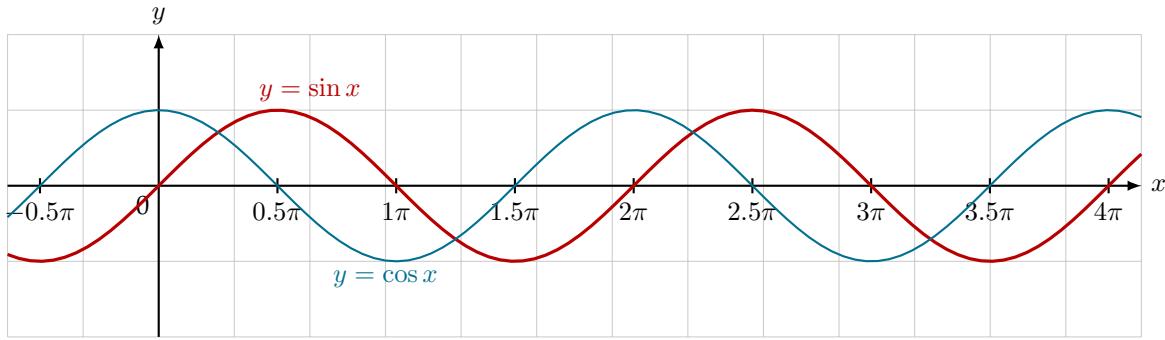


Figure 3: The graph of the functions $\sin x$ and $\cos x$.

1.2 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 $e = 2.71828 \dots$ is Euler's number.

The exponential function $\exp(x) = e^x$ is plotted in figure 4. 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.

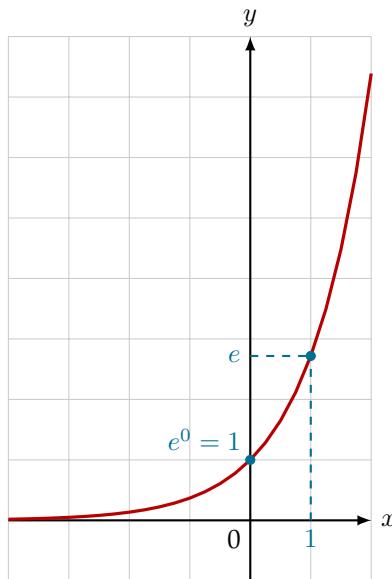


Figure 4: The graph of the function $y = \exp(x) = e^x$.

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 5. 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.

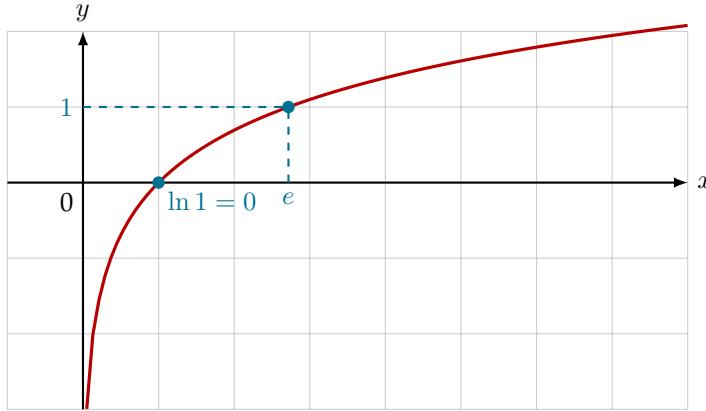


Figure 5: The graph of the function $y = \exp(x) = e^x$.

1.3 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 $f(x) \rightarrow L$ as $x \rightarrow p$, or $\lim_{x \rightarrow p} f(x) = L$.

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 6 for an illustration of this.

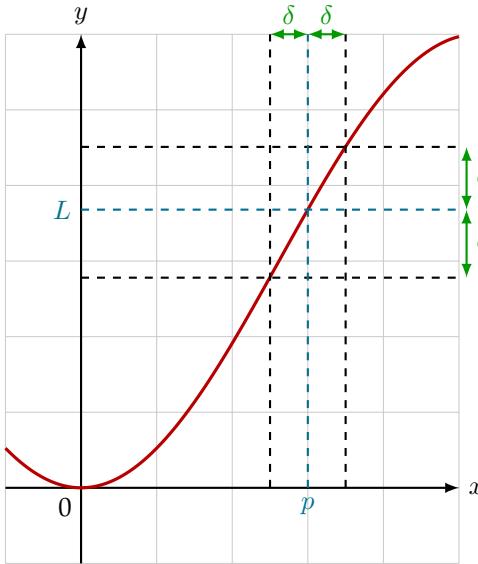


Figure 6: As x approaches p , $f(x)$ approaches L .

An example problem utilising the definition is shown below.

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

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 7.

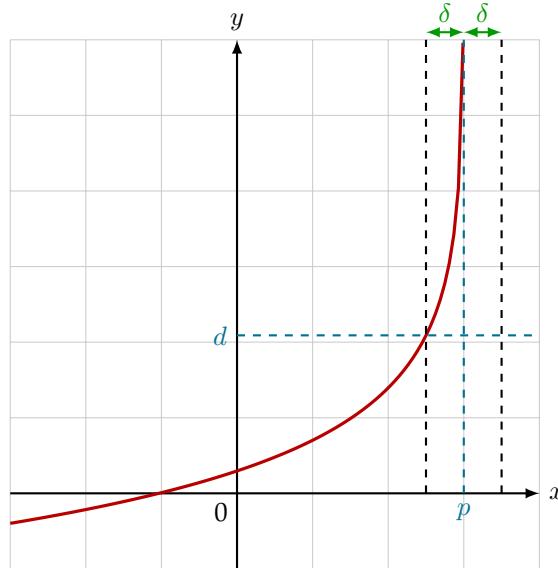


Figure 7: As x approaches p , $f(x)$ approaches infinity.

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 8.

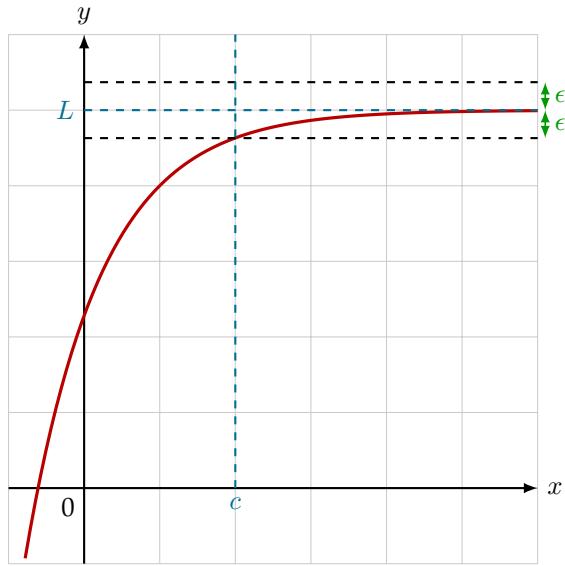


Figure 8: As x approaches infinity, $f(x)$ approaches L .

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 9.

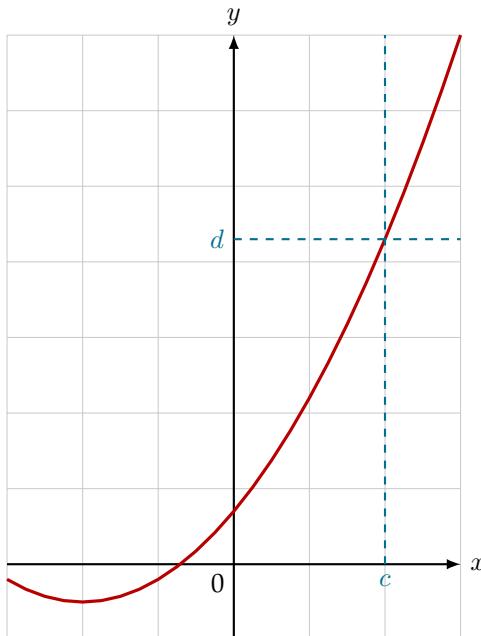


Figure 9: As x approaches infinity, so does $f(x)$.

1.4 Handling infinities and indeterminate forms

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

$$\begin{aligned}\lim_{x \rightarrow a} (f(x) + g(x)) &= \lim_{x \rightarrow a} f(x) + \lim_{x \rightarrow a} g(x) \\ \lim_{x \rightarrow a} (f(x) \cdot g(x)) &= \lim_{x \rightarrow a} f(x) \cdot \lim_{x \rightarrow a} g(x) \\ \lim_{x \rightarrow a} \frac{f(x)}{g(x)} &= \frac{\lim_{x \rightarrow a} f(x)}{\lim_{x \rightarrow a} g(x)}\end{aligned}$$

We use the following rules to handle infinities.

$$\begin{aligned}a \times \infty &= \infty \\ \frac{a}{\infty} &= 0\end{aligned}$$

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).

$$\begin{aligned}\lim_{x \rightarrow 0^+} \frac{1}{x} &= \infty \\ \lim_{x \rightarrow 0^-} \frac{1}{x} &= -\infty\end{aligned}$$

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.

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

1.5 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 too 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.

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

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

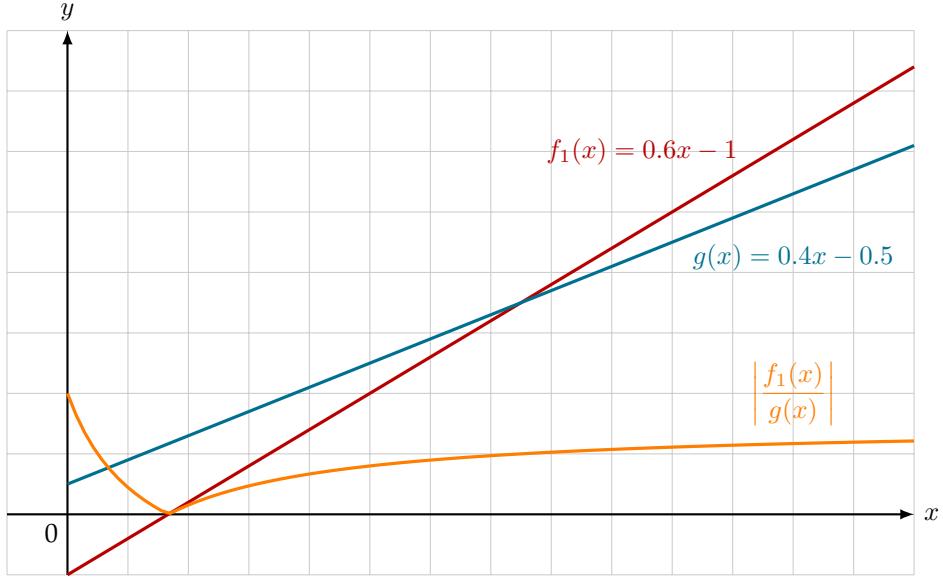


Figure 10: Graphs of a linear function $f_1(x)$, another linear function $g(x)$ and the absolute value of their quotient $|f(x)/g(x)|$.

Figure 10 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 11, 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.

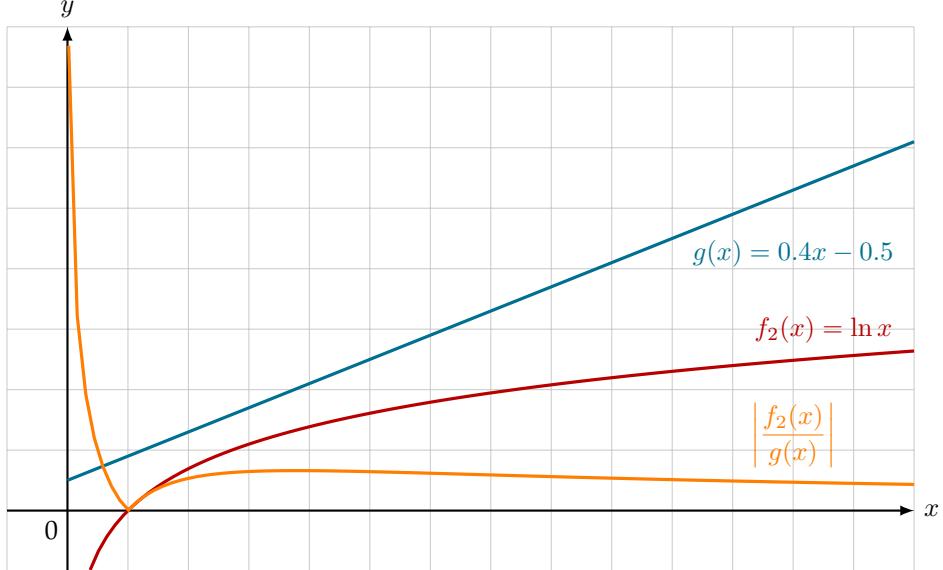


Figure 11: Graphs of a logarithmic function $f_2(x)$, a linear function $g(x)$ and the absolute value of their quotient $|f(x)/g(x)|$.

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 $\lim_{x \rightarrow \infty} \left| \frac{f(x)}{g(x)} \right|$ is bounded, i.e.

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

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

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

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 $\lim_{x \rightarrow \infty} \frac{f(x)}{g(x)} = 0$.

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 $\lim_{x \rightarrow b} \frac{f(x)}{g(x)} = 0$.

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

1.6 Continuity

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

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 12 and 13.

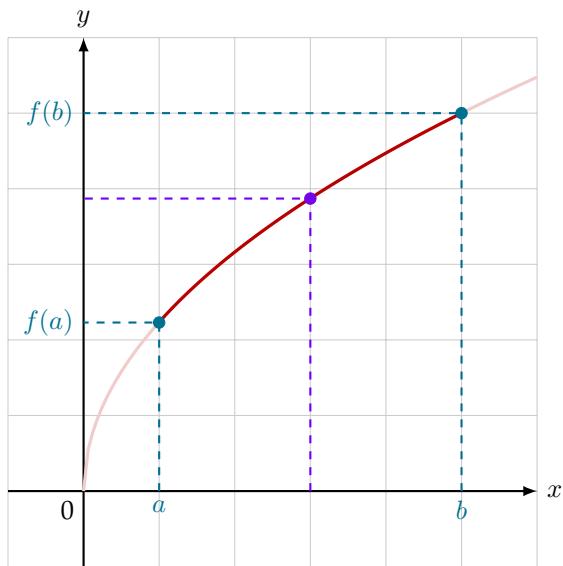


Figure 12: The intermediate value theorem.

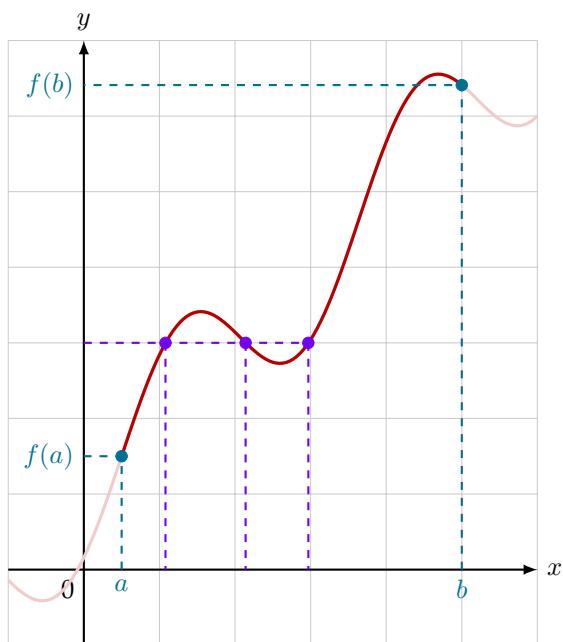


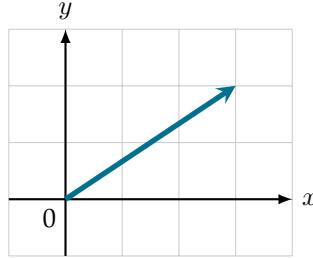
Figure 13: In the intermediate value theorem, for a given value y , the value of x does not necessarily have to be unique.

2 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!

2.1 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

$$\begin{array}{ll} \vec{u} + \vec{v} \in V & \text{(closure under vector addition)} \\ a\vec{u} \in V & \text{(closure under scalar multiplication)} \end{array}$$

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.

2.2 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:

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

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 14 and 15.

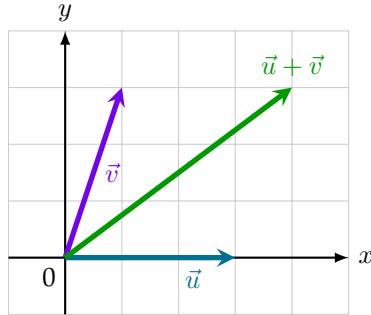


Figure 14: Vectors in \mathbb{R}^2 are closed under addition. For any given pair of vectors \vec{u} and \vec{v} both in V , their sum $\vec{u} + \vec{v}$ must also be a vector.

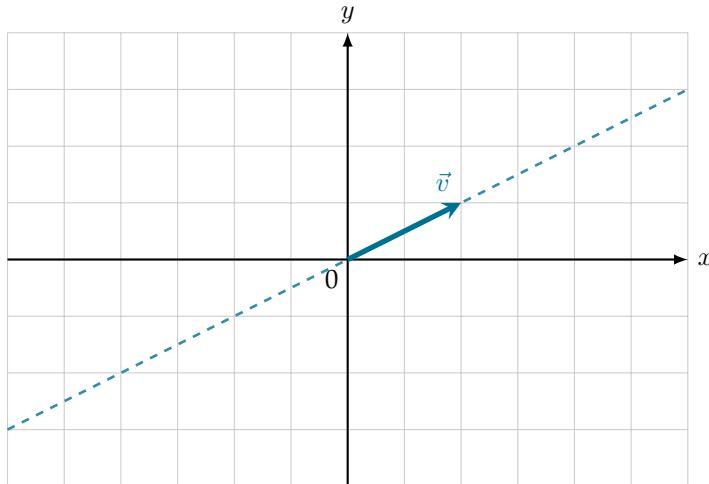


Figure 15: Vectors in \mathbb{R}^2 are closed under scalar multiplication. For any given vector $v \in \mathbb{R}^2$, the product between v and any scalar a (i.e. any vector lying on the dashed line) is also a vector.

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,

$$\begin{aligned}\alpha &= a_1x^2 + b_1x + c_1 \\ \beta &= a_2x^2 + b_2x + c_2\end{aligned}$$

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 1 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 1: 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.

2.3 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 16.

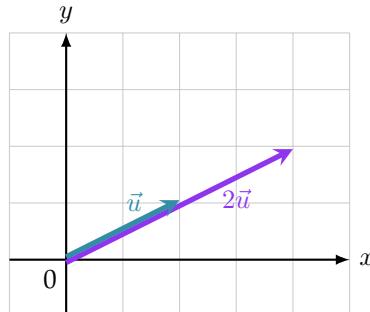


Figure 16: Two collinear vectors.

Given some vectors $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ and some scalars a_1, a_2, \dots, a_n , the sum $a_1\vec{u}_1 + a_2\vec{u}_2 + \dots + a_n\vec{u}_n$ is called a *linear combination* of those vectors. This is illustrated in figure 17.

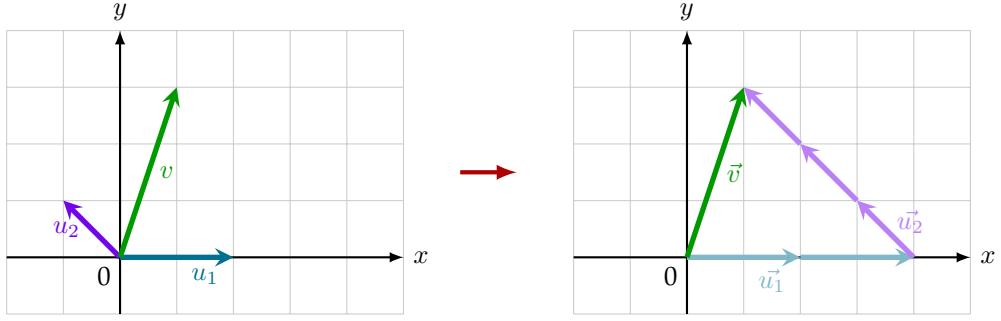


Figure 17: The green vector v can be expressed as a linear combination of the vectors u_1 and u_2 , i.e. $v = 2u_1 + 3u_2$.

Given a family of vectors $S = (\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n)$, 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 18.

- 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.

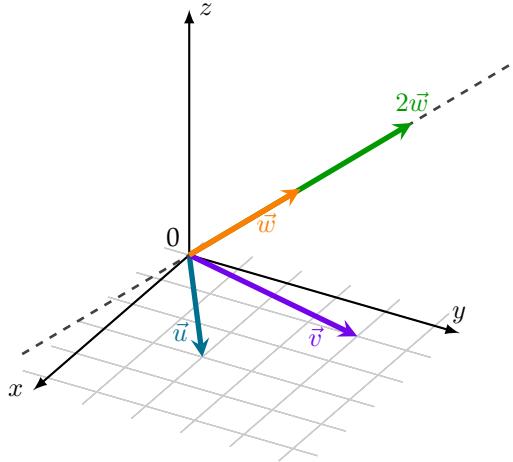


Figure 18: Three vectors in \mathbb{R}^3 . Both \vec{u} and \vec{v} lie flat on the xy -plane. The dashed line represents the span of \vec{w} and $2\vec{w}$.

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 .

2.4 Linear independence

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

Definition of linear independence.

The vectors $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$ are said to be *linearly independent* if:

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

$$a_1\vec{u}_1 + a_2\vec{u}_2 + \dots + a_n\vec{u}_n = 0. \quad (*)$$

- 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, \dots, 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, \dots, a_n are zero.)

For example, in figure 18, 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.

2.5 Basis

Let V be a vector space. A basis¹ of V is a set S of linearly independent vectors that spans the entirety of V . For instance, in figure 18, 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

¹Plural: bases.

that the xy -plane in figure 18 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 19 and 20.

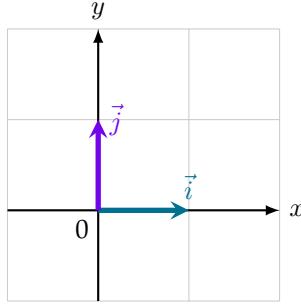


Figure 19: The canonical basis $\{\vec{i}, \vec{j}\}$ of \mathbb{R}^2 .

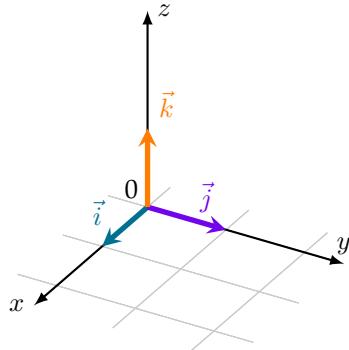


Figure 20: The canonical basis $\{\vec{i}, \vec{j}, \vec{k}\}$ of \mathbb{R}^3 .

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 .

2.6 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 $f : V \rightarrow W$ 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

$$\begin{aligned} f(\vec{u} + \vec{v}) &= f(\vec{u}) + f(\vec{v}) \\ f(a\vec{u}) &= af(\vec{u}). \end{aligned}$$

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

Alternative definition of a linear map.

The function $f : V \rightarrow W$ 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.

$$f : V \rightarrow V$$

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:

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

we can *compose* them into a new linear map

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

where f is applied first, then g .

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

2.7 Kernel and image

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

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

Both $\text{Ker}(f)$ and $\text{Im}(f)$ are vector spaces. Note that $\text{Ker}(f)$ 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 $\text{Ker}(g)$, 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

$$\text{Ker}(g) = \left\{ \begin{bmatrix} x \\ -x \end{bmatrix} \mid x \in \mathbb{R} \right\} = \text{Span} \left(\begin{bmatrix} 1 \\ -1 \end{bmatrix} \right).$$

To find the image $\text{Im}(g)$, 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 $\text{Im}(g) = \text{Span} \left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \right)$.

3 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 21 shows a diagrammatic representation of this procedure.

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

Figure 21: 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.

3.1 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 $f : U \rightarrow V$ 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.

3.1.1 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 22.

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

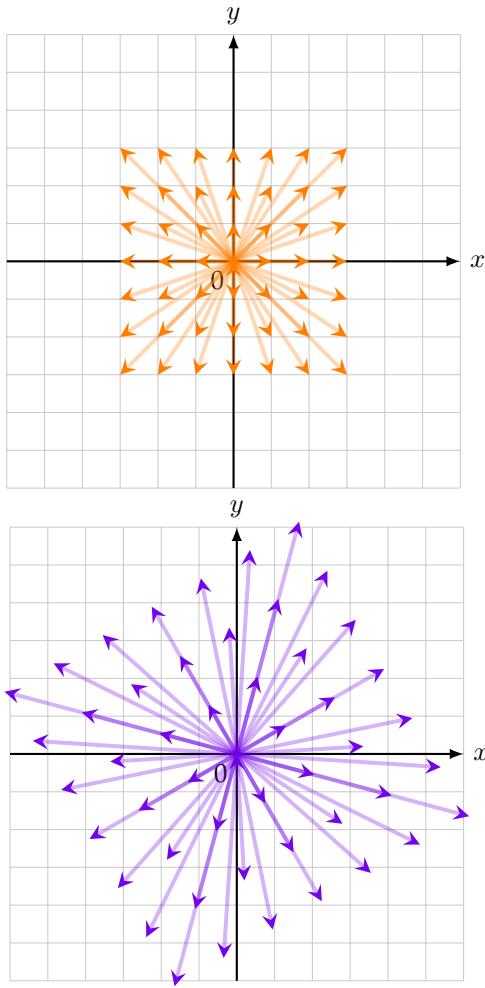


Figure 22: The linear map f takes a vector (in orange) and scales it by a factor of 1.5 before rotating it by 30° anticlockwise. This results in a new vector (in purple).

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 2×2 , 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 $\begin{bmatrix} 2 \cos 30^\circ \\ 2 \sin 30^\circ \end{bmatrix}$. 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 $\begin{bmatrix} 2 \cos 120^\circ \\ 2 \sin 120^\circ \end{bmatrix}$, or $\begin{bmatrix} -2 \sin 30^\circ \\ 2 \cos 30^\circ \end{bmatrix}$. 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

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

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 23.

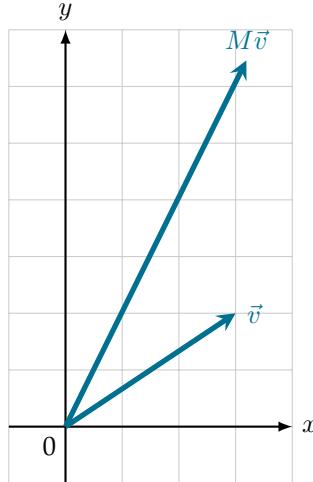


Figure 23: Verifying the equivalence between the linear map f and the matrix $M = \begin{bmatrix} \sqrt{3} & -1 \\ 1 & \sqrt{3} \end{bmatrix}$ by using an example vector $\vec{v} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$.

3.1.2 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.

3.1.3 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 24.

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.

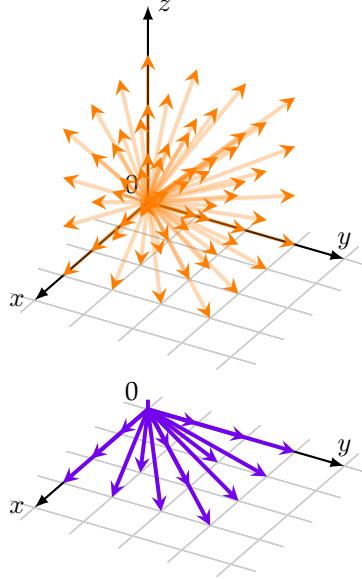


Figure 24: The linear map f projects a three-dimensional vector (in orange) onto the xy -plane. This results in a new two-dimensional vector (in purple).

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 3×2 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 .

3.1.4 Kernels and images

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

$$\begin{aligned} \text{Ker}(M) &= \text{Ker}(f) \\ \text{Im}(M) &= \text{Im}(f) \end{aligned}$$

3.2 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 $a_{i,j}$.

3.3 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.

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

They can also be multiplied by a scalar.

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

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

$$\begin{aligned} M &\in \mathcal{M}_{m,\textcolor{red}{n}}(\mathbb{R}) \\ Q &\in \mathcal{M}_{\textcolor{red}{n},k}(\mathbb{R}) \end{aligned}$$

for some integers m, n and k . See figure 25.

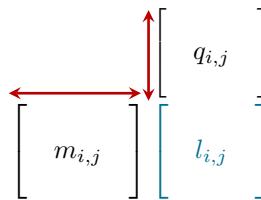


Figure 25: The product of two matrices $L = MQ$ is only valid when the number of columns in M matches the number of rows in Q .

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.

3.4 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.
- $\text{Ker}(M) = \{0\}$.

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

3.5 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.

$$M \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

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

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.

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

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

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.

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

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

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

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

3.6 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 2×2 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.

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

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

$$\begin{bmatrix} \square & \square & \square & \square & \dots \\ \square & \square & \square & \square & \dots \\ \square & \square & \square & \square & \dots \\ \square & \square & \square & \square & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

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.

4 Sequences and series

4.1 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.

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

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

We denote the sequence $u_0, u_1, u_2, u_3, \dots$ 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

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

4.2 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.

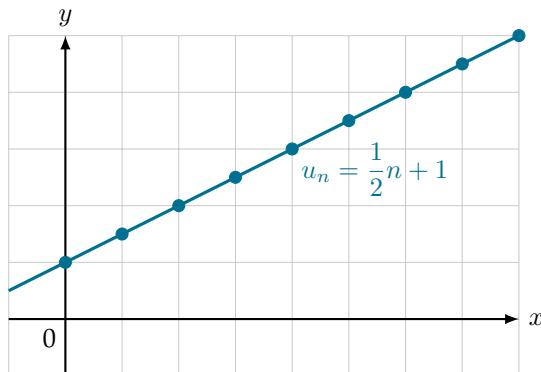


Figure 26: The graph of an arithmetic sequence.

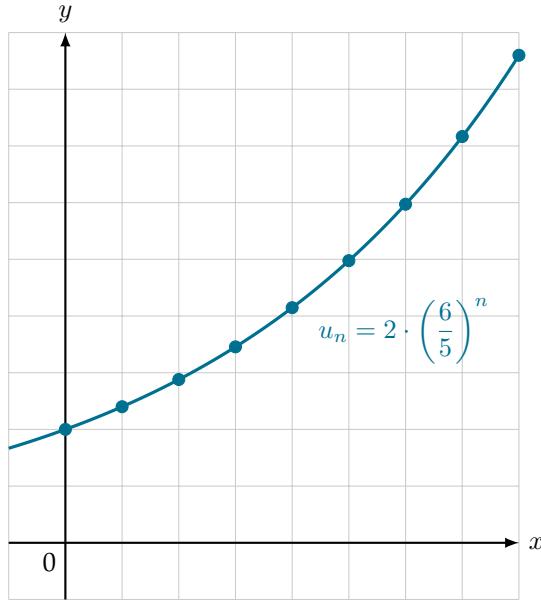


Figure 27: The graph of an geometric sequence.

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.

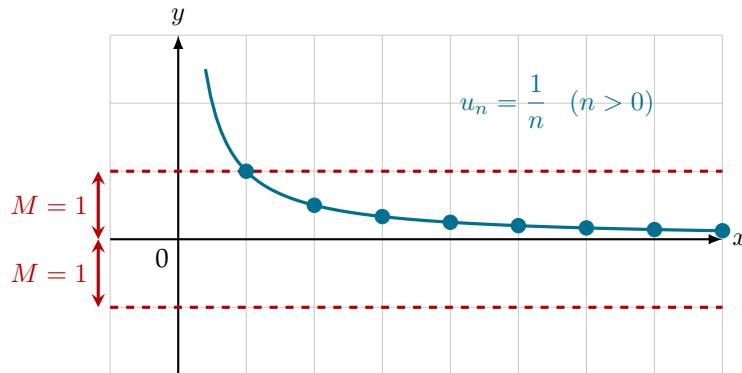


Figure 28: The graph of a decreasing and bounded sequence. Note that none of the terms have an absolute value that exceeds $M = 1$, as shown by the red dashed line.

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.

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

4.3 Convergence of a sequence

Some sequences (u_n) have a limit as n approaches infinity³. For instance, as shown in figure 28, 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.

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

Intuition. Recall the definition of a limit of a sequence.

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

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

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

³Note that as sequences are 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.

$$\begin{aligned} r &= 1 \\ -1 &< r < 1 \end{aligned}$$

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.

4.4 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:

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

4.5 Important theorems 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.

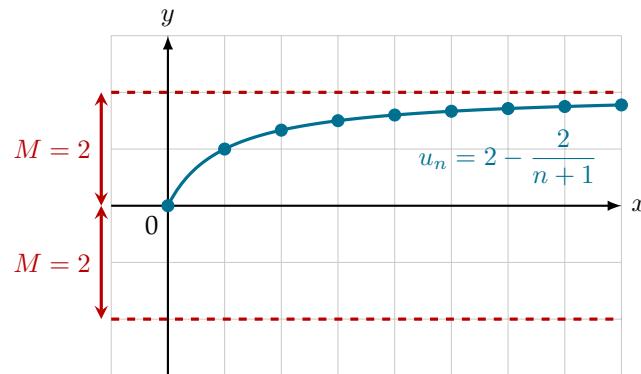


Figure 29: The graph of a increasing and bounded sequence that converges to 2.

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

$$\begin{aligned} -1 &\leq \sin n \leq 1 \\ -\frac{1}{n} &\leq \frac{\sin n}{n} \leq \frac{1}{n} \\ u_n &\leq v_n \leq w_n \end{aligned}$$

By the squeeze theorem, the sequence (v_n) also converges to zero. See figure 30.

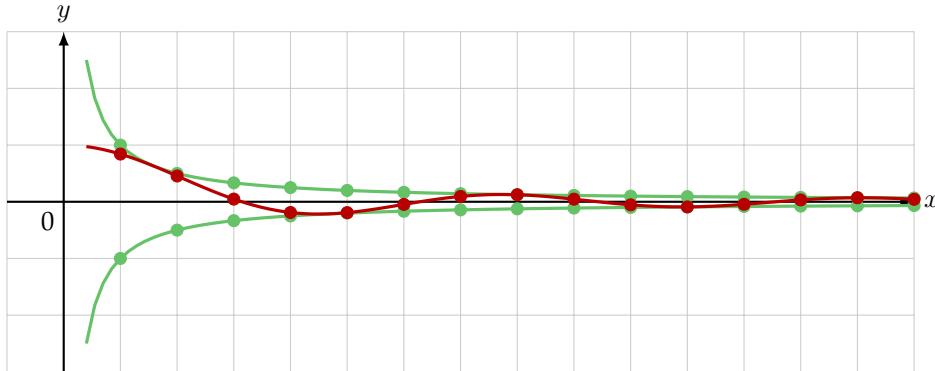


Figure 30: A demonstration of the squeeze theorem. The red curve represents the sequence $v_n = (\sin x)/x$. The green curves represent the sequences $u_n = -1/n$ and $w_n = 1/n$.

4.6 What is a series?

Consider the following sequence.

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

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

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

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

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

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

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

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 $u_0 + u_1 + u_2 + \dots + u_n$ as S_n . We can write

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

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.

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

4.7 Convergence of a series

Some series do not converge.

$$1 + 2 + 3 + 4 + \dots \rightarrow \infty$$

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

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

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.

⁴The formula can be proved as follows. Suppose we want to find $S_n = 1 + 2 + 3 + \dots + (n-2) + (n-1) + n$. This can be rewritten as $S_n = n + (n-1) + (n-2) + \dots + 3 + 2 + 1$. We can add these two equations together to get $2S_n = (n+1) + (n+1) + \dots + (n+1) = n(n+1)$. 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 $\frac{|a_{n+1}|}{|a_n|}$.

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} > 1 \implies \text{the series diverges}$$

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} < 1 \implies \text{the series converges absolutely}$$

$$\lim_{n \rightarrow \infty} \frac{|a_{n+1}|}{|a_n|} = 1 \implies \text{no conclusion reached}$$

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

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

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.