Linear Algebra 2024 Notes

Rachael Carey

TB1 2024

Contents

Introduction

Welcome to Linear Algebra 2024. These are lecture notes for the first half of the first year *Linear Algebra* course in Bristol. Changes are made from year to year, so please do let me know if you find any typos (email `r.m.carey@bristol.ac.uk`).

These notes have been written by Rachael Carey based on notes originally written by Roman Schubert and further developed by Misha Rudnev and John Mackay.

About Linear Algebra

Linear algebra developed as the study of systems of linear equations. It looks at structures which are preserved under the linear operations of addition and multiplication by scalars. In this course we will introduce vectors and matrices, examine their properties, and see how they can be used to solve systems of linear equations. We will look at how these concepts can be generalised into the idea of a vector space

Linear algebra has many applications, both in other areas of maths and in other scientific disciplines. For example, we can use concepts and techniques from linear algebra to tell us about the nature of solutions of certain differential equations, or for linear regression in statistics. It is used extensively in data science, and has many uses in engineering. In this course we will explore linear algebra in its own right, but will also develop the tools you need for applications in later areas of your degree.

About these notes

These notes cover the main material we will develop in the course, and they are meant to be used parallel to the lectures. The lectures will follow roughly the content of the notes, but sometimes in a different order, or with additional material. Likewise, sometimes there will be additional material in the notes which will be referred to in lectures.

The notes contain some exercises which will be discussed in class. Sometimes the solutions to the exercises will be included as well as they may feature useful or important results. However, you are strongly encouraged to think about the exercises before reading the solution. To help with this, the gitbook version of the notes have the solutions hidden and a button to click to expand them later.

The lectures and lecture notes should be used alongside the questions on the problem sheets and in quizzes, where you will get additional practice at understanding ideas and solving problem. Mathematics is primarily learned by doing, and so you should make sure to attempt as many of these problems as you can in order to develop your mathematical skills. You will also have tutorials to help reinforce the material, and will be able to ask for help or clarification on topics both during your tutorials or at weekly drop-in sessions. Theese drop-in sessions will have space to work, so you are welcome to come along to work independently or with friends, and then ask questions as they arise.

The lectures and lecture notes for this course cover all the material you need. However, you may want to explore a textbook alongside the course material. These will usually contain much more material than notes and gives you a broader view of the subject, as well as being sources of additional problems to try. Textbooks come in different styles and use different approaches to a subject, so you should look a bit around to find one which is to your taste (so I would strongly recommend making use of the library rather than buying one, especially initially!). The following is a selection of books about Linear Algebra which are available in our Library:

- Elementary Linear Algebra. Howard Anton
- Linear Algebra and its Applications. Gilbert Strang
- Linear Algebra, S. Lipschutz and M. Lipson
- Linear Algebra. Fraleigh/Beauregard
- Linear Algebra, an introduction. A.O. Morris
- Guide to Linear Algebra. David Towers
- Linear Algebra. Allenby
- Linear Algebra and its Applications I & II. David Griffel

The books above are either introductory or have a focus on applications. A more abstract approach is followed in:

- Abstract Linear Algebra. Curtis
- Linear Algebra. Serge Lang
- Finite-dimensional vector spaces. Paul Halmos

1. Vectors in Euclidean space

To begin our journey into Linear Algebra, we will start by introducing the idea of a vector in Euclidean space. In two or three dimensions, we often represent vectors as arrows with a certain length and direction starting from some reference point. Vectors can also be thought of as an ordered list of numbers, and it is this approach that we will mainly use in order to work more generally in n-dimensional space.

1.1 Introduction to vectors

Vectors as objects which have both length and direction arise in physics and geometry to describe quantities such as forces and velocities. As lists of numbers, they also appear in areas such as data science. In the first chapter of this course, we will consider vectors as elements of Euclidean space $\mathbb{R}^n$. Our scalars (objects with only a size) will primarily be the real numbers. In later chapters we will go on to generalise the concept of a vector and consider different sets of scalars.

Definition 1.1 (Euclidean space)**.** Let $n \in \mathbb{N}$ be a positive integer. The set $\mathbb{R}^n$ consists of all ordered n-tuples $x = (x_1, x_2, x_3, \ldots, x_n)$ where $x_1, x_2, \cdots x_n$ are real numbers, that is

$$\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{R}\}.$$

So, for example, $\mathbb{R}^2$ is the set of all ordered pairs of real numbers (x_1, x_2) where $x_1, x_2 \in \mathbb{R}$, and we call $\mathbb{R}^2$ the Euclidean plane.

Remarks:

- In this course we will often use Roman letters (commonlly u, v, w, x, y) to represent vectors and Greek letters (often lambda λ and mu μ) to represent scalars. Some texts may use explicit notation for vectors, such as bold letters, underlined letters or an arrow above the letter $\vec{v}$, but we will not use any special notation. This means that we need to be careful when we introduce new symbols to specify which sets we are taking them from.
- We will mostly write elements of $\mathbb{R}^n$ in the form $x = (x_1, x_2, x_3, \cdots, x_n)$ or as a column

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

  The two notations are equivalent: the formal difference is that row notation views $x \in \mathbb{R}^n$ as an element of the direct product of n copies of $\mathbb{R}$, while the column notation treats x as an $n \times 1$ 'matrix', a concept which we will study in Chapter 4. In this course we will move between these two notations, but from Chapter 4 onward will primarily think of our vectors as column vectors (although the row notation may sometimes be used to save space on the page).
- The order that the entries are listed in our vectors matter, as they are ordered n-tuples. So for example in $\mathbb{R}^2$ we have $(x_1, x_2) \neq (x_2, x_1)$ if $x_1 \neq x_2$.
- We refer to the element in the ith position of $x \in \mathbb{R}^n$ as the ith component of x.

- We will use the convention that a vector $x \in \mathbb{R}^n$ has ith component x_i without always explicitly stating this. This will apply in the same way to whichever letter we use to represent our vectors, so for example for $v \in \mathbb{R}^n$ we take $v = (v_1, v_2, \ldots, v_n)$, for $y \in \mathbb{R}^m$ we take $y = (y_1, y_2, \ldots, y_m)$ and so on.
- We can visualise a vector in $\mathbb{R}^2$ as a point in the plane, with the first component (also called the x-component) on the horizontal axis and the second (or y-) component on the vertical axis. Similarly, we could extend this to $\mathbb{R}^3$ by including a third (or z-) component on the z-axis. In these cases we may write, for example, $v = (x, y, z)$. These are shown in Figures 1.1 and 1.2.

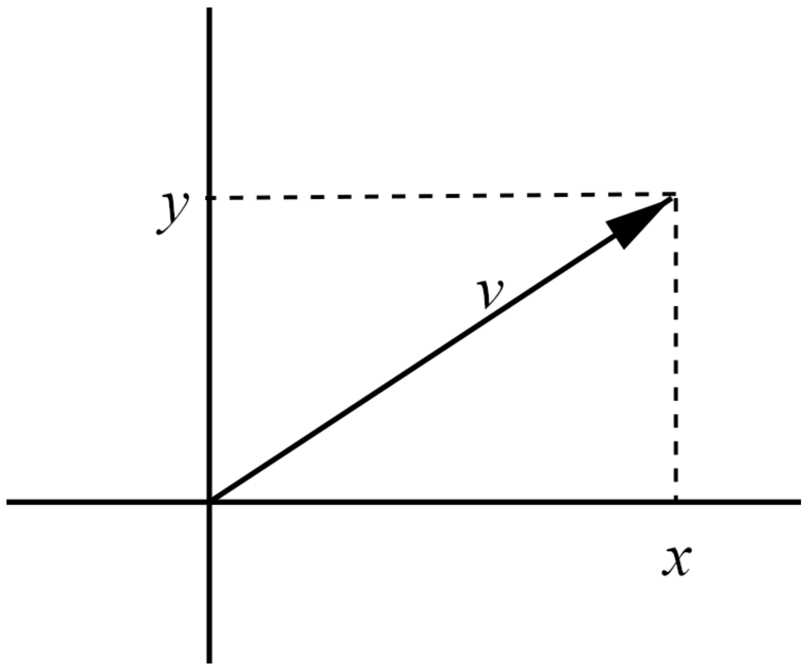

Figure 1.1: An element $v = (x, y)$ in $\mathbb{R}^2$ represented by a vector in the plane.

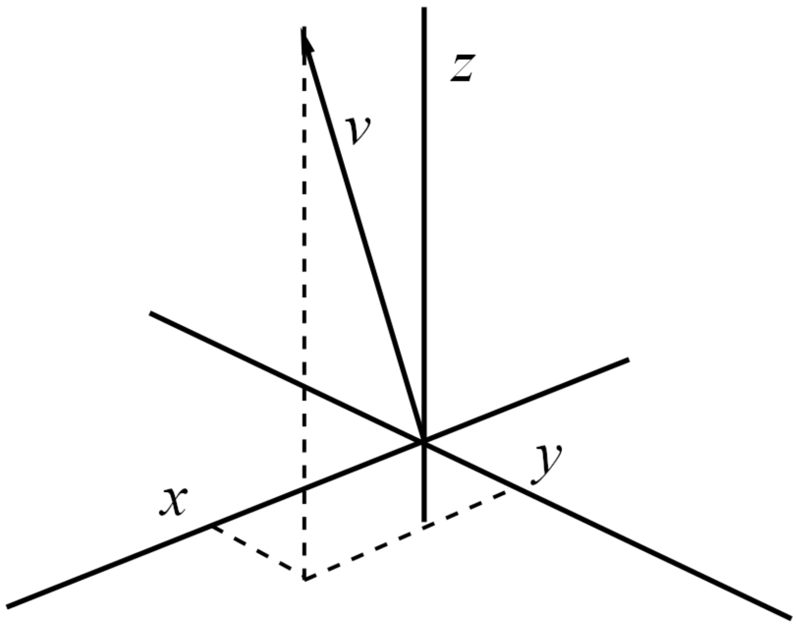

Figure 1.2: An element $v = (x, y)$ in $\mathbb{R}^3$ represented by a vector in the plane. Note the positioning of the axes relative to one another in 3D: this is the convention we follow and x and y should not be switched (as the their relative positions have consequences for later results).

We define two key operations on vectors, addition and scalar multiplication.

Definition 1.2 (Vector addition and scalar multiplication)**.** Let $x, y \in \mathbb{R}^n$, where $x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$. We define the sum of x and y, denoted $x + y$, to be the vector

$$x + y := \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{pmatrix}.$$

If $\lambda \in \mathbb{R}$ we define the multiplication of $x \in \mathbb{R}^n$ by λ by

$$\lambda x := \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{pmatrix}.$$

We may refer to the addition as 'component-wise', meaning that we are performing the operation of addition on each component individually. Note that we are **not** defining multiplication of two vectors here: we are multiplying a vector x by a scalar λ.

A simple consequence of the definition is that scalar multiplication is distributive over addition. This means that for any $x, y \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$

$$\lambda(x+y) = \lambda x + \lambda y.$$

By further combining the above operations we can form expressions like $\lambda v + \mu w$ for $v, w \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$. We call this a **linear combination** of v and w.

Example 1.3. We could take a linear combination of the vectors $(1,-1)$ and $(0,2)$ and simplify it as follows:

$$5\begin{pmatrix}1\\-1\end{pmatrix} + 6\begin{pmatrix}0\\2\end{pmatrix} = \begin{pmatrix}5\\-5\end{pmatrix} + \begin{pmatrix}0\\12\end{pmatrix} = \begin{pmatrix}5\\7\end{pmatrix}.$$

We can also consider linear combinations of k vectors $v_1, v_2, \cdots, v_k \in \mathbb{R}^n$ with coefficients $\lambda_1, \lambda_2 \cdots, \lambda_k \in \mathbb{R}$,

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \sum_{i=1}^{k} \lambda_i v_i$$

Notice that $0v = (0,0,\ldots,0)$ for any $v \in \mathbb{R}^n$ and we will denote the vector whose entries are all 0 by $\mathbf{0}$, so we have

$$v + \mathbf{0} = \mathbf{0} + v = v$$

for any $v \in \mathbb{R}^n$. Note that when we say $0v = \mathbf{0}$, the 0 on the left hand side is $0 \in \mathbb{R}$, whereas the $\mathbf{0}$ in the right hand side is $\mathbf{0} = (0,0,\cdots,0) \in \mathbb{R}^n$. We use the notation $\mathbf{0}$ to represent the zero vector regardless of n, but the length of the vector should be clear from the context.

We will also the shorthand $-v$ to denote $(-1)v$ and $w - v := w + (-1)v$. Notice that with this notation

$$v - v = \mathbf{0}$$

for all $v \in \mathbb{R}^n$.

Addition and negative vectors can also be pictured in two and three dimensions, as shown in Figure 1.3 for the $\mathbb{R}^2$ setting.

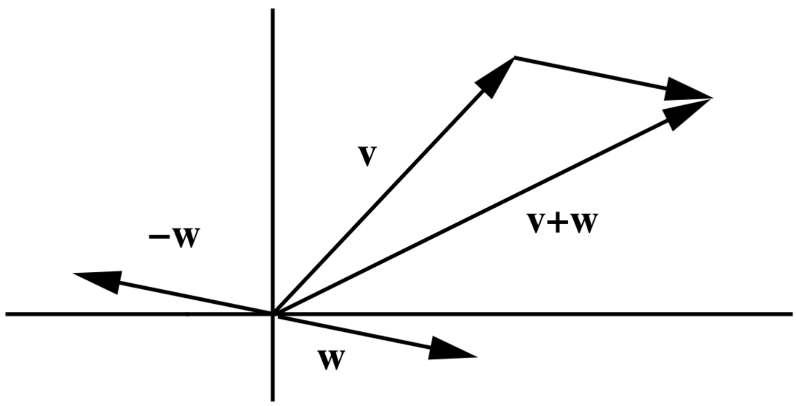

Figure 1.3: Vector addition $v + w$, and the negative $-w$.

Before moving on to look at further properties of our vectors we introduce some notation for some vectors which will come up frequently during the course.

Definition 1.4 (e_i)**.** Fix an $n \in \mathbb{N}$ and choose an $i \in \{1, 2, \dots, n\}$. Then e_i is the vector in $\mathbb{R}^n$ with ith entry 1 and all other entries zero.

Example 1.5. In $\mathbb{R}^{\nvDash}$ we have $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

Example 1.6. In $\mathbb{R}^{\nVDash}$ we have $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.

1.2 Dot products, norms and angles

When working with vectors we often want to know their length and the angle between different vectors. In two or three dimensions we can work geometrically to determine these, but in order to define these concepts more generally in $\mathbb{R}^n$ we first define another way of combining two vectors, namely the dot product.

Definition 1.7 (Dot product)**.** Let $x, y \in \mathbb{R}^n$, then the dot product of x and y is defined by

$$x \cdot y := x_1 y_1 + x_2 y_2 + \cdots x_n y_n = \sum_{i=1}^{n} x_i y_i.$$

This is sometimes known as the scalar product, as it takes two vectors and outputs a scalar.

The following theorem gives us some key properties satisfied by the dot product.

Theorem 1.8. *For all $x, y, v, w \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ we have that:*

(i) $x \cdot y = y \cdot x$.
(ii) $x \cdot (v + w) = x \cdot v + x \cdot w$ *and* $(x + y) \cdot v = x \cdot v + y \cdot v$.
(iii) $(\lambda x) \cdot y = \lambda (x \cdot y)$ *and* $x \cdot (\lambda y) = \lambda (x \cdot y)$.
(iv) $x \cdot x \geq 0$ *and* $x \cdot x = 0$ *is equivalent to* $x = \mathbf{0}$.

Proof. All these properties follow directly from the definition. So we leave most of them as an exercise, and just prove (ii) and (iv).

To prove (ii) we use the definition

$$x \cdot (v + w) = \sum_{i=1}^{n} x_i (v_i + w_i) = \sum_{i=1}^{n} x_i v_i + x_i w_i = \sum_{i=1}^{n} x_i v_i + \sum_{i=1}^{n} x_i w_i = x \cdot v + x \cdot w,$$

and the second identity in (ii) is proved the same way.

For (iv), we notice that

$$x \cdot x = \sum_{i=1}^{n} x_i^2$$

is a sum of squares, i.e.no term in the sum can be negative. Therefore, if the sum is 0, all terms in the sum must be 0, i.e., $x_i = 0$ for all i, which means that $x = \mathbf{0}$. Conversely, if $x = \mathbf{0}$ then it immediately follows from the definition that $x \cdot x = \mathbf{0} \cdot \mathbf{0} = 0$. $\square$

We can now use this to define the norm of a vector, which is a measure of its length.

Definition 1.9 (Norm)**.** The norm of a vector in $\mathbb{R}^n$ is defined as

$$\|x\| := \sqrt{x \cdot x} = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}.$$

A vector which has a norm of 1 is called a **unit vector**. For a vector $x \in \mathbb{R}^n$ we use the notation $\hat{x}$ to represent the unit vector in the direction of the vector x, and we have that $\hat{x} = \frac{1}{||x||}x$.

If $v, w \in \mathbb{R}^n$ then $\|v - w\|$ is the distance between the points v and w.

Example 1.10. The norm of the vector $v = \begin{pmatrix} 5 \\ 0 \end{pmatrix}$ is $\|v\| = 5$, the norm of $w = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$ is $\|w\| = \sqrt{9+1} = \sqrt{10}$, and the distance between v and w is $\|v - w\| = \sqrt{4+1} = \sqrt{5}$.

Note that in $\mathbb{R}^2$ the norm is just the geometric length of the distance between the point in the plane with coordinates (v_1, v_2) and the origin $\mathbf{0}$, which can be found using Pythagoras' Theorem. In fact this can be extended to a version of Pythagoras' Theorem in $\mathbb{R}^n$. In order to state this, we must first extend the concept of our vectors being at right-angles, or orthogonal, to one another.

Definition 1.11 (Orthogonal)**.** The vectors $x, y \in \mathbb{R}^n$ are called orthogonal if $x \cdot y = 0$. We often write $x \perp y$ to indicate that $x \cdot y = 0$ holds.

For example, if $x = (1, 1)$ and $y = (1, -1)$ then $x \cdot y = 1 \cdot 1 + 1 \cdot (-1) = 0$, so these are an example of orthogonal vectors.

We can now consider our more general version of Pythagoras' Theorem.

Theorem 1.12 (Pythagoras' Theorem)**.** *Let $x, y \in \mathbb{R}^n$. We have $x \cdot y = 0$ if and only if*

$$\|x + y\|^2 = \|x\|^2 + \|y\|^2.$$

The proof is left as an exercise.

A fundamental property of the dot product is the Cauchy–Schwarz inequality, which relates the dot product of two vectors to their norms.

Theorem 1.13 (Cauchy–Schwarz inequality)**.** *For any $x, y \in \mathbb{R}^n$*

$$|x \cdot y| \leq \|x\| \|y\|.$$

Proof. If $y = \mathbf{0}$ the inequality is true, so we assume $y \neq \mathbf{0}$. Notice that $v \cdot v \geq 0$ for any $v \in \mathbb{R}^n$, so let us try to use this inequality by applying it to $v = x - ty$, where t is a real number which we will choose later. First we get

$$0 \leq (x - ty) \cdot (x - ty) = x \cdot x - 2tx \cdot y + t^2 y \cdot y,$$

and we see how the dot products and the norm related in the Cauchy–Schwarz inequality appear.

Now we have to make a clever choice for t, let us try

$$t = \frac{x \cdot y}{y \cdot y},$$

this is actually the value of t for which the right hand side becomes minimal. With this choice we obtain

$$0 \leq \|x\|^2 - \frac{(x \cdot y)^2}{\|y\|^2}$$

and so $(x \cdot y)^2 \leq \|x\|^2 \|y\|^2$ which after taking the square root gives the desired result. □

To make sure that we have defined the norm in a sensible way, we consider the properties that we would expect it to have and confirm that these do hold.

Exercise 1.14. The norm of a vector is a measure of its length. Take some time to consider what properties we would want this definition to have for it to be a sensible definition.

We prove some of these properties below.

Theorem 1.15. *The norm satisfies*

(i) $\|v\| \geq 0$ *for all* $v \in \mathbb{R}^n$ *and* $\|v\| = 0$ *if and only if* $v = \mathbf{0}$.
(ii) $\|\lambda v\| = |\lambda| \|v\|$ *for all* $\lambda \in \mathbb{R}, v \in \mathbb{R}^n$.
(iii) $\|v + w\| \leq \|v\| + \|w\|$ *for all* $v, w \in \mathbb{R}^n$ *(this is known as the* ***triangle inequality****).*

Proof.

(i) This follows from the properties of the dot product in Theorem 1.8.
(ii) This follows from a direct computation:

$$\begin{aligned} \|\lambda v\| &= \sqrt{(\lambda v_1)^2 + (\lambda v_2)^2 + \cdots + (\lambda v_n)^2} \\ &= \sqrt{\lambda^2 (v_1^2 + v_2^2 + \cdots + v_n^2)} \\ &= \sqrt{\lambda^2} \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2} = |\lambda| \|v\|. \end{aligned}$$

(iii) We consider

$$\|v + w\|^2 = (v + w) \cdot (v + w) = v \cdot v + 2v \cdot w + w \cdot w = \|v\|^2 + 2v \cdot w + \|w\|^2.$$

and now applying the Cauchy–Schwarz inequality in the form $v \cdot w \leq \|v\| \|w\|$ to the right hand side gives

$$\|v + w\|^2 \leq \|v\|^2 + 2\|w\| \|v\| + \|w\|^2 = (\|v\| + \|w\|)^2,$$

and taking the square root gives the triangle inequality.

□

We can also use the dot product to define the angle between two vectors.

Definition 1.16 (Angle between vectors)**.** Let $x, y \in \mathbb{R}^n$ with $x \neq 0$ and $y \neq 0$. Then the angle θ between the two vectors is defined by

$$\cos\theta = \frac{x \cdot y}{||x||\,||y||}.$$

Notice that this definition makes sense because the Cauchy–Schwarz inequality holds, namely Cauchy–Schwarz gives us

$$-1 \leq \frac{x \cdot y}{||x||\,||y||} \leq 1$$

and therefore there exist an $\theta \in [0, \pi)$ such that

$$\cos\theta = \frac{x \cdot y}{||x||\,||y||}.$$

Notice that if our vectors are orthogonal as defined in Definition 1.11 this corresponds to $\cos\theta = \frac{\pi}{2}$ as expected.

We can now use this definition to compute the angles between vectors.

Example 1.17. If $v = (-1, 7)$ and $w = (2, 1)$, then we find $v \cdot w = 5$, $||v|| = \sqrt{50}$ and $||w|| = \sqrt{5}$, hence $\cos\theta = 5/\sqrt{250} = 1/\sqrt{10}$.

If we are working in $\mathbb{R}^2$ we can prove that this definition of the angle using the dot product does indeed coincide with the geometric method of finding the angle. The proof of this is left as an exercise.

1.3 Polar form in the Euclidean plane

In this section we will focus on vectors in $\mathbb{R}^2$, and consider a different way of representing them. As well as defining a vector in $\mathbb{R}^2$ based on its components in the x and y directions, we could also define it based on its length and its angle from the x-axis. This is known as **polar form**, and is illustrated in Figure 1.4.

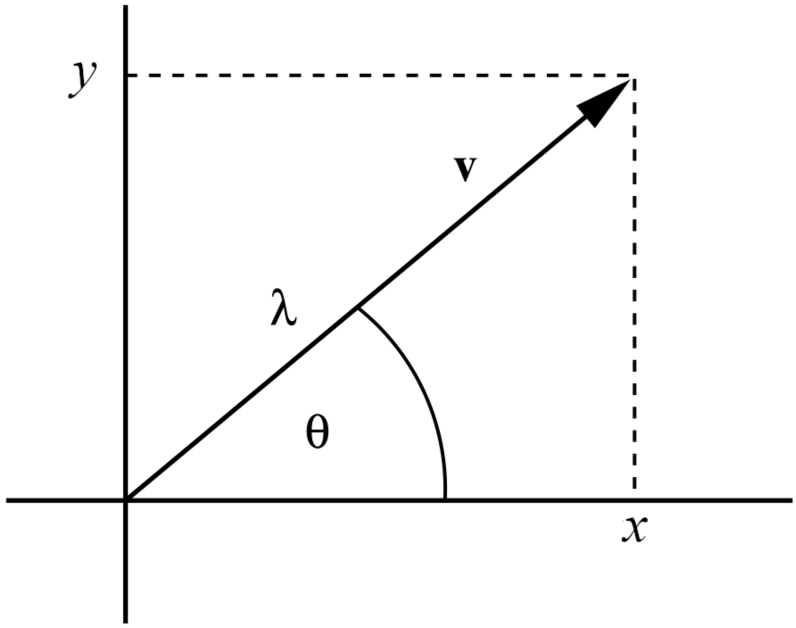

Figure 1.4: A vector v in $\mathbb{R}^2$ represented by Cartesian coordinates (x, y) or by polar coordinates λ, θ. We have $x = \lambda \cos\theta$, $y = \lambda \sin\theta$, $\lambda = \sqrt{x^2 + y^2}$ and $\tan\theta = \frac{y}{x}$.

In particular, a unit vector has length one, hence all unit vectors lie on the circle of radius one in $\mathbb{R}^2$, and a unit vector is determined solely by its angle θ with the x-axis. By elementary geometry we find that the unit vector with angle θ to the x-axis is given by

$$u(\theta) := \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}.$$

We can then multiply by a scalar order to obtain any vector in $\mathbb{R}^2$, and this gives us a unique vector. In particular, the scalar that we multiply our unit vector by is the norm of the vector.

Theorem 1.18. *For every $v \in \mathbb{R}^2$, $v \neq 0$, there exist unique $\theta \in [0, 2\pi)$ and $\lambda \in (0, \infty)$ with*

$$v = \lambda u(\theta)$$

Proof. Given $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \neq 0$ we have to find $\lambda > 0$ and $\theta \in [0, 2\pi)$ such that

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \lambda u(\theta) = \begin{pmatrix} \lambda\cos\theta \\ \lambda\sin\theta \end{pmatrix}.$$

Since $\|\lambda u(\theta)\| = \lambda \|u(\theta)\| = \lambda$ (note that $\lambda > 0$, hence $|\lambda| = \lambda$) we get immediately

$$\lambda = \|v\|.$$

To determine θ we have to solve the two equations

$$\cos\theta = \frac{v_1}{\|v\|}, \quad \sin\theta = \frac{v_2}{\|v\|},$$

which is in principle easy, but we have to be a bit careful with the signs of v_1, v_2. If $v_2 > 0$ we can divide the first by the second equation and obtain $\cos\theta/\sin\theta = v_1/v_2$, hence $\theta = \cot^{-1}\frac{v_1}{v_2}$, that is

$$\theta = \arctan\frac{v_2}{v_1} \in (0,\pi).$$

If $v_1 > 0$ and $v_2 < 0$ we have $\arctan(v_2/v_1) \in (-\frac{\pi}{2}, 0)$ and so $\theta = 2\pi + \arctan(v_2/v_1)$; analogous arguments apply in the remaining cases and this is illustrated in Figure 1.5. $\square$

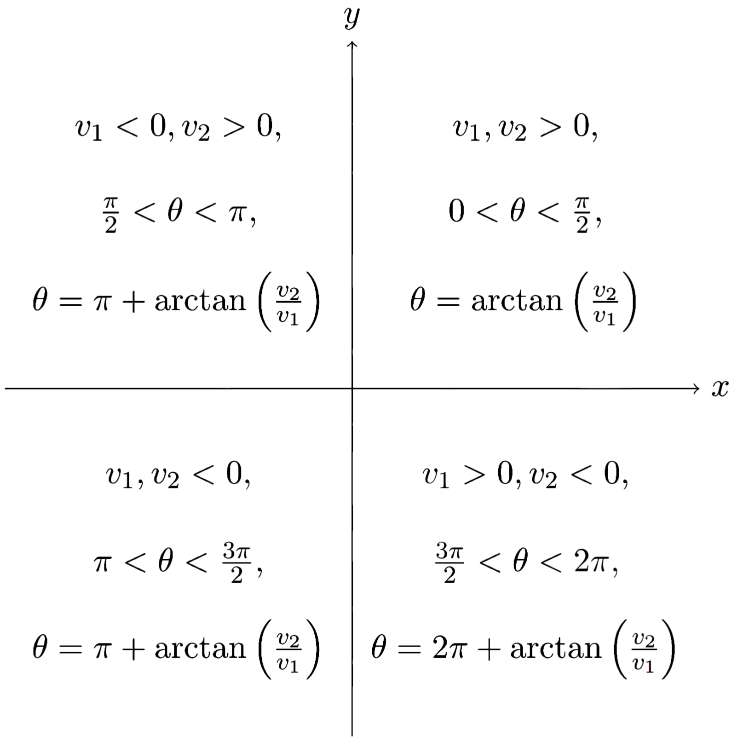

Figure 1.5: How to calculate the argument for (v_1, v_2) in each quadrant of the Euclidean plane.

The converse of this result also holds, that is given $\theta \in [0, 2\pi)$ and $\lambda \geq 0$ we get a unique vector with direction θ and length λ:

$$v = \lambda u(\theta) = \begin{pmatrix} \lambda\cos\theta \\ \lambda\sin\theta \end{pmatrix}.$$

Exercise 1.19. Let $v = \begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix}$. Find $\lambda \in (0,\infty]$ and $\theta \in [0, 2\pi)$ such that $v = \lambda u(\theta)$.

There are many practical situations where the polar form of a vector might be more useful than the Cartesian form, for example a ship navigating from a port may travel a certain distance at a given angle. In the next chapter, we will look at an extended example of vectors in $\mathbb{R}^2$, namely the complex plane, see how complex numbers can be viewed as vectors and how their polar form provides a useful insight into the geometric interpretation of multiplication.

1.4 Vector spaces

To finish this chapter, let's return to our vectors in $\mathbb{R}^n$ and summarise the properties of how these vectors interact with each other and with scalars in $\mathbb{R}^n$.

We have that:

- $u+v=v+u$ for all $u,v\in\mathbb{R}^n$ (vector addition is **commutative**).
- $(u+v)+w=u+(v+w)$ for all $u,v,w\in\mathbb{R}^n$ (vector addition is **associative**).
- there exists an element $0\in\mathbb{R}^n$ such that $v+0=v$ for all $v\in\mathbb{R}^n$ ($\mathbb{R}^n$ contains an **additive identity**).
- for every $v\in\mathbb{R}^n$ there exists an element $w\in\mathbb{R}^n$ such that $v+w=0$ (every element has an **additive inverse**, namely $w=-v$).
- $\lambda(u+v)=\lambda u+\lambda v$ for all $\lambda\in\mathbb{R}$ and all $u,v\in\mathbb{R}^n$ (multiplication is **distributive** over addition).
- $(\lambda+\mu)v=\lambda v+\mu v$ for all $\lambda,\mu\in\mathbb{R}$ and all $v\in\mathbb{R}^n$.
- $\lambda(\mu u)=(\lambda\mu)v$ for all $\lambda,\mu\in\mathbb{R}$ and all $v\in\mathbb{R}^n$.
- $1v=v$ for all $v\in\mathbb{R}^n$.

Note that the first four properties deal only with how the vectors interact with each other, and the last four deal with how our vectors and scalars interact with each other. It is straightforward to check that all of these properties follow from our definitions of vector addition and scalar multiplication in Definition 1.2.

This combination of properties means that our vectors and scalars interact with each other in a 'nice' way. It is then natural to ask whether we can find other examples of sets V where we define some type of addition and scalar multiplication in a way that we satisfy these same properties. This leads us to define a **vector space**.

Definition 1.20 (Vector space)**.** A vector space is a set V together with an underlying set of scalars[a] $\mathbb{F}$ with operations of addition, which takes vectors $u,v\in V$ and gives an element $u+v\in V$, and scalar multiplication, which takes a scalar $\lambda\in\mathbb{F}$ and vector $v\in V$ and gives an element $\lambda V\in V$, such that:

- $u+v=v+u$ for all $u,v\in V$.
- $(u+v)+w=u+(v+w)$ for all $u,v,w\in V$.
- there exists an element $0\in V$ such that $v+0=v$ for all $v\in V$.
- for every $v\in V$ there exists an element $w\in V$ such that $v+w=0$.
- $\lambda(u+v)=\lambda u+\lambda v$ for all scalars $\lambda\in\mathbb{F}$ and all $u,v\in V$.
- $(\lambda+\mu)v=\lambda v+\mu v$ for all scalars λ and $\mu\in\mathbb{F}$ and all $v\in V$.
- $\lambda(\mu u)=(\lambda\mu)v$ for all scalars λ and $\mu\in\mathbb{F}$ and all $v\in V$.
- $1v=v$ for all $v\in V$.

[a]Formally, we need our set of scalars to be an algebraic object known as a field. This is a set with addition and multiplication that are associative, commutative and distributive; we need an additive and multiplicative identity; every element needs an additve inverse; and every non-zero element needs a multiplicative inverse. For now, we will just think of this as $\mathbb{R}$.

For now, we think of our set of scalars as the real numbers, and refer to V as a vector space over $\mathbb{R}$, or a real vector space. Later in the course we will explore the requirements for our set of scalars in more detail. Our key example of a vector space is $\mathbb{R}^n$, and this will be the focus for much of the course. However, many of the definitions and theorems that we will see in the context of $\mathbb{R}^n$ can be generalised to any vector space. We will also see another core example of a vector space in our next chapter, where we will explore the complex numbers.

2. Complex numbers

Up to now, we have focused on the real numbers, but we will now consider an extension of the real number system, the set of complex numbers. One way of looking at complex numbers is to view them as elements in $\mathbb{R}^2$ which can be multiplied. In this section we will introduce complex numbers and use them as an extended development of some of the properties of vectors that we have seen so far.

The basic idea underlying the introduction of complex numbers is to extend the set of real numbers in a way that polynomial equations have solutions. The standard example is the equation

$$x^2 = -1$$

which has no solution in $\mathbb{R}$. We introduce then in a formal way a new number i with the property $\mathrm{i}^2 = -1$ which is a solution to this equation.

Definition 2.1 (Complex numbers)**.** The set of complex numbers is the set of linear combinations of multiples of i and real numbers, that is

$$\mathbb{C} := \{x + \mathrm{i}y : x, y \in \mathbb{R}\},$$

where $\mathrm{i}^2 = -1$.

We will denote complex numbers by $z = x + \mathrm{i}y$ and call $x = \operatorname{Re} z$ the **real part** of z and $y = \operatorname{Im} z$ the **imaginary part** of z.

What is amazing is that having added this one new number, all polynomial equations with real or complex coefficients have a solution in the set of complex numbers!

2.1 Operations on complex numbers

We can define addition and multiplication on the set of complex numbers in a natural way.

Definition 2.2 (Addition and multiplication of complex numbers)**.** Let $z_1 = x_1 + \mathrm{i}y_1$ and $z_2 = x_2 + \mathrm{i}y_2 \in \mathbb{C}$. Then we define

$$z_1 + z_2 := x_1 + x_2 + \mathrm{i}(y_1 + y_2)$$

and

$$z_1 z_2 := x_1 x_2 - y_1 y_2 + \mathrm{i}(x_1 y_2 + x_2 y_1).$$

Notice that the definition of multiplication just follows if we multiply $z_1 z_2$ like normal numbers and then substitute in $\mathrm{i}^2 = -1$:

$$\begin{aligned} z_1 z_2 &= (x_1 + \mathrm{i}y_1)(x_2 + \mathrm{i}y_2) \\ &= x_1 x_2 + \mathrm{i}x_1 y_2 + \mathrm{i}y_1 x_2 + \mathrm{i}^2 y_1 y_2 \\ &= x_1 x_2 - y_1 y_2 + \mathrm{i}(x_1 y_2 + x_2 y_1). \end{aligned}$$

A complex number is defined by a pair of real numbers, and so we can associate a vector in $\mathbb{R}^2$ with every complex number $z = x + \mathrm{i}y$ by $v(z) = (x, y)$. In other words, we associate a point in the plane with every complex number, and we then call this the **complex plane**, which is illustrated in Figure 2.1. For example, if $z = x$ is real, then the corresponding vector lies on the x-axis. If $z = \mathrm{i}$, then $v(\mathrm{i}) = (0,1)$, and any purely imaginary number $z = \mathrm{i}y$ lies on the y-axis. So we sometimes call these the **real** and **imaginary axes** , respectively.

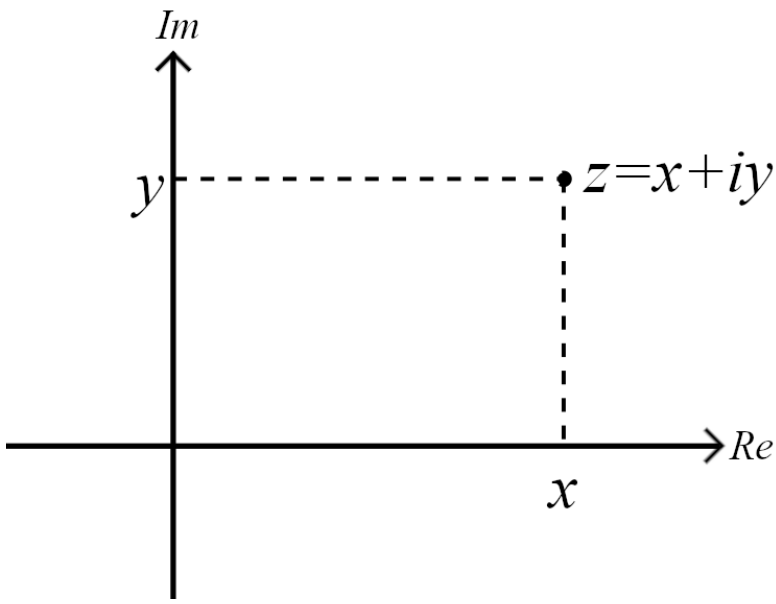

Figure 2.1: Complex numbers as points in the plane: with the complex number $z = x + \mathrm{i}y$ we associate the point $(v)(z) = (x,y) \in \mathbb{R}^2$.

The addition of vectors corresponds to addition of complex numbers as we have defined it, that is

$$v(z_1 + z_2) = v(z_1) + v(z_2),$$

but will seldom use the notation $v(z)$ in favour of just z.

Multiplication of complex numbers is a new operation which had no correspondence for vectors. Therefore we want to study the geometric interpretation of multiplication a bit more carefully. To this end let us first introduce another operation on complex numbers, **complex conjugation** .

Definition 2.3. For $z = x + \mathrm{i}y$ we define the complex conjugate of z to be

$$\bar{z} = x - \mathrm{i}y.$$

This corresponds to reflection across the x-axis. Using complex conjugation we find

$$z\bar{z} = (x + \mathrm{i}y)(x - \mathrm{i}y) = x^2 - \mathrm{i}xy + \mathrm{i}yx + y^2 = x^2 + y^2 = \|v(z)\|^2.$$

So we can find the norm of the vector representation of z by using the complex conjugate. In the complex number setting we usually call this the **modulus** of z and denote it by

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

Complex conjugation is useful when dividing complex numbers: for $z \neq 0$ we have

$$\frac{1}{z} = \frac{\bar{z}}{\bar{z}z} = \frac{\bar{z}}{|z|^2} = \frac{x}{x^2 + y^2} - \mathrm{i}\frac{y}{x^2 + y^2}.$$

and so, for example

$$\frac{z_1}{z_2} = \frac{\bar{z}_2 z_1}{|z_2|^2}.$$

Example 2.4. - We have $(2+3\mathrm{i})(4-2\mathrm{i}) = 8 - 6\mathrm{i}^2 + 12\mathrm{i} - 4\mathrm{i} = 14 + 8\mathrm{i}$.

- We have $\dfrac{1}{2+3\mathrm{i}} = \dfrac{2-3\mathrm{i}}{(2+3\mathrm{i})(2-3\mathrm{i})} = \dfrac{2-3\mathrm{i}}{4+9} = \dfrac{2}{13} - \dfrac{3}{13}\mathrm{i}$.
- We have $\dfrac{4-2\mathrm{i}}{2+3\mathrm{i}} = \dfrac{(4-2\mathrm{i})(2-3\mathrm{i})}{(2+3\mathrm{i})(2-3\mathrm{i})} = \dfrac{2-10\mathrm{i}}{4+9} = \dfrac{2}{13} - \dfrac{10}{13}\mathrm{i}$.

2.2 The polar form of complex numbers and Euler's formula

It turns out that to discuss the geometric meaning of multiplication it is useful to switch the viewpoint to the **polar coordinates** representation of complex numbers.

Take a complex number $z = x + \mathrm{i}y \neq 0$, which has associated point $v(z) = (x, y)$ in $\mathbb{R}^2$. By Theorem 1.18 we can write

$$v(z) = \lambda u(\theta) = \lambda(\cos\theta, \sin\theta)$$

where $\lambda = \sqrt{x^2+y^2} = |z|$ is the distance from (x,y) to the origin, and θ is the angle that the line connecting the origin with (x,y) makes with the x-axis. As in Theorem 1.18, to determine θ we solve $\cos\theta = \frac{x}{\sqrt{x^2+y^2}}$ and $\sin\theta = \frac{y}{\sqrt{x^2+y^2}}$. We call θ the **argument** (or phase) of z. Writing this in terms of complex numbers,

$$z = x + \mathrm{i}y = \sqrt{x^2+y^2}\left(\frac{x}{\sqrt{x^2+y^2}} + \mathrm{i}\frac{y}{\sqrt{x^2+y^2}}\right) = |z|(\cos\theta + \mathrm{i}\sin\theta).$$

Note that $z = 0$ is special: its argument is not defined.

There is a different way to view this through an important connection with the exponential function. The **exponential function** e^z for a complex number z is defined by the series

$$\mathrm{e}^z := 1 + z + \frac{1}{2}z^2 + \frac{1}{3!}z^3 + \frac{1}{4!}z^4 + \cdots = \sum_{n=0}^{\infty} \frac{1}{n!}z^n \tag{2.1}$$

This definition covers the case $z \in \mathbb{C}$, since we can compute powers z^n of z and we can add complex numbers.[1]

We will use that for arbitrary complex z_1, z_2 the exponential function satisfies

$$\mathrm{e}^{z_1}\mathrm{e}^{z_2} = \mathrm{e}^{z_1+z_2}. \tag{2.2}$$

(The proof of this fact for real numbers also works for complex numbers.)

This leads to the following important relationship between different forms of complex numbers.

Theorem 2.5. *For any $\theta \in \mathbb{R}$ we have*

$$\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta.$$

[1] We ignore the issue of convergence here, but this sum is convergent for all $z \in \mathbb{C}$.

Proof. This is basically a calculus result, we will sketch the proof, but you need more calculus to fully justify it. Similarly to (2.1), the sine function and the cosine function can be defined by the following power series

$$\sin(z) = z - \frac{1}{3!}z^3 + \frac{1}{5!}z^5 - \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)!} z^{2k+1} \quad \text{and}$$
$$\cos(z) = 1 - \frac{1}{2}z^2 + \frac{1}{4!}z^4 - \cdots = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k)!} z^{2k}.$$

Now we use (2.1) with $z = \mathrm{i}\theta$, and since $(\mathrm{i}\theta)^2 = -\theta^2$, $(\mathrm{i}\theta)^3 = -\mathrm{i}\theta^3$, $(\mathrm{i}\theta)^4 = \theta^4$, $(\mathrm{i}\theta)^5 = \mathrm{i}\theta^5$, and so on, we find by comparing the power series

$$\begin{aligned} \mathrm{e}^{\mathrm{i}\theta} &= 1 + \mathrm{i}\theta - \frac{1}{2}\theta^2 - \mathrm{i}\frac{1}{3!}\theta^3 + \frac{1}{4!}\theta^4 + \mathrm{i}\frac{1}{5!}\theta^5 + \cdots \\ &= \left[1 - \frac{1}{2}\theta^2 + \frac{1}{4!}\theta^4 + \cdots\right] + \mathrm{i}\left[\theta - \frac{1}{3!}\theta^3 + \frac{1}{5!}\theta^5 + \cdots\right] = \cos\theta + \mathrm{i}\sin\theta. \end{aligned}$$

□

Using Euler's formula, in the notation of Theorem 1.18, we see that

$$v(\mathrm{e}^{\mathrm{i}\theta}) = u(\theta).$$

In particular we can write any complex number z, $z \neq 0$, in **polar coordinate** form

$$z = \lambda \mathrm{e}^{\mathrm{i}\theta},$$

where $\lambda = |z|$ is the modulus (or amplitude) and θ is the argument (or phase).

For the multiplication of nonzero complex numbers we find then that if $z_1 = \lambda_1 \mathrm{e}^{\mathrm{i}\theta_1}$ and $z_2 = \lambda_2 \mathrm{e}^{\mathrm{i}\theta_2}$ then

$$z_1 z_2 = \lambda_1 \lambda_2 \mathrm{e}^{\mathrm{i}(\theta_1+\theta_2)} \quad \text{and} \quad \frac{z_1}{z_2} = \frac{\lambda_1}{\lambda_2} \mathrm{e}^{\mathrm{i}(\theta_1-\theta_2)},$$

so multiplication corresponds to adding the arguments and multiplying the modulus. In particular if $\lambda = 1$, then multiplying by $\mathrm{e}^{\mathrm{i}\theta}$ corresponds to rotation by θ in the complex plane.

Using the polar form of complex numbers also makes taking large powers much easier. By Equation (2.2) we have for $n \in \mathbb{N}$ that $(\mathrm{e}^{\mathrm{i}\theta})^n = \mathrm{e}^{\mathrm{i}n\theta}$, and since $\mathrm{e}^{\mathrm{i}\theta} = \cos\theta + \mathrm{i}\sin\theta$ and $\mathrm{e}^{\mathrm{i}n\theta} = \cos(n\theta) + \mathrm{i}\sin(n\theta)$ this gives us the following identity which is known as **de Moivre's Theorem** .

Theorem 2.6. *Let $n \in \mathbb{N}$ and $\theta \in \mathbb{R}$. Then we have*

$$(\cos\theta + \mathrm{i}\sin\theta)^n = \cos(n\theta) + \mathrm{i}\sin(n\theta).$$

These results have some nice applications to trigonometric functions.

- We can use de Moivre's Theorem to derive trigonometric identities. For instance, if we let $n = 2$, and multiply out the left hand side, we obtain $\cos^2\theta + 2\mathrm{i}\sin\theta\cos\theta - \sin^2\theta = \cos(2\theta) + \mathrm{i}\sin(2\theta)$ and separating real and imaginary part leads to the two angle doubling identities
$$\cos(2\theta) = \cos^2\theta - \sin^2\theta, \quad \sin(2\theta) = 2\sin\theta\cos\theta.$$
  Similar identities can be derived for larger n.

- If we use $e^{i\theta}e^{-i\varphi} = e^{i(\theta-\varphi)}$ and apply Euler's formula to both sides we obtain $(\cos\theta + i\sin\theta)(\cos\varphi - i\sin\varphi) = \cos(\theta-\varphi) + i\sin(\theta-\varphi)$ and multiplying out the left hand side gives the two relations

$$\cos(\theta-\varphi) = \cos\theta\cos\varphi + \sin\theta\sin\varphi, \quad \sin(\theta-\varphi) = \sin\theta\cos\varphi - \cos\theta\sin\varphi.$$

- Euler's formula can be used to obtain the following standard representations for the sine and cosine functions:

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i}, \quad \cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2}.$$

Polar form also lets us deduce some results about complex numbers that may feel surprising, such as below.

Exercise 2.7. What type of number is i^i?

2.3 Complex vectors

As well as thinking of complex numbers as vectors in $\mathbb{R}^2$ with some extra properties, we can also consider vectors whose components are complex numbers by defining $\mathbb{C}^n$ in a similar way to $\mathbb{R}^n$.

Definition 2.8 (The set of complex vectors)**.** Let $n \in \mathbb{N}$ be a positive integer. The set $\mathbb{C}^n$ consists of all ordered n-tuples $x = (x_1, x_2, x_3, \ldots, x_n)$ where $x_1, x_2, \cdots x_n$ are complex numbers, that is

$$\mathbb{C}^n = \{(x_1, x_2, \ldots, x_n) : x_1, x_2, \ldots, x_n \in \mathbb{C}\}.$$

We can add complex vectors componentwise just as we did for real vectors. When it comes to multiplying by scalars, we now have two options: we can either choose complex numbers or real numbers for our scalars. In either case, we can easily check that the properties from 1.20 hold, giving us two more examples of vector spaces.

Example 2.9. - We have that $\mathbb{C}^n$ over $\mathbb{C}$ is a (complex) vector space. - We have that $\mathbb{C}^n$ over $\mathbb{R}$ is a (real) vector space.

When we refer to a complex or real vector space as above, we are using this as shorthand to indicate our choice of scalars $\mathbb{F}$.

Exercise 2.10. Is $\mathbb{R}^n$ a vector space over $\mathbb{C}$?

Click for solution

No, $\mathbb{R}^n$ is **not** a vector space over $\mathbb{C}$, as we would not have $\lambda V \in \mathbb{R}^n$ for all $\lambda \in \mathbb{C}$ and all $v \in \mathbb{R}^n$. For example, $ie_1 = \begin{pmatrix} i \\ 0 \end{pmatrix} \notin \mathbb{R}^n$.

We can also generalise the dot product on complex vectors as follows.

Definition 2.11 (The dot product for complex vectors)**.** Let $u = (u_1, u_2, \ldots, u_n), v = (v_1, v_2, \ldots, v_n) \in \mathbb{C}^n$. Then the dot product is defined as

$$u \cdot v = \sum_{i=1}^{n} u_i \overline{v_i}.$$

Exercise 2.12. Do the properties of 1.8 still hold for the complex dot product?

Click for solution

Not all of the properties hold in the complex case. In particular, the dot product is not commutative for complex numbers, so property (i) does not hold. Properties (ii) and (iv) do hold (this is left as an exercise to verify). The first part of property (iii) holds but the second part must be modified since λ may be complex, giving $x \cdot (\lambda y) = \bar{\lambda}(x \cdot y)$.

We can still find the norm in the same way as for real vectors.

Definition 2.13 (Norm of a complex vector)**.** Let $v = (v_1, v_2, \ldots, v_n) \in \mathbb{C}^n$. Then the norm of V is given by

$$||v|| = \sqrt{v \cdot v}.$$

Because of the way we have defined our complex dot product, we know that $v \cdot v$ wil be a positive real number, so this is well defined. We can revisit Theorem @ref{thm:normprop2} and confirm that the properties of the norm still hold. This is left as an exercise.

Example 2.14. Let $u = (i, 1, 4+i), v = (0, 1-i, -2) \in \mathbb{C}^3$. Then

$$u \cdot v = (i)(0) + (1)(1+i) + (4+i)(-2) = 0 + 1 + i - 4 - 2i = -3 + i$$

and

$$u \cdot u = (i)(-i) + (1)(1) + (4+i)(4-i) = 1 + 1 + (16+1) = 19$$

so $||u|| = \sqrt{19}$.

We will revisit this idea and the properties of the complex dot product later in the course in its general context, which is known as an inner product.

Having discussed vectors and their properties, we next move on to look at another type of mathematical object in the next chapter, namely matrices.

3. Linear equations and Matrices

In this chapter we will introduce a new mathematical object, matrices, which are rectangular arrays of numbers. In order to motivate this, we will begin exploring systems of linear equations and will see how matrices will allow us to better understand these.

3.1 Linear equations

The simplest linear equation is an equation of the form

$$ax = b,$$

where x is an unknown number which we want to determine and a and b are known real numbers with $a \neq 0$, for example $2x = 7$. For this example we find the solution $x = 7/2$. Linear means that no powers or more complicated expressions of x occur, for instance the equations

$$3x^5 - 2x = 3 \quad \text{and} \quad x + \cos(x) = 1$$

are **nonlinear**.

But more interesting than the case of one unknown are equations where we have more than one unknown. Let us look at a couple of simple examples.

Example 3.1. Consider

$$3x - 4y = 3,$$

where x and y are two unknown real numbers. In this case the equation is satisfied for all x, y such that

$$y = \frac{3}{4}x - \frac{3}{4},$$

so instead of determining a single solution the equation defines a set of x, y which satisfy the equation. This set is a line in $\mathbb{R}^2$.

Example 3.2. We could add another equation and consider the solutions to two equations, for example

$$3x - 4y = 3 \quad \text{and} \quad 3x + y = 1.$$

In this example we again find a single solution. Subtracting the second equation from the first gives $-5y = 2$, hence $y = -2/5$ and then from the first equation $x = 1 + \frac{4}{3}y = 7/15$. Another way to look at the two equations is that they define two lines in $\mathbb{R}^2$ and the joint solution is the intersection of these two straight lines, as depicted in Figure 3.1.

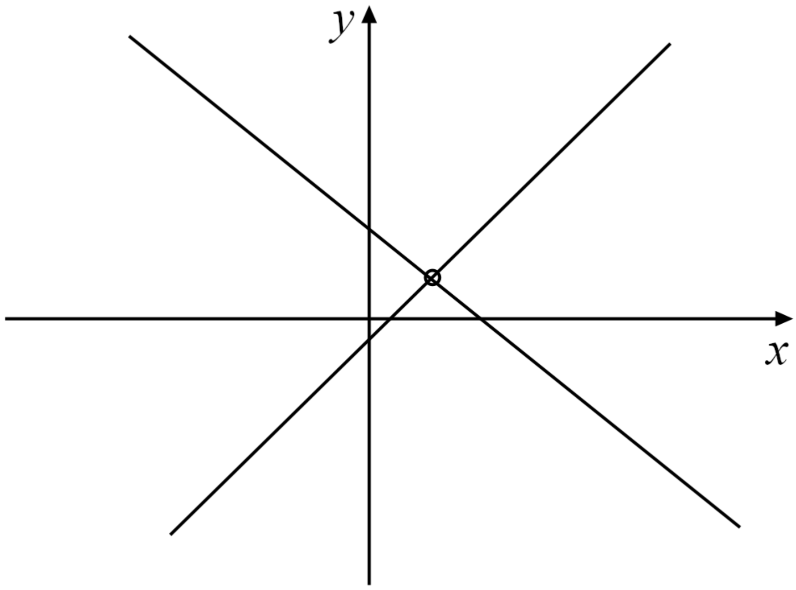

Figure 3.1: A system of two linear equations in two unknowns (x,y) which determines two lines; their intersection gives the solution.

Example 3.3. If we look instead at the slightly modified system of two equations

$$3x-4y=3 \quad \text{and} \quad -6x+8y=0,$$

then we find that these two equations have **no** solutions. To see this we multiply the first equation by -2, and then the set of two equations becomes

$$-6x+8y=6 \quad \text{and,} \quad -6x+8y=0,$$

so the two equations contradict each other and the system has no solutions. Geometrically speaking this means that the straight lines defined by the two equations have no intersection, that is they are parallel, as depicted in Figure 3.2.

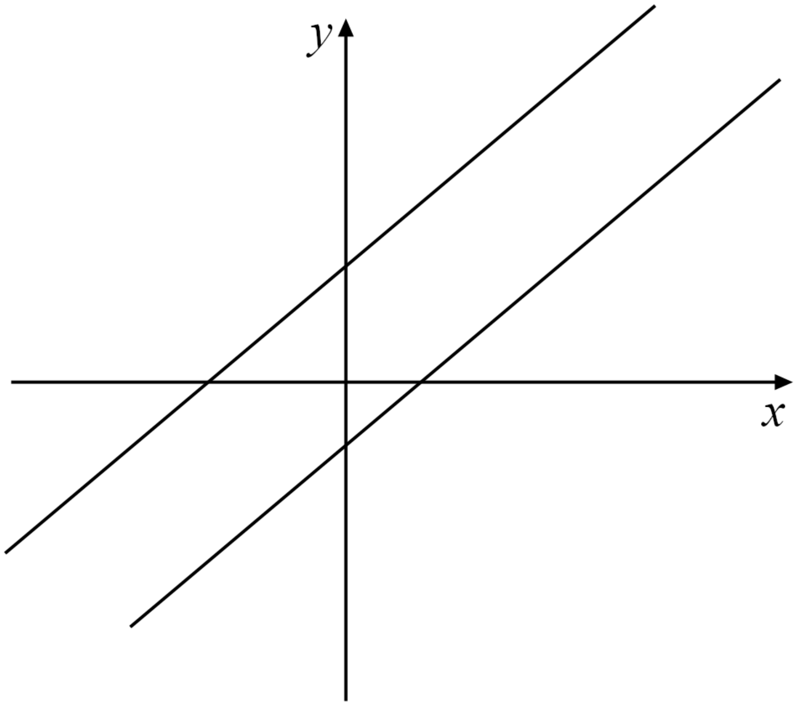

Figure 3.2: A system of two linear equations in two unknowns (x, y) where the corresponding lines have no intersection, hence the system has no solution.

So above we have found examples of systems of linear equations which have exactly one solution, many solutions, and no solutions at all. We will see in the following that these are all the possible outcomes which can occur in general. So far we have talked about linear equations but haven't really defined them in general, so we now do so below.

Definition 3.4 (Linear equation)**.** A linear equation in n variables $x_1, x_2, \cdots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

where $a_1, a_2, \cdots, a_n$ and b are given numbers. These are known as the coefficients of the equation.

In the rest of this chapter, the numbers $a_1, a_2, \cdots, a_n$ and b will be real numbers, and the solutions we are searching for will also be real numbers; however there may be other settings where the coefficients and/or solutions are taken from different sets. For example we could allow our coefficents and solutions to be complex numbers rather than real numbers, and then by replacing every instance of $\mathbb{R}$ with $\mathbb{C}$ the corresponding results would follow in the same way.

In this course we will often be interested in systems of linear equations.

Definition 3.5 (System of linear equations)**.** A system of m linear equations in n unknowns

$x_1, x_2, \cdots, x_n$ is a collection of m linear equations of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots \qquad \vdots \qquad &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned}$$

where the coefficients a_{ij} and b_j are given numbers.

When we ask for a solution $x_1, x_2, \cdots, x_n$ to a system of linear equations, then we ask for a set of numbers $x_1, x_2, \cdots, x_n$ which satisfy all m equations simultaneously.

One often looks at the set of coefficients a_{ij} defining a system of linear equations as an independent entity in its own right.

Definition 3.6 (Matrix)**.** For $m, n \in \mathbb{N}$, a $m \times n$ matrix A (an "m by n" matrix) is a rectangular array of numbers $a_{ij} \in \mathbb{R}$, $i = 1, 2, \ldots, m$ and $j = 1, \ldots, n$ of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

The numbers a_{ij} are called the **elements** of the matrix A, and we often write $A = (a_{ij})$ to denote the matrix A with elements a_{ij}. The set of all $m \times n$ matrices with real elements will be denoted by

$$M_{m,n}(\mathbb{R}),$$

and if $n = m$ we will write

$$M_n(\mathbb{R}).$$

One can similarly define matrices with elements in other sets, e.g,. $M_{m,n}(\mathbb{C})$ is the set of matrices with complex elements.

Note: The plural of matrix is **matrices**.

Example 3.7. An example of a 3×2 matrix is

$$\begin{pmatrix} 1 & 3 \\ -1 & 0 \\ 2 & 2 \end{pmatrix}.$$

An $m \times n$ matrix has m **rows** and n **columns**. The ith row of $A = (a_{ij})$ is

$$\begin{pmatrix} a_{i1} & a_{i2} & \cdots & a_{in} \end{pmatrix}$$

and is naturally identified as a **row vector** $(a_{i1}, a_{i2}, \cdots, a_{in}) \in \mathbb{R}^n$ with n components. The jth column of A is

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix},$$

which is a **column vector** in $\mathbb{R}^m$ with m components.

Example 3.8. For the matrix in Example 3.7 the first and second column vectors are

$$\begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix} \text{ and } \quad \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix},$$

respectively, and the first, second and third row vectors are

$$\begin{pmatrix} 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} 2 & 2 \end{pmatrix}.$$

There is one somewhat unpleasant notational subtlety here. Take, say a vector $(3,4) \in \mathbb{R}^2$. This vector can be written as a matrix either as a 1×2 matrix $(3\ 4)$, with just one row or a 2×1 matrix $\begin{pmatrix} 3 \\ 4 \end{pmatrix}$, with just one column. To avoid confusion, we need a convention whether we are going to identify vectors with row or column matrices, which will later lead us to the general concept of the transpose of a matrix.

The standard convention is to identify a vector $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ with a column-matrix (or column-vector)

$$x = (x_1, \ldots, x_n) = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

but bear in mind that the same quantity can also be represented by a row-matrix (or row-vector)

$$x^t = (x_1 \ \ldots \ x_n).$$

To distinguish, with the boldface notation, between row and column-matrices, representing a single vector, we will use the superscript ^{t} (to be read "transpose", this will be discussed further later in the chapter) for row-matrices. The difference between the latter two formulae is that we do not use comas to separate elements of row-matrices.

When dealing with matrices it will often be useful to write them in terms of their rows or in terms of their columns. That is, if the rows of A are $r_1^t, r_2^t, \ldots, r_m^t$ (for now, think of the superscript t as notation so that we remember they are rows) we may write

$$A = \begin{pmatrix} \cdots & r_1^t & \cdots \\ \cdots & r_2^t & \cdots \\ & \vdots & \\ \cdots & r_m^t & \cdots \end{pmatrix} \text{ or just } A = \begin{pmatrix} r_1^t \\ r_2^t \\ \vdots \\ r_m^t \end{pmatrix},$$

and if the columns of A are $c_1, \ldots, c_n$ then we may write

$$A = \begin{pmatrix} \vdots & \vdots & & \vdots \\ c_1 & c_2 & \cdots & c_n \\ \vdots & \vdots & & \vdots \end{pmatrix} \text{ or just } A = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix}.$$

In Definition 3.5, the rows of the matrix of coefficients are combined with the n unknowns to produce m numbers b_i, we will take these formulas and turn them into a definition for the action of $m \times n$ matrices on vectors with n components:

Definition 3.9. Let $A = (a_{ij})$ be an $m \times n$ matrix and $x \in \mathbb{R}^n$ with components $x = (x_1, x_2, \cdots, x_n)$, then the action of A on x is defined by

$$Ax := \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix} \in \mathbb{R}^m.$$

Note that Ax is a vector in $\mathbb{R}^m$ and if we write $y = Ax$ then the components of y are given by

$$y_i = \sum_{j=1}^{n} a_{ij x_j} \tag{3.1}$$

which is the dot-product between x and the ith row vector of A. The action of A on elements of $\mathbb{R}^n$ is a map from $\mathbb{R}^n$ to $\mathbb{R}^m$, i.e.,

$$A : \mathbb{R}^n \to \mathbb{R}^m.$$

Using the notation of matrices and their action on vectors, a system of linear equations of the form in Definition 3.5 can now be rewritten as

$$Ax = b. \tag{3.2}$$

So using matrices allows us to write a system of linear equations in a much more compact way.

Another way of looking at the action of a matrix on a vector is as follows: Let $a_1, a_2, \cdots, a_n \in \mathbb{R}^m$ be the column vectors of A, then

$$Ax = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n. \tag{3.3}$$

So Ax is a linear combination of the column vectors of A with coefficients given by the components of x. This relation follows directly from (3.1). Solving $Ax = b$ means that we want to find coefficients $(x_1, \ldots, x_n)$ so that b may written as a linear combination of the column-vectors of the matrix A. Such a linear combination may or may not exist, and if it exists may or may not be unique.

Exercise 3.10. Which matrix equation is equivalent to the system $3x + 5y - z = 0, y - 2x = 5, x + 5y + 2z = 9$?

The map $A : \mathbb{R}^n \to \mathbb{R}^m$ has the important property that it respects addition (meaning that we can add two vectors and then apply the map or apply the map separately and then add the results and get the same outcome) and scalar multiplication (similarly), as demonstrated in the following theorem.

Theorem 3.11. *Let A be an $m \times n$ matrix, then the map defined in Definition 3.9 satisfies the two properties*

- $A(x + y) = Ax + Ay$ *for all* $x, y \in \mathbb{R}^n$,
- $A(\lambda x) = \lambda Ax$ *for all* $x \in \mathbb{R}^n$ *and* $\lambda \in \mathbb{R}$.

Proof. This is most easily shown using (3.1). Let us denote the components of the vector

$A(x+y)$ by z_i, $i=1,2,\cdots,m$, i.e., $z=A(x+y)$ with $z=(z_1,z_2,\cdots,z_m)$. Then by (3.1)

$$z_i=\sum_{j=1}^{n}a_{ij}(x_j+y_j)=\sum_{j=1}^{n}a_{ij}x_j+\sum_{j=1}^{n}a_{ij}y_j,$$

and on the right hand side we have the sum of the ith components of Ax and Ay, again by (3.1). The second assertion $A(\lambda x)=\lambda Ax$ follows again directly from (3.1) and is left as a simple exercise. □

Before we start solving systems of linear equations let us study matrices in some more detail.

3.2 Matrices

The most important property of matrices is that one can multiply them under suitable conditions on the number of rows and columns. The product of matrices appears naturally if we consider a vector $y=Ax$ and apply another matrix to it, i.e., $By=B(Ax)$. The question is then if there exist a matrix C such that

$$Cx=B(Ax),$$

then we would call $C=BA$ the matrix product of B and A. If we use equation (3.3) and Theorem 3.11 we obtain

$$\begin{aligned}B(Ax)&=B(x_1a_1+x_2a_2+\cdots+x_na_n)\\&=x_1Ba_1+x_2Ba_2+\cdots+x_nBa_n.\end{aligned}\tag{3.4}$$

Hence if C is the matrix with columns $Ba_1,\cdots,Ba_n$, then, again by (3.3), we have $Cx=B(Ax)$.

We formulate this now a bit more precisely:

Theorem 3.12. *Let $A=(a_{ij})\in M_{m,n}(\mathbb{R})$ and $B=(b_{ij})\in M_{l,m}(\mathbb{R})$ then there exists a matrix $C=(c_{ij})\in M_{l,n}(\mathbb{R})$ such that for all $x\in\mathbb{R}^n$ we have*

$$Cx=B(Ax)$$

and the elements of C are given by

$$c_{ij}=\sum_{k=1}^{m}b_{ik}a_{kj}.$$

Note that c_{ij} is the dot product between the ith row vector of B and the jth column vector of A. We call $C=BA$ the product of B and A.

The theorem follows from (3.4), but to provide a different perspective we are going to give another proof.

Proof. We write $y=Ax$ and note that $y=(y_1,y_2,\cdots,y_m)$ with

$$y_k=\sum_{j=1}^{n}a_{kj}x_j\tag{3.5}$$

and similarly we write $z=By$ and note that $z=(z_1,z_2,\cdots,z_l)$ with

$$z_i=\sum_{k=1}^{m}b_{ik}y_k.\tag{3.6}$$

Now inserting the expression (3.5) for y_k into (3.6) gives

$$z_i = \sum_{k=1}^{m} b_{ik} \sum_{j=1}^{n} a_{kj} x_j = \sum_{j=1}^{n} \sum_{k=1}^{m} b_{ik} a_{kj} x_j = \sum_{j=1}^{n} c_{ij} x_j,$$

where we have exchanged the order of summation. □

Note that in order to multiply to matrices A and B, the number of rows of A must be the same as the number of columns of B in order that BA can be formed.

Exercise 3.13. Let $A = \begin{pmatrix} 1 & 0 & -4 \\ -1 & 4 & 2 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 1 \\ -2 & 0 \end{pmatrix}$. Which of the products A^2, AB, BA, B^2 are defined?

If the matrices are of appropriate sizes so that the multiplication is defined, then matrix multiplication is distributive and associative.

Theorem 3.14. *Let A, B be $m \times n$ matrices and C an $l \times m$ matrix, then*

$$C(A+B) = CA + CB.$$

Let A, B be $m \times n$ matrices and C an $n \times l$ matrix, then

$$(A+B)C = AC + BC.$$

Let A be an $m \times n$ matrix, B be an $n \times l$ matrix and C a $l \times k$ matrix, then

$$A(BC) = (AB)C.$$

The proof of this theorem will be a simple consequence of general properties of linear maps which we will discuss in Chapter 7.

Now let us look at a few examples of matrices and products of them. We say that a matrix is a **square matrix** if $m = n$. If $A = (a_{ij})$ is a $n \times n$ square matrix, then we call the elements a_{ii} the **diagonal elements** of A and a_{ij} for $i \neq j$ the **off-diagonal** elements of A. A square matrix A is called a **diagonal matrix** if all off-diagonal elements are 0.

Example 3.15. The following is a 3×3 diagonal matrix

$$\begin{pmatrix} -2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

with diagonal elements $a_{11} = -2, a_{22} = 3$ and $a_{33} = 1$.

A special role is played by the so called **unit matrix** I, also known as **identity matrix**. This is a matrix with elements

$$\delta_{ij} := \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases},$$

i.e., a diagonal matrix with all diagonal elements equal to 1:

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

The symbol δ_{ij} is often called the **Kronecker delta**. If we want to specify the size of the unit matrix we write I_n for the $n \times n$ unit matrix (although often the size will be left implicit). The unit matrix is the matrix of the identity in multiplication, so for any $m \times n$ matrix A

$$AI_n = I_m A = A.$$

Let us now look at a couple of examples of products of matrices. We will start by looking at an example of multiplying a pair of 2×2 matrices.

Example 3.16. We have that

$$\begin{aligned} \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} -5 & 1 \\ 3 & -1 \end{pmatrix} &= \begin{pmatrix} 1 \times (-5) + 2 \times 3 & 1 \times 1 + 2 \times (-1) \\ -1 \times (-5) + 0 \times 3 & -1 \times 1 + 0 \times (-1) \end{pmatrix} \\ &= \begin{pmatrix} 1 & -1 \\ 5 & -1 \end{pmatrix}, \end{aligned}$$

where we have explicitly written out the intermediate step where we write each element of the product matrix as a dot product of a row vector of the first matrix and a column vector of the second matrix.

Exercise 3.17. Is matrix multiplication commutative? That is must we have $AB = BA$ for any two matrices A and B?

Click for solution

Example 3.18. Let us compute the product from Example 3.16 the other way round:

$$\begin{pmatrix} -5 & 1 \\ 3 & -1 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} -6 & -10 \\ 4 & 6 \end{pmatrix}$$

and we see that the result is different.

So contrary to the multiplication of numbers, **the product of matrices depends on the order in which we take the product.** In other words, matrix multiplication is not commutative in general, that is

$$AB \neq BA.$$

A few other interesting matrix products are noted below.

- It is possible for the product of two non-zero matrices to be a zero matrix, for example $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$. We use 0 to denote a zero matrix (again, the size of this should be clear from the context, or can be specified using subscript notation).

- Similarly, the square of a non-zero matrix can be 0, for example $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}^2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0$.
- Let $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ then $J^2 = -I$, i.e., the square of J is $-I$, very similar to $\mathrm{i} = \sqrt{-1}$.

Exercise 3.19. Let $A, B, C \in M_n(\mathbb{R})$ be non-zero matrices. Is it true that if $AB = AC$, then $B = C$?

Click for solution

No, this is not true in general. The first two bullet points above give a counterexample. However, if a matrix is invertible, a concept we will explore towards the end of the chapter, then this will hold.

These examples show that matrix multiplication behaves very different from multiplication of numbers which we are used to.

It is also instructive to look at products of matrices which are not square matrices. Recall that by definition we can only form the product of A and B, AB, **if the number of rows of B is equal to the number of columns of A.**

Example 3.20. Consider for instance the following matrices

$$A = \begin{pmatrix} 1 & -1 & 2 \end{pmatrix}, \quad B = \begin{pmatrix} 2 \\ 0 \\ 1 \end{pmatrix}, \quad C = \begin{pmatrix} 1 & 3 & 0 \\ -2 & 1 & 3 \end{pmatrix}, \quad D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

then A is 1×3 matrix, B a 3×1 matrix, C a 2×3 matrix and D a 2×2 matrix. So we can form the following products

$$AB = 4, \quad BA = \begin{pmatrix} 2 & -2 & 4 \\ 0 & 0 & 0 \\ 1 & -1 & 2 \end{pmatrix}, \quad CB = \begin{pmatrix} 2 \\ -1 \end{pmatrix}, \quad DC = \begin{pmatrix} -2 & 1 & 3 \\ 1 & 3 & 0 \end{pmatrix}, \quad D^2 = I$$

and no others. Notice that the product of A and B is the 1×1 matrix $\begin{pmatrix} 4 \end{pmatrix}$, but this is naturally identified with the scalar 4.

There are a few types of matrices which occur quite often and therefore have special names. We will give a list of some we will encounter:

- **Triangular matrices**: These come in two types,
  - upper triangular: $A = (a_{ij})$ with $a_{ij} = 0$ if $i > j$, e.g., $\begin{pmatrix} 1 & 3 & -1 \\ 0 & 2 & 1 \\ 0 & 0 & 3 \end{pmatrix}$
  - lower triangular: $A = (a_{ij})$ with $a_{ij} = 0$ if $i < j$, e.g., $\begin{pmatrix} 1 & 0 & 0 \\ 5 & 2 & 0 \\ 2 & -7 & 3 \end{pmatrix}$
- **Symmetric matrices**: $A = (a_{ij})$ with $a_{ij} = a_{ji}$, e.g., $\begin{pmatrix} 1 & 2 & 3 \\ 2 & -1 & 0 \\ 3 & 0 & 1 \end{pmatrix}$.

- **Anti- (or skew-)symmetric matrices**: $A = (a_{ij})$ with $a_{ij} = -a_{ji}$, e.g., $\begin{pmatrix} 0 & -1 & 2 \\ 1 & 0 & 3 \\ -2 & -3 & 0 \end{pmatrix}$.

The following operation on matrices occurs quite often in applications.

Definition 3.21 (Transpose)**.** Let $A = (a_{ij}) \in M_{m,n}(\mathbb{R})$ then the transpose of A, denoted A^t, is a matrix in $M_{n,m}(\mathbb{R})$ with elements $A^t = (a_{ji})$ (the indices i and j are switched). In other words, A^t is obtained from A by exchanging the rows with the columns.

Example 3.22. For the matrices A, B, C, D we considered in Example 3.20 we obtain

$$A^t = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}, \quad B^t = \begin{pmatrix} 2 & 0 & 1 \end{pmatrix}, \quad C^t = \begin{pmatrix} 1 & -2 \\ 3 & 1 \\ 0 & 3 \end{pmatrix}, \quad D^t = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

A matrix is symmetric if $A^t = A$ and anti-symmetric if $A^t = -A$. Any square matrix $A \in M_{n,n}(\mathbb{R})$ can be decomposed into a sum of a symmetric and an anti-symmetric matrix by

$$A = \frac{1}{2}(A + A^t) + \frac{1}{2}(A - A^t).$$

One of the reasons why the transpose is important is the following relation with the dot-product.

Theorem 3.23. *Let $A \in M_{m,n}(\mathbb{R})$, then we have for any $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$*

$$y \cdot Ax = (A^t y) \cdot x.$$

Proof. The ith component of Ax is $\sum_{j=1}^{n} a_{ij} x_j$ and so

$$y \cdot Ax = \sum_{i=1}^{m} \sum_{j=1}^{n} y_i a_{ij} x_j.$$

On the other hand the jth component of $A^t y$ is $\sum_{i=1}^{m} a_{ij} y_i$ and so

$$(A^t y) \cdot x = \sum_{j=1}^{n} \sum_{i=1}^{m} x_j a_{ij} y_i.$$

Since the order of summation does not matter in a finite double sum the two expressions agree. □

One important property of the transpose which can be derived from this relation is

Theorem 3.24. *Let $A \in M_{m,n}(\mathbb{R})$ and $B \in M_{n,k}(\mathbb{R})$. Then*

$$(AB)^t = B^t A^t$$

Proof. Using Theorem 3.23 for (AB) gives $\big((AB)^t y\big)\cdot x = y\cdot(ABx)$ and now we apply Theorem 3.23 first to A and then to B which gives $y\cdot(ABx) = (A^t y)\cdot(Bx) = (B^t A^t y)\cdot x$ and so we have

$$\big((AB)^t y\big)\cdot x = (B^t A^t y)\cdot x.$$

Since this is true for any x, y we have $(AB)^t = B^t A^t$. □

There is another connection between transposes and dot products. When we have a vector $x \in \mathbb{R}^n$ and wish to view it as a matrix, the standard convention is to view it as a column vector, that is as a $n\times 1$ matrix. With this convention, multiplying an $m\times n$ matrix A and x together gives the $m\times 1$ matrix $Ax \in \mathbb{R}^m$ as in Definition 3.9.

Moreover, if $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \in \mathbb{R}^n$, then

$$x\cdot y = x_1y_1 + \cdots x_ny_n = \begin{pmatrix} x_1 & \cdots & x_n \end{pmatrix}\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = x^t y.$$

So, for example, $y\cdot Ax = y^t Ax$. Beware that xy^t is very different: it is an $n\times n$ matrix

$$xy^t = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_n \end{pmatrix} = (x_iy_j) \in M_{n,n}(\mathbb{R}).$$

In this section we have primarily focused on multiplication of matrices, as this is our most important operation. However, before we move on it is worth mentioning that we can also add matrices and multiply them by scalars, both of which are just done componentwise.

Definition 3.25 (Matrix addition)**.** For matrices $A = (a_{ij})$ and $B = (b_{ij}) \in M_{m,n}(\mathbb{R})$ we have that $A + B = C \in M_{m,n}(\mathbb{R})$ with $c_{ij} = a_{ij} + b_{ij}$.

Definition 3.26 (Scalar multiplication of matrices)**.** For a matrix$A = (a_{ij}) \in M_{m,n}(\mathbb{R})$ and $\lambda \in \mathbb{R}$ we have that $\lambda A = C \in M_{m,n}(\mathbb{R})$ with $c_{ij} = \lambda a_{ij}$.

3.3 The structure of the set of solutions

In this section we will study the general structure of the set of solutions to a system of linear equations, when it has solutions at all. In the next section we will then look at methods to actually solve a system of linear equations.

Definition 3.27 (S(A,b))**.** Let $A \in M_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$, then we set

$$S(A,b) := \{x \in \mathbb{R}^n : Ax = b\}.$$

This is a subset of $\mathbb{R}^n$ and consists of all the solutions to the system of linear equations $Ax = b$. If there are no solutions then $S(A,b) = \emptyset$.

One often distinguishes between two types of systems of linear equations based on their constant terms.

Definition 3.28 (Homogeneous and inhomogenous)**.** The system of linear equations $Ax = b$ is called homogeneous if $b = \mathbf{0}$, i.e., if it is of the form

$$Ax = \mathbf{0}.$$

If $b \neq \mathbf{0}$ the system is called inhomogeneous.

If the system is inhomogeneous, then it doesn't necessarily have a solution. But for the ones which have a solution we can determine the structure of the set of solutions. The key observation is that if we have one solution, say $x_0 \in \mathbb{R}^n$ which satisfies $Ax_0 = b$, then we can create further solutions by adding solutions of the corresponding homogeneous system, $Ax = \mathbf{0}$, since if $Ax = \mathbf{0}$

$$A(x_0 + x) = Ax_0 + Ax = b + \mathbf{0} = b,$$

and so $x_0 + x$ is another solution to the inhomogeneous system.

Theorem 3.29. *Let $A \in M_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$ and assume there exists $x_0 \in \mathbb{R}^n$ with $Ax_0 = b$. Then*

$$S(A,b) = \{x_0\} + S(A,\mathbf{0}) := \{x_0 + x : x \in S(A,\mathbf{0})\}$$

Proof. As we noticed above, if $x \in S(A,\mathbf{0})$, then $A(x_0 + x) = b$, hence $\{x_0\} + S(A,\mathbf{0}) \subseteq S(A,b)$.

On the other hand, if $y \in S(A,b)$ then $A(y - x_0) = Ay - Ax_0 = b - b = \mathbf{0}$, and so $y - x_0 \in S(A,\mathbf{0})$. Therefore $S(A,b) \subseteq \{x_0\} + S(A,\mathbf{0})$ and so $S(A,b) = \{x_0\} + S(A,\mathbf{0})$. □

Remarks:

- The structure of the set of solutions is often described as follows: The general solution of the inhomogeneous system $Ax = b$ is given by a special solution x_0 to the inhomogeneous system plus a general solution to the corresponding homogeneous system $Ax = \mathbf{0}$.
- The case that there is unique solution to $Ax = b$ corresponds to $S(A,\mathbf{0}) = \{\mathbf{0}\}$, in which case $S(A,b) = \{x_0\}$.

At first sight the definition of the set $\{x_0\} + S(A,\mathbf{0})$ seems to depend on the choice of the particular solution x_0 to $Ax_0 = b$. But this is not so; another choice y_0 just corresponds to a different labelling of the elements of the set.

Example 3.30. Let us look at an example of three equations with three unknowns:

$$\begin{aligned} 3x + z &= 0, \\ y - z &= 1, \\ 3x + y &= 1. \end{aligned}$$

This set of equations corresponds to

$$A = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 1 & -1 \\ 3 & 1 & 0 \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}.$$

To solve this set of equations we try to simplify it: if we subtract the first equation from the third the third equation becomes $y - z = 1$ which is identical to the second equation. Hence the initial system of three equations is equivalent to the following system of two equations:

$$3x + z = 0, \quad y - z = 1.$$

In the first one we can solve for x as a function of z and in the second for y as a function of z, hence

$$x = -\frac{1}{3}z, \quad y = 1 + z. \tag{3.7}$$

So z is arbitrary, but once z is chosen, x and y are fixed, and the set of solutions is given by

$$S(A,b) = \{(-z/3, 1+z, z) : z \in \mathbb{R}\}.$$

A similar computation for the corresponding homogeneous system of equations

$$\begin{aligned} 3x + z &= 0, \\ y - z &= 0, \\ 3x + y &= 0 \end{aligned}$$

gives us the solutions $x = -z/3$, $y = z$, and $z \in \mathbb{R}$ arbitrary, hence

$$S(A,\mathbf{0}) = \{(-z/3, z, z) : z \in \mathbb{R}\}.$$

A particular solution to the inhomogeneous system is given by choosing $z = 0$ in (3.7), i.e., $x_0 = (0,1,0)$, and then the relation

$$S(A,b) = \{x_0\} + S(A,\mathbf{0})$$

can be seen directly, since for $x = (-z/3, z, z) \in S(A,\mathbf{0})$ we have $x_0 + x = (0,1,0) + (-z/3, z, z) = (-z/3, 1+z, z)$ which was the general form of an element in $S(A,b)$. But what happens if we choose another element of $S(A,b)$? Let $\lambda \in \mathbb{R}$, then $x_\lambda := (-\lambda/3, 1+\lambda, \lambda)$ is in $S(A,b)$ and we again have

$$S(A,b) = \{x_\lambda\} + S(A,\mathbf{0}),$$

since $x_\lambda + x = (-\lambda/3, 1+\lambda, \lambda) + (-z/3, z, z) = (-(\lambda+z)/3, 1+(\lambda+z), (\lambda+z))$. Then if z runs through $\mathbb{R}$ we again obtain the whole set $S(A,b)$, independent of which λ we chose initially. The choice of λ only determines the way in which we label the elements in $S(A,b)$.

Finally we should notice that the set $S(A,\mathbf{0})$ is spanned by one vector, namely we have $(-z/3, z, z) = z(-1/3, 1, 1)$ and hence with $v = (-1/3, 1, 1)$ we have $S(A,\mathbf{0}) = \operatorname{span}\{v\}$ and

$$S(A,b) = \{x_\lambda\} + \operatorname{span}\{v\}.$$

In the next section we will develop systematic methods to solve large systems of linear equations.

3.4 Solving systems of linear equations

To solve a system of linear equations we will introduce a systematic way to simplify it until we can read off directly if it is solvable and compute the solutions easily. Again it will be useful to write the system of equations in matrix form and then observe that operations on equations boil down to those on rows of the matrix. The procedure in question is often referred to as **Gaussian**

elimination.

Let us return for a moment to the original way of writing a set of m linear equations in n unknowns,

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1, \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2, \\ &\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m. \end{aligned}$$

We can perform the following operations on the set of equations without changing the solutions:

- multiply an equation by a non-zero constant,
- add a multiple of any equation to one of the other equations,
- exchange two of the equations.

It is clear that operations (i) and (iii) don't change the set of solutions, and to see that operation (ii) doesn't change the set of solutions we can argue as follows: If $x \in \mathbb{R}^n$ is a solution to the system of equations and we change the system by adding λ times equation i to equation j then x is clearly a solution of the new system, too. But if $x' \in \mathbb{R}^n$ is a solution to the new system we can return to the old system by subtracting λ times equation i from equation j, and so x' must be a solution to the old system, too. Hence both systems have the same set of solutions.

The way to solve a system of equations is to use the above operations to simplify a system of equations systematically until we can basically read off the solutions. It is useful to formulate this using the matrix representation of a system of linear equations

$$Ax = b.$$

Definition 3.31 (Augmented matrix)**.** Let $Ax = b$ be a system of linear equations, the augmented matrix associated with this system is

$$\begin{pmatrix} A & b \end{pmatrix}.$$

It is obtained by adding b as the final column to A, hence it is an $m \times (n+1)$ matrix if the system has n unknowns and m equations.

Now we translate the above operations on systems of equations into operations on the augmented matrix.

Definition 3.32 (Elementary row operations)**.** An elementary row operation (ERO) is one of the following operations on matrices:

- multiply a row by a non-zero number (row $i \to \lambda \times$ row$i$)
- add a multiple of one row to another row (row $i \to$ row $i + \lambda \times$ row$j$)
- exchange two rows (row $i \leftrightarrow$ row j)

Theorem 3.33. *Let $A \in M_{m,n}(\mathbb{R})$, $b \in \mathbb{R}^m$. If the augmented matrix $(A'\ b')$ is obtained from $(A\ b)$ by a sequence of ERO's, then the corresponding system of linear equation has the same solutions, i.e.,*

$$S(A,b) = S(A',b').$$

Proof. If we apply these operations to the augmented matrix of a system of linear equations then they clearly correspond to the three operations (i), (ii), and (iii) we introduced above, hence the system corresponding to the new matrix has the same set of solutions. □

We want to use these operations to systematically simplify the augmented matrix. Let us look at an example to get an idea which type of simplification we can achieve.

Example 3.34. Consider the following system of equations

$$\begin{aligned} x+y+2z &= 9, \\ 2x+4y-3z &= 1, \\ 3x+6y-5z &= 0. \end{aligned}$$

This is of the form $Ax = b$ with

$$A = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 4 & -3 \\ 3 & 6 & -5 \end{pmatrix} \quad b = \begin{pmatrix} 9 \\ 1 \\ 0 \end{pmatrix},$$

hence the corresponding augmented matrix is given by

$$\begin{pmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{pmatrix}.$$

Applying elementary row operations gives

$$\begin{aligned} \begin{pmatrix} 1 & 1 & 2 & 9 \\ 2 & 4 & -3 & 1 \\ 3 & 6 & -5 & 0 \end{pmatrix} &\xrightarrow{\text{row } 2-2\times\text{row } 1, \quad \text{row } 3-3\times\text{row } 1} \\ \begin{pmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 3 & -11 & -27 \end{pmatrix} &\xrightarrow{\text{row } 3-\text{row } 2} \\ \begin{pmatrix} 1 & 1 & 2 & 9 \\ 0 & 2 & -7 & -17 \\ 0 & 1 & -4 & -10 \end{pmatrix} &\xrightarrow{\text{row } 3\leftrightarrow\text{row } 2} \\ \begin{pmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -4 & -10 \\ 0 & 2 & -7 & -17 \end{pmatrix} &\xrightarrow{\text{row } 3-2\times\text{row } 2} \\ \begin{pmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -4 & -10 \\ 0 & 0 & 1 & 3 \end{pmatrix} & \end{aligned}$$

where we have written next to the matrix which elementary row operations we applied in order to arrive at the given line from the previous one. The system of equations corresponding to the last matrix is

$$\begin{aligned} x+y+2z &= 9, \\ y-4z &= -10, \\ z &= 3 \end{aligned}$$

so we have $z = 3$ from the last equation. Then substituting this in the second equation gives $y = -10+4z = -10+12 = 2$ and substituting this in the first equation gives $x = 9-y-2z = 9-2-6 = 1$. So we see that if the augmented matrix is in the above triangular like form we can solve the system of equations easily by what is sometimes called back-substitution.

But we can also continue applying elementary row operations and find

$$\begin{pmatrix} 1 & 1 & 2 & 9 \\ 0 & 1 & -4 & -10 \\ 0 & 0 & 1 & 3 \end{pmatrix} \xrightarrow{\text{row 1}-\text{row 2}}$$
$$\begin{pmatrix} 1 & 0 & 6 & 19 \\ 0 & 1 & -4 & -10 \\ 0 & 0 & 1 & 3 \end{pmatrix} \xrightarrow{\text{row 1}-6\times\text{row 3}, \quad \text{row 2}+4\times\text{row 3}}$$
$$\begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \end{pmatrix}.$$

Now the corresponding system of equations is of even simpler form

$$x = 1,$$
$$y = 2,$$
$$z = 3$$

and gives the solution directly.

The different forms into which we brought the matrix by elementary row operations are of a special type.

Definition 3.35 (Row echelon form and reduced row echelon form)**.** A matrix M is in row echelon form (REF) if

- each row is either all zeros, or the leftmost non-zero number is 1 (this is called the **leading** 1 in that row), and
- if row i is above row j, then the leading 1 of row i is to the left of row j; any zero rows are below any non-zero rows.

A matrix is in reduced row echelon form (RREF) if, in addition to (i) and (ii), it satisfies

- in each column which contains a leading 1, all other numbers are 0.

Example 3.36. The following matrices are in row echelon form

$$\begin{pmatrix} \mathbf{1} & 4 & 3 & 2 \\ 0 & \mathbf{1} & 6 & 2 \\ 0 & 0 & \mathbf{1} & 5 \end{pmatrix}, \qquad \begin{pmatrix} \mathbf{1} & 1 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & \mathbf{1} & 2 & 6 & 0 \\ 0 & 0 & \mathbf{1} & -1 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{1} \end{pmatrix}$$

and these ones are in reduced row echelon form:

$$\begin{pmatrix} \mathbf{1} & 0 & 0 \\ 0 & \mathbf{1} & 0 \\ 0 & 0 & \mathbf{1} \end{pmatrix}, \quad \begin{pmatrix} \mathbf{1} & 0 & 0 & 4 \\ 0 & \mathbf{1} & 0 & 7 \\ 0 & 0 & \mathbf{1} & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & \mathbf{1} & -2 & 0 & 1 \\ 0 & 0 & 0 & \mathbf{1} & 3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

In Example 3.36 we have marked the leading 1's in bold; we will see later that their distribution determines the nature of the solutions of the corresponding system of equations.

Exercise 3.37. Is the matrix $\begin{pmatrix} 1 & 0 & -1 & 2 \\ 0 & 1 & -1 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ in row echelon form? Is it in reduced row echelon form?

The reason for introducing these definitions is that elementary row operations can be used to bring any matrix to these forms.

Theorem 3.38. *Any matrix M can by a finite number of elementary row operations, called* ***Gaussian elimination****, be brought to*

- *row echelon form, or*
- *reduced row echelon form.*

Reduction to reduced row echelon form is sometimes called ***Gauss-Jordan elimination*** *to distinguish it from the first case.*

Proof. Let $M = (m_1, m_2, \cdots, m_n)$ where $m_i \in \mathbb{R}^m$ are the column vectors of M. Take the leftmost column vector which is non-zero, say this is m_j, and exchange rows until the first entry in that vector is non-zero, and divide the first row by that number. Now the matrix is $M' = (m'_1, m'_2, \cdots, m'_n)$ and the leftmost non-zero column vector is of the form

$$m'_j = \begin{pmatrix} 1 \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}.$$

Now we can subtract multiples of the first row from the other rows until all numbers in the jth column below the top 1 are 0; more precisely, we subtract from row i a_{ij} times the first row. We have transformed the matrix now to the form

$$\begin{pmatrix} 0 & 1 & \cdots \\ 0 & 0 & \tilde{M} \end{pmatrix}$$

and now we apply the same procedure to the matrix $\tilde{M}$. Eventually we arrive at row echelon form. To arrive at reduced row echelon form we start from row echelon form and use the leading 1's to clear out all non-zero elements in the columns containing a leading 1. □

Example 3.34 is an illustration on how the reduction to row echelon and reduced row echelon form works.

Let us now turn to the question what the row echelon form tells us about the structure of the set of solutions to a system of linear equations. The key information lies in the distribution of the leading 1's.

Theorem 3.39. *Let $Ax = b$ be a system of equations in n unknowns, and M be the row echelon form of the associated augmented matrix. Then*

- *the system has no solutions if and only if the last column of M contains a leading 1,*
- *the system has a unique solution if and only if every column except the last one of contains a leading 1,*
- *the system has infinitely many solutions if and only if the last column of M does not contain a leading 1 and there are less than n leading 1's. Then there $n-k$ unknowns which can chosen arbitrarily, where k is the number of leading 1' s of M.*

Proof. Let us first observe that the leading 1's of the reduced row echelon form of a system are the same as the leading 1's of the row echelon form. Therefore we can assume the system is in reduced row echelon form, which makes the arguments slightly simpler. Let us start with the last non-zero row, that is the row with the rightmost leading 1, and consider the corresponding equation. If the leading 1 is in the last column, then this equation is of the form

$$0x_1 + 0x_2 + \cdots + 0x_n = 1,$$

and so we have the contradiction $0 = 1$ and therefore there is no $x \in \mathbb{R}^n$ solving the set of equations. This is case (i) of the theorem.

If the last column does not contain a leading 1, but all other columns contain leading 1's then the reduced row echelon form is

$$\begin{pmatrix} 1 & 0 & \cdots & 0 & b'_1 \\ 0 & 1 & \cdots & 0 & b'_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & b'_n \\ & & \cdots & & \end{pmatrix}$$

and if $m > n$ there are $m-n$ rows with only 0's. The corresponding system of equations is then

$$x_1 = b'_1, \quad x_2 = b'_2, \quad \cdots \quad x_n = b'_n$$

and so there is a unique solution. This is case (ii) in the theorem.

Finally let us consider the case that there are k leading 1's with $k < n$ and none of them is in the last column. Let us index the column with leading 1's by $j_1, j_2, \cdots, j_k$. Then the system of equations corresponding to the reduced row echelon form is of the form

$$\begin{aligned} x_{j_1} + \sum_{i \text{ not leading}} a_{j_1 i} x_i &= b'_{j_1}, \\ x_{j_2} + \sum_{i \text{ not leading}} a_{j_2 i} x_i &= b'_{j_2}, \\ &\cdots \\ x_{j_k} + \sum_{i \text{ not leading}} a_{j_k i} x_i &= b'_{j_k} \end{aligned}$$

where the sums contain only those x_i whose index is not labelling a column with a leading 1. These are $n-k$ unknowns x_i whose value can be chosen arbitrarily and once their value is fixed, the remaining k unknowns are determined uniquely. This proves part (iii) of the theorem.

Note that in each case we have proved the 'only if' part of the statement, and the 'if' part follows from the fact that these three cases are mutually exclusive and considering the contrapositive of each statement. □

Exercise 3.40. The system with augmented matrix $\begin{pmatrix} 1 & 0 & 0 & 0 & 2 & 5 \\ 0 & 0 & 1 & 0 & 3 & 6 \\ 0 & 0 & 0 & 1 & 4 & 7 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$ has how many solutions?

Let us consider one simple consequence of this general theorem which we will use in a couple of proofs later on.

Exercise 3.41. Let $A \in M_{m,n}(\mathbb{R})$ and assume that $S(A,\mathbf{0}) = \{\mathbf{0}\}$, i.e., the only solution to $Ax = \mathbf{0}$ is $x = \mathbf{0}$. What can we conclude about the relationship between m and n?

Click for solution

This gives a rigorous basis to the intuitive idea that if you have n unknowns, you need at least n linear equations to determine the unknowns uniquely, and we state the result, which also holds for complex systems, formally below.

Corollary 3.42. *Let $A \in M_{m,n}(\mathbb{F})$ and assume that $S(A,\mathbf{0}) = \{\mathbf{0}\}$, i.e., the only solution $x \in \mathbb{F}^n$ to $Ax = \mathbf{0}$ is $x = \mathbf{0}$. Then $m \geq n$.*

Proof. This follows from part (ii) of the previous theorem. If every column has a leading one then there are at least as many rows as columns, i.e., $m \geq n$. □

3.5 Elementary matrices and inverting a matrix

We now want to discuss inverses of matrices in some more detail.

Definition 3.43 (Invertible and singular matrices)**.** A matrix $A \in M_{n,n}(\mathbb{R})$ is called invertible, or non-singular, if there exists a matrix $A^{-1} \in M_{n,n}(\mathbb{R})$ such that

$$A^{-1}A = I.$$

If A is not invertible then it is called singular.

We will first give some properties of inverses, namely that a left inverse is as well a right inverse, and that the inverse is unique. Note that these properties are direct consequences of the corresponding properties of linear maps, a concept we will explre in Chapter 7. (In fact, they also are implied by Theorem 3.53 below.) Recall that $M_n(\mathbb{R}) = M_{n,n}(\mathbb{R})$.

Theorem 3.44. *Let $A \in M_n(\mathbb{R})$ be invertible with inverse A^{-1}, then*

- $AA^{-1} = I$.
- *If $BA = I$ for some matrix $B \in M_n(\mathbb{R})$ then $B = A^{-1}$.*
- *If $B \in M_n(\mathbb{R})$ is also invertible, then AB is invertible with $(AB)^{-1} = B^{-1}A^{-1}$.*

We will prove this theorem later in the chapter, once we have introduced some further concepts which will help us do so.

The third property implies that arbitrary long products of invertible matrices are invertible too. The first property can can be interpreted as saying that A^{-1} is invertible too, and has the inverse A, i.e.,

$$(A^{-1})^{-1} = A.$$

[1] We will now turn to the question of how to compute the inverse of a matrix. This will involve similar techniques as for solving systems of linear equations, in particular the use of elementary row operations. The first step will be to show that elementary row operations can be performed using matrix multiplication. To this end we introduce a special type of matrix.

Definition 3.45. A matrix $E \in M_n(\mathbb{R})$ is called an elementary matrix if it is obtained from the identity matrix I_n by precisely one elementary row operation.

Example 3.46. - Switching rows in I_2 gives $E = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

- Multiplying row 1 by $\lambda \in \mathbb{R}$ in I_2 gives $E = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix}$.
- Adding row 2 to row 1 in I_2 gives $E = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$.
- Switching row 3 and row 5 in I_5 gives $\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}$.

Exercise 3.47. What is the elementary matrix corresponding to adding three times row 2 to row 3 in a 3×3 matrix?

Exercise 3.48. Are $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ elementary matrices?

The important property of elementary matrices is the following, which says that multiplying an arbitrary matrix A on the left by an elementary matrix E is the same as applying the row operation

[1] These results establish that the set of invertible $n \times n$ matrices forms a ‘group’ under multiplication — we define what this means exactly later in the course — which is called the general linear group over $\mathbb{R}^n$,

$$GL_n(\mathbb{R}) := \{A \in M_n(\mathbb{R}) : A \text{ is invertible}\}.$$

that was used to generate E from the identity to A. Before looking at this theorem, let's consider an example to see this.

Example 3.49. We consider the effect of multiplying a general 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by some of the elementary matrices from Example 3.46. We find

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} c & d \\ a & b \end{pmatrix}, \quad \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b \\ c & d \end{pmatrix}$$

and

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a+c & b+d \\ c & d \end{pmatrix},$$

which correspond indeed to the associated elementary row operations.

So we now state this as a general theorem:

Theorem 3.50. *Let $A \in M_{m,n}(\mathbb{R})$ and assume B is obtained from A by an elementary row operation with corresponding elementary matrix E. Then*

$$B = EA.$$

Proof. Let $c_1, \ldots, c_n \in \mathbb{R}^m$ be the columns of A, and $f_1^t, \ldots, f_m^t \in \mathbb{R}^m$ the rows of E, then the matrix B has rows $b_1^t, \cdots, b_m^t$ with

$$b_i^t = (f_i \cdot c_1 \;\; f_i \cdot c_2 \;\; \cdots \;\; f_i \cdot c_n).$$

Since E is an elementary matrix, there are only four possibilities for f_i:

- if the elementary row operation didn't change row i, then $f_i^t = e_i^t$ and $b_i^t = a_i^t$;
- if the elementary row operation exchanged row i and row j, then $f_i^t = e_j^t$ and $b_i^t = a_j^t$;
- if the elementary row operation multiplied row i by λ, then $f_i^t = \lambda e_i^t$ and $b_i^t = \lambda a_i^t$;
- if the elementary row operation added row j to row i then $f_i^t = e_i^t + e_j^t$ and $b_i^t = a_i^t + a_j^t$.

So we see that in all possible cases the multiplication of A by E has the same effect as applying an elementary row operation to A. □

Exercise 3.51. What can we deduce about the invertibility of elementary matrices?

Click for solution

Since any elementary row operation can be undone by other elementary row operations we immediately obtain the following

Corollary 3.52. *Any elementary matrix is invertible.*

Now let us see how we can use these elementary matrices. Assume we can find a sequence of N elementary row operations which transform a matrix A into the diagonal matrix I and let $E_1, \cdots, E_N$ be the elementary matrices associated with these elementary row operations, then repeated application of Theorem 3.50 gives $I = E_N \cdots E_2 E_1 A$, hence

$$A^{-1} = E_N \cdots E_2 E_1 .$$

So we have found a representation of A^{-1} as a product of elementary matrices, but we can simplify this even further. Since $E_N \cdots E_2 E_1 = E_N \cdots E_2 E_1 I$ we can again invoke Theorem 3.50 to conclude that A^{-1} is obtained by applying the sequence of elementary row operations to the identity matrix I. This means that we don't have to compute the elementary matrices, nor their product.

On the other hand, since A is an $n \times n$ matrix if the reduced row echelon form is not I it must have strictly less then n leading 1's, so Theorem 3.39 implies that the equation $Ax = 0$ has infinitely many solutions, hence A can't be invertible.

What we found is summarised in the following theorem.

Theorem 3.53. *Let $A \in M_n(\mathbb{R})$, if A can be transformed by successive elementary row operations into the identity matrix, then A is invertible and the inverse is obtained by applying the same sequence of elementary row operations to I.*

This leads to the following algorithm, known as the ***Gauss-Jordan*** *procedure:*

- *Form the $n \times 2n$ matrix $\begin{pmatrix} A & I \end{pmatrix}$ and apply elementary row transformations until A is in reduced row echelon form C, i.e, $\begin{pmatrix} A & I \end{pmatrix}$ is transformed to $\begin{pmatrix} C & B \end{pmatrix}$ for some matrix B.*
- *If the reduced row echelon form of A is I, i.e., $C = I$, then $B = A^{-1}$.*
- *If the reduced row echelon form is not I, then A is not invertible.*

Let us look at a few examples to see how this algorithm works.

Example 3.54. Let $A = \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix}$, then $\begin{pmatrix} A & I \end{pmatrix} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{pmatrix}$ and consecutive elementary row operations give

$$\begin{pmatrix} 0 & 2 & 1 & 0 \\ 1 & -1 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 & 1 \\ 0 & 2 & 1 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & -1 & 0 & 1 \\ 0 & 1 & 1/2 & 0 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 1/2 & 1 \\ 0 & 1 & 1/2 & 0 \end{pmatrix},$$

and so $A^{-1} = \begin{pmatrix} 1/2 & 1 \\ 1/2 & 0 \end{pmatrix}$.

Example 3.55. Let $A = \begin{pmatrix} 2 & -4 \\ -1 & 2 \end{pmatrix}$, then adding 2 times the second row to the first gives $\begin{pmatrix} 0 & 0 \\ -1 & 2 \end{pmatrix}$ and the reduced row echelon form of this matrix is $\begin{pmatrix} 1 & -2 \\ 0 & 0 \end{pmatrix}$ hence A is not invertible.

Example 3.56. Let $A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & 2 \\ 2 & 1 & -1 \end{pmatrix}$, then elementary row operations give

$$\begin{aligned}
\begin{pmatrix} 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 2 & 1 & -1 & 0 & 0 & 1 \end{pmatrix} &\to \begin{pmatrix} 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & -1 & -1 & 0 & 1 \end{pmatrix} \\
&\to \begin{pmatrix} 2 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -2 & 1 & 2 \\ 0 & 0 & -1 & -1 & 0 & 1 \end{pmatrix} \\
&\to \begin{pmatrix} 2 & 0 & 0 & 3 & -1 & -2 \\ 0 & 1 & 0 & -2 & 1 & 2 \\ 0 & 0 & -1 & -1 & 0 & 1 \end{pmatrix} \\
&\to \begin{pmatrix} 1 & 0 & 0 & 3/2 & -1/2 & -1 \\ 0 & 1 & 0 & -2 & 1 & 2 \\ 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix}
\end{aligned}$$

and so $A^{-1} = \begin{pmatrix} 3/2 & -1/2 & -1 \\ -2 & 1 & 2 \\ 1 & 0 & -1 \end{pmatrix}$.

For a general 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we find

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

which is well defined if

$$ad - bc \neq 0.$$

We can now sketch the proof of Theorem 3.44 above.

Proof of Theorem 3.44. To address (iii) we note that if A, B are both invertible, then

$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I,$$

so by definition of the inverse, $B^{-1}A^{-1}$ is the inverse of AB.

If A is invertible, we find elementary matrices so that $I = E_N \cdots E_2E_1A$, hence $A^{-1} = E_N \dots E_2E_1$. Moreover, it is clear that all E_i's are invertible (with inverses also elementary matrices). Multiplying $I = E_N \cdots E_2E_1A$ on the left by E_N^{-1}, then E_{N-1}^{-1}, and so on, one gets that

$$E_1^{-1}E_2^{-1} \cdots E_N^{-1} = E_1^{-1}E_2^{-1} \cdots E_N^{-1}E_N \cdots E_2E_1A = A.$$

So $AA^{-1} = E_1^{-1}E_2^{-1} \cdots E_N^{-1}E_N \cdots E_2E_1 = I$, giving (i).

Finally, if $I = BA$ then $IA^{-1} = BAA^{-1}$ so by (i) $A^{-1} = B$, giving (ii). □

Note that matrix inverses also provide another way of solving a system of equations. If we have a system $Ax = b$ where A is an invertible matrix then we have that $x = A^{-1}b$. As it is usually harder to invert the matrix than solve the system using Gaussian elimination we will not often use this method, but it can be useful in the case where you already know the inverse of the matrix.

Having explored matrices and systems of linear equations, we next look at a important function on matrices, which has links to invertibility of matrices and the number of solutions to a system of equations. This is the determinant function.

4. Determinants

When we computed the inverse of a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ we saw that it is invertible if $ad - bc \neq 0$. This combination of numbers has a name, it is called the determinant of A,

$$\det A = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} := ad - bc. \tag{4.1}$$

The determinant is a single number we can associate with a square matrix, and it is very useful, since many properties of the matrix are reflected in that number.

The determinant also has a geometric meaning: $|\det A|$ is the volume of the parallelogram (or higher-dimensional equivalent) with edges that correspond to the column vectors of A; in fact, $\det A$ can be thought of as a "signed area (or volume)". (For this reason determinants appear in multi-variable calculus.)

While determinants have many useful and simple properties, explicit formulas for determinants are more complicated. So in our treatment of determinants of $n \times n$ matrices for $n > 2$ we will use an axiomatic approach, i.e., we will single out a few properties of the determinant and use these to define what a determinant should be, and then derive other properties from them. While this approach is conceptually clear, it has a slight disadvantage by being rather abstract at the beginning, before we eventually arrive at some explicit formulas. But along the way we will encounter some key mathematical ideas which are of wider use.

4.1 Definition and basic properties

To begin with, we will use the axiomatic approach to define the determinant of a 2×2 matrix, and show that it gives the formula (4.1). We will generalise this to $n \times n$ matrices later.

We will write the determinant as a function of the column vectors of a matrix[1] which will take the column vectors as input and output a real number. Recall that for the 2×2 matrix $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ the two column vectors are $a_1 = \begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}$ and $a_2 = \begin{pmatrix} a_{12} \\ a_{22} \end{pmatrix}$.

Definition 4.1 (2-determinant function)**.** An 2-determinant function $d_2(a_1, a_2)$ is a function

$$d_2 : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R},$$

which satisfies the following three conditions:

- **Multilinearity**: The function is linear in each argument, that is
  - $d_2(\lambda a_1 + \mu b_1, a_2) = \lambda d_2(a_1, a_2) + \mu d_2(b_1, a_2)$ for all $\lambda, \mu \in \mathbb{R}$ and $a_1, a_2, b_1 \in \mathbb{R}^2$, and
  - $d_2(a_1, \lambda a_2 + \mu b_2) = \lambda d_2(a_1, a_2) + \mu d_2(a_1, b_2)$ for all $\lambda, \mu \in \mathbb{R}$ and $a_1, a_2, b_2 \in \mathbb{R}^2$.

[1] One could instead choose row vectors to define a determinant; both approaches give the same result. This is later expressed in the theorem that $\det A = \det A^t$.

- **Alternating**: The function is antisymmetric under exchange of arguments, so

$$d_2(a_2,a_1) = -d_2(a_1,a_2)$$

for all $a_1, a_2 \in \mathbb{R}^2$
- **Normalisation**: $d_2(e_1,e_2) = 1$.

These three conditions prescribe what happens to the determinant if we manipulate the columns of a matrix, e.g., (A) says that exchanging columns changes the sign. In particular we can rewrite (A) as

$$d_2(a_1,a_2) + d_2(a_2,a_1) = 0,$$

and so if $a_1 = a_2 = a$, then

$$d_2(a,a) = 0. \tag{4.2}$$

That means if the two columns in a matrix are equal, then the determinant is 0.

The first condition can be used to find out how a determinant function behaves under elementary column operations on the matrix[2]. Say if we multiply column 1 by λ, then

$$d_2(\lambda a_1, a_2) = \lambda d_2(a_1,a_2),$$

and if we add λ times column 2 to column 1 we get

$$d_2(a_1 + \lambda a_2, a_2) = d_2(a_1,a_2) + \lambda d_2(a_2,a_2) = d_2(a_1,a_2),$$

by (4.2).

Now let us see how much the conditions in the definition restrict the function d_2. If we write $a_1 = a_{11}e_1 + a_{21}e_2$ and $a_2 = a_{12}e_1 + a_{22}e_2$, then we can use multilinearity to obtain

$$\begin{aligned} d_2(a_1,a_2) &= d_2(a_{11}e_1 + a_{21}e_2, a_{12}e_1 + a_{22}e_2) \\ &= a_{11}d_2(e_1, a_{12}e_1 + a_{22}e_2) + a_{21}d_2(e_2, a_{12}e_1 + a_{22}e_2) \\ &= a_{11}a_{12}d_2(e_1,e_1) + a_{11}a_{22}d_2(e_1,e_2) + a_{21}a_{12}d_2(e_2,e_1) + a_{21}a_{22}d_2(e_2,e_2). \end{aligned}$$

This means that the function is completely determined by its values on the standard basis vectors e_i. Now (4.2) implies that

$$d_2(e_1,e_1) = d_2(e_2,e_2) = 0,$$

and by antisymmetry $d_2(e_2,e_1) = -d_2(e_1,e_2)$, hence

$$d_2(a_1,a_2) = (a_{11}a_{22} - a_{21}a_{12})d_2(e_1,e_2).$$

Finally the normalisation $d_2(e_1,e_2) = 1$ means that d_2 is actually uniquely determined and

$$d_2(a_1,a_2) = a_{11}a_{22} - a_{21}a_{12} = \det A,$$

as it was defined by formula (4.1). Let us also note that if we just invoke the axioms (ML), (A), we get

$$d_2(a_1,a_2) = \det A \cdot d_2(e_1,e_2),$$

that such is a function d_2, without the (N) axiom equals $\det A$ times a number which is $d_2(e_1,e_2)$, its value on the identity matrix.

[2] Elementary column operations are defined the same way as elementary row operations. They can also be viewed as matrix multiplication by elementary matrices, but on the right, rather than on the left.

But with all the three axioms in play, there is only **one** determinant function d_2 satisfying the three properties (ML), (A), (N), and it coincides with the expression (4.1), i.e.,

$$d_2(a_1,a_2) = a_{11}a_{22} - a_{21}a_{12} = \det\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}.$$

The determinant was originally not introduced this way, it emerged from the study of systems of linear equation as a combination of coefficients which seemed to be indicative of solvability.

The conditions in the definition probably seem to be a bit ad hoc. If we think of $d_2(a,b)$ as a "signed area" they seem less strange: consider the area of the parallelogram with sides a,b, with sign $+$ if the direction of b is obtained from that of a by rotating up to π radians anti-clockwise, and with sign $-$ otherwise. So for e_1,e_2 we get a signed area of 1 (this is property (N)), while for $e_1,-e_2$ we get a signed area of -1. Swapping a,b to b,a switches the sign of the signed area, giving (A). Finally, on drawing a picture it is easy to see that a,b and $a,b+\lambda a$ give the same signed areas, and (ML) can be shown similarly.

We extend the definition now to $n \times n$ matrices:

Definition 4.2 (n-determinant function)**.** An n-determinant $d_n(a_1,a_2,\cdots,a_n)$ is a function

$$d_n : \underbrace{\mathbb{R}^n \times \mathbb{R}^n \times \cdots \times \mathbb{R}^n}_{n \text{ times}} \to \mathbb{R},$$

which satisfies

- **Multilinearity**: The function is linear in each column, that is for any j and any $a_j, b_j \in \mathbb{R}^n$, $\lambda, \mu \in \mathbb{R}$

$$d_n(\cdots,\lambda a_j + \mu b_j,\cdots) = \lambda d_n(\cdots,a_j,\cdots) + \mu d_n(\cdots,b_j,\cdots),$$

  where the $\cdots$ mean the other $n-1$ vectors stay fixed.
- **Alternating**: The function is antisymmetric in each pair of arguments, that is whenever we exchange two vectors we pick up a factor -1, so if $i \neq j$ then

$$d_n(\cdots,a_i,\cdots,a_j,\cdots) = -d_n(\cdots,a_j,\cdots,a_i,\cdots).$$

- **Normalisation**: $d_n(e_1,e_2,\cdots,e_n) = 1$.

We will sometimes call these three properties the axioms of the determinant. We have formulated the determinant function as function of vectors; to connect it to matrices we take these vectors to be the column vectors of a matrix. The properties (ML) and (A) then correspond to column operations in the same way as we discussed after the definition of a 2-determinant. The property (N) means that the unit matrix has determinant 1.

Before proceeding to the proof that there is only one n-determinant let us consider some properties of the determinant.

Exercise 4.3. Let $A \in M_n(\mathbb{R})$ and $\lambda \in \mathbb{R}$. Does $d_n(\lambda A) = \lambda d_n(A)$?

We next consider a generalisation of (4.2).

Proposition 4.4. *Let $d_n(a_1, a_2, \cdots, a_n)$ be an n-determinant, then*

(i) whenever one of the vectors $a_1, a_2, \cdots, a_n$ is $\mathbf{0}$ then

$$d_n(a_1, a_2, \cdots, a_n) = 0.$$

(ii) whenever two of the vectors $a_1, a_2, \cdots, a_n$ are equal, then

$$d_n(a_1, a_2, \cdots, a_n) = 0.$$

Proof.

(i) We use multilinearity. We have for any a_j that $d_n(\cdots, \lambda a_j, \cdots) = \lambda d_n(\cdots, a_j, \cdots)$ for any $\lambda \in \mathbb{R}$, and setting $\lambda = 0$ gives $d_n(\cdots, \mathbf{0}, \cdots) = 0$.

(ii) We rewrite condition (A) in the definition as

$$d_n(\cdots, a_i, \cdots, a_j, \cdots) + d_n(\cdots, a_j, \cdots, a_i, \cdots) = 0,$$

and so if $a_i = a_j$, then $2d_n(\cdots, a_i, \cdots, a_j, \cdots) = 0$.

□

As a direct consequence we obtain the following useful property: we can add to a column any multiple of one of the other columns without changing the value of the determinant function.

Corollary 4.5. *We have for any $j \neq i$ and $\lambda \in \mathbb{R}$ that*

$$d_n(a_1, \cdots, a_i + \lambda a_j, \cdots a_n) = d_n(a_1, \cdots, a_i, \cdots a_n).$$

Proof. By multilinearity we have

$$d_n(a_1, \cdots, a_i + \lambda a_j, \cdots a_n) = d_n(a_1, \cdots, a_i, \cdots a_n) + \lambda d_n(a_1, \cdots, a_j, \cdots a_n),$$

but in the second term two of the vectors in the arguments are the same, hence the term is 0. □

Exercise 4.6. What effect will the elementary column operations

(i) swapping two columns,
(ii) multiplying a column by $\lambda \in \mathbb{R}$,
(iii) adding a multiple of one column to another

have on the determinant of a matrix?

Click for solution

We have that

(i) alters the sign of the value of d_n,
(ii) multiplies the determinant by λ, and
(iii) does not checange the determinant.

We can now compute the determinant function for some special classes of matrices.

Example 4.7. For diagonal matrices, i.e, $A = (a_{ij})$ with $a_{ij} = 0$ if $i \neq j$, the columns are $a_1 = a_{11}e_1, a_2 = a_{22}e_2, \cdots, a_n = a_{nn}e_n$ and using multilinearity in each argument and normalisation we get

$$\begin{aligned} d_n(a_{11}e_1, a_{22}e_2, \cdots, a_{nn}e_n) &= a_{11}d_n(e_1, a_{22}e_2, \cdots, a_{nn}e_n) \\ &= a_{11}a_{22}d_n(e_1, e_2, \cdots, a_{nn}e_n) \\ &= a_{11}a_{22}\cdots a_{nn}d_n(e_1, e_2, \cdots, e_n) = a_{11}a_{22}\cdots a_{nn} \end{aligned} \tag{4.3}$$

Example 4.8. Using the properties of a determinant function we know by now we can actually already compute them, though not in a very efficient way. If we repeat what we've done for d_2 in the case of 3 vectors in $\mathbb{R}^3$ we get the slightly cumbersome formula

$$\det\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31} - a_{33}a_{12}a_{21}, \tag{4.4}$$

but we'll see an easier way to find such determinants later (that works for $n > 3$ too).

We can now show that there exist only one n-determinant function.

Theorem 4.9. *There exists one, and only one, n-determinant function.*

Non-examinable. We only give a sketch of the proof. Let us expand the vectors a_j, $j = 1, \dots, n$, by writing them as a sum of multiples of the vectors $e_1, \dots, e_n$[a]:

$$a_j = \sum_{i=1}^{n} a_{ij}e_i = \sum_{i_j=1}^{n} a_{i_j j}e_{i_j}.$$

The reason for the second expression is bookkeeping: we will have to use different indices i_j in the expansion for every vector a_j involved. So the second subscript j is just indicating that this is the vector e_i in the expansion for a_j.

Insert these expansions into $d_n(a_1, a_2, \cdots, a_n)$. Doing this for a_1 and using multilinearity gives

$$d_n(a_1, a_2, \cdots, a_n) = d_n\left(\sum_{i_1=1}^{n} a_{i_1 1}e_{i_1}, a_2, \cdots, a_n\right) = \sum_{i_1=1}^{n} a_{i_1 1}d_n(e_{i_1}, a_2, \cdots, a_n).$$

Repeating the same step for a_2, a_3, etc., gives us

$$d_n(a_1, a_2, \cdots, a_n) = \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_n=1}^{n} a_{i_1 1}a_{i_2 2}\cdots a_{i_n n}d_n(e_{i_1}, e_{i_2}, \cdots, e_{i_n}). \tag{4.5}$$

This formula tells us that the function d_n is determined by its value on the vectors $e_1, \dots, e_n$. There are n^n choices of ordered n-tuples $(e_{i_1}, e_{i_2}, \cdots, e_{i_n})$. By Proposition 4.4 whenever at least two of the vectors $e_{i_1}, e_{i_2}, \cdots, e_{i_n}$ are equal then $d_n(e_{i_1}, e_{i_2}, \cdots, e_{i_n}) = 0$. Hence, there are only at most $n!$ non-zero terms in the sum coming from the number of different ways to reorder the indices $1, 2, \ldots, n$.

If the vectors $e_{j_1}, \ldots, e_{j_n}$ are all different, we can swap pairs of vectors and after finitely many steps we can reach $e_1, \ldots, e_n$, picking up a $-$ sign each time by (A), so if there are k swaps we get $d_n(e_{j_1}, \ldots, e_{j_n}) = (-1)^k d_n(e_1, \ldots, e_n) = (-1)^k$ by (N). So we have determined the unique value of the sum in (4.5).

What we haven't shown is existence: perhaps no such function d_n can be defined. We don't prove this here as it uses a little group theory, namely permutations. The rearrangement of $e_{j_1}, \ldots, e_{j_n}$ to $e_1, \ldots, e_n$ is a permutation σ of the set $\{1, \ldots, n\}$. While the number of swaps k is not well-defined, its **sign**, which is the function $\operatorname{sign}(\sigma) = (-1)^k$, is well-defined, and we can define the determinant using the **Leibniz formula**:

$$d_n(a_1, a_2, \cdots, a_n) := \sum_{\text{permutations } \sigma} \operatorname{sign}\sigma a_{\sigma(1)1} a_{\sigma(2)2} \cdots a_{\sigma(n)n} .$$

Using group theory we can show that this function satisfies the three axioms. □

[a] This is known as a linear combination, an idea we will explore further in Chapters 5 and 6.

Given this result, we define the determinant of a matrix by applying d_n to its column vectors $a_1, \ldots, a_n$, so

$$\det A := d_n(a_1, a_2, \cdots, a_n). \tag{4.6}$$

From now on we have two equivalent notations for the determinant function: det and d_n. We will mostly use the former one, as far as square matrices are concerned, but may return to d_n when we view the determinant as the function of the columns of the matrix.

In fact, the argument from the proof of Theorem 4.9 shows more than just the uniqueness of the determinant. It also shows the more general theorem below, which tells us that if we have a function which is multilinear and alternating then it it will scale the determinant function based on the value of the function on the identity matrix.

Theorem 4.10. *Let f_n be a real-valued function of n vectors $a_1, a_2, \cdots, a_n \in \mathbb{R}^n$, satisfying just (ML) and (A). Then*

$$f_n(a_1, a_2, \cdots, a_n) = C \cdot \det A,$$

for some constant $C = f_n(e_1, \cdots, e_n)$ and A being the matrix whose columns are $a_1, a_2, \cdots, a_n$.

Let us now continue with computing some determinants. We learned in (4.3) that the determinant of a diagonal matrix is just the product of the diagonal elements. The same is true for upper triangular matrices.

Theorem 4.11. *Let $A = (a_{ij}) \in M_n(\mathbb{R})$ be upper triangular, i.e., $a_{ij} = 0$ if $i > j$, and let $a_1, a_2, \cdots, a_n$ be the column vectors of A. Then we have*

$$\det A = a_{11} a_{22} \cdots a_{nn},$$

i.e., the determinant is the product of the diagonal elements.

Proof. Let us first assume that all the diagonal elements a_{ii} are nonzero. The matrix A is of the form

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

In the first step we subtract multiples of column 1 from the other columns to remove the entries in the first row, so $a_2 - a_{12}/a_{11}a_1$, $a_3 - a_{13}/a_{11}a_1$, etc. By Corollary 4.5 these operations do not change the determinant and hence we have

$$\det A = \det \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

In the next step we repeat the same procedure with the second row, i.e., subtract suitable multiples of the second column from the other columns, and then we continue with the third row, etc. At the end we arrive at a diagonal matrix and then by (4.3)

$$\det A = \det \begin{pmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{pmatrix} = a_{11}a_{22}a_{33}\cdots a_{nn}.$$

If one of the diagonal matrix elements is 0, then we can follow the same procedure until we arrive at the first column where the diagonal element is 0. But this column will be entirely 0 then and so by Proposition 4.4 the determinant is 0. □

Example 4.12. We have that $\det \begin{pmatrix} 1 & 4 & 7 \\ 0 & 1 & 3 \\ 0 & 0 & -2 \end{pmatrix} = -2$.

Let us now collect a few important properties of the determinant.

Theorem 4.13. *Let A and B be $n \times n$ matrices, then*

$$\det(AB) = \det(A)\det(B).$$

Proof. The theorem follows from Theorem 4.10. Let $B = (b_1 \ \ldots \ b_n)$. The columns of AB are $Ab_1, \ldots, Ab_n$. So $\det(AB) = d_n(Ab_1, \ldots, Ab_n)$, where the proven to be unique function d_n satisfies the axioms (ML), (A), and (N).

Now consider $\det(AB)$ as a function of just $b_1, \ldots, b_n$, that is $\det(AB) = f_n(b_1, \ldots, b_n)$, for some function f_n. This f_n satisfies (ML), because multiplication on the left by A is linear in b_j, as well as (A), for swapping b_i and b_j we swap Ab_i and Ab_j. So by Theorem 4.10 the value of this function equals $\det B$ times some constant C, where C is defined by equating

$b_j = e_j$, for $j = 1, \ldots, n$. But in the latter case we get $Ae_1 = a_1, \ldots, Ae_n = a_n$. In other words, if $b_j = e_j$, for $j = 1, \ldots, n$, then $\det(AB) = \det(AI) = \det A$. Hence, $C = \det A$, and $\det(AB) = \det(A)\det(B)$. $\square$

This multiplicative property of determinants is very important. Note that generally $AB \neq BA$, however according to this theorem we do have that $\det(AB) = \det(BA) = \det(A)\det(B)$.

Exercise 4.14. Is the determinant linear? Must we have $\det(A+B) \neq \det A + \det B$?

Click for solution

The determinant is **not** linear, so in general

$$\det(A+B) \neq \det A + \det B.$$

For example if $A = B = I_2$ then $\det A = \det B = 1$ but $\det(A+B) = 4$.

One of the consequences of Theorem 4.13 is that if A is invertible, i.e., there exists an A^{-1} such that $A^{-1}A = I$, then $\det A^{-1} \det A = 1$, and hence $\det A \neq 0$ and

$$\det A^{-1} = \frac{1}{\det A}.$$

So if A is invertible, then $\det A \neq 0$. This is an important result, and one we will return to later in Chapter 6.

Finally, we have the following theorem, which we will not prove.

Theorem 4.15. *Let A be an $n \times n$ matrix, then*

$$\det A = \det A^t.$$

Let us comment on the meaning of this result. We defined the determinant of a matrix in two steps, we first defined the determinant function $d_n(a_1, a_2, \cdots, a_n)$ as a function of n vectors, and then we related it to a matrix A by choosing for $a_1, a_2, \cdots, a_n$ the column vectors of A. We could have instead chosen the row vectors of A, that would have been an alternative definition of a determinant. The theorem tells us that both ways we get the same result.

Properties (ML) and (A) from the basic Definition 4.2 tell us what happens to determinants if we manipulate the columns by linear operations, in particular they tell us what happens if we apply elementary column operations to the matrix. But using $\det A^t = \det A$ we get the same properties for elementary row operations, as the following theorem states.

Theorem 4.16. *Let A be an $n \times n$ matrix, then we have:*

- *If A' is obtained from A by exchanging two rows, then* $\det A' = -\det A$ *and if E is the elementary matrix corresponding to the row exchange, then* $\det E = -1$.
- *If A' is obtained from A by adding λ times row j to row i, ($i \neq j$), then* $\det A = \det A'$ *and the corresponding elementary matrix satisfies* $\det E = 1$.
- *If A' is obtained from A by multiplying row i by $\lambda \in \mathbb{R}$, then* $\det A' = \lambda \det A$ *and the corresponding elementary matrix satisfies* $\det E = \lambda$.

An interesting consequence of this result is that it shows, independently from Theorem 4.13, that $\det EA = \det E \det A$ for any elementary matrix. We see that by just computing the left and the right hand sides for each case in Theorem 4.16. This observation can be used to give a different proof of the multiplicative property $\det(AB) = \det A \det B$; the main idea is to write A as a product of elementary matrices, which turns out to be possible if A is non-singular, and then use that we have the multiplicative property for elementary matrices.

Similarly, the results of Proposition 4.4 is true for rows as well.

4.2 Computing determinants

We know already how to compute the determinants of general 2×2 matrices, and for 3×3 one can use formula (4.4) which is just the Leibniz formula with 6 terms, corresponding to the $3! = 6$ elements of S_3. Let us look at determinants of larger matrices. There is a convention to denote the determinant of a matrix by replacing the brackets by vertical bars, for example

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} := \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

We will now discuss some systematic methods to compute determinants. The first is Laplace expansion. As a preparation we need some definitions.

Definition 4.17 (Minors of a matrix)**.** Let $A \in M_n(\mathbb{R})$, then we define the following:

- $\hat{A}_{ij} \in M_{n-1}(\mathbb{R})$ is the matrix obtained from A by removing row i and column j.
- $\det \hat{A}_{ij}$ is called the minor associated with a_{ij}.
- $A_{ij} := (-1)^{i+j} \det \hat{A}_{ij}$ is called the signed minor associated with a_{ij}.

Example 4.18. Let $A = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 3 \\ 0 & 1 & 0 \end{pmatrix}$. Then $\hat{A}_{11} = \begin{pmatrix} 1 & 3 \\ 1 & 0 \end{pmatrix}$, and $A_{11} = -3$.

Exercise 4.19. Find $\hat{A}_{12}, \hat{A}_{13}$, and $\hat{A}_{32}$, and the corresponding signed minors.

Click for solution

We have$\hat{A}_{12} = \begin{pmatrix} 2 & 3 \\ 0 & 0 \end{pmatrix}$, $\hat{A}_{13} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$, $\hat{A}_{32} = \begin{pmatrix} 1 & -1 \\ 2 & 3 \end{pmatrix}$ and so on. For the signed minors we find that $A_{12} = 0$, $A_{13} = 2$, $A_{32} = -5$.

Example 4.20. Let $A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then $\hat{A}_{11} = 4$, $\hat{A}_{12} = 3$, $\hat{A}_{21} = 2$, $\hat{A}_{22} = 1$ (where these 1×1 matrices are identified with the corresponding real number as usual) and $A_{11} = 4$, $A_{12} = -3$, $A_{21} = -2$, $A_{22} = 1$.

Practically, determinants are usually calculated recursively, using the so-called **Laplace (or cofactor) expansion**.

Theorem 4.21 (Laplace expansion). *Let $A \in M_n(\mathbb{R})$, then we can rewrite the determinant of A in the following ways:*

- *For any row $(a_{i1}, a_{i2}, \cdots, a_{in})$ we have*

$$\det A = a_{i1}A_{i1} + a_{i2}A_{i2} + \cdots + a_{in}A_{in} = \sum_{j=1}^{n} a_{ij}A_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det \hat{A}_{ij}.$$

  This is known as ***expanding by row*** *i.*

- *For any column $\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$ we have*

$$\det A = a_{1j}A_{1j} + a_{2j}A_{2j} + \cdots + a_{nj}A_{nj} = \sum_{i=1}^{n} a_{ij}A_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det \hat{A}_{ij}.$$

  This is known as ***expanding by column*** *j.*

We will discuss the proof of this result later, but let us first see how we can use it. The main idea is that Laplace expansion gives an expression for a determinant of an $n \times n$ matrix as a sum over n determinants of smaller $(n-1) \times (n-1)$ matrices, and so we can iterate this, the determinants of $(n-1) \times (n-1)$ matrices can then be expressed in terms of determinants of $(n-2) \times (n-2)$ matrices, and so on, until we arrive at, say 2×2 matrices whose determinants we can compute directly.

Now let us look at some examples to see how to use this result. It is useful to visualise the sign-factors $(-1)^{i+j}$ by looking at the corresponding matrix

$$\begin{pmatrix} + & - & + & - & \cdots \\ - & + & - & + & \cdots \\ + & - & + & - & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

which has a chess board pattern of alternating $+$ and $-$ signs. So if, for instance, we want to expand A into the second row we get

$$\det A = -a_{21} \det \hat{A}_{21} + a_{22} \det \hat{A}_{22} - a_{23} \det \hat{A}_{23} + \cdots$$

and the pattern of signs in front of the terms is the same as the second row of the above sign-matrix.

Example 4.22. Let $A = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 1 & 3 \\ 0 & 1 & 0 \end{pmatrix}$. We can find the determinant using Laplace expansion in different ways, choosing a row or column to expand by.

- Expansion in the first row gives:

$$\begin{vmatrix} 1 & 0 & -1 \\ 2 & 1 & 3 \\ 0 & 1 & 0 \end{vmatrix} = 1 \begin{vmatrix} 1 & 3 \\ 1 & 0 \end{vmatrix} - 0 \begin{vmatrix} 2 & 3 \\ 0 & 0 \end{vmatrix} + (-1) \begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} = -3 - 2 = -5.$$

- Expansion in the last column gives:

$$\begin{vmatrix} 1 & 0 & -1 \\ 2 & 1 & 3 \\ 0 & 1 & 0 \end{vmatrix} = (-1)\begin{vmatrix} 2 & 1 \\ 0 & 1 \end{vmatrix} - 3\begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} + 0\begin{vmatrix} 1 & 0 \\ 2 & 1 \end{vmatrix} = -2-3 = -5.$$

- Expansion in the last row gives:

$$\begin{vmatrix} 1 & 0 & -1 \\ 2 & 1 & 3 \\ 0 & 1 & 0 \end{vmatrix} = -\begin{vmatrix} 1 & -1 \\ 2 & 3 \end{vmatrix} = -5$$

(note that in the second step we omitted the terms where $a_{3j} = 0$).

Note that regardless of which row or column we choose to expand we obtain the same result.

Example 4.23. Let $A = \begin{pmatrix} 2 & 3 & 7 & 0 & 1 \\ -2 & 0 & 3 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ -10 & 1 & 0 & -1 & 3 \\ 0 & 2 & -2 & 0 & 0 \end{pmatrix}$. We expand in the 3rd row, then the 2nd row, then 2nd column (omitting writing any terms which would be multiplied by a minor of 0), so

$$\det A = \begin{vmatrix} 2 & 3 & 0 & 1 \\ -2 & 0 & 0 & 0 \\ -10 & 1 & -1 & 3 \\ 0 & 2 & 0 & 0 \end{vmatrix} = -(-2)\begin{vmatrix} 3 & 0 & 1 \\ 1 & -1 & 3 \\ 2 & 0 & 0 \end{vmatrix} = -2\begin{vmatrix} 3 & 1 \\ 2 & 0 \end{vmatrix} = 4.$$

This expansion works similarly for larger matrices, but it becomes rather long. As the example shows, one can use the freedom of choice of rows or columns for the expansion to chose one which contains as many 0's as possible; this reduces the computational work one has to do.

We showed already in Theorem 4.11 that determinants of triangular matrices are easy to calculate. Let us derive this as well from using Laplace expansion.

Proposition 4.24. *Let $A \in M_n(\mathbb{R})$.*

(i) If A is upper triangular, i.e., $a_{ij} = 0$ if $i > j$, then

$$\det A = a_{11}a_{22}\cdots a_{nn}.$$

(ii) If A is lower triangular, i.e., $a_{ij} = 0$ if $i < j$, then

$$\det A = a_{11}a_{22}\cdots a_{nn}.$$

Proof. We will only prove (i); part (ii) will be left as exercise. Since A is upper triangular its first column is $\begin{pmatrix} a_{11} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, hence expanding by that column gives $\det A = a_{11}A_{11}$. But $\hat{A}_{11}$ is again upper triangular with first column $\begin{pmatrix} a_{22} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ and so iterating this argument gives $\det A = a_{11}a_{22}\cdots a_{nn}$. □

This implies that a triangular matrix is invertible if, and only if, all its diagonal elements are non-zero.

This result will be the starting point for the second method we can use to compute determinants. Whenever we have a triangular matrix we can compute the determinant easily. In Section 7.1 we discussed how elementary row or column operations affected the determinant. So combining this with this result about determinants of triangular matrices gives us the following strategy: First use elementary row or column operations to bring a matrix to triangular form (which can always be done, as a consequence of Theorem 3.38), and then use the above result to compute the determinant of that triangular matrix.

When applying this strategy, we need to keep track of how each operation will change the determinant. Recall that:

(i) adding multiples of one column (or row) to another does not change the determinant.
(ii) multiplying a column (or row) by a real number λ will multiply the determinant by λ also. Hence to find the determinant of the original matrix we need to multiply by a factor of $\frac{1}{\lambda}$.
(iii) switching two columns (or two rows) changes the sign of the determinant, so we must multiply by a factor of -1.

Example 4.25. We have that:

- $\det \begin{pmatrix} 2 & 0 & 3 \\ -1 & 2 & 0 \\ 2 & 0 & 0 \end{pmatrix} = -\det \begin{pmatrix} 3 & 0 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 2 \end{pmatrix} = -12.$
  Here we have switched columns 1 and 3, so must multiply the determinant by -1.
- $\det \begin{pmatrix} 3 & 2 & 1 \\ 2 & 2 & 1 \\ 2 & 1 & 1 \end{pmatrix} = \det \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix} = 1.$
  Here we have subtracted column 3 from column 2 and 2 times column 3 from column 1, neither of which will change the determinant.
- $$\det \begin{pmatrix} 1 & 0 & 3 & 4 \\ 0 & 2 & 1 & -1 \\ 0 & 2 & -1 & 0 \\ 2 & 1 & 1 & -1 \end{pmatrix} = \det \begin{pmatrix} 9 & 4 & 7 & 4 \\ -2 & 1 & 0 & -1 \\ 0 & 2 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = \det \begin{pmatrix} 9 & 18 & 7 & 4 \\ -2 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$
  $$= \det \begin{pmatrix} 45 & 18 & 7 & 4 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} = 45.$$
  In the first step we used column 4 to remove all non-zero entries in row 4 except the last.

Then we used column 3 to simplify column 2 and finally we used column 2 to simplify column 1. None of these operations change the determinant.

- $\begin{vmatrix} 1/2 & 3/2 \\ 2 & 0 \end{vmatrix} = \frac{1}{2}\begin{vmatrix} 1 & 3 \\ 2 & 0 \end{vmatrix} = -\frac{1}{2}\begin{vmatrix} 2 & 0 \\ 1 & 3 \end{vmatrix} = -\frac{1}{2} \times 2 \times 3 = -3$. Here we have multiplied row 1 by 2, so must balance this with a factor of $\frac{1}{2}$, and then have switched rows 1 and 2, which introduces a factor of -1.

We haven't yet given a proof of the Laplace expansion formulas. We will sketch one now.

Proof of Theorem 4.21, Non-examinable. Since $\det A = \det A^t$ it is enough to prove either the expansion formula for rows, or for columns. Let's do it for the jth column a_j of A. We have

$$\det A = d_n(a_1, a_2, \cdots, a_j, \cdots, a_n) = (-1)^{j-1} d_n(a_j, a_1, a_2, \cdots, a_n),$$

where we have exchanged column j with column $j-1$, then with column $j-2$, and so on until column j is the new column 1 and the other columns follow in the previous order. We need $j-1$ switches of columns to do this so we picked up the factor $(-1)^{j-1}$. Now we use linearity applied to $a_j = \sum_{i=1}^n a_{ij} e_i$, so

$$d_n(a_j, a_1, a_2, \cdots, a_n) = \sum_{i=1}^n a_{ij} d_n(e_i, a_1, a_2, \cdots, a_n)$$

and we have to determine $d_n(e_i, a_1, a_2, \cdots, a_n)$. Now we observe that since $\det A = \det A^t$ we can also exchange rows in a matrix and change the corresponding determinant by a sign. Switching the ith row upwards until it is the first row gives

$$d_n(e_i, a_1, a_2, \cdots, a_n) = (-1)^{i-1} d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)})$$

where $a_1^{(i)} = (a_{i1}, a_{11}, a_{21}, \cdots, a_{n1})^t$, and so on, are the original column vectors with the ith component moved to the first place. We now claim that

$$d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)}) = \det \hat{A}_{ij}.$$

This follows from two observations,

- First, $d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)})$ does not depend on $a_{i1}, a_{i2}, \cdots, a_{in}$, since by Theorem 4.16, part (c), one can add arbitrary multiples of e_1 to all other arguments without changing the value of $d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)})$. This means $d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)})$ depends only on $\hat{A}_{ij}$ (recall that we removed column j already).
- The function $d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)})$ is by construction a multilinear and alternating function of the columns of $\hat{A}_{ij}$ and furthermore if $\hat{A}_{ij} = I$, then $d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)}) = 1$, hence by Theorem 4.9 we have $d_n(e_1, a_1^{(i)}, a_2^{(i)}, \cdots, a_n^{(i)}) = \det \hat{A}_{ij}$.

So collecting all formulas we have found

$$\det A = \sum_{j=1}^n (-1)^{i+j-2} a_{ij} \det \hat{A}_{ij}$$

and since $(-1)^{i+j-2} = (-1)^{i+j}$ the proof is complete. □

4.3 Some applications of determinants

In this section we will collect a few applications of determinants. Let us also note that everything we've done about determinants has not used specifically that we were dealing with reals, so all the above results hold if one wishes to replace $\mathbb{R}$ by, say $\mathbb{C}$ or $\mathbb{Z}$. In particular, note that the determinant of an integer square matrix is an integer.

4.3.1 Determinants and systems of equations

Recall that system of m linear equations in n unknowns can be written in the form

$$Ax = b,$$

where $A \in M_{m,n}(\mathbb{R})$ and $b \in \mathbb{R}^m$, and $x \in \mathbb{R}^n$ is the vector of the unknowns. If $m = n$, i.e., the system has as many equations as unknowns, then A is a square matrix and so we can ask if it is invertible.

Exercise 4.26. If A is invertible, what can we say about the solutions to the system?

Click for solution

We know that A is invertible if and only if $\det A \neq 0$, and then we find

$$x = A^{-1}b.$$

So $\det A \neq 0$ means the system has a unique solution.

Exercise 4.27. If A is not invertible, what can we say about the solutions to the system?

Click for solution

In this case we could have either infinitely many solutions or no solutions

If $\det A = 0$ then the row echelon form of A must have at least one row of zeros. Then whether the system has no solutions or infinitely many solutions will depend on the row echelon form of the augmented matrix $(A\,b)$. Then by Theorem 3.39 this has the same number of zero rows then there will be infinitely many solutions; if it has fewer then there will be no solutions.

If $\det A \neq 0$ one can go even further and use the determinant to compute an inverse and the unique solution to $Ax = b$.

Definition 4.28 (Adjugate matrix)**.** Let $A \in M_n(\mathbb{R})$ and let A_{ij} be the signed minors of A. The matrix $\tilde{A} = (A_{ij})$, which has the minors as elements, lets us define the adjugate (also known as the classical adjoint)

$$\operatorname{adj} A := \tilde{A}^t = (A_{ji}) \in M_n(\mathbb{R}).$$

So to calculate the adjugate we find the signed minors of the original matrix, put these in as the entries in the corresponding place in the matrix, and then take the transpose.

This matrix is useful. Firstly, it gives us an explicit formula for A^{-1}.

Theorem 4.29. *Let $A \in M_n(\mathbb{R})$, with $\det A \neq 0$. Then*

$$A^{-1} = \frac{1}{\det A} \operatorname{adj} A.$$

The following related result is called **Cramer's rule**, and gives the explicit formula for solutions of systems of linear equations.

Theorem 4.30. *Let $A \in M_n(\mathbb{R})$, with $\det A \neq 0$. Let $b \in \mathbb{R}^n$ and A_j be the matrix obtained from A by replacing its jth column by b. Then the unique solution $x = (x_1, x_2, \cdots, x_n)$ to $Ax = b$ is given by*

$$x_j = \frac{\det A_j}{\det A}, j = 1, 2, \cdots, n.$$

Both results can be proven by playing around with Laplace expansion and other basic properties of determinants. These results are mainly of theoretical use since in practice Gaussian/Gauss-Jordan elimination is more efficient at solving systems of equations or finding matrix inverses as it is a more scalable process.

4.3.2 The cross product

The determinant can be used to define the cross product of two vectors in $\mathbb{R}^3$, which will be another vector in $\mathbb{R}^3$. If we recall Laplace expansion in the first row for a 3×3 matrix,

$$\det A = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}, \tag{4.7}$$

then we can interpret this as the dot product between the first row-vector of A and the vector (A_{11}, A_{12}, A_{13}) whose components are the signed minors associated with the first column. If we denote the first column by $z = (z_1, z_2, z_3)$ and the second and third by $x = (x_1, x_2.x_3)$ and $y = (y_1, y_2, y_3)$, then the above formula reads

$$\det \begin{pmatrix} z_1 & x_1 & y_1 \\ z_2 & x_2 & y_2 \\ z_3 & x_3 & y_3 \end{pmatrix} = z_1(x_2y_3 - x_3y_2) + z_2(x_3y_1 - x_1y_3) + z_3(x_1y_2 - x_2y_1),$$

and we therefore use this to define the cross product.

Definition 4.31 (Cross product)**.** Let $x, y \in \mathbb{R}^3$. Then their cross product is

$$x \times y := \begin{pmatrix} x_2y_3 - x_3y_2 \\ x_3y_1 - x_1y_3 \\ x_1y_2 - x_2y_1 \end{pmatrix}.$$

Example 4.32. We have that

$$\begin{pmatrix} 2 \\ -2 \\ 3 \end{pmatrix} \times \begin{pmatrix} -1 \\ 3 \\ 5 \end{pmatrix} = \begin{pmatrix} -19 \\ -13 \\ 4 \end{pmatrix}.$$

The formula (4.7) then becomes

$$\det \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ z_1 & z_2 & z_3 \end{pmatrix} = \det \begin{pmatrix} z_1 & x_1 & y_1 \\ z_2 & x_2 & y_2 \\ z_3 & x_3 & y_3 \end{pmatrix} = z \cdot (x \times y). \tag{4.8}$$

The cross product, and notions derived from it, appear in many applications: mechanics, vector calculus, geometry. Let us collect now a few properties.

Theorem 4.33. *The cross product is a map $\mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3$ which satisfies the following properties for any $x, y, z \in \mathbb{R}^3$ and $\alpha, \beta \in \mathbb{R}$:*

- ***Antisymmetry**: $y \times x = -x \times y$ and $x \times x = \mathbf{0}$.*
- ***Bilinearity**: $(\alpha x + \beta y) \times z = \alpha(x \times z) + \beta(y \times z)$.*
- *$x \cdot (x \times y) = y \cdot (x \times y) = 0$.*
- *$\|x \times y\|^2 = \|x\|^2 \|y\|^2 - (x \cdot y)^2$.*
- *$x \times (y \times z) = (x \cdot z)y - (x \cdot y)z$.*

We will leave this as an exercise. The first three properties follow easily from the relation (4.8) and properties of the determinant, and the remaining two can be verified by direct, although not so easy computations.

Property (iii) means that $x \times y$ is orthogonal to the plane spanned by x and y, and (iv) gives us the length as

$$\|x \times y\|^2 = \|x\|^2 \|y\|^2 |\sin\theta|^2, \tag{4.9}$$

where θ is the angle between x and y (since $(x \cdot y)^2 = \|x\|^2 \|y\|^2 \cos^2\theta$) . Let n be the unit vector (i.e., $\|n\| = 1$) orthogonal to x and y chosen according to the **right hand rule**: if x points in the direction of the thumb, y in the direction of the index finger, then n points in the direction of the middle finger. For example, if $x = e_1$, $y = e_2$ then $n = e_3$, whereas $x = e_2$, $y = e_1$ gives $n = -e_3$. Then we have the following result.

Theorem 4.34. *The cross product satisfies*

$$x \times y = \|x\| \|y\| |\sin\theta| n.$$

Physicists often take this as the definition of the cross product. We omit the proof.

Property (v) of Theorem (4.7) implies that the cross product is not associative, i.e., in general $(x \times y) \times z \neq x \times (y \times z)$. Instead the so called Jacobi identity holds (which can be verified directly using (v) in Theorem 4.33):

$$x \times (y \times z) + y \times (z \times x) + z \times (x \times y) = \mathbf{0}.$$

In summary, we have the following geometric interpretations,:

- $x \cdot x$ is the square of the length of the vector x;
- $\|x \times y\|$ is the area of the parallelogram spanned by x and y;
- $x \cdot (y \times z)$ is the oriented volume of the parallelepiped spanned by x, y, z.

More details on all this are on problem sheets. Finally, note that although the dot product makes sense for pairs of vectors in any $\mathbb{R}^n$, the cross product is defined for vectors in $\mathbb{R}^3$ only.

As we have seen, the determinant function has a range of applications, and as we continue through the course we will encounter some additional ones.

Having explored matrices and determinants, we are now ready to explore the concept of vector spaces and their properties in more detail.

5. Linear subspaces and spans

When we defined our vectors in Euclidean space, we saw how to perform the operations of addition and scalar multiplication. A common theme of linear algebra is to study first and foremost these two key operations and try to give them intuitive or geometric meaning. Having seen how these operations are defined on $\mathbb{R}^n$, it is natural to next turn to thinking about the behaviour of subsets of $\mathbb{R}^n$, and consider the interaction between subsets and these operations.

Therefore, in this chapter we want to study the following two closely related questions:

- Which type of subsets of $\mathbb{R}^n$ (or $\mathbb{C}^n$, or a vector space in general) stay invariant under these two operations?
- Which type of subsets of $\mathbb{R}^n$ (or $\mathbb{C}^n$, or a vector space in general) can be generated by using these two operations?

Again, we will focus on $\mathbb{R}^n$ and $\mathbb{C}^n$ in this chapter, but these results can be generalised to other vector spaces as well.

5.1 Subspaces

We begin by considering our first question, about when subsets of a vector space stay invariant under the operations of our vector space, and define a linear subspace.

Definition 5.1 (Linear Subspace)**.** Let W be a vector space over $\mathbb{F}$. A subset $V \subseteq W$ is called a linear subspace of W if and only if the following three conditions hold:

(i) $V \neq \emptyset$, i.e., V is non-empty.
(ii) for all $v, w \in V$, we have $v + w \in V$, i.e., V is closed under addition.
(iii) for all $\lambda \in \mathbb{F}$, $v \in V$, we have $\lambda v \in V$, i.e., V is closed under multiplication by scalars.

Note that this definition is equivalent to saying that V is itself a vector space over $\mathbb{F}$ under the same operations as W. The proof of this fact is left as an exercise.

When referring to linear subspaces we will often just say 'subspace' for short.

In order to determine whether a subset is a subspace, we must show that the subset satisfies all of the properties from the definition.

Example 5.2. For any vector space W here are two trivial examples: $V = \{\mathbf{0}\}$, the set containing only $\mathbf{0}$, is a subspace, and $V = W$ itself satisfies the conditions for a linear subspace.

Example 5.3. Consider the real vector space $\mathbb{C}^n$ over $\mathbb{R}$. Then $\mathbb{R}^n$ is a subspace since

(i) $\nvdash \in \mathbb{R}^n$, so $\mathbb{R}^n$ is non-empty.
(ii) If $u, v \in \mathbb{R}^n$ then $u + v \in \mathbb{R}^n$.
(iii) If $v \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$ then $\lambda v \in \mathbb{R}^n$.

Exercise 5.4. Consider the complex vector space $\mathbb{C}^n$ over $\mathbb{C}$. Is $\mathbb{R}^n$ a subspace?

Example 5.5. Let $v \in \mathbb{R}^n$ be non-zero vector and let us take the set of all multiples of v, i.e.,

$$V := \{\lambda v : \lambda \in \mathbb{R}\}$$

We show that this is a subspace by checking each of the conditions in the definition:

(i) We have that $V \neq \emptyset$, since, for example, $v \in V$.
(ii) If $x, y \in V$ then there are $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $x = \lambda_1 v$ and $y = \lambda_2 v$ (this follows from the definition of V). Hence $x + y = \lambda_1 v + \lambda_2 v = (\lambda_1 + \lambda_2)v \in V$, since $\lambda_1 + \lambda_2 \in R$.
(iii) If $x \in V$ and $\lambda \in \mathbb{R}$, then there exists $\lambda_1 \in \mathbb{R}$ such that $x = \lambda_1 v$ and $\lambda x = \lambda\lambda_1 v \in V$, since $\lambda\lambda_1 \in \mathbb{R}$.

In geometric terms V is a straight line through the origin, e.g., if $n = 2$ and $v = (1, 1)$, then V is just the diagonal in $\mathbb{R}^2$, as shown in Figure 5.1.

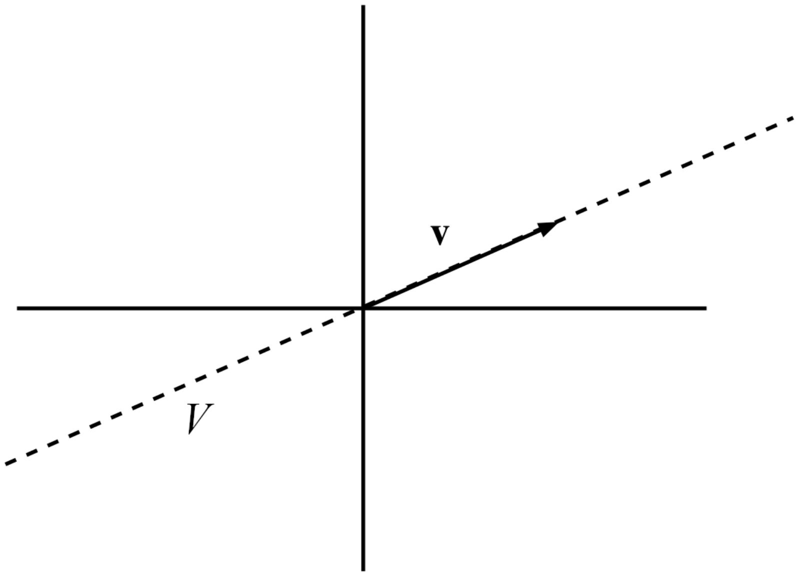

Figure 5.1: The subspace $V \subseteq \mathbb{R}^2$ (a line) generated by a vector $v \in \mathbb{R}^2$.

Theorem 5.6. *Let $A \in M_{m,n}(\mathbb{R})$. Then $S(A, \mathbf{0}) \subseteq \mathbb{R}^n$ is a linear subspace.*

Before proving this theorem let us consider in detail what this will mean for a homogeneous set of linear equations $Ax = \mathbf{0}$:

- There is always at least one solution, namely $x = \mathbf{0}$.
- The sum of any two solutions is again a solution.
- Any multiple of a solution is again a solution.

Proof. We check the conditions of the definition and find that

(i) $S(A, \mathbf{0})$ is nonempty since $A\mathbf{0} = \mathbf{0}$, hence $\mathbf{0} \in S(A, \mathbf{0})$,
(ii) if $x, y \in S(A, \mathbf{0})$, then $A(x + y) = Ax + Ay = \mathbf{0} + \mathbf{0} = \mathbf{0}$, hence $x + y \in S(A, \mathbf{0})$, and finally
(iii) if $x \in S(A, \mathbf{0})$ then $A(\lambda x) = \lambda Ax = \lambda\mathbf{0} = \mathbf{0}$ and therefore $\lambda x \in S(A, \mathbf{0})$.

□

Example 5.7. Consider the set V of vectors in $\mathbb{R}^2$ of the form $(x,1)$, i.e. the second coordinate being fixed as 1. Is this a linear subspace? To answer this question we have to check the three properties in the definition:

(i) Since, for instance, $(1,1) \in V$ we have $V \neq \emptyset$.
(ii) Choose two elements in V, e.g., $(1,1)$ and $(2,1)$, then $(1,1)+(2,1)=(3,2) \notin V$, hence the condition (ii) is not fulfilled and V is not a subspace.

Note that as soon as one of the three conditions of the definition fails then this set cannot possibly satisfy the definition, and so we do not need to check the third condition.

Example 5.8. Now, for comparison to the previous example, choose $V = \{(x,0) : x \in \mathbb{R}\}$. Then:

(i) $V \neq \emptyset$, since $(0.0) \in V$.
(ii) For any $(x,0),(y,0) \in V$ we have that $(x,0)+(y,0)=(x+y,0) \in V$ and so V is closed under addition.
(iii) For any $(x,0) \in V$ and $\lambda \in \mathbb{R}$ we have that $\lambda(x,0)=(\lambda x,0) \in V$, so V is closed under scalar multiplication.

Hence V satisfies all three conditions of the definition and is a linear subspace.

Exercise 5.9. Consider the vector space $\mathbb{C}^3$ over $\mathbb{C}$. Is $V = \{(x,iy,0) : x,y \in \mathbb{R}\}$ a linear subspace?

Exercise 5.10. Consider the vector space $\mathbb{C}^3$ over $\mathbb{R}$. Is $V = \{(x,iy,0) : x,y \in \mathbb{R}\}$ a linear subspace?

The following theorem provides a quick way of showing that something is not a linear subspace.

Theorem 5.11. *Let V be a linear subspace of a vector space W. Then we must have $\mathbf{0} \in V$.*

The proof is left as an exercise. This theorem means that we can immediately say a subset is not a subspace if it does not contain the zero vector.

Exercise 5.12. Does the converse hold? That is if $0 \in V$ must V be a subspace?

Click for solution

No, this is not the case, for example $V = \{(x,x^2) : x \in \mathbb{R}\}$ is not a subspace of $\mathbb{R}^2$ as $(1,1) \in V$ but $2(1,1)=(2,2) \notin V$. However we do have that $(0,0) \in V$.

This means we can use Theorem 5.11 as a way to prove something is **not** a subspace, but it does not help us to show that something **is** a subspace.

5.2 Spans

Example 5.5 is related to the second question we asked at the start of the chapter. Here we fixed one vector and took all its multiples, and that gave us a straight line. Generalising this idea to two and more vectors and taking sums as well into account leads us to the following definition:

Definition 5.13 (Span)**.** Let V be a vector space over $\mathbb{F}$ and $x_1, x_2, \cdots, x_k \in V$ be k vectors. The span of this set of vectors is defined as

$$\mathrm{span}_{\mathbb{F}}\{x_1, x_2, \ldots, x_k\} := \{\lambda_1 x_1 + \lambda_2 x_2 + \cdots \lambda_k x_k : \lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{F}\}.$$

When $\mathbb{F} = \mathbb{R}$ we call this an $\mathbb{R}$-span and when $\mathbb{F} = \mathbb{C}$ we call this a $\mathbb{C}$-span.

If the field that we are using is clear from context then we will often just say span, and omit the $\mathbb{F}$ from the subscript.

We will call an expression like

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots \lambda_k x_k$$

a **linear combination** of the vectors $x_1, \cdots, x_k$ with coefficients $\lambda_1, \cdots, \lambda_k$. We may specify where the scalars in our linear combination come from by referring to an $\mathbb{R}$-linear combination (meaning that the scalars come from $\mathbb{R}$) or a $\mathbb{C}$-linear combination (meaning that the scalars come from $\mathbb{C}$).

So the span of a set of vectors is the set generated by taking all linear combinations of the vectors from the set. We have seen one example already in Example 5.5, where we took linear combinations of just one vector, but if we take for instance two vectors $x_1, x_2 \in \mathbb{R}^3$, and if they do point in different directions, then their span is a plane through the origin in $\mathbb{R}^2$. The geometric picture associated with a span in Euclidean space is that it is a generalisation of lines and planes through the origin in $\mathbb{R}^2$ and $\mathbb{R}^3$ to $\mathbb{R}^n$.

Example 5.14. Consider the vector space $\mathbb{R}^3$ over $\mathbb{R}$. Let $x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $x_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. Then

$$\mathrm{span}\{x_1, x_2\} = \left\{ \lambda_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} : \lambda_1, \lambda_2 \in \mathbb{R} \right\} = \left\{ \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ 0 \end{pmatrix} : \lambda_1, \lambda_2 \in \mathbb{R} \right\}.$$

So in this case the span of these two vectors gives us the xy-plane.

Example 5.15. Consider the vector space $\mathbb{C}^3$ over $\mathbb{C}$. Let $x_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and $x_2 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. Then

$$\mathrm{span}\{x_1, x_2\} = \left\{ \lambda_1 \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} : \lambda_1, \lambda_2 \in \mathbb{C} \right\} = \left\{ \begin{pmatrix} \lambda_1 \\ \lambda_1 \\ \lambda_2 \end{pmatrix} : \lambda_1, \lambda_2 \in \mathbb{C} \right\}.$$

Exercise 5.16. Consider the vector space $\mathbb{R}^3$ over $\mathbb{R}$. Let $x_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $x_2 = \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix}$ and $x_3 = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}$.

What is $\mathrm{span}\{x_1, x_2\}$?

It turns out that Example 5.5 being a subspace was not a fluke; a span of vectors always gives us a subspace.

Theorem 5.17. *For a vector space V over $\mathbb{F}$ and $x_1, x_2, \cdots x_k \in V$ then* $\text{span}\{x_1, x_2, \cdots, x_k\}$ *is a linear subspace of V.*

Proof. The set is clearly non-empty. Now assume $v, w \in \text{span}\{x_1, x_2, \cdots, x_k\}$, i.e., there exist $\lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{F}$ and $\mu_1, \mu_2, \cdots, \mu_k \in \mathbb{F}$ such that

$$v = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k \quad \text{and} \quad w = \mu_1 x_1 + \mu_2 x_2 + \cdots + \mu_k x_k.$$

Therefore

$$v + w = (\lambda_1 + \mu_1)x_1 + (\lambda_2 + \mu_2)x_2 + \cdots + (\lambda_k + \mu_k)x_k \in \text{span}\{x_1, x_2, \cdots, x_k\},$$

and

$$\lambda v = \lambda\lambda_1 x_1 + \lambda\lambda_2 x_2 + \cdots + \lambda\lambda_k x_k \in \text{span}\{x_1, x_2, \cdots, x_k\},$$

for all $\lambda \in \mathbb{F}$. So $\text{span}\{x_1, x_2, \cdots, x_k\}$ is closed under addition and multiplication by numbers, hence it is a subspace. □

In fact we have that $\text{span}\{x_1, x_2, \cdots, x_k\}$ will be the smallest linear subspace containing the vectors $x_1, x_2, \cdots, x_k \in \mathbb{R}^n$. The proof of this is left as an exercise.

We may want to check if a certain vector is in a given span. To do this we will once again need to solve systems of linear equations.

Example 5.18. Consider $\mathbb{R}^n$ over $\mathbb{R}$ and let $V = \text{span}\{x_1, x_2\} \subseteq \mathbb{R}^n$ with $x_1 = (1,1,1)$ and $x_2 = (2,0,1)$.

The span is the set of all vectors of the form

$$\lambda_1 x_1 + \lambda_2 x_2,$$

where $\lambda_1, \lambda_2 \in \mathbb{R}$ can take arbitrary values. For instance if we set $\lambda_2 = 0$ and let λ_1 run through $\mathbb{R}$ we obtain the line through x_1, similarly by setting $\lambda_1 = 0$ we obtain the line through x_2. The set V is now the plane containing these two lines, as shown in Figure 5.2. To check if a vector is in this plane, i.e, in V, we have to see if it can be written as a linear combination of x_1 and x_2.

Let us check if $(1,0,0) \in V$. We have to find λ_1, λ_2 such that

$$(1,0,0) = \lambda_1 x_1 + \lambda_2 x_2 = (\lambda_1 + 2\lambda_2, \lambda_1, \lambda_1 + \lambda_2).$$

This gives us three equations, one for each component:

$$1 = \lambda_1 + 2\lambda_2, \quad \lambda_1 = 0, \quad \lambda_1 + \lambda_2 = 0.$$

From the second equation we get $\lambda_1 = 0$, then the third equation gives $\lambda_2 = 0$ but the first equation then becomes $1 = 0$, hence there is a contradiction and $(1,0,0) \notin V$.

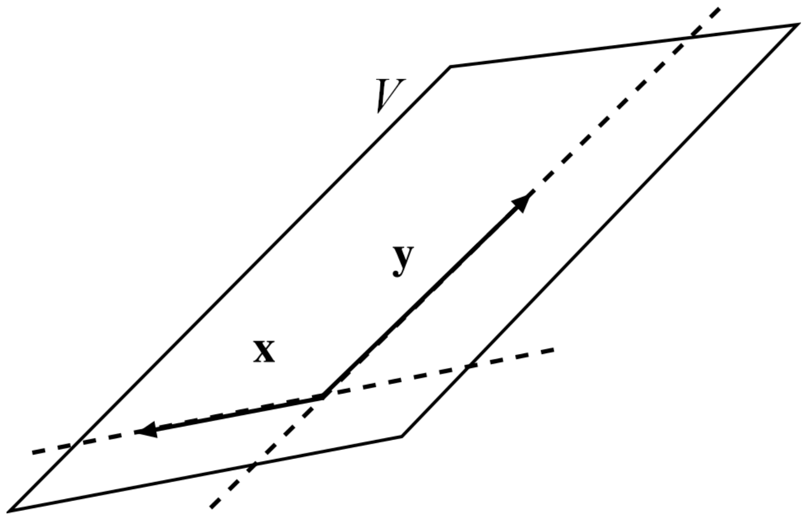

Figure 5.2: The subspace $V \subseteq \mathbb{R}^3$ (a plane) generated by vectors $x, y \in \mathbb{R}^3$. It contains the lines through x and y, and is spanned by these.

Exercise 5.19. Continue Example 5.18. Is $(0,2,1) \in V$?

Note that when checking whether a vector is in a span, we are solving a system of linear equations to find the coefficients λ_i.To do this, we can make use of the methods we developed in Chapter 3.

Example 5.20. Consider the vector space $\mathbb{C}^2$ over $\mathbb{C}$ and the subspace $V = \operatorname{span}\left\{\begin{pmatrix} i \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1+i \end{pmatrix}\right\}$.

We want to see if $v = \begin{pmatrix} -3 \\ -1 \end{pmatrix} \in V$, so we want to see if there exist $\lambda_1, \lambda_2 \in \mathbb{C}$ such that $v = \lambda_1 \begin{pmatrix} i \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ 1+i \end{pmatrix}$.

This means we want to solve the system

$$\begin{aligned} i\lambda_1 + 2\lambda_2 &= -3 \\ \lambda_1 + (1+i)\lambda_2 &= -1. \end{aligned}$$

This gives the augmented matrix

$$\begin{pmatrix} i & 2 & -3 \\ 1 & 1+i & 2 \end{pmatrix}.$$

Performing row operations (multiplying row 1 by $-i$ then subtracting row 1 from row 2) gives

$$\begin{pmatrix} 1 & -2i & 3i \\ 0 & 1+i & -1-3i \end{pmatrix},$$

then back-substitution gives $\lambda_2 = -1$ and $\lambda_1 = 2i\lambda_2 + 3i = i$.

It also turns out that we can write any subspace of a vector space as a span of a finite set of vectors. We refer to such sets as spanning sets.

Definition 5.21 (Spanning set)**.** If $\text{span}\{v_1, v_2, \cdots, v_k\} = W$ then we call $\{v_1, v_2, \cdots, v_k\}$ a spanning set for W and say that $v_1, v_2, \cdots, v_k$ span W.

Note that this means that if we have a spanning set for a linear subspace, then we can construct any vector in the subspace from our spanning set and that taking linear combinations of our spanning set will give us only the vectors in our subspace and no others.

Example 5.22. We have that $\left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \right\}$ is a spanning set for $\left\{ \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ 0 \end{pmatrix} : \lambda_1, \lambda_2 \in \mathbb{R} \right\} \subset \mathbb{R}$.

Exercise 5.23. Find a spanning set for $W = \left\{ \begin{pmatrix} 2ia \\ a+b \\ b \end{pmatrix} : a, b \in \mathbb{C} \right\} \subseteq \mathbb{C}^3$.

5.3 Further examples of subspaces

Now that we have seen that once way of forming a subspace is taking the span of a set of vectors we will consider some other ways that we can build subspaces.

Exercise 5.24. Consider $\mathbb{R}^n$ over $\mathbb{R}$. Is every non-zero linear subspace infinite?

Click for solution

Yes, this is true for any subspace V of $\mathbb{R}^n$, since as soon as we have some non-zero vector $v \in V$ then $\lambda V \in V$ for any $\lambda \in \mathbb{R}$.

Similarly, subspaces of $\mathbb{C}^n$ over $\mathbb{C}$ or $\mathbb{R}$ will be infinite. However, this is not true for **all** vector spaces and later in the course we will see examples of finite choices for $\mathbb{F}$ which give rise to finite vector (sub)spaces.

Another way to create a subspace is by giving conditions on the vectors contained in it. For instance let us chose a vector $a \in \mathbb{R}^n$ and let us look at the set W_a of vectors x in $\mathbb{R}^n$ which are orthogonal to a, i.e., which satisfy $x \cdot a = 0$ i.e.,

$$W_a := \{x \in \mathbb{R}^n : x \cdot a = 0\}.$$

This is shown in Figure 5.3.

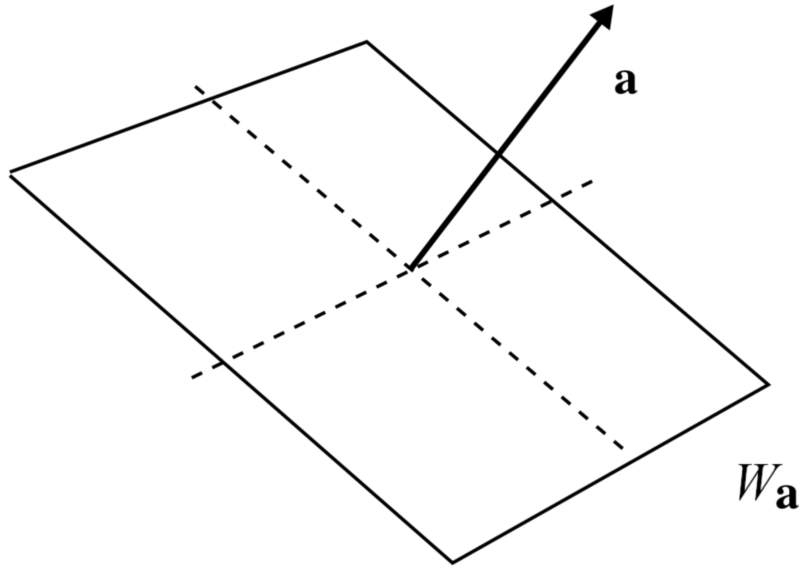

Figure 5.3: The plane orthogonal to a non-zero vector a is a subspace W_a.

Theorem 5.25. *Let $a \in \mathbb{R}^n$. Then $W_a := \{x \in \mathbb{R}^n : x \cdot a = 0\}$ is a subspace of $\mathbb{R}^n$.*

Proof. Clearly $\mathbf{0} \in W_a$, so $W_a \neq \emptyset$. Let $x, y \in W_a$ and $\lambda \in \mathbb{R}^n$. So $x \cdot a = 0$ and $y \cdot a = 0$, and we have that $(x+y) \cdot a = x \cdot a + y \cdot a = 0$ and $(\lambda x) \cdot a = \lambda x \cdot a = 0$. Hence $x + y \in W_a$ and $\lambda x \in W_a$ so W_a is a linear subspace. $\square$

For instance if $n = 2$, then W_a is the line perpendicular to a (if $a \neq \mathbf{0}$, otherwise $W_a = \mathbb{R}^2$) and if $n = 3$, then W_a is a plane perpendicular to a (if $a \neq \mathbf{0}$, otherwise $W_a = \mathbb{R}^3$).

There can be different vectors a which determine the same subspace; in particular notice that since for $\lambda \neq 0$ we have $x \cdot a = 0$ if and only if $x \cdot (\lambda a) = 0$ we get $W_a = W_{\lambda a}$ for $\lambda \neq 0$. In terms of the subspace $V = \operatorname{span}\{a\}$ this means

$$W_a = W_b, \quad \text{for all} \quad b \in V \backslash \{\mathbf{0}\},$$

and so W_a is actually perpendicular to the whole subspace V spanned by a. This motivates the following definition:

Definition 5.26 (Orthogonal complement)**.** Let V be a subspace of $\mathbb{R}^n$, then the orthogonal complement $V^\perp$ is defined as

$$V^\perp = \{x \in \mathbb{R}^n : x \cdot y = 0 \text{ for all } y \in V\}.$$

So the orthogonal complement consists of all vectors $x \in \mathbb{R}^n$ which are perpendicular to all vectors in V. So, for instance, if V is a plane in $\mathbb{R}^3$, then $V^\perp$ is the line perpendicular to it.

Theorem 5.27. *Let V be a subspace of $\mathbb{R}^n$. Then $V^\perp$ is a subspace of $\mathbb{R}^n$.*

Proof. Clearly $\mathbf{0} \in V^\perp$, so $V^\perp \neq \emptyset$. If $x \in V^\perp$, then for any $v \in V$, $x \cdot v = 0$ and therefore $(\lambda x) \cdot v = \lambda x \cdot v = 0$ and so $\lambda x \in V^\perp$, so $V^\perp$ is closed under multiplication by scalars. Finally if

$x, y \in V^{\perp}$, then $x \cdot v = 0$ and $y \cdot v = 0$ for all $v \in V$ and hence $(x+y) \cdot v = x \cdot v + y \cdot v = 0$ for all $v \in V$, therefore $x + y \in V^{\perp}$. □

Note that if $V = \text{span}\{a_1, \cdots, a_k\}$ is spanned by k vectors, then $x \in V^{\perp}$ means that the k conditions

$$\begin{aligned} a_1 \cdot x &= 0, \\ a_2 \cdot x &= 0, \\ &\vdots \\ a_k \cdot x &= 0 \end{aligned}$$

hold simultaneously.

To define an orthogonal complement, we have made use of the dot product as defined on vectors in $\mathbb{R}^n$. We could also define such a structure for $\mathbb{C}^n$ but using the dot product as defined on complex vectors instead. It is left as an excercise to convince yourself that the above proof will still hold. We will see later in the course how we can generalise these dot products to something called an **inner product** on a vector space in order to define these orthogonal complements more generally.

Given multiple subspaces of the same vector space, we can also combine them to make a new subspace.

Theorem 5.28. *Assume U, V are subspaces of a vector space W, then*

- *$U \cap V = \{w \in W : w \in U \text{ and } w \in V\}$ is a subspace of W*
- *$U + W := \{u + v : u \in U \text{ and } v \in V\}$ is a subspace of W.*

The proof of this result will be left as an exercise, as will the following generalisation:

Theorem 5.29. *Let $W_1, W_2, \cdots W_m$ be subspaces of $\mathbb{R}^n$, then*

$$W_1 \cap W_2 \cap \cdots \cap W_m$$

is also a subspace of $\mathbb{R}^n$.

Exercise 5.30. Is the union $U \cup V = \{w \in W : w \in U \text{ or } w \in V\}$ of two subspaces a subspace?

Click for solution

No, in general the union of two subspaces is not another subspace.

5.4 Direct sums

Another useful concept related to subspaces is the notion of a **direct sum**, which is a stronger condition than the sum defined in Theorem 5.28.

Definition 5.31 (Direct sum)**.** A subspace $V \subseteq W$ is said to be the direct sum of two subspaces $V_1, V_2 \subseteq W$, with the notation $V = V_1 \oplus V_2$ if

- $V = V_1 + V_2$, and

- $V_1 \cap V_2 = \{\mathbf{0}\}$.

Example 5.32. Consider $V_1 = \text{span}\{e_1\}$, $V_2 = \text{span}\{e_2\}$ with $e_1 = (1,0), e_2 = (0,1) \in \mathbb{R}^2$, then

$$\mathbb{R}^2 = V_1 \oplus V_2.$$

If a subspace V is the sum of two subspaces V_1, V_2, every element of V can be written as a sum of two elements of V_1 and V_2, and if V is a direct sum this decomposition is unique.

Theorem 5.33. *Let V_1 and V_2 be subspaces of a vector space. If $W = V_1 \oplus V_2$, then for any $w \in W$ there exist unique $v_1 \in V_1, v_2 \in V_2$ such that $w = v_1 + v_2$.*

Proof. It is clear that there exist v_1, v_2 with $w = v_1 + v_2$ from the sum part of the direct sum definition. So what we have to show is uniqueness. So Let us assume there is another pair $v_1' \in V_1$ and $v_2' \in V_2$ such that $w = v_1' + v_2'$, then we can subtract the two different expressions for w and obtain

$$\mathbf{0} = (v_1 + v_2) - (v_1' + v_2') = v_1 - v_1' - (v_2' - v_2)$$

and therefore $v_1 - v_1' = v_2' - v_2$. But in this last equation the left hand side is a vector in V_1, the right hand side is a vector in V_2 and since they have to be equal, they lie in $V_1 \cap V_2 = \{\mathbf{0}\}$, so $v_1' = v_1$ and $v_2' = v_2$. □

A trivial but useful example is as follows.

Example 5.34. Consider the subspaces

$$V_1 = \{(x, y, 0) : x, y \in \mathbb{R}\}$$

and

$$V_2 = \{(0, 0, z) : z, \in \mathbb{R}\}$$

of $\mathbb{R}^3$. Then for any $v \in \mathbb{R}^3$, there is a unique decomposition

$$v = (x, y, z) = (x, y, 0,) + (0, 0, z)$$

as a sum of two elements from V_1 and V_2. In other words, we can think of $\mathbb{R}^3$ as the direct sum of the xy-plane and the z-axis.

Exercise 5.35. Let $V_1 = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_1 + x_2 = x_3\}$ and $V_2 = \{(y_1, y_2, y_3) \in \mathbb{R}^3 : y_1 - y_2 = y_3\}$. Does $\mathbb{R}^3 = V \oplus W$?

Once again, we may find ourselves in a situation where we have a system of linear equations to solve when trying to determine whether we have a direct sum, and the techiniques from Chapter 3 come in useful.

While exploring the idea of subspace, we have seen that it is possible to take a set of vectors and use it to construct a subspace. A natural question is how can we do this in the most ‘efficient’ way, that is using the smallest number of vectors possible. This is what we will explore in the next chapter.

6. Linear independence, bases and dimension

The idea of dimension is one we have come across informally when thinking about objects in two or three dimensions in the world around us. We will now formalise this concept for linear subspaces by thinking about how we can build a given subspace by taking linear combinations of a minimal set of vectors.

6.1 Linear dependence and independence

How do we characterise a vector space V? One possibility is to choose a set of vectors $v_1, v_2, \cdots, v_k \in V$ which span V, i.e., such that

$$V = \operatorname{span}\{v_1, v_2, \cdots, v_k\}.$$

In order do this in an efficient way, we want to choose the minimum number of vectors necessary. If one vector from the set can be written as a linear combination of the others, it is redundant. This leads to the following definition.

Definition 6.1 (Linear dependence)**.** Let V be a vector space over $\mathbb{F}$. The vectors $v_1, v_2, \cdots, v_k \in V$ are called linearly dependent if there exist $\lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{F}$, not all 0, such that

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \mathbf{0}.$$

Example 6.2. If $k = 1$ then $\lambda_1 v_1 = \mathbf{0}$ with $\lambda_1 \neq 0$ means $v_1 = \mathbf{0}$.

Example 6.3. If $k = 2$, then if two vectors v_1, v_2 are linearly dependent, it means there are λ_1, λ_2 which are not both simultaneously zero, such that

$$\lambda_1 v_1 + \lambda_2 v_2 = \mathbf{0}. \tag{6.1}$$

Now it could be that at least one of the vectors is 0, say for instance $v_1 = \mathbf{0}$, then $\lambda_1 v_1 + 0 v_2 = \mathbf{0}$ for any λ_1, so v_1, v_2 are indeed linearly dependent. But this is a trivial case: whenever in a finite set of vectors at least one of the vectors is $\mathbf{0}$, then the set of vectors is linearly dependent. So assume $v_1 \neq \mathbf{0}$ and $v_2 \neq \mathbf{0}$, then in (6.1) both λ_1 and λ_2 must be non-zero and hence

$$v_1 = \lambda v_2, \quad \text{with} \quad \lambda = -\lambda_2/\lambda_1$$

so one vector is just a multiple of the other.

Exercise 6.4. If we have a set of three vectors that is linearly dependent, what can we say about the relationship between these vectors?

Click for solution

A similar analysis to the above shows that 3 cases can occur:

(i) at least one of them is $\mathbf{0}$,
(ii) two of them are proportional to each other, or

(iii) one of them is a linear combination of the other two.

As the examples illustrate, when $v_1, \cdots, v_k$ are linearly dependent, then we can write one of the vectors as a linear combination of the others.

If a set of vectors is not linearly dependent they are called linearly independent.

Definition 6.5 (Linear independence)**.** Let V be a vector space over $\mathbb{F}$. The vectors $v_1, v_2, \cdots, v_k \in V$ are called linearly independent if

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \mathbf{0}$$

implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_k = 0$.

So if the vectors are linearly independent the only way to get $\mathbf{0}$ as a linear combination is to choose all the coefficients to be 0.

Example 6.6. Assume we want to know if the two vectors $x = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $y = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \in \mathbb{R}^n$ are linearly independent or not. So we have to see if we can find $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $\lambda_1 x + \lambda_2 y = \mathbf{0}$, but since $\lambda_1 x + \lambda_2 y = \begin{pmatrix} \lambda_1 2 + \lambda_2 \\ \lambda_1 3 - \lambda_2 \end{pmatrix}$ this translates into the two equations

$$2\lambda_1 + \lambda_2 = 0 \quad \text{and} \quad 3\lambda_1 - \lambda_2 = 0.$$

Adding the two equations gives $5\lambda_1 = 0$, hence $\lambda_1 = 0$ and then $\lambda_2 = 0$. Therefore the two vectors are linearly independent.

Example 6.7. Consider the two vectors $x = \begin{pmatrix} 2 \\ 3 \end{pmatrix}$ and $y = \begin{pmatrix} -8 \\ -12 \end{pmatrix} \in \mathbb{R}^n$. Again we look for $\lambda_1, \lambda_2 \in \mathbb{R}$ with

$$\lambda_1 x + \lambda_2 y = \begin{pmatrix} 2\lambda_1 - 8\lambda_2 \\ 3\lambda_1 - 12\lambda_2 \end{pmatrix} = \mathbf{0},$$

which leads to the two equations $2\lambda_1 - 8\lambda_2 = 0$ and $3\lambda_1 - 12\lambda_2 = 0$. Dividing the first by 2 and the second by 3 reduces both equations to the same one, $\lambda_1 - 4\lambda_2 = 0$, and this is satisfied whenever $\lambda_1 = 4\lambda_2$, hence the two vectors are linearly dependent.

What these examples showed is that questions about linear dependence or independence lead to linear systems of equations. So the question of whether a set of vectors is linearly independent is the same as asking whether the corresponding system of equations has a unique solution or not.

Theorem 6.8. *Let $v_1, v_2, \cdots, v_k \in \mathbb{F}^n$ and let $A \in M_{n,k}(\mathbb{F})$ be the matrix which has $v_1, v_2, \cdots, v_k$ as its columns. Then the vectors $v_1, v_2, \cdots, v_k$ are linearly independent if*

$$S(A, \mathbf{0}) = \{\mathbf{0}\}$$

and linearly dependent otherwise.

Proof. By the definition of A we have for $x = (\lambda_1, \lambda_2, \cdots, \lambda_k)$

$$Ax = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k$$

(this follows from (3.3)). So if $S(A, \mathbf{0}) = \{\mathbf{0}\}$ then $v_1, v_2, \cdots, v_k$ are linearly independent, and otherwise are not. □

As a consequence of this result we can use Gaussian elimination to determine if a set of vectors is linearly dependent or linearly independent. We consider the matrix A whose column vectors are the set of vectors under investigation and apply elementary row operations until it is in row-echelon form. If every column has a leading one the vectors are linearly independent, otherwise they are linearly dependent.

Example 6.9. Take the three vectors $v_1 = (1,2,3)$, $v_2 = (-1,2,1)$ and $v_3 = (0,0,1) \in \mathbb{R}^n$, then

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 2 & 0 \\ 3 & 1 & 1 \end{pmatrix}$$

and after a couple of elementary row operations (row 2$-$2$\times$row 1, row 3$-$3$\times$row 1, row 3$-$row 2, row 2$\rightarrow$ (row 2)/4) we find the following row echelon form

$$\begin{pmatrix} 1 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

and so the vectors are linearly independent.

Example 6.10. On the other hand side, if we take $v_1 = (1,2,3)$, $v_2 = (-1,2,1)$ and $v_3 = (2,0,2) \in \mathbb{R}^n$, then

$$A = \begin{pmatrix} 1 & -1 & 2 \\ 2 & 2 & 0 \\ 3 & 1 & 2 \end{pmatrix}$$

and after a couple of elementary row operations (row 2$-$2$\times$row 1, row 3$-$3$\times$row 1, row 3$-$row 2, row 2$\rightarrow$ (row 2)/4) we find the following row echelon form

$$\begin{pmatrix} 1 & -1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

and so the vectors are linearly dependent.

Exercise 6.11. Let $v_1 = (i, 2)$, $v_2 = (1, -2i) \in \mathbb{C}^n$. Are these linearly independent over $\mathbb{C}$? Are these linearly independent over $\mathbb{R}$?

Click for solution

Over $\mathbb{C}$, we have that $v_1 = iv_2$, so they are linearly independent. However, over $\mathbb{R}$ they are linearly independent as there are no non-zero real solutions to $\lambda_1 v_1 + \lambda_2 v_2 = \mathbf{0}$.

So this example shows that it is important to know exactly which vector space we are working in when considering linear dependence.

As a consequence of this relation to systems of linear equations we have the following fundamental result.

Corollary 6.12. *If $v_1, v_2, \cdots, v_k \in \mathbb{F}^n$ over $\mathbb{F}$ are linearly independent, then $k \leq n$. So any set of linearly independent vectors in $\mathbb{F}^n$ over $\mathbb{F}$ can contain at most n elements.*

Proof. Let A be the $n \times k$ matrix consisting of the columns $v_1, v_2, \cdots, v_k$, then by Theorem 6.8 the vectors are linearly independent if $S(A, \mathbf{0}) = \{\mathbf{0}\}$, but by Corollary 3.42 this gives $k \leq n$. □

The corresponding result also holds for vectors in $\mathbb{C}^n$ over $\mathbb{C}$ (but not for $\mathbb{C}^n$ over $\mathbb{R}$ - we will explore the differences between these further in the next section).

In the case where we have n vectors in $\mathbb{R}^n$ we can also use our determinant to see whether they are linearly independent, thanks to a theorem that connects the determinant, the invertibility of a matrix, and the linear independence of its columns (or rows).

Theorem 6.13. *Let $A \in M_n(\mathbb{F})$, then the following properties are equivalent:*

(1) $\det A \neq 0$,
(2) A is invertible,
(3) the column vectors of A are linearly independent over $\mathbb{F}$.
(4) the row vectors of A are linearly independent over $\mathbb{F}$.

Proof. Let us first show that (1) implies (3). Let $a_1, a_2, \cdots, a_n$ be the column vectors of A. If $a_1, a_2, \cdots, a_n$ are linearly dependent, then there exists an $1 \leq i \leq n$ and some $\lambda_j \in \mathbb{F}$ such that

$$a_i = \sum_{j \neq i} \lambda_j a_j,$$

and using linearity in the ith component we get

$$\begin{aligned} d_n(a_1, \cdots, a_i, \cdots, a_n) &= d_n\left(a_1, \cdots, \sum_{j \neq i} \lambda_j a_j, \cdots, a_n\right) \\ &= \sum_{j \neq i} \lambda_j d_n(a_1, \cdots, a_j, \cdots, a_n) = 0, \end{aligned}$$

where in the last step we used that there are always at least two equal vectors in the argument of $d_n(a_1, \cdots, a_j, \cdots, a_n)$, so by part (ii) of Proposition 4.4 we get 0.

Now we show that (3) implies (2). If the column vectors are linearly independent, then the corresponding homogeneous equation has a unique solution. This means that the reduced row echelon form of A is the identity matrix, and hence A is invertible.

Then (2) implies (1) since if $A^{-1}A = I$ we get by the product formula for determinants that $\det A^{-1} \det A = 1$, hence $\det A \neq 0$, as we have previously seen in 4.

Finally, we have that (3) and (4) are equivalent since $\det A = \det A^t$ so if $\det A \neq 0$ then $\det A^t \neq 0$, and the columns of A^t are the rows of A. □

This is one of the most important results about determinants and it is often used when one needs a criterion for invertibility or linear independence.

In particular, note that we can use this for checking linear independence of vectors in $\mathbb{R}^n$ over $\mathbb{R}$ and $\mathbb{C}^n$ over $\mathbb{C}$.

Exercise 6.14. Are the vectors $v_1 = (1,0,1), V_2 = (0,2,1), v_3 = (0,1,4) \in \mathbb{R}^n$ linearly independent?

6.2 Bases and dimension

In Section 5.2 we explored how we can build a space from a set of vectors by taking their span, and the concept of linear independence now gives us a way to prevent any redundancy in our choice of vectors in the spanning set.

So if a collection of vectors span a subspace and are linearly independent, then they deserve a special name.

Definition 6.15. Let $V \subseteq W$ be a linear subspace of a vector space W. A **basis** of V is a set of vectors $v_1, v_2 \cdots, v_k \in V$ such that

- $\operatorname{span}\{v_1, v_2 \cdots, v_k\} = V$, and
- the vectors $v_1, v_2 \cdots, v_k \in V$ are linearly independent.

The plural of basis is bases.

So a basis of V is a set of vectors in V which generate the whole subspace V, but with the minimal number of vectors necessary.

Example 6.16. The set $e_1 = (1,0,0,\cdots,0)$, $e_2 = (0,1,0,\cdots,0)$, $e_3 = (0,0,1,\cdots,0)$, ... , $e_n = (0,0,0,\cdots,1) \in \mathbb{R}^n$ is a basis for $\mathbb{R}^n$ over $\mathbb{R}$. For any $v \in \mathbb{R}^n$ we have

$$v = (x_1, x_2, \cdots, x_n) = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n,$$

so this is a spanning set for $\mathbb{R}^n$ and is linearly independent since the corresponding matrix is $I \in M_n(\mathbb{R})$.

Definition 6.17 (Standard basis)**.** The set $e_1 = (1,0,0,\cdots,0)$, $e_2 = (0,1,0,\cdots,0)$, $e_3 = (0,0,1,\cdots,0)$, ... , $e_n = (0,0,0,\cdots,1) \in \mathbb{R}^n$is called the **standard basis** for $\mathbb{R}^n$ over $\mathbb{R}$.

Exercise 6.18. Is $e_1, \ldots, e_n \in \mathbb{C}^n$ a basis for $\mathbb{C}^n$ over $\mathbb{C}$? What about over $\mathbb{R}$?

Click for solution

Yes, this is a basis for $\mathbb{C}^n$ over $\mathbb{C}$ by a similar argument to the example above (and we may similarly refer to this as the standard basis for $\mathbb{C}^n$ over $\mathbb{C}$). However it is not a basis for $\mathbb{C}^n$ over $\mathbb{R}$ as it is not a spanning set, since there is no $\mathbb{R}$-linear combination of these vectors that give $(i,0,\ldots,0)$ for example.

An example of a basis for $\mathbb{C}^n$ over $\mathbb{R}$ would be $e_1, \ldots, e_n, ie_1, \ldots, ie_n$.

The definition a basis means that every vector can be written as a linear combination of basis vectors in a unique way.

Theorem 6.19. *Let W be a vector space over $\mathbb{F}$, $V \subseteq W$ be a linear subspace, and $v_1, v_2 \cdots, v_k \in V$ a basis of V. Then for any $v \in V$ there exist a unique set of scalars $\lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{F}$ such that*

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k.$$

Proof. Since the vectors $v_1, v_2 \cdots, v_k$ span V there exist for any $v \in V$ scalars $\lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{F}$ such that

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k. \tag{6.2}$$

We have to show that these scalars are unique. So let us assume there is another, possibly different, set of scalars $\mu_1, \mu_2, \cdots, \mu_k$ with

$$v = \mu_1 v_1 + \mu_2 v_2 + \cdots + \mu_k v_k, \tag{6.3}$$

then subtracting (6.2) from (6.3) gives

$$0 = (\mu_1 - \lambda_1)v_1 + (\mu_2 - \lambda_2)v_2 + \cdots + (\mu_k - \lambda_k)v_k$$

but since we assumed that the vectors $v_1, v_2, \cdots, v_k$ are linearly independent we get that $\mu_1 - \lambda_1 = \mu_2 - \lambda_2 = \cdots = \mu_k - \lambda_k = 0$ and hence

$$\mu_1 = \lambda_1, \quad \mu_2 = \lambda_2, \ldots, \quad \mu_k = \lambda_k.$$

□

Bases are not unique, that is our vector spaces will have more than one choice of basis. The standard basis defined in 6.17 is a useful basis as it is very straightforward to see how we can write any given vector in terms of this basis. However, there are often good reasons for choosing different bases as well. We will see more about this later in the course, but for now we will look at some further examples of bases and some useful properties.

Example 6.20. Consider $v_1 = (1,1)$ and $v_2 = (-1,1)$, and let us see if they form a basis of $\mathbb{R}^2$. To check if they span $\mathbb{R}^2$, we take an arbitrary $(x_1, x_2) \in \mathbb{R}^2$ and have to find $\lambda_1, \lambda_2 \in \mathbb{R}$ such that

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \lambda_1 v_1 + \lambda_2 v_2 = \begin{pmatrix} \lambda_1 - \lambda_2 \\ \lambda_1 + \lambda_2 \end{pmatrix}.$$

This is just a system of two linear equations for λ_1, λ_2 and can be easily solved to give

$$\lambda_1 = \frac{x_1 + x_2}{2}, \quad \lambda_2 = \frac{-x_1 + x_2}{2},$$

hence the two vectors span $\mathbb{R}^2$. Furthermore if we set $x_1 = x_2 = 0$ we see that the only solution to $\lambda_1 v_1 + \lambda_2 v_2 = \mathbf{0}$ is $\lambda_1 = \lambda_2 = 0$, so the vectors are also linearly independent, and hence we do have a basis for $\mathbb{R}^2$.

Theorem 6.19 tells us that we can write any vector in a unique way as a linear combination of the vectors in a basis, so we can interpret the basis vectors as giving us a **coordinate system**, and the coefficients λ_i in an expansion $v = \sum_i \lambda_i v_i$ are the **coordinates** of v. See Figure 6.1 for an illustration of this idea.

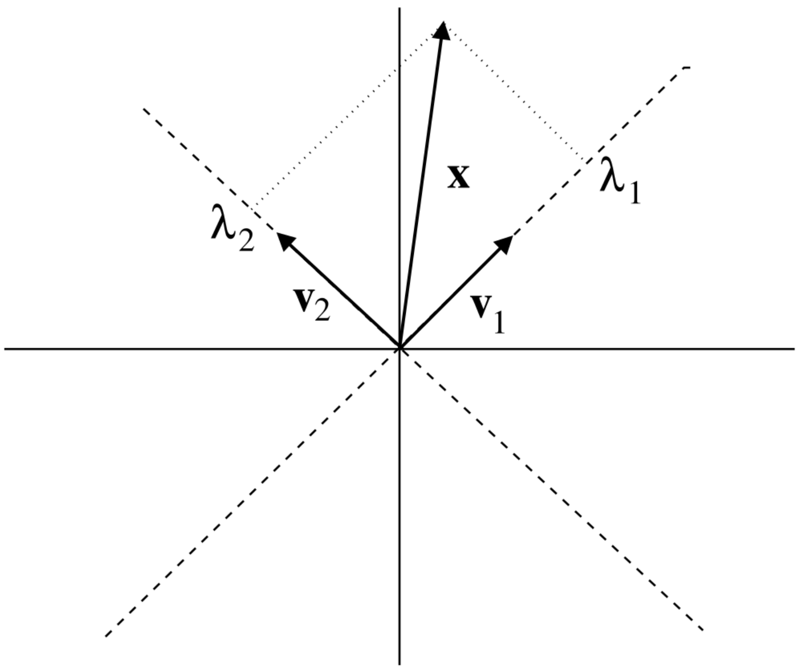

Figure 6.1: Illustrating how a basis v_1, v_2 of $\mathbb{R}^2$ acts as a coordinate system: the dashed lines are the new coordinate axes spanned by v_1, v_2, and λ_1, λ_2 are the coordinates of $v = \lambda_1 v_1 + \lambda_2 v_2$.

Notice that in the standard basis $e_1, \cdots, e_n$ of $\mathbb{R}^n$ the expansion coefficients of v are $x_1, \cdots, x_n$, the usual Cartesian coordinates.

Exercise 6.21. Is $B = \{e_1, e_2, e_3\}$ a basis for $V = \{(x, 2x, z) : x, z \in \mathbb{R}\} \subseteq \mathbb{R}^3$?

Given a basis $v_1, v_2 \cdots, v_k$ of V it is not always straightforward to compute the expansion of a vector v in that basis, i.e., to find the numbers $\lambda_1, \lambda_2, \cdots, \lambda_k \in \mathbb{F}$. In general this leads to a system of linear equations for the $\lambda_1, \lambda_2, \cdots, \lambda_k$.

Example 6.22. Let us consider the set of vectors $v_1 = (1,2,3)$, $v_2 = (-1,2,1)$ and $v_3 = (0,0,1) \in \mathbb{R}^n$. We know from Example 6.9 that they are linearly independent, so they form a good candidate for a basis of $V = \mathbb{R}^3$ and we just have to show that they span $\mathbb{R}^3$. Let $x = (x,y,z) \in \mathbb{R}^3$, and then we have to find $\lambda_1, \lambda_2, \lambda_3$ such that

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = x.$$

If we write out the components this gives a system of three linear equations for three unknowns $\lambda_1, \lambda_2, \lambda_3$ and the corresponding augmented matrix is

$$\begin{pmatrix} A & x \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 & x \\ 2 & 2 & 0 & y \\ 3 & 1 & 1 & z \end{pmatrix}$$

and after a couple of elementary row operations (row 2$-$2$\times$row 1, row 3$-$3$\times$row 1, row 3$-$row

2, row 2$\to$ (row 2)/4) we find the following row echelon form:

$$\begin{pmatrix} 1 & -1 & 0 & x \\ 0 & 1 & 0 & y/4-x/2 \\ 0 & 0 & 1 & z-y-x \end{pmatrix}.$$

Back-substitution then gives

$$\lambda_3 = z-y-x, \quad \lambda_2 = \frac{y}{4}-\frac{x}{2} \quad \lambda_1 = \frac{y}{4}+\frac{x}{2}.$$

This means that given an arbitrary vector (x,y,z) we have shown how to find coefficients $\lambda_1, \lambda_2, \lambda_3$ in order to write our vector in terms of v_1, v_2, v_3 and so these vectors do span $\mathbb{R}^3$. Therefore the the vectors form a basis and the expansion of an arbitrary vector $x = (x,y,z) \in \mathbb{R}^3$ in that basis is given by

$$x = \left(\frac{y}{4}+\frac{x}{2}\right)v_1 + \left(\frac{y}{4}-\frac{x}{2}\right)v_2 + (z-y-x)v_3.$$

We now want to show that any subspace of the vector spaces that we have been exploring will indeed have a basis. To do this we first introduce some terminology.

Definition 6.23 (Finite-dimensional vector space)**.** A vector space is called finite-dimensional if it can be spanned by a finite number of vectors.

So $\mathbb{R}^n$ over $\mathbb{R}$, $\mathbb{C}^n$ over $\mathbb{R}$ and $\mathbb{C}^n$ over $\mathbb{C}$ are all finite-dimensional

So we want to show any subspace of a finite-dimensional vector space must have a basis. This will be a consequence of the following result which says that any set of linearly independent vectors in a subspace V is either already a basis of V, or can be extended to a basis of V.

Theorem 6.24. *Let V be a linear subspace of a finite-dimensional vector space over $\mathbb{F}$, and $v_1, v_2, \cdots, v_r \in V$ be a set of linearly independent vectors. Then either $v_1, \cdots, v_r$ are a basis of V, or there exist a finite number of further vectors $v_{r+1}, \cdots, v_k \in V$ such that $v_1, \cdots, v_k$ is a basis of V.*

Proof. Let us set

$$V_r := \operatorname{span}\{v_1, v_2, \cdots, v_r\},$$

which is a subspace with basis $v_1, v_2, \cdots, v_r$ and $V_r \subseteq V$.

Now if $V_r = V$, then we are done. Otherwise, $V_r \neq V$ and then there exists a vector $v_{r+1} \neq \mathbf{0}$ with $v_{r+1} \in V$ but $v_{r+1} \notin V_r$. We claim that $v_1, v_2, \cdots, v_r, v_{r+1}$ are linearly independent.

To show this assume

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_r v_r + \lambda_{r+1} v_{r+1} = \mathbf{0},$$

and then if $\lambda_{r+1} \neq 0$ we get

$$v_{r+1} = -\frac{\lambda_1}{\lambda_{r+1}} v_1 - \frac{\lambda_2}{\lambda_{r+1}} v_2 - \cdots - \frac{\lambda_r}{\lambda_{r+1}} v_r \in V_r$$

which contradicts our assumption $v_{r+1} \notin V_r$. Hence $\lambda_{r+1} = 0$ but then all the other λ_i's must be 0, too, since $v_1, v_2, \cdots, v_r$ are linearly independent.

So we set

$$V_{r+1} := \mathrm{span}\{v_1, v_2, \cdots, v_r, v_{r+1}\},$$

which is again a subspace with basis $v_1, v_2, \cdots, v_r, v_{r+1}$, and proceed as before: Either $V_{r+1} = V$, or we can find a another linearly independent v_{r+2}, etc. In this way we find a chain of subspaces

$$V_r \subseteq V_{r+1} \subseteq \cdots \subseteq V$$

which are strictly increasing. But by Corollary 6.12 there can be at most n linearly independent vectors in $\mathbb{F}^n$, and therefore there must be a finite k such that $V_k = V$, and then $v_1, \cdots, v_k$ is a basis of V. □

Exercise 6.25. Consider $\mathbb{C}^3$ over $\mathbb{C}$ and let $v_1 = (i, 0, 0)$ and $v_2 = (0, 2, 0)$. Is this a basis for $\mathbb{C}^3$ over $\mathbb{C}$? If not, can we extend it to a basis?

Corollary 6.26. *Any linear subspace V of a finite dimensional vector space has a basis.*

Proof. If $V = \{\mathbf{0}\}$ then this has basis $\emptyset$. Now assume $V \neq \{\mathbf{0}\}$, and so there exists at least one $v \neq \mathbf{0}$ with $v \in V$ and by Theorem 6.24 this can be extended to a basis. □

We found above that $\mathbb{F}^n$ over $\mathbb{F}$ can have at most n linearly independent vectors, and we now extend this result to subspaces, that is we see that the number of linearly independent vectors is bounded by the number of elements in a basis.

Theorem 6.27. *Let V be linear subspace of a vector space over $\mathbb{F}$ and $v_1, \cdots, v_k \in V$ be a basis of V. Then if $w_1, \cdots w_r \in V$ are a set of linearly independent vectors we have $r \leq k$.*

Proof. Since $v_1, \cdots, v_k$ is a basis we can write each vector w_i, $i = 1, \cdots, r$ in terms of that basis, giving

$$w_i = \sum_{j=1}^{k} a_{ji} v_j,$$

where the $a_{ji} \in \mathbb{R}$ are the expansion coefficients. Now the assumption that $w_1, \cdots w_r$ are linearly independent means that $\sum_{i=1}^{r} \lambda_i w_i = 0$ implies that $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$. But with the expansion of the w_i in the basis we can rewrite this equation as

$$0 = \sum_{i=1}^{r} \lambda_i w_i = \sum_{i=1}^{r} \sum_{j=1}^{k} a_{ji} v_j \lambda_i = \sum_{j=1}^{k} \left(\sum_{i=1}^{r} a_{ji} \lambda_i \right) v_j.$$

Now we use that the vectors $v_1, \cdots, v_k$ are linearly independent, and therefore we find

$$\sum_{i=1}^{r} a_{1i} \lambda_i = 0, \quad \sum_{i=1}^{r} a_{2i} \lambda_i = 0, \quad \cdots \quad , \quad \sum_{i=1}^{r} a_{ki} \lambda_i = 0.$$

This is system of k linear equations for the r unknowns $\lambda_1, \cdots, \lambda_r$, and in order that the only solution to this system is $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 0$ we must have that $k \geq r$ by Corollary 3.42. □

Exercise 6.28. What does this tell us about the number of elements in a basis of a vector space?

Click for solution

Corollary 6.29. *Let V be a linear subspace of a finite-dimensional vector space. Then any basis of V has the same number of elements.*

So the number of elements in a basis does not depend on the choice of the basis, it is an attribute of the subspace V, which can be viewed as an indicator of its size. Hence we give this attribute a name.

Definition 6.30 (Dimension)**.** Let V be a vector space. The dimension of V, written $\dim V$, is the minimal number of vectors needed to span V, which is the number of elements in a basis of V.

Note that this does indeed align with the intuitive understanding of dimension in Euclidean space we will have previously encountered.

Exercise 6.31. What is the dimension of:

- $\mathbb{R}^n$ over $\mathbb{R}$?
- $\mathbb{C}^n$ over $\mathbb{C}$?
- $\mathbb{C}^n$ over $\mathbb{R}$?

Click for solution

We have that
- $\mathbb{R}^n$ over $\mathbb{R}$ has dimension n, since the standard basis has n vectors.

- $\mathbb{C}^n$ over $\mathbb{C}$ has dimension n, since again $e_1, \dots, e_n$ is a basis.
- $\mathbb{C}^n$ over $\mathbb{R}$ has dimension 2n, since $e_1, \dots, e_n, ie_1, \dots, ie_n$ is a basis.

Example 6.32. Let us use the dimension to classify the types of linear subspaces of $\mathbb{R}^n$ for $n = 1, 2, 3$.

- If $n = 1$, the only linear subspaces of $\mathbb{R}$ are $V = \{\mathbf{0}\}$ and $V = \mathbb{R}$. We have $\dim\{\mathbf{0}\} = 0$ and $\dim \mathbb{R} = 1$.
- If $n = 2$:
  - When $\dim V = 0$ the only possible subspace is $\{\mathbf{0}\}$.
  - If $\dim V = 1$, we need one vector v to span V, hence every one dimensional subspace is a line through the origin.
  - If $\dim V = 2$ then $V = \mathbb{R}^2$.
- If $n = 3$:
  - When $\dim V = 0$ the only possible subspace is $\{\mathbf{0}\}$.
  - If $\dim V = 1$, we need one vector v to span V, hence every one dimensional subspace is a line through the origin.
  - If $\dim V = 2$ we need two vectors to span V, so we obtain a plane through the origin. So two dimensional subspaces of $\mathbb{R}^3$ are planes through the origin.
  - If $\dim V = 3$, then $V = \mathbb{R}^3$.

Exercise 6.33. Find a basis of $V = \{(x,y,x+y,-y) : x,y \in \mathbb{R}\}$.

If we know the dimension of a space this can make it easier to check whether a given set is a basis, thanks to the following result.

Theorem 6.34. *Let W be a vector space over $\mathbb{F}$. Let $V \subseteq W$ be a subspace with* $\dim V = k$.

(i) If $v_1, \cdots, v_k \in V$ are linearly independent then they form a basis.
(ii) If $v_1, \cdots, v_k \in V$ span V then they form a basis.

The proof of this is left as an exercise. Note that this means if we have a subspace where of dimension k and a set of k vectors, showing it is linearly independent or a spanning set is enough to show it is a basis. However, if we don't know the dimension already we will need to check both properties.

Having explored how we can build our subspaces, we next turn to consider the effects of applying maps to them, and in particular we are interested in maps which preserve our two key operations of addition and scalar multiplication.

7. Linear Maps

So far we have studied addition and multiplication by numbers of elements of vector spaces, and the structures which are generated by these two operations. Now we turn our attention to maps. In general a map T from a set V to a set W is a rule which assigns to each element of V an element of W. For example, $T(x,y) := (x^3y-4, \cos(xy))$ is a map from $\mathbb{R}^2$ to $\mathbb{R}^2$.

In Linear Algebra we focus on a special class of maps, namely linear maps – the ones which respect our fundamental operations, addition of vectors and multiplication by scalars. Some texts call these linear transformations, and in the case of $V = W$ we may call this a linear operator.

Definition 7.1 (Linear map)**.** Let V and W be vector spaces over $\mathbb{F}$. A map $T: V \to W$ is called an $\mathbb{F}$-linear map if

(i) $T(x+y) = T(x) + T(y)$, for all $x, y \in V$,
(ii) $T(\lambda x) = \lambda T(x)$, for all $\lambda \in \mathbb{F}$ and $x \in V$.

If the field $\mathbb{F}$ is clear from the context then we will often drop the $\mathbb{F}$ and just refer to a linear map. Note that the addition and multiplication on the left hand sides of the expressions above are taking place in W and the addition and multiplication on the right hand sides are taking place in V.

Let us note two immediate consequences of the definition.

Proposition 7.2. *Let V and W be vector spaces over $\mathbb{F}$ and $T: V \to W$ be a linear map, then*

(i) $T(\mathbf{0}) = \mathbf{0}$,
(ii) For arbitrary $x_1, \cdots, x_k \in V$ and $\lambda_1, \cdots, \lambda_k \in \mathbb{F}$ we have

$$T(\lambda_1 x_1 + \cdots + \lambda_k x_k) = \lambda_1 T(x_1) + \cdots + \lambda_k T(x_k).$$

Proof.

(i) This follows from $T(\lambda x) = \lambda T(x)$ if one sets $\lambda = 0$.
(ii) This is obtained by applying the two conditions from the definition of a linear map repeatedly.

□

Note that we can write the second property using the summation sign as well as

$$T\left(\sum_{i=1}^{k} \lambda_i x_i\right) = \sum_{i=1}^{k} \lambda_i T(x_i).$$

Linearity is a strong condition on a map. Let us consider some examples of what a linear map will look like.

Exercise 7.3. Let V and W be vector spaces over $\mathbb{F}$. Does there exist a map $T: V \to W$ such that $T(\mathbf{0}) = \mathbf{0}$ but T is not linear?

Let us now have a look at some examples of linear maps.

Example 7.4. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_3 \\ 4x_2 \end{pmatrix}$. Then for any $x, y \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$ we have that

$$\begin{aligned} T(x+y) = T\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ x_3+y_3 \end{pmatrix} &= \begin{pmatrix} (x_1+y_1)+(x_3+y_3) \\ 4(x_2+y_2) \end{pmatrix} \\ &= \begin{pmatrix} x_1+x_3 \\ 4x_2 \end{pmatrix} + \begin{pmatrix} y_1+y_3 \\ 4y_2 \end{pmatrix} = T(x)+T(y) \end{aligned}$$

and

$$\begin{aligned} T(\lambda x) = T\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \end{pmatrix} &= \begin{pmatrix} \lambda x_1 + \lambda x_3 \\ 4\lambda x_2 \end{pmatrix} \\ &= \lambda \begin{pmatrix} x_1+x_3 \\ 4x_2 \end{pmatrix} = \lambda T(x). \end{aligned}$$

So T is an $\mathbb{R}$-linear map.

Example 7.5. Let $T : \mathbb{R}^2 \to \mathbb{C}$ be defined by $T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x_1 + ix_2$. Then for any $x, y \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$ we have that

$$\begin{aligned} T(x+y) = T\begin{pmatrix} x_1+y_1 \\ x_2+y_2 \end{pmatrix} &= (x_1+y_1)+i(x_2+y_2) \\ &= (x_1+ix_2)+(y_1+iy_2) = T(x)+T(y) \end{aligned}$$

and

$$\begin{aligned} T(\lambda x) = T\begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix} &= \lambda x_1 + i\lambda x_2 \\ &= \lambda(x_1+ix_2)\lambda T(x). \end{aligned}$$

So T is an $\mathbb{R}$-linear map. This formalises the idea we saw in Chapter 2 about how we can identify complex numbers with vectors in $\mathbb{R}^2$.

Exercise 7.6. Consider $T : \mathbb{C}^2 \to \mathbb{C}^3$ defined by $T\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} (2+i)x_3 \\ i \\ x_1+x_2 \end{pmatrix}$. IS T a $\mathbb{C}$-linear map?

Example 7.7. The action of a matrix $A \in M_n(\mathbb{R})$ is a linear map by Theorem 3.11.

7.1 Properties of linear maps

In this section we will study some properties of linear maps and develop some of the related structures. We will later see that this connects to the exploration of matrices that we looked at earlier in the course.

7.1.1 Combinations of linear maps

Throughout this section we assume that U, V, W and X are vector spaces over $\mathbb{F}$.

We first notice that we can add linear maps if they relate the same spaces, and multiply them by scalars:

Definition 7.8 (Addition and scalar multiplication of linear maps)**.** Let $S : V \to W$ and $T : V \to W$ be linear maps, and $\lambda \in \mathbb{F}$. Then we define

- $(\lambda T)(x) := \lambda T(x)$ and
- $(S+T)(x) := S(x) + T(x)$.

Exercise 7.9. Do you think that $S+T$ and λT will also be linear maps?

Click for solution

Theorem 7.10. *Let $\lambda \in \mathbb{F}$ and $T, S : V \to W$ be linear maps. The maps λT and $S+T$ are linear maps from V to W.*

The proof follows directly from the definitions.

We can also compose maps in general, and for linear maps we find that the composition (if defined) is linear.

Theorem 7.11. *Let $T : U \to V$ and $S : V \to W$ be linear maps, then the composition*

$$S \circ T(x) := S(T(x))$$

is a linear map $S \circ T : U \to V$.

Proof. We consider first the action of $S \circ T$ on λx: Since T is linear we have $S \circ T(\lambda x) = S(T(\lambda x)) = S(\lambda T(x))$ and since S is linear, too, we get $S(\lambda T(x)) = \lambda S(T(x)) = \lambda S \circ T(x)$. Now we apply $S \circ T$ to $x+y$:

$$\begin{aligned} S \circ T(x+y) &= S(T(x+y)) \\ &= S(T(x) + T(y)) \\ &= S(T(x)) + S(T(y)) = S \circ T(x) + S \circ T(y). \end{aligned}$$

□

In a similar way one can prove the following.

Theorem 7.12. *Let $T : U \to V$ and $R, S : V \to W$ be linear maps, then*

$$(R+S) \circ T = R \circ T + S \circ T$$

and if $S,T : U \to V$ and $R : V \to W$ be linear maps, then

$$R \circ (S+T) = R \circ S + R \circ T.$$

Furthermore if $T : U \to V$, $S : V \to W$ and $R : W \to X$ are linear maps then

$$(R \circ S) \circ T = R \circ (S \circ T).$$

Note that the first and third properties hold for any functions, but the second relies on the linearity of the functions.

7.1.2 Image and kernel

Let us define two subsets related naturally to each linear map.

Definition 7.13 (Image and kernel)**.** Let $T : V \to W$ be a linear map, then we define

- the image of T to be
$$\operatorname{Im} T = \{y \in W : \text{there exists } x \in V \text{ with } T(x) = y\} \subseteq W,$$
- the kernel of T to be
$$\ker T = \{x \in V : T(x) = \mathbf{0}\} \subseteq V.$$

Example 7.14. Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be defined by $T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_3 \\ 4x_2 \end{pmatrix}$. We have seen that this is a linear map. Then

$$\operatorname{Im} T = \left\{ \begin{pmatrix} x_1 + x_3 \\ 4x_2 \end{pmatrix} : x_1, x_2, x_3 \in \mathbb{R} \right\} = \mathbb{R}^2$$

and

$$\begin{aligned} \ker T &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} x_1 + x_3 \\ 4x_2 \end{pmatrix} = \mathbf{0} \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : x_1 + x_3 = 0 \text{ and } 4x_2 = 0 \right\} \\ &= \left\{ \begin{pmatrix} x_1 \\ 0 \\ -x_1 \end{pmatrix} : x_1 \in \mathbb{R} \right\} \\ &= \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \right\}. \end{aligned}$$

Example 7.15. Let $A \in M_{m,n}(\mathbb{R})$ and let $a_1, a_2, \cdots, a_n \in \mathbb{R}^m$ be the column vectors of A, and set $T_A x := Ax$ which is a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$. Then since $T_A x = \mathbf{0}$ means $Ax = \mathbf{0}$ we have

$$\ker T_A = S(A, \mathbf{0}),$$

and from the relation $T_A x = Ax = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$ we see that

$$\operatorname{Im} T_A = \operatorname{span}\{a_1, a_2, \cdots, a_n\}.$$

In these examples we see that the image and kernel are actually subspaces rather than just subsets. This was not a coincidence, as the following theorem shows.

Theorem 7.16. *Let $T : V \to W$ be a linear map, then $\operatorname{Im} T$ is a linear subspace of W and $\ker T$ is a linear subspace of V.*

The proof is left as an exercise.

Now let us relate the image and the kernel to the some general properties of a map.

Recall that if A, B are sets and $f : A \to B$ is a map (not necessarily linear), then f is called

- **surjective**, if for any $b \in B$ there exists an $a \in A$ such that $f(a) = b$.
- **injective**, if whenever $f(a) = f(a')$ then $a = a'$.
- **bijective**, if f is injective and surjective, that is if for any $b \in B$ there exist exactly one $a \in A$ with $f(a) = b$.

Theorem 7.17. *If $f : A \to B$ is bijective, then there exists a unique map $f^{-1} : B \to A$ with $f \circ f^{-1}(b) = b$ for all $b \in B$, $f^{-1} \circ f(a) = a$ for all $a \in A$ and f^{-1} is bijective, too.*

Proof. Let us first show existence: For any $b \in B$, there exists an $a \in A$ such that $f(a) = b$, since f is surjective. Since f is injective, this a is unique, i.e., if $f(a') = b$, then $a' = a$, so we can set

$$f^{-1}(b) := a.$$

By definition this map satisfies $f \circ f^{-1}(b) = f(f^{-1}(b)) = f(a) = b$ and $f^{-1}(f(a)) = f^{-1}(b) = a$. From these we also get that f^{-1} is bijective. □

In the special case of linear maps we can connect these general properties to the image and kernel of our map.

Theorem 7.18. *Let $T : V \to W$ be a linear map, then*

(i) T is surjective if and only if $\operatorname{Im} T = W$,
(ii) T is injective if and only if $\ker T = \{\mathbf{0}\}$, and
(iii) T is bijective if and only if $\operatorname{Im} T = W$ and $\ker T = \{\mathbf{0}\}$.

Proof.

(i) That surjective is equivalent to $\operatorname{Im} T = W$ follows directly from the definitions of surjectivity and $\operatorname{Im} T$.
(ii) Notice that we always have $\mathbf{0} \in \ker T$, since $T(\mathbf{0}) = \mathbf{0}$. Now assume T is injective and let $x \in \ker T$, i.e, $T(x) = \mathbf{0}$. But then $T(x) = T(\mathbf{0})$ and injectivity of T gives $x = \mathbf{0}$, hence $\ker T = \{\mathbf{0}\}$. For the converse, let $\ker T = \{\mathbf{0}\}$, and assume there are $x, x' \in V$ with $T(x) = T(x')$. Using linearity of T we then get $\mathbf{0} = T(x) - T(x') = T(x - x')$ and hence $x - x' \in \ker T$, and since $\ker T = \{\mathbf{0}\}$ this means that $x = x'$ and hence T is injective.
(iii) This follows immediately from the previous two parts of the theorem.

□

Exercise 7.19. Is $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_3 \\ 4x_2 \end{pmatrix}$ injective? Is it surjective?

An important property of linear maps with $\ker T = \{\mathbf{0}\}$ is the following.

Theorem 7.20. *Let $x_1, x_2, \cdots, x_k \in V$ be linearly independent, and $T : V \to W$ be a linear map with $\ker T = \{\mathbf{0}\}$. Then $T(x_1), T(x_2), \cdots, T(x_k)$ are linearly independent.*

Proof. Assume $T(x_1), T(x_2), \cdots, T(x_k)$ are linearly dependent, i.e., there exist $\lambda_1, \lambda_2, \cdots, \lambda_k$, not all 0, such that

$$\lambda_1 T(x_1) + \lambda_2 T(x_2) + \cdots + \lambda_k T(x_k) = \mathbf{0}.$$

But since T is linear we have

$$T(\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k) = \lambda_1 T(x_1) + \lambda_2 T(x_2) + \cdots + \lambda_k T(x_k) = \mathbf{0},$$

and hence $\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k \in \ker T$. But since $\ker T = \{\mathbf{0}\}$ it follows that

$$\lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_k x_k = \mathbf{0},$$

which means that the vectors $x_1, x_2, \cdots, x_k$ are linearly dependent, and this contradicts the assumption. Therefore $T(x_1), T(x_2), \cdots, T(x_k)$ are linearly independent. □

Notice that this result implies that if T is bijective, it maps a basis of V to a basis of W, hence $V = W$.

We saw that a bijective map has an inverse; we now show that if T is linear, then the inverse is linear, too.

Theorem 7.21. *Let $T : V \to V$ be a linear map and assume T is bijective. Then $T^{-1} : V \to V$ is also linear.*

Proof. Let $y, y' \in V$. We want to show $T^{-1}(y + y') = T^{-1}(y) + T^{-1}(y')$. Since T is bijective we know that there are unique $x, x' \in \mathbb{R}^n$ with $y = T(x)$ and $y' = T(x')$, therefore

$$y + y' = T(x) + T(x') = T(x + x')$$

and applying T^{-1} to both sides of this equation gives

$$T^{-1}(y + y') = T^{-1}(T(x + x')) = x + x' = T^{-1}(y) + T^{-1}(y').$$

The second property of a linear map is shown in a similar way and this is left as an exercise. □

7.1.3 Rank and nullity

As the image and kernel of a linear map are subspaces we can find bases for these and hence determine their dimension. We give the dimension of the image and the kernel names as they are important concepts that can tell us a lot about a map.

Definition 7.22 (Nullity and rank)**.** Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map, then we define the nullity of T as

$$\operatorname{nullity} T := \dim \ker T,$$

and the rank of T as

$$\operatorname{rank} T := \dim \operatorname{Im} T.$$

Example 7.23. Let $T : \mathbb{R}^2 \to \mathbb{R}^2$, $T(x) = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_2 \\ 0 \end{pmatrix}$. This is a linear map (left as an exercise). Then $x \in \ker T$ means $x_2 = 0$, hence $\ker T = \operatorname{span}\{e_1\}$, and $\operatorname{Im} T = \operatorname{span}\{e_1\}$. Therefore we find $\operatorname{rank} T = 1$ and $\operatorname{nullity} T = 1$.

Exercise 7.24. Find the rank and nullity of the linear map $T : \mathbb{R}^3 \to \mathbb{R}^2$, $T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x_1 + x_3 \\ 4x_2 \end{pmatrix}$

So in view of our discussion in the previous subsection we have that a map $T : V \to W$ is injective if $\operatorname{nullity} T = 0$ and surjective if $\operatorname{rank} T = \dim W$. It turns out that rank and nullity are actually related; this is the content of the Rank Nullity Theorem.

Theorem 7.25 (Rank-Nullity Theorem)**.** *Let $T : V \to W$ be a linear map, then*

$$\operatorname{rank} T + \operatorname{nullity} T = \dim V.$$

We will explore this theorem and its proof further later in the course. For now, let us just consider a few consequences of this.

Exercise 7.26. Suppose that $T : \mathbb{R}^n \to \mathbb{R}^m$ is invertible. What can we say about n andm m?

Click for solution

We must have that $n = m$. In fact we can state this as a result about subspaces in general.

Corollary 7.27. *If the linear map $T : V \to W$ is invertible then* $\dim V = \dim W$.

Proof. Let $\dim V = n$ and $\dim W = m$. Since T is invertible we have that $\operatorname{rank} T = m$ and $\operatorname{nullity} T = 0$. Hence by the Rank-Nullity Theorem we have that $m = \operatorname{rank} T = n$. □

Corollary 7.28. *Let $T : V \to W$ be a linear map, then*

- *if* $\operatorname{rank} T = \dim V$, *then T is invertible,*
- *if* $\operatorname{nullity} T = 0$, *then T is invertible.*

Proof. We have that T is invertible if $\operatorname{rank} T = \dim V$ and $\operatorname{nullity} T = 0$, but by the Rank-Nullity Theorem $\operatorname{rank} T + \operatorname{nullity} T = \dim V$, hence any one of the conditions implies the other. □

We have already seen a couple of examples where we have defined linear maps using matrices.

However, the connection between matrices and linear maps is even stronger, as we will now explore.

7.2 Linear maps and matrices

For most of this section we will focus on linear maps in Euclidean space, that is maps from $\mathbb{R}^n$ to $\mathbb{R}^m$ for some $m,n \in \mathbb{N}$.It turns out that such maps correspond to matrices in $M_{m,n}(\mathbb{R})$.

Note: although we focus on real matrices, we could again replace every $\mathbb{R}$ by $\mathbb{C}$ and proceed in the same way.

We have already seen that any matrix can be thought of as a linear map, so we need only consider how an arbitrary linear map can be represented as a matrix. Consider a linear map $T : \mathbb{R}^n \to \mathbb{R}^m$. In order for this map to be linear, each of the components must be a linear equation with zero constant term.The proof of this is left as an exercise.

Now if we think about how this map acts on on $x = (x_1,\dots,x_n) \in \mathbb{R}^n$ we must have that

$$T\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n} \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

for some coefficients $a_{ij} \in \mathbb{R}$. But this is exactly the same as the action of the matrix $A = (a_{ij}) \in M_{m,n}(\mathbb{R})$.

In cases where the map is presented as linear equations in each component it is straightforward to read off the matrix as above. We have also seen some examples where the form of the linear map is less obvious. In general, we can use the fact that the ith column of our matrix will be given by $T(e_i)$ for all of our standard basis vectors.

Definition 7.29 (Matrix of a linear map)**.** Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear map. The matrix corresponding to this linear map in the standard basis is the matrix $M_T \in M_{m,n}(\mathbb{R})$ with ith column is $T(e_i)$.

Note that the matrix corresponding to a linear map is not unique, but instead depends on our choice of basis for the domain and co-domain. For now we will assume that we use the standard bases for both, but later in the course we will explore how to deal with different choices of basis, and some reasons why we may prefer to use alternative bases.

Example 7.30. Consider $T : \mathbb{R}^3 \to \mathbb{R}^3$ given by $T(x) = (x \cdot v)x$ for $v = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$. This is a linear map (the proof of this is left as an exercise). Then we have $T(e_1) = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $T(e_2) = \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix}$ and $T(e_3) = \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$ so the corresponding matrix is

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

We previously introduced some operations for linear maps: addition, multiplication by a scalar, and composition. We want to study now how these translate to matrices.

Theorem 7.31. *Let $S,T:\mathbb{R}^n\to\mathbb{R}^m$ be linear maps with corresponding matrices $M_T=(t_{ij}), M_S=(s_{ij})$, and let $\lambda\in\mathbb{R}$. Then the matrices corresponding to the maps λT and $S+T$ are given by*

$$M_{\lambda T}=\lambda M_T=(\lambda t_{ij}) \quad \textit{and} \quad M_{S+T}=(s_{ij}+t_{ij})=M_S+M_T.$$

Proof. Let $R=S+T$. The matrix associated with R is by Definition 7.29 given by $M_R=(r_{ij})$ with ith column $R(e_i)$, but since $R(e_i)=(S+T)(e_i)=S(e_i)+T(e_i)$ we have that the ith column of M_R is the ith column of M_S plus the ith column of M_T, and so $M_R=M_S+M_T$.

Similarly we find that $M_{\lambda T}$ has ith column $\lambda T(e_i)$ and so $M_{\lambda T}=\lambda M_T$. □

So the above theorem tells us that when adding linear maps/multiplying them by a scalar we just add the corresponding matrix elements/multiply them by a scalar. Note that this extends to expressions of the form

$$M_{\lambda S+\mu T}=\lambda M_S+\mu M_T.$$

and these expressions actually define the addition of matrices.

The composition of maps leads to multiplication of matrices.

Theorem 7.32. *Let $T:\mathbb{R}^n\to\mathbb{R}^m$ and $S:\mathbb{R}^m\to\mathbb{R}^l$ be linear maps with corresponding matrices $M_T=(t_{ij})$ and $M_S=(s_{ij})$, where M_T is $m\times n$ and M_S is $l\times m$. Then the matrix $M_{S\circ T}=(r_{ik})$ corresponding to the composition $R=S\circ T$ of T and S has elements*

$$r_{ik}=\sum_{j=1}^{m}s_{ij}t_{jk}$$

and is an $l\times n$ matrix.

The proof of this is left as an exercise. However, note that it follows from our definitions, and in fact it is for this reason that we defined matrix multiplication in the way that we did.

We can think about the formula for matrix multiplication as r_{ik} being the dot product of the ith row vector of M_S and the kth column vector of M_T. This formula defines a product of matrices by

$$M_SM_T:=M_{S\circ T}.$$

So we have now used the notions of addition and composition of linear maps to define addition and products of matrices. The results about maps then immediately imply corresponding results for matrices.

Theorem 7.33. *Let A,B be $m\times n$ matrices and C an $l\times m$ matrix, then*

$$C(A+B)=CA+CB.$$

Let A,B be $l\times m$ matrices and C an $m\times n$ matrix, then

$$(A+B)C=AC+BC.$$

Let C be an $m \times n$ matrix, B be an $l \times m$ matrix and A a $k \times l$ matrix, then

$$A(BC) = (AB)C.$$

Proof. We saw in Theorem 3.11 that matrices define linear maps, and in Theorem 7.12 the above properties were shown for linear maps. □

The first two properties mean that matrix multiplication is distributive over addition, and the last one is called associativity. In particular associativity would be quite cumbersome to prove directly for matrix multiplication, whereas the proof for linear maps is very simple. This shows that often an abstract approach simplifies proofs a lot. The price one pays for this is that it takes sometimes longer to learn and understand the material in a more abstract language.

Having identified our matrices with linear maps, we can now consider concepts like the image, kernel, rank and nullity of a matrix. Once again we can make use of Gaussian elimination, this time to find the rank and nullity of a matrix and hence of the corresponding linear map.

Theorem 7.34. *Let $A \in M_{m,n}(\mathbb{R})$ and assume that the row echelon form of A has k leading 1' s, then* $\operatorname{rank} A = k$ *and* $\operatorname{nullity} A = n - k$.

So in order to find the rank of a matrix we use elementary row operations to bring it to row echelon form and then we just count the number of leading 1's. The proof will be left as an exercise.

Example 7.35. Consider $A = \begin{pmatrix} 1 & -1 & 3 \\ 2 & 0 & 4 \\ 0 & 3 & 1 \\ -1 & -1 & 1 \end{pmatrix} \in M_{4,3}(\mathbb{R})$. We want to find the rank and nullity of this matrix. Using row operations, we find that a row echelon form of the matrix is $\begin{pmatrix} 1 & -1 & 3 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}$

Hence we have that the rank $= 3$ and nullity $= 1$.

We also previously saw that we could use the determinant to tell us about whether a system of n equations in n unknowns had a unique solution, but we can now go a step further using the image of a matrix to determine the outcome when the determinant is zero.

Theorem 7.36. *The system of linear equations $Ax = b$, with $A \in M_n(\mathbb{R})$, has a unique solution if and only if* $\det A \neq 0$. *If* $\det A = 0$ *and* $b \notin \operatorname{Im} A$ *no solution exists, and if* $\det A = 0$ *and* $b \in \operatorname{Im} A$ *then infinitely many solutions exist.*

Proof. We know that A is invertible if and only if $\det A \neq 0$, and then we find

$$x = A^{-1}b.$$

So $\det A \neq 0$ means the system has a unique solution. If $\det A = 0$, then $\operatorname{nullity} A > 0$ and $\operatorname{rank} A < n$, so a solution exists only if $b \in \operatorname{Im} A$, and if a solution x_0 exists, then all vectors in $\{x_0\} + \ker A$ are solutions, too, hence there are infinitely many solutions. □

In the next chapter, we will continue to explore linear maps and different ways we can represent them.