MA150 Algebra 2: Linear Algebra

Linear Algebra is one the most powerful and user-friendly parts of mathematics. It ties together a host of calculations that you probably know in some form already, but refines and strengthens them so they we may apply them across a huge range of different situations. Mathematics is about solving equations. Linear Algebra takes the ideas that solve simultaneous linear equations and builds a theoretical superstructure in which they are far more powerful and precise.

Perhaps the most difficult thing is merely that there are several starting points that are useful to have at our fingertips before we discuss the bigger unifying ideas. You probably already know:

(i) basic arithmetic of column vectors: how to add them together $\underline{v} + \underline{w}$ and multiply them by scalars $\lambda\underline{v}$ where $\underline{v}, \underline{w} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$.

(ii) how to multiply a column vector $\underline{v} \in \mathbb{R}^n$ by an $m \times n$ matrix A to get a new column vector $A\underline{v} \in \mathbb{R}^m$. We will get a lot of profit by treating $\underline{v} \mapsto A\underline{v}$ as a map $\mathbb{R}^n \to \mathbb{R}^m$.

(iii) the dot product $\underline{v} \cdot \underline{w} \in \mathbb{R}$ of two column vectors $\underline{v}, \underline{w} \in \mathbb{R}^n$, and how to use it to compute the length of a vector and the angle between two vectors, or to project one vector onto another.

(iv) how to solve a system of m linear equations in n unknowns (for example by forming combinations of equations to eliminate variables, giving equations in fewer unknowns and then backsolving).

We will review all this briefly, but you should make sure whatever is already familiar is ready to roll. Having said that, do not be tricked into thinking that this is all Linear Algebra is about: after we have set the scene with these calculations, we will state general definitions and prove theorems with a much more formal flavour: you need to stay *en garde*.

In broad terms, we will develop this collection of ideas in different ways:

(i) As a suite of basic calculations that, with care, we can perform algorithmically without error or confusion again and again forever.

(ii) A formal structure that binds together a host of examples and particular cases we may encounter.

(iii) A collection of clean and simple proofs that tie calculations to general theoretical ideas.

If you can recognise the substance of the calculations even when things get more formal, then you will have intuition for most results. It is important that you keep in touch with what you already know well, and see clearly how it ties in with our new much broader and more powerful viewpoints.

You may ignore all side remarks (shaded in yellow) if they are distractions. The definitions (shaded in red, including places that flag idiomatic use of language or conventions) will become part of your DNA by overuse, if that's how DNA works (which it isn't). The real point is to recognise how the theorems capture the essence of calculations in a general setting.

Linear Algebra in the formal way we present it is relatively new. Although a range of calculations have been available for some centuries, the first formal modern treatment I know of is in Birkhoff and Maclane's 1942 book on Algebra. I have no idea whether at the time it seemed completely natural or gratuitous hocus pocus, but today it reads like a standard, and only slightly dated, approach to the subject. For a more applied view, Strang's book is great.

Gavin Brown — Mathematics Institute, University of Warwick
December 2023

Contents

Chapter 1

Column vectors in $\mathbb{R}^n$

We work very frequently in **the plane**, which we may also call **2-space**,

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} a \\ b \end{pmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}$$

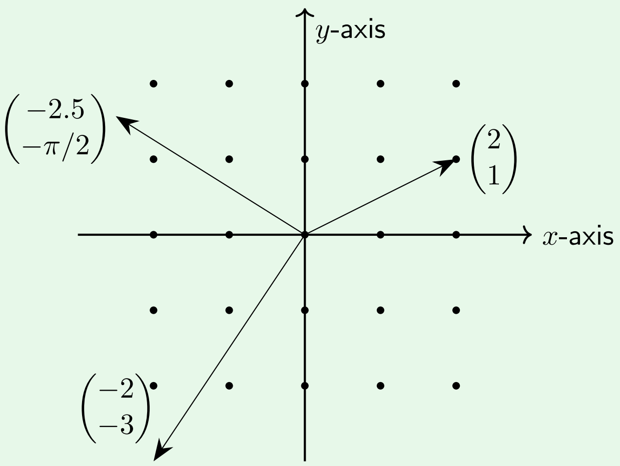
and in 3-dimensional space, which we call **3-space**,

$$\mathbb{R}^3 = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \;\middle|\; a_1, a_2, a_3 \in \mathbb{R} \right\}$$

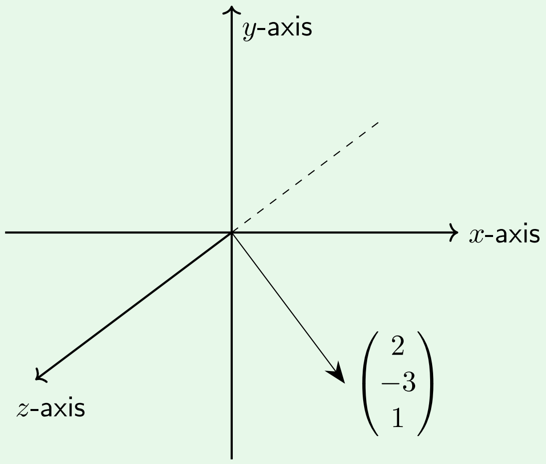
It is important that these are wholly familiar to you, and that you are very good at doing simple calculations with vectors in them. We will review everything we need, but it will be useful if you can revise this now and forever maintain your black belt ninja vector powers at all times.

Example

We usually draw $\mathbb{R}^2$ as a planar picture, with an x-axis and a y-axis, so that it is easy to visualise a square grid and plot vectors (with real coordinates, not only integers) relative to that.



We also try to draw $\mathbb{R}^3$ as a 3-dimensional picture, with the x- and y-axes in the plane and an optical illusion of the z-axis pointing out of the plane: I hope you can see the x-y plane lying flat on the page (or blackboard), with the positive z-axis pointing out in our direction and its negative part behind the page (which the dotted line tries to indicate). We have some intuition for this picture, but it's no help at all for the 17-eyed mathematicians who just landed from some multi-dimensional interstellar void, and it won't help with our formal proofs later either.



As this sketch illustrates, it is almost impossible to draw particular vectors in 3-space in this picture: you have to watch the picture being drawn, think about and trust the labels, and listen carefully to what the artist says they are trying to illustrate – and then draw it again for yourself.

More generally, it is important that we are just as comfortable working in the space

$$\mathbb{R}^n = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \middle| \, a_i \in \mathbb{R} \text{ for all } i = 1, \ldots, n \right\}$$

for any given $n \geq 1$. This is called n-**space**. It is futile trying to imagine what it looks like, unless $n = 1, 2$ or 3: you have to rely on the algebra and any intuition from 3-space that seems to help.

Language 1.1

We refer to each of the spaces $\mathbb{R}^n$ as a **vector space** (and we sometimes say 'over the real numbers' if we wish to emphasise $\mathbb{R}$). The elements of any of these spaces are called **vectors**. We write elements of $\mathbb{R}^n$ as **column** vectors; being systematic about this helps us later. To save space on the page, we may sometimes write $(a_1, a_2, \ldots, a_n)^T$ to denote the column vector

$$(a_1, a_2, \ldots, a_n)^T = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}.$$

The entries, a_1, a_2, and so on, of a vector $\underline{v} = (a_1, a_2, \ldots, a_n)^T$ are called the **components** of $\underline{v}$, and we say that a_i is the i**th component of** $\underline{v}$. The components are real numbers.

Remark

You might wonder why we didn't allow $n = 0$ to give us the vector space $\mathbb{R}^0$. In fact we do, but in this notation it's a bit confusing to contemplate. We will work it out properly later, but for now if you need to think about it just treat $\mathbb{R}^0$ as the set $\{0\}$ with just zero in it.

For this module, vectors have some vital properties that we summarise as:

(i) We can add vectors together (and subtract them) and multiply them by real numbers.

(ii) We can multiply vectors by matrices.

(iii) We can define a notion of length of a vector and angle between vectors.

Those statements alone are just slogans and make no precise sense by themselves. We will work out exactly what we mean by each one.

1.1 Linear combinations and the standard basis

We discuss the fundamental algebraic operations of the vector spaces $\mathbb{R}^n$, and the basic substantial problem that needs solving.

Add, subtract and multiply by scalars

Two vectors $\underline{v}, \underline{w} \in \mathbb{R}^n$ in the same space may be added together componentwise to make a third vector in the same space: if $\underline{v} = (a_1, a_2, \ldots, a_n)^T$ and $\underline{w} = (b_1, b_2, \ldots, b_n)^T$ then

$$\underline{v} + \underline{w} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix} \in \mathbb{R}^n$$

With this operation $+$, you easily see that $\mathbb{R}^n$ is an abelian group with

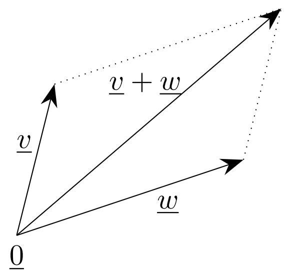
identity element $\underline{0} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathbb{R}^n$ and **inverse** $-\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_n \end{pmatrix}$

In particular

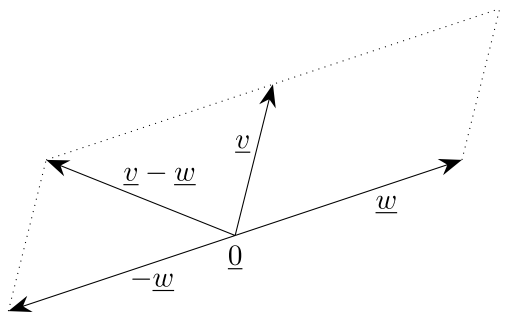
$$\underline{v} - \underline{w} = \begin{pmatrix} a_1 - b_1 \\ a_2 - b_2 \\ \vdots \\ a_n - b_n \end{pmatrix} \in \mathbb{R}^n$$

We also refer to $\underline{0} \in \mathbb{R}^n$ as the **zero vector** or the **origin** of $\mathbb{R}^n$.

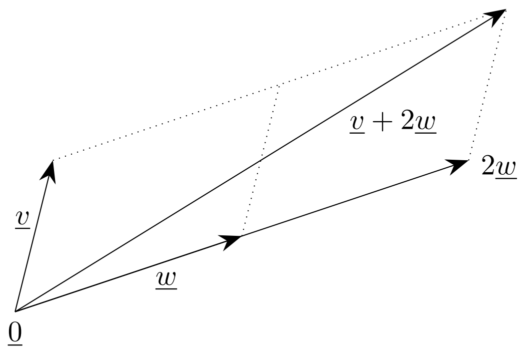
It is usual to think of this addition of vectors as completing a parallelogram with a vertex at the origin and $\underline{v}$ and $\underline{w}$ as two adjacent sides there.



You can do this with more complicated linear combinations too: for example $\underline{v} - \underline{w}$ and $\underline{v} + 2\underline{w}$ appear in the picture above as



and



Convention 1.2

When giving names to vectors in $\mathbb{R}^n$ we usually underline them, writing $\underline{v} \in \mathbb{R}^n$ rather than $v \in \mathbb{R}^n$, to remind ourselves that these are column vectors of real numbers. Of course it doesn't really matter, and there may be exceptions, but it will help later if we fix this convention now.

When we write some vector $\underline{v} \in \mathbb{R}^n$ explicitly as a column vector $\underline{v} = (a_1, \ldots, a_n)^T$, we often say that we are **writing $\underline{v}$ in coordinates**.

Remark

When drawing pictures, it is very useful to draw vectors as arrows, as we have above. But remember that the vector $\underline{v} = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$ is really that single point of $\mathbb{R}^n$ that sits at the tip of the arrow. Sometimes people refer to these as **position vectors** to distinguish them from, for example, velocity vectors or force vectors. In this module we do not need to make the distinction: we happily draw vectors as arrows and remember that they refer to the endpoints.

In particular, we won't draw the kind of phase portrait pictures you might have seen in differential equations, which have a vector based at every point of the picture and you imagine flowing along those lines of velocity: our vectors are based at the origin (except on occasion when it's clearer in a picture to move them around a bit). I can explain the connection at the end.

That's all fine, but $\mathbb{R}^n$ is more than just an abelian group with operation $+$. We can also multiply vectors componentwise by real numbers: if $\lambda \in \mathbb{R}$ and $\underline{v} = (a_1, a_2, \ldots, a_n)^T \in \mathbb{R}^n$ then

$$\lambda\underline{v} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix} \in \mathbb{R}^n$$

Lemma 1.3

Let $\underline{v}, \underline{w} \in \mathbb{R}^n$ be vectors and $\lambda, \mu \in \mathbb{R}$ be scalars. Then

(i) $\underline{v} + \underline{w} = \underline{w} + \underline{v}$

(ii) $\lambda(\underline{v} + \underline{w}) = \lambda\underline{v} + \lambda\underline{w}$ and $(\lambda + \mu)\underline{v} = \lambda\underline{v} + \mu\underline{v}$

(iii) $\lambda(\mu\underline{v}) = (\lambda\mu)\underline{v}$

(iv) $0\underline{v} = \underline{0}$, $1\underline{v} = \underline{v}$, $(-1)\underline{v} = -\underline{v}$, $\underline{v} + \underline{v} = 2\underline{v}$ and $\underline{v} - \underline{v} = \underline{v} + (-\underline{v}) = \underline{0}$.

The proof illustrates how useful it can be to consider each component of a vector separately: a vector in $\mathbb{R}^n$ is nothing more or less than the data of its components in that order.

Proof. We prove (ii). Let a_i be the ith component of $\underline{v}$ and b_i be the ith component of $\underline{w}$. Then the ith component of $\lambda(\underline{v} + \underline{w})$ is, by definition, $\lambda(a_i + b_i)$, while the ith component of $\lambda\underline{v} + \mu\underline{v}$ is, by definition, $\lambda a_i + \mu b_i$. These are evidently equal, since all quantities are real numbers. Since every component of the vectors $\lambda(\underline{v} + \underline{w})$ and $\lambda\underline{v} + \lambda\underline{v}$ are the same, these two vectors are the same. The other parts of the lemma are similar. □

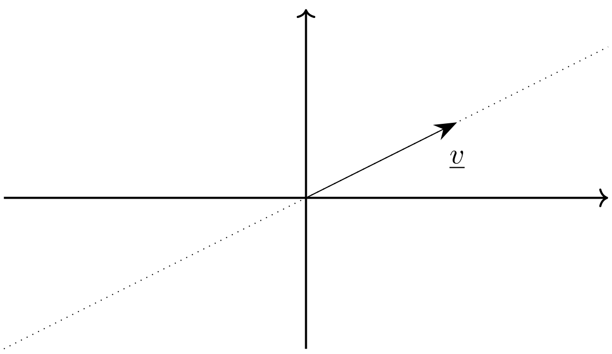
The calculations in part (iv) may look confusing at first, but they are clear if you unpick them carefully. For example, in the first one, on the left there is a vector $\underline{v}$ multiplied by the scalar zero, $0 \in \mathbb{R}$, while on the right is the zero vector $\underline{0} \in \mathbb{R}^n$. That's just what you expect: if you scale a vector by zero, you get the zero vector – but it does need checking (just this once) to confirm that our expectations do indeed match the definitions, and so do the other points.

Convention 1.4

We refer to real numbers as **scalars**: they are used to scale vectors, after all. The point is that later we may use other scalars (complex numbers, for example). So from here on we will refer to 'multiplication by scalars' or 'scalar multiplication' when we multiply a vector by a real number.

Linear combinations

For any nonzero vector $\underline{v} \in \mathbb{R}^n$, we may consider the straight line along it that passes through the origin. This line consists exactly of all vectors of the form $\lambda\underline{v}$ as λ varies through all elements of $\mathbb{R}$. In the picture below, this line is indicated by the dotted line (which you should imagine continuing indefinitely in both directions): you can imagine drawing arrows for each of the vectors $3\underline{v}$, $-2\underline{v}$, $\frac{3}{2}\underline{v}$, $\frac{\pi}{7}\underline{v}$ and so on (even including $0\underline{v} = \underline{0}$), and they would all lie along the dotted line.



With that in mind, we say that two vectors are collinear if they lie on the same line **through the origin**. Of course the formal definition does not use the idea of the picture: it uses scalar multiples.

Definition 1.5

Two vectors $\underline{v}$ and $\underline{w} \in \mathbb{R}^n$ are **collinear** if either $\underline{v} = \lambda \underline{w}$ or $\underline{w} = \lambda \underline{v}$ for some $\lambda \in \mathbb{R}$.

This definition is fine but a little clumsy. It has to cope with the fact that either $\underline{v}$ or $\underline{w}$ or both might be the zero vector $\underline{0} \in \mathbb{R}^n$, which is why it offers the scaling by λ both ways round. Note again that this notion of collinearity is referring only to lines that pass **though the origin**.

The ideas of addition and multiplication by scalars immediately produce more complicated expressions: linear combinations of vectors. If $\underline{v}, \underline{w} \in \mathbb{R}^n$ and $\lambda, \mu \in \mathbb{R}$, then the expression (the vector)

$$\lambda \underline{v} + \mu \underline{w} \in \mathbb{R}^n$$

is called a **linear combination** of $\underline{v}$ and $\underline{w}$, with (scalar) coefficients λ and μ respectively. The same notion works with any finite number of vectors.

Language 1.6

For any vectors $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_s \in \mathbb{R}^n$ and scalars $\lambda_1, \lambda_2, \ldots, \lambda_s \in \mathbb{R}$, the expression

$$\lambda_1 \underline{v}_1 + \lambda_2 \underline{v}_2 + \ldots + \lambda_s \underline{v}_s \in \mathbb{R}^n.$$

is called a **linear combination of** $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_s \in \mathbb{R}^n$. As usual, we may use the summation symbol to abbreviate this linear combination as

$$\sum_{i=1}^{s} \lambda_i \underline{v}_i$$

but it's often worth continuing to write out such summations in the long form with $+ \ldots +$ until you're wholly comfortable with the abbreviation.

We call such expressions a **nontrivial linear combination** if at least one of the λ_i is not zero, and conversely when all the $\lambda_i = 0$ we call the expression the **trivial linear combination**.

This idea has a particularly simple but important case.

Definition 1.7

The **standard basis** of $\mathbb{R}^n$ is the collection of n vectors

$$\underline{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ \underline{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \ \ldots, \ \underline{e}_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \in \mathbb{R}^n$$

That is $\underline{e}_i$ is the vector whose components are all zero except the ith component which is 1.

You will use the standard basis all the time. The point is that any other vector may be written uniquely (up to the order in which you write the sum) as a linear combination of the standard basis:

$$\text{if } \underline{v} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n \text{ then } \underline{v} = a_1\underline{e}_1 + \ldots + a_n\underline{e}_n = \sum_{i=1}^{n} a_i\underline{e}_i$$

The fundamental problem

The basic problem in the subject is this: given some vectors $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$ and a **target vector** $\underline{b} \in \mathbb{R}^n$, can you find scalars $\lambda_1, \ldots, \lambda_s \in \mathbb{R}$ for which

$$\lambda_1\underline{v}_1 + \ldots + \lambda_s\underline{v}_s = \underline{b}$$

or can you prove that no such scalars exist? Furthermore, if you can find one solution, can you go on to find all possible ways this can be done?

Example

Let $\underline{v}_1 = (2, 1, 3)^T$ and $\underline{v}_2 = (1, -1, 1)^T \in \mathbb{R}^3$. We consider three different target vectors:
(1) $\underline{b} = \underline{0}$ (2) $\underline{b} = (1, 5, 3)^T$ (3) $\underline{b} = \underline{e}_1 = (1, 0, 0)^T$.

For (1), we look for $\lambda_1, \lambda_2 \in \mathbb{R}$ such that

$$\lambda_1 \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Considering the three components separately, this is exactly the same as solving the following three linear equations simultaneously:

$$\begin{aligned} 2\lambda_1 + \lambda_2 &= 0 \\ \lambda_1 - \lambda_2 &= 0 \\ 3\lambda_1 + \lambda_2 &= 0 \end{aligned}$$

You can be ad hoc or systematic or mysterious or magical about how you solve such equations – you probably have your own favourite methods, and how you do it probably depends on exactly what the equations are. In this case you could say that the second equation says that $\lambda_1 = \lambda_2$, and then substituting into either of the other two equations shows that they are both zero. Thus, there is exactly one solution $\lambda_1 = \lambda_2 = 0$. Later we will say that $\underline{v}_1$ and $\underline{v}_2$ are **linearly independent** because of this.

For (2), we look for $\lambda_1, \lambda_2 \in \mathbb{R}$ such that

$$\lambda_1 \begin{pmatrix} 2 \\ 1 \\ 3 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 5 \\ 3 \end{pmatrix} \tag{1.1}$$

Considering the three components separately, this is exactly the same as solving

$$\begin{aligned} 2\lambda_1 + \lambda_2 &= 1 \\ \lambda_1 - \lambda_2 &= 5 \\ 3\lambda_1 + \lambda_2 &= 3 \end{aligned}$$

The equations are a bit harder to solve, but not much. For example, adding the first and second equations together gives $3\lambda_1 = 6$ so if there is any solution at all it must have $\lambda_1 = 2$. Plugging that value of λ_1 into the first equation gives $4+\lambda_2 = 1$, so if there is any solution at all it must have $\lambda_2 = -3$. Finally we check that the pair of values $(\lambda_1, \lambda_2) = (2, -3)$ satisfies all three equations: it does, so again we have a **unique solution** – and the wise mathematician quickly checks that this really does solve the original problem (1.1) to avoid any daft mistakes.

We approach (3) in the same way. After considering components, we get

$$\begin{aligned} 2\lambda_1 + \lambda_2 &= 1 \\ \lambda_1 - \lambda_2 &= 0 \\ 3\lambda_1 + \lambda_2 &= 0 \end{aligned}$$

We attempt to solve them as before: adding the first and second equations together gives $3\lambda_1 = 1$ so if there is any solution at all it must have $\lambda_1 = 1/3$. Plugging that value of λ_1 into the first equation gives $2/3 + \lambda_2 = 1$, so if there is any solution at all it must have $\lambda_2 = 1/3$. Finally we check whether the pair of values $(\lambda_1, \lambda_2) = (1/3, 1/3)$ satisfies all three equations: but now while it is a solution of the first two equations (necessarily, given the way we found it), it is not a solution of the third. In this case there is **no solution at all**.

It is useful to think of this exercise in geometrical terms, without being too precise about things. The collection of all possible vectors $\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2$ describes a (flat, linear) plane W through the origin in $\mathbb{R}^3$. The origin is clearly on the plane since we may choose $\lambda_1 = \lambda_2 = 0$, and this is what happens in case (1): the only question is how many different solutions are there, and in this example there was only one.

In case (2), the vector $\underline{b} = (1, 5, 3)^T$ happens to lie on this plane W, which we reveal by finding values for λ_1 and λ_2. The question is then how many different solutions are there, and the answer is the same as in the previous case: there was precisely one solution then, and so there is also precisely one solution in this case (necessarily, as we shall prove later).

In case (3), the vector $\underline{e}_1$ does not lie on W: we discover this when we find that there are no solutions at all for the pair λ_1, λ_2. If we could have seen the picture and spotted that $\underline{e}_1 \notin W$, then we would not have had to do any work at all to say that there cannot be any solutions.

As you see, this problem is exactly the same problem as solving systems of simultaneous linear equations. We will build a powerful machine for solving all three of these problems in Chapter 2 below, but for now it is useful practice to work out solutions with our bare hands.

Example

Consider $\underline{v}_1 = (2, -1)^T$, $\underline{v}_2 = (1, 1)^T$ and $\underline{v}_3 = (3, 1)^T \in \mathbb{R}^2$ and two different target vectors:
(1) $\underline{b} = \underline{0}$ (2) $\underline{b} = (7, -1)^T$.

For (1), we look for $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ such that

$$\lambda_1 \begin{pmatrix} 2 \\ -1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Considering the two components separately gives

$$\begin{aligned} 2\lambda_1 + \lambda_2 + 3\lambda_3 &= 0 \\ -\lambda_1 + \lambda_2 + \lambda_3 &= 0 \end{aligned}$$

To solve this, we might add twice the second equation to the first, **eliminating λ_1 from the equations**, to get $3\lambda_2 + 5\lambda_3 = 0$. This has lots of solutions: for any value of $\lambda_3 \in \mathbb{R}$, simply choose $\lambda_2 = -(5/3)\lambda_3$. Then we could use the second equation to calculate

$$\lambda_1 = \lambda_2 + \lambda_3 = -\frac{2}{3}\lambda_3$$

Finally we need to check whether these solutions also satisfy the first equation: they do, since

$$2\lambda_1 + \lambda_2 + 3\lambda_3 = 2 \times \frac{-2}{3}\lambda_3 - \frac{5}{3}\lambda_3 + 3\lambda_3 = \underline{0}$$

Thus, there is a whole 1-dimensional set of solutions: we have one degree of freedom to choose $\lambda_3 \in \mathbb{R}$ just as we please, and then suitable values for λ_1 and λ_2 are determined by the equations. For example, choosing $\lambda_3 = -3$ gives $\lambda_1 = 2$ and $\lambda_2 = 5$. Later we will say that $\underline{v}_1$, $\underline{v}_2$ and $\underline{v}_3$ are **linearly dependent**: they satisfy a non-trivial linear relation $2\underline{v}_1 + 5\underline{v}_2 - 3\underline{v}_3 = \underline{0}$, to use the solution we just picked, or any (nonzero) multiple of that.

For (2), we proceed as before: after considering components we obtain the simultaneous linear equations

$$\begin{aligned} 2\lambda_1 + \lambda_2 + 3\lambda_3 &= 7 \\ -\lambda_1 + \lambda_2 + \lambda_3 &= -1 \end{aligned}$$

which we solve as before to get $\lambda_2 = -(5/3)\lambda_3 + (5/3)$ and then

$$\lambda_1 = \lambda_2 + \lambda_3 + 1 = -(2/3)\lambda_3 + (8/3)$$

for any value of $\lambda_3 \in \mathbb{R}$ (please check this). Again we can get a particular solution by picking any value of λ_3. For example $\lambda_3 = 1$ gives $\lambda_1 = 2$ and $\lambda_2 = 0$. Again there is a one degree of freedom in the choice of solution $\lambda_3 \in \mathbb{R}$: given that we have a solution at all, the number of solutions is the same as in the case $\underline{b} = \underline{0}$.

The machine we build in Chapter 2 makes this kind of calculation efficient and systematic: it does not break the problem down into components, but abstracts it to working on matrices.

1.2 Dot product: length, angle, orthonormal vectors

This section has a different flavour. The vector space $\mathbb{R}^n$ has another operation that you will know well: the **dot product**. It's good to note that this is an additional structure, beyond merely the linear combinations that make $\mathbb{R}^n$ into a vector space, but for now we can safely bundle it all together.

Definition 1.8

For $\underline{v} = (a_1, \ldots, a_n)^T$, $\underline{w} = (b_1, \ldots, b_n)^T \in \mathbb{R}^n$, the **dot product** of $\underline{v}$ and $\underline{w}$, denoted $\underline{v} \cdot \underline{w}$, is the scalar

$$\underline{v} \cdot \underline{w} = a_1 b_1 + a_2 b_2 + \ldots + a_n b_n = \sum_{i=1}^{m} a_i b_i \in \mathbb{R}$$

This is also called the **scalar product**, and we will use the two terms interchangeably.

Example

The scalar product in $\mathbb{R}^2$ of $(1,-3)^T$ and $(5,2)^T$ is the scalar

$$\begin{pmatrix}1\\-3\end{pmatrix}\cdot\begin{pmatrix}5\\2\end{pmatrix} = 1\times 5 + (-3)\times 2 = 5-6 = -1$$

Proposition 1.9

The scalar product of vectors in $\mathbb{R}^n$ satisfies the following:

(i) $\underline{v}\cdot\underline{w} = \underline{w}\cdot\underline{v}$ for any $\underline{v},\underline{w}\in\mathbb{R}^n$

(ii) $(\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2)\cdot\underline{w} = \lambda_1(\underline{v}_1\cdot\underline{w}) + \lambda_2(\underline{v}_2\cdot\underline{w})$

(iii) For any $\underline{v}\in\mathbb{R}^n$, $\underline{v}\cdot\underline{v}\geq 0$, and furthermore $\underline{v}\cdot\underline{v}=0$ if and only if $\underline{v}=\underline{0}$ is the zero vector.

Properties (i) and (ii) are often referred to by saying that scalar product is **bilinear**: you can expand out linear combinations in the first factor, and by switching the factors around using the first property you can also expand out linear combinations in the second factor.

Proof. Write $\underline{v} = (a_1,\ldots,a_n)^T$ and $\underline{w} = (c_1,\ldots,c_n)^T$ in coordinates. Part (i) is then immediate since $a_ic_i = c_ia_i$ for each i.

In the same notation, part (iii) is also clear: $\underline{v}\cdot\underline{v} = a_1^2 + \ldots + a_n^2$ is a sum of squares, so cannot be negative. Furthermore, the sum can only be zero if each $a_i^2 = 0$, and that only happens if each $a_i = 0$, which is the claim.

Part (ii) is almost immediate too: write $\underline{v}_1 = (a_1,\ldots,a_n)^T$ and $\underline{v}_2 = (b_1,\ldots,b_n)^T$, with $\underline{w}$ as before, and then the ith component on the left-hand side is

$$(\lambda_1a_i + \lambda_2b_i)c_i = \lambda_1a_ic_i + \lambda_2b_ic_i$$

which equals the ith component of the right-hand side. □

Lengths of vectors

Since by Proposition 1.9(iii) the dot product of any vector with itself is not negative, we may always form the square root $\sqrt{\underline{v}\cdot\underline{v}}\in\mathbb{R}$ as a real number. Therefore the following definition makes sense.

Definition 1.10 (Length of a vector)

We define the **length** of a vector $\underline{v}\in\mathbb{R}^n$, denoted $\|\underline{v}\|$, to be

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

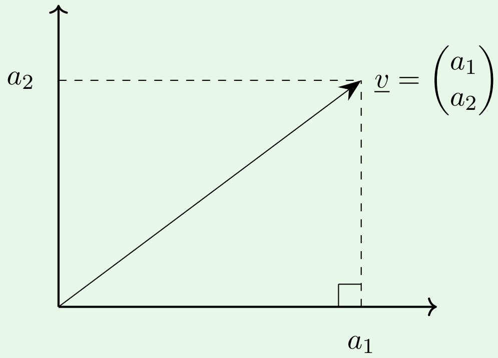
which is a non-negative real number. (Notice the double lines in the notation.)

Example 1.11

In $\mathbb{R}^2$, this definition of length of a vector is what you know from Pythagoras's theorem: if $\underline{v} = (a_1, a_2)^T$, then

$$\|\underline{v}\| = \sqrt{a_1^2 + a_2^2}$$

is the length of the hypotenuse of a right-angled triangle:



Proposition 1.12

The length of a vector determines a function $\mathbb{R}^n \to \mathbb{R}$, given by $\underline{v} \mapsto \|\underline{v}\|$, that satisfies the following properties: for any $\underline{v} \in \mathbb{R}^n$ and any $\lambda \in \mathbb{R}$,

(i) $\|\underline{v}\| \geq 0$, and equality holds if and only if $\underline{v} = \underline{0}$ is the zero vector

(ii) $\|\lambda\underline{v}\| = |\lambda|\|\underline{v}\|$

(iii) If $\underline{v} \in \mathbb{R}^n \setminus \{\underline{0}\}$ is a nonzero vector, then

$$\hat{\underline{v}} = \frac{1}{\|\underline{v}\|}\underline{v}$$

is a vector of length 1 that is collinear with $\underline{v}$.

Convention 1.13

The vector $\hat{\underline{v}}$ in part (iii) is often referred to as the **unit vector in the direction of** $\underline{v}$. It is characterised by having length 1 and being a positive multiple of $\underline{v}$. The convention of putting a hat (or circumflex if you prefer) over the vector to indicate this unit vector is standard.

Proof. (i) is Proposition 1.9(iii) expressed in the language of $\|\underline{v}\|$.

For (ii), write $\underline{v} = (a_1, \ldots, a_n)$ in coordinates. Then

$$\|\lambda\underline{v}\|^2 = (\lambda a_1)^2 + \ldots + (\lambda a_n)^2 = \lambda^2(a_1^2 + \ldots + a_n^2) = \lambda^2\|\underline{v}\|^2$$

and taking (positive) square roots (hence the modulus sign for $|\lambda|$) proves (ii). Finally (iii) follows from (ii) by setting $\lambda = 1/\|\underline{v}\| > 0$. □

The angle between vectors

Most people find that definition of length acceptable – it agrees with what we already know, and has the properties we expect of lengths (including the triangle inequality in Proposition 1.16 below).

The definition of angle between vectors takes more absorbing. Let's just state it for now and think about it afterwards – though you might ask yourself what your own definition of angle between vectors is, and perhaps realise you don't have one (you probably don't even own a protractor any more).

Definition 1.14 (Angle between vectors)

Let $\underline{v}, \underline{w} \in \mathbb{R}^n \setminus \{\underline{0}\}$ be nonzero vectors. We define the **angle between** $\underline{v}$ **and** $\underline{w}$, denoted $\angle \underline{v}\underline{w}$, to be the real number

$$\angle \underline{v}\underline{w} = \cos^{-1}\left(\frac{\underline{v} \cdot \underline{w}}{\|\underline{v}\|\|\underline{w}\|}\right)$$

where we take the principal preimage of $\cos$, so that $\angle \underline{v}\underline{w}$ lies in the interval $[0, \pi]$.

It is sometimes simpler or comforting to write $\vartheta = \angle \underline{v}\underline{w}$ and express the complicated formula above in rearranged form as

$$\underline{v} \cdot \underline{w} = \|\underline{v}\|\|\underline{w}\| \cos(\vartheta) \quad \text{or even (assuming } \|\underline{v}\|\|\underline{w}\| \neq 0) \quad \hat{\underline{v}} \cdot \hat{\underline{w}} = \cos(\vartheta)$$

but remember that this is not a formula we have derived from anything: it is the very definition of the angle itself.

Example

Let $\underline{v} = (4, -2)^T$ and $\underline{w} = (1, 2)^T \in \mathbb{R}^2$. First note that they are both nonzero vectors. We calculate the angle ϑ between these vectors thus: $\underline{v} \cdot \underline{w} = 0$, so we must have $\cos(\vartheta) = 0$, and so $\vartheta = \pi/2$ (or 90°, if you prefer degrees to radians). That is, the vectors are at right angles to one another, which matches what you see when you draw the picture.

This works in $\mathbb{R}^3$ too, indeed in any $\mathbb{R}^n$. For example $(1, 1, 1)^T$ and $(a, b, c)^T \in \mathbb{R}^3$ are at right angles whenever their dot product is zero, that is whenever $a + b + c = 0$.

You will have noticed the terrible hole in our definition of angle: we need to know that the scalar quantity $(\underline{v} \cdot \underline{w})/(\|\underline{v}\|\|\underline{w}\|)$ lies in the interval $[-1, 1]$, otherwise it does not have a preimage under $\cos$. This is what the famous Cauchy–Schwartz inequality does for us: in absolute value, the numerator is no bigger than the denominator – phew!

Proposition 1.15 (Cauchy–Schwartz inequality)

For any $\underline{v}, \underline{w} \in \mathbb{R}^n$,

$$|\underline{v} \cdot \underline{w}| \leq \|\underline{v}\|\|\underline{w}\|$$

and furthermore equality is achieved only when $\underline{v}$ and $\underline{w}$ are collinear (recall Definition 1.5).

Proof. It is enough to prove that

$$(\underline{v} \cdot \underline{w})^2 \leq \|\underline{v}\|^2\|\underline{w}\|^2 \tag{1.2}$$

since taking the (positive) square root gives the result (including the modulus sign on the left).

Denoting the ith component of $\underline{v}$ as a_i and that of $\underline{w}$ as b_i, we calculate the right-hand side minus the left-hand side:

$$\begin{aligned}\|\underline{v}\|^2\|\underline{w}\|^2 - (\underline{v}\cdot\underline{w})^2 &= \left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right) - \left(\sum_{i=1}^n a_i b_i\right)^2 \\ &= \sum_{i=1}^n \sum_{j=i+1}^n (a_i b_j - a_j b_i)^2 \qquad (!!)\end{aligned}$$

which is a sum of squares, so is clearly non-negative, as claimed.

We must check the last claim. If $\underline{v} = \underline{0}$, then there is nothing to prove, so suppose without loss of generality that $a_1 \neq 0$. In that case, since $a_1 b_j - a_j b_1 = 0$, we have that each $b_j = a_j(b_1/a_1)$: that is, $\underline{w} = (b_1/a_1)\underline{v}$, and the vectors are collinear as claimed. □

That's fine, but hands up anyone who understood the line at (!!) in that proof: it is true, but you have to be able to read summation signs like Shakespeare on steroids or prove the equality for yourself somehow (by induction, say). At the very least, you should write out what it means when $n = 2$ and $n = 3$. If you do wish to prove it by induction, in fact it's not so bad:

$$\begin{aligned}\left(\sum_{i=1}^n a_i^2\right)\left(\sum_{i=1}^n b_i^2\right) &- \left(\sum_{i=1}^n a_i b_i\right)^2 \\ &= \text{terms involving only subscripts} \le n-1 \text{ that you handle by induction} \\ &\quad + a_n^2\left(b_1^2 + \ldots + b_{n-1}^2\right) + \left(a_1^2 + \ldots + a_{n-1}^2\right) b_n + \cancel{a_n^2 b_n^2} \\ &\quad - 2\left(a_1 b_1 + \ldots + a_{n-1} b_{n-1}\right) a_n b_n - \cancel{a_n^2 b_n^2}\end{aligned}$$

Now consider where we're trying to get to: for each $i = 1, \ldots, n-1$,

$$(a_i b_n - a_n b_i)^2 = a_i^2 b_n^2 + a_n^2 b_i^2 - 2(a_i b_i)(a_n b_n)$$

which accounts for all the terms in the big equation above that have a subscript i; so we are done.

If you don't like (!!) or that inductive check, here's another proof of the main point (1.2).

Proof. Certainly (1.2) holds if $\underline{v} = \underline{0}$: both sides are zero, and $\underline{v} = 0\underline{w}$ is collinear with $\underline{w}$. So suppose $\underline{v} \neq \underline{0}$. Let $\hat{\underline{v}} = \underline{v}/\|\underline{v}\|$ and define $\underline{u} = \underline{w} - \mu\hat{\underline{v}}$ where $\mu = \hat{\underline{v}}\cdot\underline{w}$. Then, since $\hat{\underline{v}}\cdot\hat{\underline{v}} = 1$,

$$\begin{aligned}0 \le \|\underline{u}\|^2 = \underline{u}\cdot\underline{u} &= \underline{w}\cdot\underline{w} - \underline{w}\cdot\mu\hat{\underline{v}} - \mu\hat{\underline{v}}\cdot\underline{w} + \mu^2 \\ &= \|\underline{w}\|^2 - 2\mu(\hat{\underline{v}}\cdot\underline{w}) + \mu^2 \\ &= \|\underline{w}\|^2 - (\hat{\underline{v}}\cdot\underline{w})^2 \\ &= \|\underline{w}\|^2 - \frac{1}{\|\underline{v}\|^2}(\underline{v}\cdot\underline{w})^2\end{aligned}$$

and the result follows by multiplying through by $\|\underline{v}\|^2 > 0$. □

And here's a more stylish proof of (1.2) if you don't like that one.

Proof. Again it suffices to consider the case $\underline{v} \neq \underline{0}$. Consider the polynomial (in a variable x)

$$f(x) := (a_1 x + b_1)^2 + \ldots + (a_n x + b_n)^2 = Ax^2 + Bx + C$$

where $A = \sum_{i=1}^n a_i^2 = \|\underline{v}\|^2 > 0$, $B = 2\sum_{i=1}^n a_i b_i = 2\underline{v}\cdot\underline{w}$ and $C = \sum_{i=1}^n b_i^2 = \|\underline{w}\|^2$.

For any $x \in \mathbb{R}$, the value of $f(x)$ is by definition a sum of squares, so it is zero if and only if each summand $(a_i x + b_i)^2$ is zero. Thus $f(x)$ can have at most one real root, namely $x = -b_i/a_i$ if $a_i \neq 0$, and moreover all such values must agree (and moreover[-squared] we must have $b_i = 0$ whenever $a_i = 0$).

Now $f(x)$ is a quadratic polynomial, so we know all about its real roots: it has two distinct real roots unless $B^2 - 4AC \leq 0$. But (up to a redundant factor of 4) that inequality is exactly what we need to prove – it is what (!!) does for us above – so we are done. □

The Cauchy–Schwartz inequality rescues our definition of angle, and it also proves the triangle inequality for lengths.

Proposition 1.16 (Triangle inequality)

The length $\| \ \|$ of a vector satisfies the triangle inequality: for any $\underline{v}, \underline{w} \in \mathbb{R}^n$,

$$\|\underline{v} + \underline{w}\| \leq \|\underline{v}\| + \|\underline{w}\|$$

and equality holds if and only if $\underline{v}$ and $\underline{w}$ are collinear.

This follows from the definition of length, the bilinearity of scalar product (Proposition 1.9(i–ii)) and Cauchy–Schwartz: once you see the first line, you can follow your nose.

Proof. We compute the length-squared of $\underline{v} + \underline{w}$:

$$\begin{aligned} \|\underline{v} + \underline{w}\|^2 &= (\underline{v} + \underline{w}) \cdot (\underline{v} + \underline{w}) \\ &= \underline{v} \cdot \underline{v} + \underline{v} \cdot \underline{w} + \underline{w} \cdot \underline{v} + \underline{w} \cdot \underline{w} \\ &\leq \|\underline{v}\|^2 + |\underline{v} \cdot \underline{w}| + |\underline{w} \cdot \underline{v}| + \|\underline{w}\|^2 \\ &\leq \|\underline{v}\|^2 + 2\|\underline{v}\|\|\underline{w}\| + \|\underline{w}\|^2 \\ &= (\|\underline{v}\| + \|\underline{w}\|)^2. \end{aligned}$$

Thus $\|\underline{v} + \underline{w}\|^2 = (\|\underline{v}\| + \|\underline{w}\|)^2$, and taking the (positive) square root gives the result. □

Orthonormal sets of vectors and orthogonal projection

The ideas of length and angle give a first indication why the standard basis is so useful. (Later you may think that this discussion is tautological, but for now let's go with it.) First recall the standard delta function.

Definition 1.17

The Kronecker delta function δ_{ij} is defined by

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

The definition is coy about what i and j actually are, but in our context they will always be integers in some specified range such as $i, j \in \{1, \ldots, s\}$.

Definition 1.18

A set of vectors $\underline{v}_1, \dots, \underline{v}_s \in \mathbb{R}^n$ is **orthonormal** if and only $\underline{v}_i \cdot \underline{v}_j = \delta_{ij}$ for each $i, j = 1, \dots, s$.

Spelling that definition out, $v_1, \dots, v_s$ are orthonormal if and only if they all have length 1 and they are pairwise at right angles to one another.

Example

The standard basis is the fundamental example of an orthonormal set of vectors.

As a more exotic example in $\mathbb{R}^2$, for any fixed $\vartheta \in \mathbb{R}$ consider

$$\underline{v}_1 = \begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix}, \quad \underline{v}_2 = \begin{pmatrix} -\sin\vartheta \\ \cos\vartheta \end{pmatrix}$$

It is useful to draw this picture: you will see the standard basis of $\mathbb{R}^2$ rotated by ϑ. Can you persuade yourself that these are the only orthonormal sets of two vectors in $\mathbb{R}^2$ (not considering the order in which you write them)?

There is a lot more to say about dot products and orthonormal vectors, but for now let's consider just one more point: the orthogonal projection of one vector onto another. First a seeming triviality.

Example 1.19

Although it seems ridiculous to say it in this context, dot product with the standard basis is a formal way of finding the ith component of a vector: if $\underline{v} = (a_1, \dots, a_n)^T \in \mathbb{R}^n$ and $\underline{e}_1, \dots, \underline{e}_n$ is the standard basis of $\mathbb{R}^n$, then

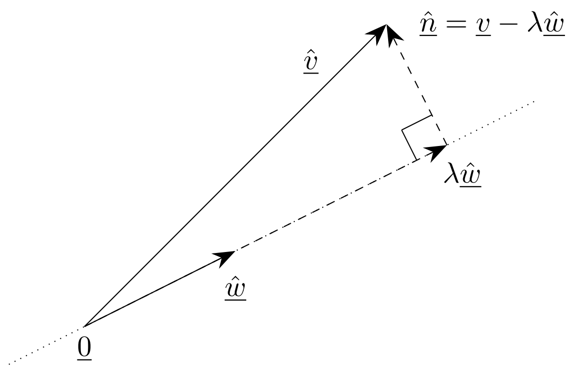
$$a_i = \underline{e}_i \cdot \underline{v} \qquad \text{for each } i = 1, \dots, n$$

Put differently, even if nobody told you what the components a_i of $\underline{v}$ were, you could still find them out at once using this formula.

When we consider orthogonal vectors in Euclidean spaces more generally later, this idea will seem more subtle, even though it is exactly the same.

Pictorially, we may regard the ith component of $\underline{v}$ as the length of the projection of $\underline{v}$ onto $\underline{e}_i$: compare the picture in Example 1.11. The calculation in the example above then gives us a method of calculating this length.

This idea works more generally. Consider the following picture: the vector $\lambda\hat{\underline{w}}$ is the **orthogonal projection** of $\underline{v}$ onto the line through the unit vector $\hat{\underline{w}}$, though the required scalar multiple λ is yet to be calculated.



To calculate λ, we define $\underline{\hat{n}} = \underline{v} - \lambda\underline{\hat{w}}$ and simply notice that

$$\underline{n} \cdot \underline{\hat{w}} = \underline{v} \cdot \underline{\hat{w}} - \lambda\|\underline{\hat{w}}\|^2 = \underline{v} \cdot \underline{\hat{w}} - \lambda$$

since $\|\underline{\hat{w}}\| = 1$. Therefore $\lambda\underline{\hat{w}}$ is orthogonal to $\underline{\hat{n}}$ (as in the picture) if and only if $\lambda = \underline{v} \cdot \underline{\hat{w}}$.

Definition 1.20

Let $\underline{v}, \underline{w} \in \mathbb{R}^n$ with $\underline{w} \neq \underline{0}$ and let $\underline{\hat{w}} = \underline{w}/\|\underline{w}\|$ be the unit vector in the direction of $\underline{w}$. Then the scalar quantity $\underline{v} \cdot \underline{\hat{w}}$ is called **component of $\underline{v}$ in the direction of $\underline{w}$**, and the vector $(\underline{v} \cdot \underline{\hat{w}})\underline{\hat{w}}$ is the **orthogonal projection of $\underline{v}$ in the direction of $\underline{w}$**.

Notice that the definition uses any nonzero vector $\underline{w}$, while the calculation uses the unit vector $\underline{\hat{w}}$: they both serve the role of pinpointing the line that $\underline{v}$ is projected onto, but it is important that $\underline{\hat{w}}$ has length 1 for the calculation – or you could write $((\underline{v} \cdot \underline{w})/\|\underline{w}\|^2)\underline{w}$ instead, if you don't like hats.

Since the vectors of the standard basis have length 1, the trivial calculation of Example 1.19 is computing exactly the component of $\underline{v}$ in the directioin of each of the standard basis elements, and it confirms that these are simply the usual components of $\underline{v}$. The more general calculation in Definition 1.20 is just as simple to use.

1.3 Geometry of lines and planes in $\mathbb{R}^3$

First warm up in the plane $\mathbb{R}^2$. Everyone knows that the equation of a line $L \subset \mathbb{R}^2$ in the plane (by which we always mean a straight line, extending indefinitely in both directions) is of the form

$$L\colon (y = mx + c) \subset \mathbb{R}^2$$

for suitable $m, c \in \mathbb{R}$. Well, that's not quite true, since it doesn't cover vertical lines $(x = a)$, so to overcome this prejudice about y being better than x (and whatever it is that m and c are supposed to stand for) let's write it instead as

$$L\colon (ax + by = c) \subset \mathbb{R}^2$$

where $a, b, c \in \mathbb{R}$ and we insist that $(a, b) \neq (0, 0)$. It is clear that the line L passes through the origin if and only if $c = 0$.

Example

What is the equation of the line $L_{PQ} \subset \mathbb{R}^2$ through the points $P = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$ and $Q = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$?

First a sanity check: these two points are distinct, so we know from geometry (or experience) that there is indeed a unique straight line passing through them (though this will also follow from the calculation below). The equation of that line must be $ax+by=c$ for some $a,b,c \in \mathbb{R}$, and we note at once that these are not uniquely determined, so your answer may be different from mine, since multiplying them all by a nonzero scalar $\lambda \neq 0$ does not change the line.

With that all in mind, we treat a,b,c as unknowns, substitute the two points into the unknown equation and solve:

$$\begin{array}{lrcl} \text{at } P\colon & -2a+b & = & c \\ \text{at } Q\colon & 4a+3b & = & c \end{array}$$

which has solution $a = -(1/5)\lambda$, $b = (3/5)\lambda$ and $c = \lambda$ for any $\lambda \in \mathbb{R}$ (one degree of freedom, as we observed above), and we simply choose any nonzero solution such as $(a,b,c) = (-1,3,5)$ (setting $\lambda = 5$), giving

$$L_{PQ}\colon (-x+3y=5) \subset \mathbb{R}^2.$$

Of course we check at once that P and Q really do lie on L_{PQ} to avoid any daft error.

Remark

We worked out the example above having our usual presentation of linear equations in mind. If you had to do this quickly on the bus, you'd probably say something more like this. The equation of L_{PQ} must be of the form

$$L_{PQ}\colon \lambda(x-4) = \mu(y-3)$$

for suitable $\lambda, \mu \in \mathbb{R}$ (4 and 3 were chosen to ensure that Q lies on this line). Plugging in the coordinates of P gives $-6\lambda = -2\mu$, which has nontrivial solution $\lambda = 1$, $\mu = 3$, so as before

$$L_{PQ}\colon x-4 = 3(y-3) \qquad \text{or } -x+3y=5, \text{ if you prefer.}$$

Now to $\mathbb{R}^3$. The solution set of a single linear equation in $\mathbb{R}^3$ describes a plane.

Example

Find the plane $\Pi_{PQR} \subset \mathbb{R}^3$ through $P = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $Q = \begin{pmatrix} 1 \\ -1 \\ 2 \end{pmatrix}$ and $R = \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}$.

Its equation is of the form

$$ax+by+cz=d \qquad \text{for } a,b,c,d \in \mathbb{R} \text{ with } (a,b,c) \neq (0,0,0).$$

Treating a,b,c,d as unknowns and substituting the three points into this equation in turn gives a system of simultaneous linear equations:

$$\begin{array}{lrcl} \text{at } P\colon & a+b+c & = & d \\ \text{at } Q\colon & a-b+2c & = & d \\ \text{at } R\colon & 2b+3c & = & d \end{array}$$

You can solve this to find a nontrivial solution $(a, b, c, d) = (5, 1, 2, 8)$ (there is one degree of freedom, but we have chosen a particular solution), so that

$$\Pi_{PQR} \colon (5x + y + 2z = 8) \subset \mathbb{R}^3$$

As a routine professional courtesy, of course you check that the three points really do satisfy this equation.

Remark

Perhaps you know the vector calculus way of expressing a plane $\Pi \subset \mathbb{R}^3$. Suppose $\underline{\hat{n}}$ is a normal vector to Π, then the equation of Π has the form

$$\Pi \colon \underline{\hat{n}} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = d \quad \text{for some } d \in \mathbb{R}$$

and if $\underline{\hat{n}}$ is a unit vector, as the notation suggests, then d is the height of Π above the origin.

The equation we found in the example above is also of this form: it is simply

$$\Pi_{PQR} \colon \underline{\hat{n}} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \frac{8}{\sqrt{30}} \quad \text{with } \underline{\hat{n}} = \frac{1}{\sqrt{30}} \begin{pmatrix} 5 \\ 1 \\ 2 \end{pmatrix}$$

where we divided through by the length of the vector $(5, 1, 2)^T$ to get the unit vector $\underline{\hat{n}}$. Nothing much has changed, though we now see that the height of Π_{PQR} above the origin is $8/\sqrt{30}$.

However, this point of view does give us a idea. Choosing a different value of d also defines a plane, but a different one that is parallel to Π: the normal vector $\underline{\hat{n}}$ determines what we think of as the slope of the plane, while the value of d determines its distance from the origin.

Now what about lines $L \subset \mathbb{R}^3$? These are defined by **two** independent linear equations. This makes intuitive sense: we may consider the line as being the intersection of two distinct planes, each of those is defined by a linear equation, and so we need both to define the line. Let's do a typical calculation.

Example 1.21

What are the equations of the line $L_{PQ} \subset \mathbb{R}^3$ through the points $P = \begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix}$ and $Q = \begin{pmatrix} 4 \\ 3 \\ 2 \end{pmatrix}$?

Again, first a sanity check: these are two distinct points, so we know from experience that there is a unique line passing through them. As before, that doesn't mean we expect the equations to be unique: different equations may define the same line.

In the first place, let's consider a single linear equation with unknown coefficients $a, b, c, d \in \mathbb{R}$:

$$ax + by + cz = d. \tag{1.3}$$

Substituting the components of P and Q into this equation in turn gives the system of simultaneous linear equations

$$\begin{aligned} \text{at } P: \quad -2a + b + 3c &= d \\ \text{at } Q: \quad 4a + 3b + 2c &= d \end{aligned} \tag{1.4}$$

Adding twice the first to the second gives

$$5b + 8c = 3d \qquad \text{or, in other words,} \qquad b = -\frac{8}{5}c + \frac{3}{5}d$$

This equation by itself has lots of solutions: for any $c, d \in \mathbb{R}$, we can use this to define b. Plugging that expression for b back into the first equation gives

$$-2a + \left(-\frac{8}{5}c + \frac{3}{5}d\right) + 3c = d \qquad \text{or, in other words,} \qquad a = \frac{7}{10}c - \frac{1}{5}d$$

You can check that these expressions for a, b in terms of c, d really do satisfy both equations (1.4).

Now, with these expressions for a, b, for any values of $c, d \in \mathbb{R}$ not both zero, equation (1.3) is

$$\left(\frac{7}{10}c - \frac{1}{5}d\right)x + \left(-\frac{8}{5}c + \frac{3}{5}d\right)y + cz = d.$$

You could say that all of these (infinitely many) equations together define the line L_{PQ}, and you'd be right. But as we said at the outset, it is enough to choose two independent equations from this huge collection. To do that, simply choose two different pairs (c, d) that are not collinear (this last condition is to ensure that the equations are independent). For example,

$$\begin{aligned} \text{choosing } c = 10,\ d = 0 \text{ gives:} \quad 7x - 16y + 10z &= 0 \\ \text{choosing } c = 0,\ d = 5 \text{ gives:} \quad -x + 3y &= 5 \end{aligned}$$

and this pair of equations taken together defines $L_{PQ} \subset \mathbb{R}^3$. (As ever, to minimise errors we check again that P and Q really do lie on L_{PQ}, that is, they really do satisfy both equations.)

Remark

Perhaps you know the vector calculus way of expressing a line $L \subset \mathbb{R}^3$. Suppose $\underline{w} \in \mathbb{R}^3$ is a nonzero vector and $P \in \mathbb{R}^3$ is a point. (Remember that for us a point is the same thing as a vector based at zero, so P is also a vector.) Then the line L through P that is parallel to $\underline{w}$ is described as the set

$$L = \{P + \lambda \underline{w} \mid \lambda \in \mathbb{R}\}.$$

This is a **parametrised** way of describing the line – it lists all the points of the line – whereas our method above, writing down two equations and saying that the line is the solution set of them, is an **implicit** way of describing the line. Both ways are useful, and it is good to be able to translate between the two.

For problems such as the one in Example 1.21, we can choose $\underline{w} = Q - P$, since after all this choice of $\underline{w}$ is a vector along the line L_{PQ}, so it is certainly parallel to it, and then

$$L_{PQ} = \{P + \lambda \underline{w} \mid \lambda \in \mathbb{R}\} = \{(1-\lambda)P + \lambda Q \mid \lambda \in \mathbb{R}\}.$$

With P, Q as in Example 1.21, the points of L_{PQ} are then exactly the set

$$\{(1-\lambda)P + \lambda Q \mid \lambda \in \mathbb{R}\} = \left\{ \begin{pmatrix} -2 + 6\lambda \\ 1 + 2\lambda \\ 3 - \lambda \end{pmatrix} \,\middle|\, \lambda \in \mathbb{R} \right\} \subset \mathbb{R}^3$$

and you easily check that these points satisfy the two equations we derived earlier.

Chapter 2

Linear systems and matrices

You probably already know how to solve systems of simultaneous linear equations in several unknowns. In this chapter we review a basic naive approach with an example, and then describe a really effective way to do this that you can use forever more in all but the most trivial situations. We discuss:

§2.1 Systems of linear equations: how to solve them naively, and also express them using matrices; if you know this already, you only need to check the terminology we use.

§2.2 Algebra of matrices: addition and multiplication, including summation notation for the operations; you may wish only to skim this to be sure it accords with how you think of the material.

§2.3 Reduced row echelon form: this is the crucial section; it gives a systematic and efficient approach to understand and solve systems of linear equations. Put most of your effort here.

§2.4 How to compute the inverse of a square matrix, if it has one, using reduced row echelon form.

2.1 Systems of linear equations

Example 2.1

Determine the solutions (if any) to the following equations:

$$3x + y - 2z = -2 \qquad x + y + z = 2 \qquad 2x + 4y + z = 0.$$

We can substitute $z = 2 - x - y$ from the second equation into the first and third equations to find

$$\begin{aligned} 3x + y - 2(2 - x - y) &= 5x + 3y - 4 = -2 \qquad &\Rightarrow \quad 5x + 3y = 2 \\ 2x + 4y + (2 - x - y) &= x + 3y + 2 = 0 \qquad &\Rightarrow \quad x + 3y = -2 \end{aligned}$$

Subtracting the second of these equations from the first gives $4x = 4$ and so we see:

$$x = 1 \qquad y = (-2 - x)/3 = -1 \qquad z = 2 - x - y = 2.$$

Thus there is a unique solution, $(x, y, z) = (1, -1, 2)$.

Definition 2.2

A **linear system of equations** is a set of m simultaneous equations in n variables $x_1, x_2, \ldots, x_n$ which are of the form

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{m1}x_1 & + & a_{m2}x_2 & + & \cdots & + & a_{mn}x_n & = & b_m \end{array} \tag{2.1}$$

where the a_{ij} and b_k are constants.

We can write the linear system of equations in matrix form:

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

Example 2.3

Returning to Example 2.1, we have:

$$\begin{array}{ccccccc} 3x & + & y & - & 2z & = & -2 \\ x & + & y & + & z & = & 2 \\ 2x & + & 4y & + & z & = & 0 \end{array}$$

Written as matrices this becomes:

$$\begin{pmatrix} 3 & 1 & -2 \\ 1 & 1 & 1 \\ 2 & 4 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} -2 \\ 2 \\ 0 \end{pmatrix}$$

Definition 2.4

Any vector $(x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ which satisfies (2.1) is called a **solution** to the linear system. If the linear system has one or more solutions then it is said to be **consistent**. The **general solution** to the system is any description of all the solutions of the system.

We will see later that any linear system will have either zero, one, or infinitely many solutions.

Example 2.5

Example 2.1 has a unique solution given by the vector $(1, -1, 2)^T$. That this is a solution can be easily verified (and you should do this!). That this solution is unique follows from our working in Example 2.1, however this is not so simple to check without repeating the work we did above.

Definition 2.6

We will often write the linear system (2.1) as the **augmented matrix** $(A \mid \underline{b})$, where

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \qquad \underline{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

Example 2.7

Again returning to Example 2.1, the augmented matrix is:

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

The advantage of using the augmented matrix is that we will be able to progress systematically towards a solution. This is a methodical approach, unlike our solution in Example 2.1. This process is called **row reduction**. It relies on three types of operation, called *elementary row operations* (or *EROs* for short). The important point is that EROs do not affect the set of solutions of a linear system.

Definition 2.8

Given a matrix, an **elementary row operation** (or **ERO**) is one of the following:

(i) Two rows may be swapped;

(ii) A row can be multiplied by a non-zero scalar;

(iii) A multiple of one row may be added to a second row.

These three elementary row operations can be understood in terms of operations on the linear equations:

(i) The order of two equations may be swapped; for example, rather than writing

$$\begin{aligned} 3x + y - 2z &= -2 \\ x + y + z &= 2 \\ 2x + 4y + z &= 0 \end{aligned}$$

we may change the order (for example, swapping the first and second equations) and write

$$\begin{aligned} x + y + z &= 2 \\ 3x + y - 2z &= -2 \\ 2x + 4y + z &= 0 \end{aligned}$$

(ii) An equation may be multiplied by a non-zero scalar; for example, we might replace the equation $3x + y - 2z = -2$ with $x + \frac{1}{3}y - \frac{2}{3}z = -\frac{2}{3}$ (i.e. multiply by $\frac{1}{3}$).

(iii) A multiple of one equation may be added to another equation; for example, given the two equations

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

we can add $-\frac{1}{3}$ of the first equation to the second equation to get

$$\begin{aligned} 3x + y - 2z &= -2 \\ \frac{2}{3}y + \frac{5}{3}z &= \frac{8}{3} \end{aligned}$$

These three operations describe the steps a person would take when attempting to solve a linear system of equations.

Example 2.9

Again returning to Example 2.1, we will solve the system slowly using the three elementary row operations. First we do this by working directly with the equations:

$$\begin{array}{rcrcrcr} 3x &+& y &-& 2z &=& -2 \\ x &+& y &+& z &=& 2 \\ 2x &+& 4y &+& z &=& 0 \end{array}$$

Swap the first and second equations to get:

$$\begin{array}{rcrcrcr} x &+& y &+& z &=& 2 \\ 3x &+& y &-& 2z &=& -2 \\ 2x &+& 4y &+& z &=& 0 \end{array}$$

Add -3 times the first equation to the second equation to get:

$$\begin{array}{rcrcrcr} x &+& y &+& z &=& 2 \\ && -2y &-& 5z &=& -8 \\ 2x &+& 4y &+& z &=& 0 \end{array}$$

Add -2 times the first equation to the third equation to get:

$$\begin{array}{rcrcrcr} x &+& y &+& z &=& 2 \\ && -2y &-& 5z &=& -8 \\ && 2y &-& z &=& -4 \end{array}$$

Add the second equation to the third equation to get:

$$\begin{array}{rcrcrcr} x &+& y &+& z &=& 2 \\ && -2y &-& 5z &=& -8 \\ && && -6z &=& -12 \end{array}$$

Multiply the second equation by $-\frac{1}{2}$ to get:

$$\begin{array}{rcrcrcr} x &+& y &+& z &=& 2 \\ && y &+& \frac{5}{2}z &=& 4 \\ && && -6z &=& -12 \end{array}$$

Add -1 times the second equation to the first equation to get:

$$\begin{array}{rcrcrcr} x && &-& \frac{3}{2}z &=& -2 \\ && y &+& \frac{5}{2}z &=& 4 \\ && && -6z &=& -12 \end{array}$$

Multiply the third equation by $-\frac{1}{6}$ to get:

$$\begin{aligned} x \quad - \tfrac{3}{2}z &= -2 \\ y + \tfrac{5}{2}z &= 4 \\ z &= 2 \end{aligned}$$

Add $\frac{3}{2}$ times the third equation to the first equation to get:

$$\begin{aligned} x \qquad &= 1 \\ y + \tfrac{5}{2}z &= 4 \\ z &= 2 \end{aligned}$$

Add $-\frac{5}{2}$ times the third equation to the second equation to get:

$$\begin{aligned} x \qquad &= 1 \\ y \quad &= -1 \\ z &= 2 \end{aligned}$$

Example 2.10

Again returning to Example 2.1, we will again solve the system slowly using the three elementary row operations. This time we do this by working with the augmented matrix:

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

Swap the first and second rows to get:

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

Add -3 times the first row to the second row to get:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 2 & 4 & 1 & 0 \end{array}\right)$$

Add -2 times the first row to the third row to get:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 0 & 2 & -1 & -4 \end{array}\right)$$

Add the second row to the third row to get:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 0 & 0 & -6 & -12 \end{array}\right)$$

Multiply the second row by $-\frac{1}{2}$ to get:

$$\left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 1 & \frac{5}{2} & 4 \\ 0 & 0 & -6 & -12 \end{array}\right)$$

Add -1 times the second row to the first row to get:

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & -2 \\ 0 & 1 & \frac{5}{2} & 4 \\ 0 & 0 & -6 & -12 \end{array}\right)$$

Multiply the third row by $-\frac{1}{6}$ to get:

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

Add $\frac{3}{2}$ times the third row to the first row to get:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & \frac{5}{2} & 4 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

Add $-\frac{5}{2}$ times the third row to the second row to get:

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

From this we can read off the solution:

$$x = 1 \qquad y = -1 \qquad z = 2$$

Language 2.11

We introduce some notation to help us talk about elementary row operations. Note that this is not standard notation, however it is convenient to have.

(i) Let S_{ij} denote the elementary row operation which swaps rows i and j.

(ii) Let $M_i(\lambda)$ denote the elementary row operation which multiples row i by $\lambda \neq 0$.

(iii) Let $A_{ij}(\lambda)$, where $i \neq j$, denote the elementary row operation which adds λ times row i to row j.

Example 2.12

We solve Example 2.1 once more:

$$\left(\begin{array}{ccc|c} 3 & 1 & -2 & -2 \\ 1 & 1 & 1 & 2 \\ 2 & 4 & 1 & 0 \end{array}\right) \xrightarrow{S_{12}} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 3 & 1 & -2 & -2 \\ 2 & 4 & 1 & 0 \end{array}\right) \xrightarrow{A_{12}(-3)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 2 & 4 & 1 & 0 \end{array}\right)$$

$$\xrightarrow{A_{13}(-2)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 0 & 2 & -1 & -4 \end{array}\right) \xrightarrow{A_{23}(1)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & -2 & -5 & -8 \\ 0 & 0 & -6 & -12 \end{array}\right) \xrightarrow{M_2(-1/2)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 2 \\ 0 & 1 & \frac{5}{2} & 4 \\ 0 & 0 & -6 & -12 \end{array}\right)$$

$$\xrightarrow{A_{21}(-1)} \left(\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & -2 \\ 0 & 1 & \frac{5}{2} & 4 \\ 0 & 0 & -6 & -12 \end{array}\right) \xrightarrow{M_3(-1/6)} \left(\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & -2 \\ 0 & 1 & \frac{5}{2} & 4 \\ 0 & 0 & 1 & 2 \end{array}\right) \xrightarrow{A_{31}(3/2)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & \frac{5}{2} & 4 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

$$\xrightarrow{A_{32}(-5/2)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 2 \end{array}\right)$$

We see that:

$$x = 1 \qquad y = -1 \qquad z = 2$$

Example 2.13

Consider the linear system of equations

$$\begin{aligned} x_1 - x_2 + x_3 + 3x_4 &= 2 \\ 2x_1 - x_2 + x_3 + 2x_4 &= 4 \\ 4x_1 - 3x_2 + 3x_3 + 8x_4 &= 8 \end{aligned}$$

We solve this system by working with the augmented matrix:

$$\left(\begin{array}{cccc|c} 1 & -1 & 1 & 3 & 2 \\ 2 & -1 & 1 & 2 & 4 \\ 4 & -3 & 3 & 8 & 8 \end{array}\right) \xrightarrow{A_{12}(-2), A_{13}(-4)} \left(\begin{array}{cccc|c} 1 & -1 & 1 & 3 & 2 \\ 0 & 1 & -1 & -4 & 0 \\ 0 & 1 & -1 & -4 & 0 \end{array}\right)$$

$$\xrightarrow{A_{21}(1), A_{23}(-1)} \left(\begin{array}{cccc|c} 1 & 0 & 0 & -1 & 2 \\ 0 & 1 & -1 & -4 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

The third row of the augmented matrix has become equal to zero. This indicates that there was redundancy in the original system of equations: in this case notice that the third equation $4x_1 - 3x_2 + 3x_3 + 8x_4 = 8$ can be deduced from the first two equations $x_1 - x_2 + x_3 + 3x_4 = 2$ and $2x_1 - x_2 + x_3 + 2x_4 = 4$ (add two times the first equation to the second equation) and so provides no additional information.

Our sequence of elementary row operations above resulted in the two equations

$$x_1 - x_4 = 2 \qquad x_2 - x_3 - 4x_4 = 0 \tag{2.2}$$

However there are four variables x_1, x_2, x_3, and x_4. Thus it is impossible for this system to have a unique solution. Instead we assign parameters to the two columns (equivalently, variables) which fail to contain a leading entry: in this case, the third and forth columns representing x_3 and x_4. Setting $x_3 = s$ and $x_4 = t$ in (2.2) and rearranging slight gives a two-dimensional family of solutions:

$$x_1 = 2 + t \qquad x_2 = s + 4t \qquad x_3 = s \qquad x_4 = t$$

Equivalently, we could write this as

$$\begin{aligned}(x_1, x_2, x_3, x_4) &= (2+t, s+4t, s, t)\\ &= (2,0,0,0) + s(0,1,1,0) + t(1,4,0,1)\end{aligned}$$

We see that the solutions form a two-dimensional plane in $\mathbb{R}^4$ parameterised by s and t. This plane is parallel to the vectors $(0,1,1,0)$ and $(1,4,0,1)$, and contains the point $(2,0,0,0)$.

Example 2.14

Consider the linear system of equations

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

Applying elementary row operations to the augmented matrix we obtain:

$$\left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4\\ 2 & 3 & -2 & -3 & 1\\ 1 & 0 & 5 & 6 & 1\end{array}\right) \xrightarrow{A_{12}(-2),A_{13}(-1)} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4\\ 0 & 1 & -4 & -5 & -7\\ 0 & -1 & 4 & 5 & -3\end{array}\right)$$

$$\xrightarrow{A_{23}(1)} \left(\begin{array}{cccc|c} 1 & 1 & 1 & 1 & 4\\ 0 & 1 & -4 & -5 & -7\\ 0 & 0 & 0 & 0 & -10\end{array}\right)$$

Notice that the third row gives the equation

$$0x + 0y + 0z + 0w = -10$$

There is clearly no solution to this equation; hence there are no solutions to the original system of equations.

Remark

Examples 2.12, 2.13, and 2.14 illustrate the following important observation:

A linear system of equations can have no solution, one solution, or infinitely many solutions.

Example 2.15

Consider the linear system of equations in x, y, and z

$$\begin{aligned} x + z &= -5\\ 2x + \alpha y + 3z &= -9\\ -x - \alpha y + \alpha z &= \alpha^2.\end{aligned}$$

Here α is a constant. We can apply elementary row operations to the augmented matrix:

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 2 & \alpha & 3 & -9 \\ -1 & -\alpha & \alpha & \alpha^2 \end{array}\right) \xrightarrow{A_{12}(-2), A_{13}(1)} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & \alpha & 1 & 1 \\ 0 & -\alpha & \alpha+1 & \alpha^2-5 \end{array}\right)$$
$$\xrightarrow{A_{23}(1)} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & \alpha & 1 & 1 \\ 0 & 0 & \alpha+2 & \alpha^2-4 \end{array}\right)$$

How we proceed depends on the value of α. Ideally we would like to divide the second row by α, and divide the third row by $\alpha+2$. But the first of these operations requires $\alpha \neq 0$, and the second of these operations requires $\alpha \neq -2$.

Let us assume for now that $\alpha \neq 0, -2$. Then:

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & \alpha & 1 & 1 \\ 0 & 0 & \alpha+2 & \alpha^2-4 \end{array}\right) \xrightarrow{M_2(1/\alpha), M_3(1/(\alpha+2))} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 1 & \frac{1}{\alpha} & \frac{1}{\alpha} \\ 0 & 0 & 1 & \alpha-2 \end{array}\right)$$
$$\xrightarrow{A_{31}(-1), A_{32}(-1/\alpha)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -\alpha-3 \\ 0 & 1 & 0 & \frac{3}{\alpha}-1 \\ 0 & 0 & 1 & \alpha-2 \end{array}\right)$$

We see that there is a unique solution

$$x = -\alpha - 3 \qquad y = \frac{3}{\alpha} - 1 \qquad z = \alpha - 2$$

Now suppose that $\alpha = 0$. We have:

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & -4 \end{array}\right) \xrightarrow{A_{23}(-2), M_3(-1/6)} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{array}\right)$$

We see that the system is inconsistent (because the final row gives $0x + 0y + 0z = 1$), and so there is no solution.

Finally, suppose that $\alpha = 2$. Then:

$$\left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & -2 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{M_2(-1/2)} \left(\begin{array}{ccc|c} 1 & 0 & 1 & -5 \\ 0 & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

Since there is no leading entry in the third column, we assign a free parameter t to z and obtain infinitely many solutions:

$$x = -5 - t \qquad y = \frac{t-1}{2} \qquad z = t$$

2.2 Matrices and matrix algebra

A *matrix* is a two-dimensional array of numbers. We say that a matrix is an $m \times n$ matrix if it has m rows and n columns. For example

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \qquad \begin{pmatrix} 1 \\ \pi \\ \sqrt{2} \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

are all matrices. The first example is a 2×3 matrix, the second example is a 3×1 matrix, and the third example is a 2×2 matrix.

Definition 2.16

Let m and n be positive integers. An $m \times n$ **matrix** is an array of mn numbers arranged into m rows and n columns. The numbers in a matrix are called its **entries**. In the contexts we use matrices here, all entries are scalars, which for us means they lie in $\mathbb{R}$.

Remark

We may treat a vector in $\mathbb{R}^m$ (meaning a **column vector** as usual) as an $m \times 1$ matrix. Conversely, if A is a $m \times n$ matrix, then we may regard each of its columns as a vector in $\mathbb{R}^m$.

We may also consider **row vectors**: we write $\begin{pmatrix} c_1 & c_2 & \dots & c_n \end{pmatrix} \in \mathbb{R}^n_{\text{row}}$, where the subscript indicates that we mean row vectors. We may treat row vectors in $\mathbb{R}^n_{\text{row}}$ as $1 \times n$ matrices, and conversely if A is a $m \times n$ matrix, then we may regard its rows as m row vectors in $\mathbb{R}^n_{\text{row}}$. (We will always be explicit when we mean *row* vectors by writing $\mathbb{R}^n_{\text{row}}$.)

Example 2.17

Let

$$A = \begin{pmatrix} 1 & -3 & 7 \\ 3 & 2 & 1 \end{pmatrix}$$

The second column of A is equal to the (column) vector

$$\begin{pmatrix} -3 \\ 2 \end{pmatrix} \in \mathbb{R}^2$$

and the second row of A is equal to the **row** vector

$$\begin{pmatrix} 3 & 2 & 1 \end{pmatrix} \in \mathbb{R}^3$$

Python 2.18

Many of the calculations we perform in these notes can be replicated in Python by using the NumPy package. All code cells in these notes will require you to have run the following line of code beforehand to import the NumPy package. We give `numpy` the alias `np`.

```
import numpy as np
```

The basic data type in NumPy is the `array` object which has type `ndarray`. These are often used to simulate matrices. They are initialised using the function `np.array`, and by specifying the matrix elements to be stored as a list in the function input. The following code stores the matrix A from Example 2.17 in a variable.

```
A = np.array([[1,-3,7],[3,2,1]])
print(A)
```

Individual elements in a NumPy array can be accessed using square brackets `[` and `]`. In particular, for our matrix A above, the code `A[m,n]` will extract the element in the row indexed by `m`, and the column indexed by `n`. Remember counting in Python starts from zero, so the first row/column is indexed by 0, the second by 1, and so on. Line 1 below prints 7, and line 2 prints 2.

```
1 print(A[0,2])
2 print(A[1,1])
```

If you want to extract a row of our NumPy array, then we can use the notation `A[m]` where `m` is the index of the row we want to extract. The following code prints the second row of A.

```
print(A[1])
```

If you want to extract a column of our NumPy array, then we can use the notation `A[:,n]` where `n` is the index of the column we want to extract. The following code prints the third column of A.

```
print(A[:,2])
```

Definition 2.19

The set of all real $m \times n$ matrices is denoted $\mathrm{Mat}_{m\times n}(\mathbb{R})$, or by the abbreviation $\mathrm{Mat}_{mn}(\mathbb{R})$ or just Mat_{mn} when it is clear.

As remarked after Definition 2.16, we may regard $\mathrm{Mat}_{m1} = \mathbb{R}^m$ and $\mathrm{Mat}_{1n} = \mathbb{R}^n_{\mathrm{row}}$.

Language 2.20

We frequently write let $A = (a_{ij}) \in \mathrm{Mat}_{mn}$ as shorthand for the matrix

$$m \text{ rows} \left\{ \overbrace{\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}}^{n \text{ columns}} \right. \in \mathrm{Mat}_{mn}$$

with entry a_{ij} in the ith row and the jth column. Here the limits $1 \le i \le m$ and $1 \le j \le n$ are implicit, since $A \in \mathrm{Mat}_{mn}$. In particular the ith row of A is

$$\begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix} \in \mathbb{R}^n_{\mathrm{row}}$$

and the jth column of A is

$$\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \in \mathbb{R}^m$$

Example 2.21

Let

$$A = \begin{pmatrix} 1 & 3 & 5 \\ 0 & 2 & 4 \end{pmatrix} \in \mathrm{Mat}_{23} .$$

Then $a_{11} = 1$, $a_{12} = 3$, and $a_{23} = 4$. In fact, in this example we have the formula $a_{ij} = 2j - i$.

The simple transpose operation on matrices explains our notation $(a_1, \ldots, a_n)^T$ for column vectors.

Definition 2.22

Let $A = (a_{ij}) \in \mathrm{Mat}_{mn}$. Then the **transpose of** A, also referred to as A **transposed** and denoted A^T, is the matrix $A^T = (a_{ji}) \in \mathrm{Mat}_{nm}$.

That definition is nice and concise, but you have to notice carefully that the i and j have switched roles to make the transpose: a_{ij} became a_{ji}.

Example 2.23

Let

$$A = \begin{pmatrix} 1 & 3 & 5 \\ 0 & 2 & 4 \end{pmatrix} \in \mathrm{Mat}_{23} .$$

Then

$$A^T = \begin{pmatrix} 1 & 0 \\ 3 & 2 \\ 5 & 4 \end{pmatrix} \in \mathrm{Mat}_{32} .$$

In the transposed matrix the rows have become the columns, or equally the columns have become the rows.

There are three important operations that can be performed with matrices:

(i) matrix addition,

(ii) scalar multiplication, and

(iii) matrix multiplication.

Two matrices can only be added together if they have the same number of rows, and the same number of columns.

Definition 2.24 (Matrix addition)

Let $A = (a_{ij})$ and $B = (b_{ij})$ be two $m \times n$ matrices. Then $C = A + B$ is the $m \times n$ matrix with entries

$$c_{ij} = a_{ij} + b_{ij} \qquad \text{for } 1 \leq i \leq m \text{ and } 1 \leq j \leq n.$$

Example 2.25

Let

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Then

$$A + B = \begin{pmatrix} 2 & 2 & 0 \\ 0 & 2 & 2 \end{pmatrix} = B + A.$$

Python 2.26

We can perform matrix addition in Python. The following code emulates the calculation of Example 2.25.

```
A = np.array([[1,2,-1],[0,1,2]])
B = np.array([[1,0,1],[0,1,0]])
print(A+B)
```

If you attempt to add two NumPy arrays that are not of the same size you will receive an error '`ValueError: operands could not be broadcast together with shapes`'.

Remark

In general, matrix addition is *commutative*. That is, for $A, B \in \mathrm{Mat}_{mn}$ we have that

$$A + B = B + A.$$

Furthermore, matrix addition is *associative*. That is, for $A, B, C \in \mathrm{Mat}_{mn}$ we have that

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

Definition 2.27

The $m \times n$ matrix whose entries are all 0 is called the **zero matrix**, and is denoted by 0_{mn}. Given any $A \in \mathrm{Mat}_{mn}$ we have that

$$A + 0_{mn} = A.$$

Example 2.28

Let
$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad 0_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$
It is trivial to check that $A + 0_{22} = A$.

Definition 2.29 (Scalar multiplication)

Let $A = (a_{ij})$ be an $m \times n$ matrix, and let $k \in \mathbb{R}$. Then $C = kA$ is the $m \times n$ matrix with entries
$$c_{ij} = ka_{ij} \qquad \text{for } 1 \leq i \leq m \text{ and } 1 \leq j \leq n.$$

Example 2.30

Let
$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \qquad B = \begin{pmatrix} -1 & 0 \\ -3 & 0 \end{pmatrix}$$
We shall show that $3(A + B) = 3A + 3B$. First, notice that
$$A + B = \begin{pmatrix} 0 & 2 \\ 0 & 4 \end{pmatrix} \qquad \text{and hence} \qquad 3(A + B) = 3\begin{pmatrix} 0 & 2 \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 6 \\ 0 & 12 \end{pmatrix}.$$
On the other hand,
$$3A = \begin{pmatrix} 3 & 6 \\ 9 & 12 \end{pmatrix} \qquad \text{and} \qquad 3B = \begin{pmatrix} -3 & 0 \\ -9 & 0 \end{pmatrix}, \qquad \text{hence} \qquad 3A + 3B = \begin{pmatrix} 0 & 6 \\ 0 & 12 \end{pmatrix}.$$

Python 2.31

We can perform scalar multiplication of a matrix in Python. The following code emulates the calculation of Example 2.30.

```
A = np.array([[1,2],[3,4]])
B = np.array([[-1,0],[-3,0]])
print(3*A)
print(3*B)
print(3*A+3*B)
```

Definition 2.32

Scalar multiplication is *distributive*: for $A, B \in \mathrm{Mat}_{mn}$ and $k \in \mathbb{R}$ we have that
$$k(A + B) = kA + kB.$$

Remark

Let $A, B, C \in \mathrm{Mat}_{mn}$ and let $k, s \in \mathbb{R}$ be scalars. The following identities hold:

(i) $A + 0_{mn} = A$

(ii) $A + B = B + A$

(iii) $0A = 0_{mn}$

(iv) $A + (-A) = 0_{mn}$

(v) $(A + B) + C = A + (B + C)$

(vi) $1A = A$

(vii) $(k + s)A = kA + sA$

(viii) $k(A + B) = kA + kB$

(ix) $k(sA) = (ks)A$

(We shall see later that these identities show that Mat_{mn} is a *vector space*, giving a more exotic example than the vector spaces $\mathbb{R}^n$ we have seen so far.)

Matrix multiplication is very different from matrix addition and scalar multiplication. At first the definition may seem strange, however we shall see later that it is natural in the context of matrices representing linear maps.

Definition 2.33 (Matrix multiplication)

Let $A = (a_{ij})$ be an $m \times n$ matrix, and let $B = (b_{ij})$ be an $n \times \ell$ matrix. Then $C = AB$ is the $m \times \ell$ matrix with entries

$$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \qquad \text{for } 1 \leq i \leq m \text{ and } 1 \leq j \leq \ell.$$

Equivalently, let $\mathbf{r}_1, \ldots, \mathbf{r}_m$ denote the rows of A, and let $\mathbf{c}_1, \ldots, \mathbf{c}_\ell$ denote the columns of B. Then

$$c_{ij} = \mathbf{r}_i^T \cdot \mathbf{c}_j \qquad \text{for } 1 \leq i \leq m \text{ and } 1 \leq j \leq \ell,$$

where $\mathbf{r}_i^T \cdot \mathbf{c}_j$ denotes the scalar product.

Example 2.34

Let

$$A = \begin{pmatrix} 1 & 2 \\ -1 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$$

Let us work slowly through the calculation of the matrix product AB.

$$\begin{pmatrix}\boxed{1} & \boxed{2}\\ -1 & 0\end{pmatrix}\begin{pmatrix}\boxed{1} & -1\\ \boxed{1} & -1\end{pmatrix} = \begin{pmatrix}\boxed{1\times 1+2\times 1} & ?\\ ? & ?\end{pmatrix} = \begin{pmatrix}\boxed{3} & ?\\ ? & ?\end{pmatrix}$$

$$\begin{pmatrix}\boxed{1} & \boxed{2}\\ -1 & 0\end{pmatrix}\begin{pmatrix}1 & \boxed{-1}\\ 1 & \boxed{-1}\end{pmatrix} = \begin{pmatrix}3 & \boxed{1\times(-1)+2\times(-1)}\\ ? & ?\end{pmatrix} = \begin{pmatrix}3 & \boxed{-3}\\ ? & ?\end{pmatrix}$$

$$\begin{pmatrix}1 & 2\\ \boxed{-1} & \boxed{0}\end{pmatrix}\begin{pmatrix}\boxed{1} & -1\\ \boxed{1} & -1\end{pmatrix} = \begin{pmatrix}3 & -3\\ \boxed{(-1)\times 1+0\times 1} & ?\end{pmatrix} = \begin{pmatrix}3 & -3\\ \boxed{-1} & ?\end{pmatrix}$$

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

Hence

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

Example 2.35

Let A and B be as in Example 2.34. We wil calculate the matrix product BA.

$$\begin{pmatrix}\boxed{1} & \boxed{-1}\\ 1 & -1\end{pmatrix}\begin{pmatrix}\boxed{1} & 2\\ \boxed{-1} & 0\end{pmatrix} = \begin{pmatrix}\boxed{1\times 1+(-1)\times(-1)} & ?\\ ? & ?\end{pmatrix} = \begin{pmatrix}\boxed{2} & ?\\ ? & ?\end{pmatrix}$$

$$\begin{pmatrix}\boxed{1} & \boxed{-1}\\ 1 & -1\end{pmatrix}\begin{pmatrix}1 & \boxed{2}\\ -1 & \boxed{0}\end{pmatrix} = \begin{pmatrix}2 & \boxed{1\times 2+(-1)\times 0}\\ ? & ?\end{pmatrix} = \begin{pmatrix}2 & \boxed{2}\\ ? & ?\end{pmatrix}$$

$$\begin{pmatrix}1 & -1\\ \boxed{1} & \boxed{-1}\end{pmatrix}\begin{pmatrix}\boxed{1} & 2\\ \boxed{-1} & 0\end{pmatrix} = \begin{pmatrix}2 & 2\\ \boxed{1\times 1+(-1)\times(-1)} & ?\end{pmatrix} = \begin{pmatrix}2 & 2\\ \boxed{2} & ?\end{pmatrix}$$

$$\begin{pmatrix}1 & -1\\ \boxed{1} & \boxed{-1}\end{pmatrix}\begin{pmatrix}1 & \boxed{2}\\ -1 & \boxed{0}\end{pmatrix} = \begin{pmatrix}2 & 2\\ 2 & \boxed{1\times 2+(-1)\times 0}\end{pmatrix} = \begin{pmatrix}2 & 2\\ 2 & \boxed{2}\end{pmatrix}$$

Hence

$$BA = \begin{pmatrix}1 & -1\\ 1 & -1\end{pmatrix}\begin{pmatrix}1 & 2\\ -1 & 0\end{pmatrix} = \begin{pmatrix}2 & 2\\ 2 & 2\end{pmatrix}$$

Python 2.36

We can perform matrix multiplication in Python. The following code emulates the calculations in Examples 2.34 and 2.35.

```
A = np.array([[1,2],[-1,0]])
B = np.array([[1,-1],[1,-1]])
print(A.dot(B))
print(B.dot(A))
```

If you attempt to multiply two NumPy arrays that do not have compatible sizes – that is the number of columns of the first matrix is not equal the number of rows of the second matrix – you will receive an error '`ValueError: shapes (*,*) and (*,*) not aligned`'.

Remark

Examples 2.34 and 2.35 show that, even when both the matrix multiplications AB and BA make sense, we can have that

$$AB \neq BA$$

That is, matrix multiplication is *not commutative*.

Example 2.37

Let

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \in \mathrm{Mat}_{23} \qquad B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix} \in \mathrm{Mat}_{32}$$

Then

$$AB = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix} = \begin{pmatrix} 1-2 & 2-1 \\ 4+0 & 1+2 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 4 & 3 \end{pmatrix} \in \mathrm{Mat}_{22}$$

We shall also calculate the matrix produce BA:

$$BA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \end{pmatrix} = \begin{pmatrix} 1+0 & 2+0 & -1+0 \\ 0+0 & 0+1 & 0+2 \\ 2+0 & 4+1 & -2+2 \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 2 \\ 2 & 5 & 0 \end{pmatrix} \in \mathrm{Mat}_{33}$$

Example 2.38

Let

$$A = \begin{pmatrix} 2 & 4 & 6 \\ 1 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Then

$$AB = \begin{pmatrix} 2 & 4 & 6 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 2+0+0 & 4+4+0 & 6+0+6 \\ 1+0+0 & 2+0+0 & 3+0+0 \end{pmatrix} = \begin{pmatrix} 2 & 8 & 12 \\ 1 & 2 & 3 \end{pmatrix} \in \mathrm{Mat}_{23}$$

Notice that in this example, asking for the matrix multiplication BA makes no sense – they are of incompatible sizes.

$$BA = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 4 & 6 \\ 1 & 0 & 0 \end{pmatrix} \text{ which makes no sense.}$$

Example 2.39

Let

$$A = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix}$$

Then

$$AA = \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 1-1 & -1+1 \\ 1-1 & -1+1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = 0_{22}$$

Remark

An important consequence of Example 2.39 is the following. Let $A \in \mathrm{Mat}_{mn}$ and $B \in \mathrm{Mat}_{n\ell}$ be such that $AB = 0_{m\ell}$. It does not follow that either $A = 0_{mn}$ or $B = 0_{m\ell}$.

Definition 2.40

The $n \times n$ **identity matrix** I_n is the $n \times n$ matrix with 1s on the diagonal and 0s elsewhere. That is,

$$I_n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & 1 \end{pmatrix} \in \mathrm{Mat}_{nn}$$

Equivalently, using Definition 1.17, the (i, j)th entry of I_n is given by δ_{ij}, for $1 \leq i, j \leq n$.

Example 2.41

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Remark

Let $a_1, \dots, a_n \in \mathbb{R}$, and fix $1 \leq k \leq n$. Then

$$\sum_{i=1}^{n} a_i \delta_{ik} = a_k \tag{2.3}$$

since $\delta_{ik} = 0$ when $i \neq k$, and $\delta_{kk} = 1$. Thus the sum (2.3) selects the kth element a_k.

Proposition 2.42

(i) Let $A \in \mathrm{Mat}_{mn}$ and $\ell, p \in \mathbb{Z}_{>0}$. Then

$$A0_{np} = 0_{mp} \qquad 0_{\ell m}A = 0_{\ell n} \qquad AI_n = A \qquad I_mA = A$$

(ii) Matrix multiplication is *associative*: for matrices $A \in \mathrm{Mat}_{mn}$, $B \in \mathrm{Mat}_{n\ell}$, and $C \in \mathrm{Mat}_{\ell p}$ we have

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

(iii) Matrix multiplication is *distributive*: whenever the following products and sums make sense, we have

$$A(B+C) = AB + AC \qquad (A+B)C = AC + BC$$

Proof. (i): To find an entry of the produce $A0_{np}$ we dot a row of A with a zero column of 0_{np}, which will always give zero. Similarly for $0_{\ell m}A$. By (2.3) we have

$$\text{the } (i,j)\text{th entry of } AI_n = \sum_{k=1}^{n} a_{ik}\delta_{kj} = a_{ij}$$
$$\text{the } (i,j)\text{th entry of } I_m A = \sum_{k=1}^{m} \delta_{ik}a_{kj} = a_{ij}.$$

(ii): Given $1 \le i \le m$, $1 \le j \le p$ we have

$$\text{the } (i,j)\text{th entry of } (AB)C = \sum_{r=1}^{\ell}\left(\sum_{s=1}^{n} a_{is}b_{sr}\right)c_{rj}$$
$$\text{the } (i,j)\text{th entry of } A(BC) = \sum_{s=1}^{n} a_{is}\left(\sum_{r=1}^{\ell} b_{sr}c_{rj}\right).$$

These are equal since the order of finite sums may be swapped without changing the result.

(iii): Left as an exercise. □

Because matrix multiplication is not commutative we need to be clear what we mean when we say something like "multiply the matrix A by the matrix B". Do we mean AB or BA? Sometimes we can deduce which is meant from the sizes of A and B; sometimes the context makes this clear. But sometimes we need to use more precise language.

Definition 2.43

Let A and B be matrices.

(i) To *premultiply* B by A is to perform the matrix multiplication AB, i.e. multiplication on the left.

(ii) To '*postmultiply* B by A is to perform the matrix multiplication BA, i.e. multiplication on the right.

Definition 2.44

Let $A \in \mathrm{Mat}_{mm}$ be a square matrix. We write A^2 for the product AA. Similarly, for any $n \in \mathbb{Z}_{>0}$ we write A^n for the product

$$\overbrace{AA\cdots A}^{n \text{ times}}$$

We define $A^0 = I_m$.

Remark

For any square matrix A and for any $n, m \in \mathbb{Z}_{\geq 0}$ we have $A^m A^n = A^{m+n}$.

Example 2.45

Let

$$A = \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix}$$

Then

$$A^2 = \begin{pmatrix} \cos^2\vartheta + \sin^2\vartheta & \cos\vartheta\sin\vartheta - \sin\vartheta\cos\vartheta \\ \sin\vartheta\cos\vartheta - \cos\vartheta\sin\vartheta & \sin^2\vartheta + \cos^2\vartheta \end{pmatrix} = I_2$$

In particular, $A^2 = I_2$ for any choice of ϑ.

Example 2.46

Let

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Let us assume for a contradiction that there exists a matrix

$$B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

such that $B^2 = A$. Then

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^2 = \begin{pmatrix} a^2 + bc & b(a+d) \\ c(a+d) & d^2 + bc \end{pmatrix}$$

Since $c(a+d) = 0$ we see that either $c = 0$ of $a+d = 0$. But if $a+d = 0$ then $1 = b(a+d) = 0$, which is a contradiction. Hence $c = 0$. But then $0 = a^2 + bc = a^2$ and so $a = 0$, and $0 = d^2 + bc = d^2$, and so $d = 0$. Once again we conclude that $1 = b(a+d) = 0$, which is a contradiction. Hence no such matrix B exists.

Remark

Examples 2.45 and 2.46 show that the idea of a square root of a square matrix is much more complicated that for real or complex numbers. A square matrix may have no square roots, many square roots, or even infinitely many square roots.

Example 2.47

Consider the system of equations

$$ax + by = e \qquad cx + dy = f \tag{2.4}$$

Rearranging to solve for x and y we obtain

$$x = \frac{de - bf}{ad - bc} \qquad y = \frac{af - ce}{ad - bc} \tag{2.5}$$

This, however, assumes that $ad - bc \neq 0$ (otherwise we have divided through by zero).

We can represent this calculation using matrices. Equation (2.4) becomes

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} e \\ f \end{pmatrix}$$

and equation (2.5) becomes

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} e \\ f \end{pmatrix}$$

Notice that

$$\begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = (ad - bc)I_2 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \tag{2.6}$$

If we have that $ad - bc \neq 0$ and we define

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

then we have just seen that $BA = I_2 = AB$. In other words, B is the *inverse* of the matrix A.

Example 2.48

Let

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

Notice that $1 \times 0 - (-2) \times 3 = 6 \neq 0$. Write

$$B = \frac{1}{6} \begin{pmatrix} 0 & 2 \\ -3 & 1 \end{pmatrix}$$

Then

$$\begin{aligned} AB &= \frac{1}{6} \begin{pmatrix} 1 & -2 \\ 3 & 0 \end{pmatrix} \begin{pmatrix} 0 & 2 \\ -3 & 1 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 0+6 & 2-2 \\ 0+0 & 6+0 \end{pmatrix} = I_2 \\ BA &= \frac{1}{6} \begin{pmatrix} 0 & 2 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 \\ 3 & 0 \end{pmatrix} = \frac{1}{6} \begin{pmatrix} 0+6 & 0+0 \\ -3+3 & 6+0 \end{pmatrix} = I_2 \end{aligned}$$

Definition 2.49

Let $A \in \mathrm{Mat}_{nn}$ be a square matrix. We say the B is an **inverse** of A if $BA = I_n = AB$. If A has an inverse then we say that A is **invertible**, otherwise we say that A is **singular**.

Proposition 2.50 (Properties of inverses)

(i) If $A \in \mathrm{Mat}_{nn}$ has an inverse, then it is *unique*. We write A^{-1} for this inverse.

(ii) If $A, B \in \mathrm{Mat}_{nn}$ are invertible then AB is invertible with $(AB)^{-1} = B^{-1}A^{-1}$.

(iii) If $A \in \mathrm{Mat}_{nn}$ is invertible then A^{-1} is invertible with $(A^{-1})^{-1} = A$.

Proof. (i): Suppose that $B, C \in \mathrm{Mat}_{nn}$ are inverses for A. Then

$$C = I_n C = (BA)C = B(AC) = BI_n = B$$

Hence $C = B$.

(ii): Notice that

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n$$
$$(B^{-1}A^{-1})(AB) = B^{-1}(A^{-1}A)B = B^{-1}I_nB = B^{-1}B = I_n$$

and so $(AB)^{-1} = B^{-1}A^{-1}$ by uniqueness of inverses.

(iii): Note that

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

and so $(A^{-1})^{-1} = A$ by uniqueness of inverses. □

Definition 2.51

If $A \in \mathrm{Mat}_{mn}$ and $BA = I_n$ then B is said to be a **left inverse**. If C satisfies $AC = I_m$ then C is said to be a **right inverse**.

Proposition 2.52

The matrix

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

has an inverse if and only if $ad - bc \neq 0$. If $ad - bc \neq 0$ then

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

Proof. We saw in (2.6) that if $ad - bc \neq 0$ then $AA^{-1} = I_2 = A^{-1}A$. If, however, $ad - bc = 0$ then

$$B = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

satisfied $BA = 0_{22}$. If an inverse C for A exists, then

$$0_{22} = 0_{22}C = (BA)C = B(AC) = BI_2 = B$$

Hence $a = b = c = d = 0$ and so $A = 0_{22}$, which contradicts $AC = I_2$. □

Remark

The scalar $ad - bc$ is called the **determinant** of A, and denoted $\det A$. There is a generalisation of $\det A$ for square matrices $A \in \mathrm{Mat}_{nn}$ and, in general, A is invertible if and only if $\det A \neq 0$. We get to this later.

Python 2.53

We can calculate the determinant of matrices in Python. The following code prints the determinant of the matrix A from Example 2.48.

```
A = np.array([[1,-2],[3,0]])
print(np.linalg.det(A))
```

Python also allows us to calculate the inverse of matrices using NumPy arrays. The following code prints the inverse of A.

```
print(np.linalg.inv(A))
```

2.3 Reduced row echelon form

We will begin by showing that the set of solutions of a linear system of equations does not change under the application of an elementary row operation. Applying an elementary row operation to a linear system $(A \mid \underline{b})$ is equivalent to premultiplying by an invertible elementary matrix E to obtain $(EA \mid E\underline{b})$. It is precisely because E is invertible that the set of solutions remains unchanged.

Proposition 2.54

Let $A \in \mathrm{Mat}_{mn}$. Applying any of the elementary row operations S_{IJ}, $M_I(\lambda)$, or $A_{IJ}(\lambda)$ is equivalent to premultiplying A by matrices which we also denote, respectively, by S_{IJ}, $M_I(\lambda)$, or $A_{IJ}(\lambda)$. These matrices are defined as follows:

$$\text{the } (i,j)\text{th entry of } S_{IJ} = \begin{cases} 1, & i = j, i \neq I, i \neq J; \\ 1, & i = J, j = I; \\ 1, & i = I, j = J; \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{the } (i,j)\text{th entry of } M_I(\lambda) = \begin{cases} 1, & i = j, i \neq I; \\ \lambda, & i = j = I; \\ 0, & \text{otherwise.} \end{cases}$$

$$\text{the } (i,j)\text{th entry of } A_{IJ}(\lambda) = \begin{cases} 1, & i = j; \\ \lambda, & i = J, j = I; \\ 0, & \text{otherwise.} \end{cases}$$

We call these the **elementary matrices**.

Example 2.55

Set $m = 3$. Then

$$S_{21} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad M_3(5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{pmatrix} \qquad A_{31}(5) = \begin{pmatrix} 1 & 0 & 5 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Example 2.56

Set $m = 4$. Then

$$S_{14} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \qquad M_2(-2) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad A_{24}\left(\frac{1}{2}\right) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & \frac{1}{2} & 0 & 1 \end{pmatrix}$$

Remark

Notice that the elementary matrices are given by applying the corresponding elementary row operation to the identity matrix I_m.

Proposition 2.57

The elementary matrices are invertible.

Proof. Just observe that

$$\begin{aligned} (S_{ij})^{-1} &= S_{ji} = S_{ij} \\ (M_i(\lambda))^{-1} &= M_i\left(\frac{1}{\lambda}\right) \\ (A_{ij}(\lambda))^{-1} &= A_{ij}(-\lambda) \end{aligned}$$

□

Corollary 2.58

Let $(A \mid \underline{b})$ be a linear system of m equations and let $E \in \mathrm{Mat}_{mm}$ be an elementary matrix. Then $\underline{x}$ is a solution of $(A \mid \underline{b})$ if and only if $\underline{x}$ is a solution of $(EA \mid E\underline{b})$.

Proof. If $A\underline{x} = \underline{b}$ then, by premultiplying by E, we have $EA\underline{x} = E\underline{b}$. If $EA\underline{x} = E\underline{b}$ then, by premultiplying by E^{-1}, we have $A\underline{x} = \underline{b}$. □

Remark

Corollary 2.58 tells us that applying elementary row operations does not alter the set of solutions of a linear system of equations.

Definition 2.59

A matrix A is said to be in **reduced row echelon form** (or **RREF**) if:

(i) the first (i.e. leftmost) non-zero entry of any non-zero row is 1 (this is referred to as the **leading 1** of the row, or as a **pivot**);

(ii) the leading 1 of a non-zero row appears (strictly) to the right of the leading 1s of the nonzero rows above it;

(iii) any zero rows appear below the non-zero rows;

(iv) in a column that contains the leading 1 of some row, all other entries of that column are zero.

If only (i–iii) hold, we say the matrix is in **row echelon form**; this is also useful, but we will always use the RREF here.

Example 2.60

The following three matrices are in reduced row echelon form

$$\begin{pmatrix} 0 & 1 & 2 & 0 & -3 \\ 0 & 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$$

Example 2.61

The following two matrices are *not* in reduced row echelon form

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

The first matrix contains a leading 1 in the second column, but not all other entries of that column are 0. The second matrix fails because the leading entry of the third row is not 1.

Example 2.62

Look once again at Example 2.12. Notice that we solved the linear system of equations by placing the augmented matrix in reduced row echelon form.

Proposition 2.63

Let $(A \mid \underline{b})$ be a matrix in reduced row echelon form which represents a linear system $A\underline{x} = \underline{b}$

of m equations in n variables. Then:

(i) The system has no solutions if and only if the last non-zero row of $(A \mid \underline{b})$ is

$$\left(\begin{array}{cccc|c} 0 & 0 & \cdots & 0 & 1 \end{array} \right)$$

(ii) The system has a unique solution if and only if the non-zero rows of A form the identity matrix I_n. In particular, this case is only possible if $m \geq n$.

(iii) The system has infinitely many solutions if $(A \mid \underline{b})$ has as many non-zero rows as A, and not every column of A contains a pivot (i.e. a leading 1 of some nonzero row). The set of solutions can be described with k parameters, where k is the number of columns not containing a pivot.

Proof. If $(A \mid \underline{b})$ contains the row $(0\ 0\ \cdots\ 0 \mid 1)$ then the system is inconsistent as no $\underline{x}$ satisfies

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

Since $(A \mid \underline{b})$ is in reduced row echelon form, this is the only way in which $(A \mid \underline{b})$ can have more non-zero rows than A. We will show that whenever $(A \mid \underline{b})$ has as many non-zero rows as A then the system $(A \mid \underline{b})$ is consistent.

Suppose that both $(A \mid \underline{b})$ and A have r non-zero rows, so that there are r leading 1s within these rows and we have $k = n - r$ columns without leading 1s. By reordering the numbers of the variables $x_1, \ldots, x_n$ if necessary, we can assume that the leading 1s appear in the first r columns. So, ignoring any zero rows and remembering that the system is in reduced row echelon form, the system corresponds to the r equations

$$\begin{aligned} x_1 + a_{1(r+1)}x_{r+1} + \cdots + a_{1n}x_n &= b_1 \\ x_2 + a_{2(r+1)}x_{r+1} + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ x_r + a_{r(r+1)}x_{r+1} + \cdots + a_{rn}x_n &= b_r \end{aligned}$$

We can see that if we assign $x_{r+1}, \ldots, x_n$ the k parameters $s_{r+1}, \ldots, s_n$, then we can read off from the r equations the values for $x_1, x_2, \ldots, x_r$:

$$\begin{aligned} x_1 &= b_1 - a_{1(r+1)}s_{r+1} - \cdots - a_{1n}s_n \\ x_2 &= b_2 - a_{2(r+1)}s_{r+1} - \cdots - a_{2n}s_n \\ &\vdots \\ x_r &= b_r - a_{r(r+1)}s_{r+1} - \cdots - a_{rn}s_n \end{aligned}$$

So, for any values of the parameters, we have a solution $\underline{x}$. Conversely, if $\underline{x} = (x_1, x_2, \ldots, x_n)$ is a solution, then it appears amongst the solutions we have just found when we assign values $s_{r+1} = x_{r+1}, \ldots, s_n = x_n$ to the parameters. Thus we see that we have an infinite set of solutions associated with $k = n - r$ independent parameters when $n > r$, and a unique solution when $n = r$ (in which case the non-zero rows of A are the identity matrix I_n). □

Remark

We have just shown that:

(i) a system $(A \mid \underline{b})$ in reduced row echelon form is consistent if and only if $(A \mid \underline{b})$ has as many non-zero rows as A;

(ii) all the solutions of a consistent system can be found by assigning parameters to the variables corresponding to the columns without pivots (leading 1s of each nonzero row).

Example 2.64

Consider the linear system of equations

$$\begin{aligned} 2x + 3y - z &= 1 \\ 10x - z &= 2 \\ 4x - 9y + 3z &= 5 \end{aligned}$$

We will solve this system by first writing down the augmented matrix, and then placing it into reduced row echelon form by performing elementary row operations.

$$\left(\begin{array}{ccc|c} 2 & 3 & -1 & 1 \\ 10 & 0 & -1 & 2 \\ 4 & -9 & 3 & 5 \end{array}\right) \xrightarrow{A_{12}(-5), A_{13}(-2)} \left(\begin{array}{ccc|c} 2 & 3 & -1 & 1 \\ 0 & -15 & 4 & -3 \\ 0 & -15 & 5 & 3 \end{array}\right) \xrightarrow{M_1(1/2)} \left(\begin{array}{ccc|c} 1 & \frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & -15 & 4 & -3 \\ 0 & -15 & 5 & 3 \end{array}\right)$$

$$\xrightarrow{A_{23}(-1)} \left(\begin{array}{ccc|c} 1 & \frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & -15 & 4 & -3 \\ 0 & 0 & 1 & 6 \end{array}\right) \xrightarrow{M_2(-1/15)} \left(\begin{array}{ccc|c} 1 & \frac{3}{2} & -\frac{1}{2} & \frac{1}{2} \\ 0 & 1 & -\frac{4}{15} & \frac{1}{5} \\ 0 & 0 & 1 & 6 \end{array}\right) \xrightarrow{A_{21}(-3/2)} \left(\begin{array}{ccc|c} 1 & 0 & -\frac{1}{10} & \frac{1}{5} \\ 0 & 1 & -\frac{4}{15} & \frac{1}{5} \\ 0 & 0 & 1 & 6 \end{array}\right)$$

$$\xrightarrow{A_{31}(1/10), A_{32}(4/15)} \left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{4}{5} \\ 0 & 1 & 0 & \frac{9}{5} \\ 0 & 0 & 1 & 6 \end{array}\right)$$

Hence there is a unique solution given by

$$x = \frac{4}{5} \qquad y = \frac{9}{5} \qquad z = 6$$

Python 2.65

It is possible to solve a linear system of equations in Python using NumPy. The following code verifies the solution of Example 2.64.

```
A = np.array([[2,3,-1],[10,0,-1],[4,-9,3]])
b = np.array([1,2,5])
print(np.linalg.solve(A,b))
```

Example 2.66

The following augmented matrices are in reduced row echelon form.

(i) No solutions – notice the final row.

$$\left(\begin{array}{cccc|c} 1 & -2 & 0 & 2 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right)$$

(ii) A unique solution given by $x_1 = 2$, $x_2 = -1$, and $x_3 = 3$.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & 2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

(iii) A one-parameter family of solutions (assigning the parameter s to the second column) given by $\underline{x} = (3 - 2s, s, 2, 1)$.

$$\left(\begin{array}{cccc|c} 1 & 2 & 0 & 0 & 3 \\ 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 1 \end{array}\right)$$

(iv) A two-parameter family of solutions (assigning the parameter s to the second column and the parameter t to the fourth column) given by $\underline{x} = (3 + 2s - 2t, s, -2 - t, t)$.

$$\left(\begin{array}{cccc|c} 1 & -2 & 0 & 2 & 3 \\ 0 & 0 & 1 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

Theorem 2.67

Every matrix can be reduced by elementary row operations to a matrix in reduced row echelon form.

Proof. Let $A \in \mathrm{Mat}_{mn}$. We will proceed by induction on the number of rows, m.

First suppose $m = 1$. Notice that a $1 \times n$ matrix is either zero, or can be put into reduced row echelon form by dividing through by the leading entry.

Now suppose that the inductive hypothesis holds for any matrix with fewer than m rows. If $A = 0_{mn}$ then it is already in reduced row echelon form. So suppose A is non-zero. Let $\mathbf{c}_j$ be first column in A containing a non-zero entry α. By using elementary row operations we can swap the row containing α with the first row, and then divide the first row by $\alpha \neq 0$. Thus the $(1, j)$th entry now equals 1 and our matrix takes the form

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & \overline{a}_{1(j+1)} & \cdots & \overline{a}_{1n} \\ 0 & \cdots & 0 & \overline{a}_{2j} & \overline{a}_{2(j+1)} & \cdots & \overline{a}_{2n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & \overline{a}_{mj} & \overline{a}_{m(j+1)} & \cdots & \overline{a}_{mn} \end{pmatrix}$$

for some entries $\overline{a}_{IJ}$. Applying the row operations $A_{12}(-\overline{a}_{2j}), A_{13}(-\overline{a}_{3j}), \ldots, A_{1m}(-\overline{a}_{mj})$ transforms column $\mathbf{c}_j$ to $\mathbf{e}_1^T$. Thus our matrix becomes

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & \overline{a}_{1(j+1)} & \cdots & \overline{a}_{1n} \\ 0 & \cdots & 0 & 0 & & & \\ \vdots & & \vdots & \vdots & & B & \\ 0 & \cdots & 0 & 0 & & & \end{pmatrix}$$

By induction, the $(m-1)\times(n-j)$ matrix B can be placed in reduced row echelon form by applying elementary row operations. Applying those same elementary row operations to the bottom $m-1$ rows of the above matrix would reduce A to

$$\begin{pmatrix} 0 & \cdots & 0 & 1 & \overline{a}_{1(j+1)} & \cdots & \overline{a}_{1n} \\ 0 & \cdots & 0 & 0 & & & \\ \vdots & & \vdots & \vdots & & RREF(B) & \\ 0 & \cdots & 0 & 0 & & & \end{pmatrix}$$

(Here $RREF(B)$ denotes the reduced row echelon form of B.) To transform this matrix into reduced row echelon form, we need to zero-out any of the $\overline{a}_{1(j+1)}, \ldots, \overline{a}_{1n}$ which are above a leading 1 in $RREF(B)$. If $\overline{a}_{1k}$ is the first entry to lie above a leading 1 in row ℓ then $A_{\ell 1}(-\overline{a}_{1k})$ will transform the $(1,k)$th entry to 0. Thus we can place A in reduced row echelon form via elementary row operation. □

Definition 2.68

The process of applying elementary row operations to transform a matrix into reduced row echelon form is called **row reduction** (or just **reduction**) or **Gauss elimination** or **Gauss–Jordan elimination** or . . . (OK, maybe that's all).

Corollary 2.69

If $A \in \mathrm{Mat}_{mn}$ and $m < n$, then there is a nontrivial solution $\underline{v} \neq \underline{0}$ to $A\underline{v} = \underline{0}$.

Proof. By Gauss elimination, Theorem 2.67, the matrix A has reduced row echelon form. Moreover, the RREF of the (augmented) matrix $(A \mid \underline{0})$ is simply that of A augmented by a zero column, since the final column remains $\underline{0}$ throughout the reduction process.

Thus the reduced row echelon form of $(A \mid \underline{0})$ has the same number of rows as that of A, and since the number of pivots is at most m, not all of the $n > m$ columns can contain a pivot. Therefore there are infinitely many solutions by Proposition 2.63, and so there is a nontrivial one. □

It was enough until now simply to perform row operations in some order, and record the matrix at each stage. But we gain a lot next by recalling from Proposition 2.54 that each row operation may be performed by premultiplying A by the corresponding elementary matrix S_{IJ}, $M_I(\lambda)$ or $A_{IJ}(\lambda)$.

2.4 Inverse matrix

In Proposition 2.52 we saw a formula for the inverse of a 2×2 matrix. A similar formula holds for the inverse of a 3×3 matrix, but it is very messy. Instead, we will use elementary row operations to efficiently determine wither an $n\times n$ matrix is invertible and, if so, how to find the inverse.

Proposition 2.70

Let $A \in \mathrm{Mat}_{nn}$. Form the augmented $n \times 2n$ matrix $(A \mid I_n)$ given by placing A side-by-side with the identity matrix I_n. There are elementary row operations that will reduce A to a matrix $R \in \mathrm{Mat}_{nn}$ in reduced row echelon form. We simultaneously apply these elementary row operations to both sides of $(A \mid I_n)$ until we arrive at $(R \mid P)$, for some $P \in \mathrm{Mat}_{nn}$. Then:

(i) if $R = I_n$ then A is invertible with $A^{-1} = P$;

(ii) if $R \neq I_n$ then A is singular.

Proof. Let $E_1, E_2 \dots, E_k$ be a sequence of elementary matrices that reduce A to R. So $(A \mid I_n)$ becomes

$$(E_k E_{k-1} \cdots E_1 A \mid E_k E_{k-1} \cdots E_1) = (R \mid P)$$

Hence $P = E_k E_{k-1} \cdots E_1$ and $R = PA$. If $R = I_n$ then

$$(E_k E_{k-1} \cdots E_1)A = I_n \qquad \text{and so} \qquad A^{-1} = E_k E_{k-1} \cdots E_1 = P$$

since by Proposition 2.57 elementary matrices are invertible. If $R \neq I_n$ then, since R is in reduced row echelon form and is square, R must have at least one zero row. It follows that (possibly after reordering the rows of PA)

$$(1, 0, \dots, 0)(PA) = \mathbf{0}$$

Since P is invertible, if A were invertible then we could postmultiply by $A^{-1}P^{-1}$ to obtain

$$(1, 0, \dots, 0) = \mathbf{0}$$

which is a contradiction. Hence A is singular. □

Remark

The proof of Proposition 2.70 tells us that as soon as a zero row appears when reducing A then we know that A is singular.

It is best to treat the remaining examples in the rest of this chapter as a sequence of exercises that come with worked solutions: try to compute the inverse in each case (or prove that it does not exist), and don't forget to check your answer at the end.

Example 2.71

Let

$$A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 3 & 1 \end{pmatrix}$$

Compute A^{-1} – if it exists – using Proposition 2.70.

$$(A \mid I_3) = \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ 1 & 3 & 1 & 0 & 0 & 1 \end{array}\right) \xrightarrow{A_{12}(-2), A_{13}(-1)} \left(\begin{array}{ccc|ccc} 1 & 2 & 1 & 1 & 0 & 0 \\ 0 & -3 & -2 & -2 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{A_{31}(-2), A_{32}(3)} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 3 & 0 & -2 \\ 0 & 0 & -2 & -5 & 1 & 3 \\ 0 & 1 & 0 & -1 & 0 & 1 \end{array}\right) \xrightarrow{S_{23}} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 3 & 0 & -2 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & -2 & -5 & 1 & 3 \end{array}\right)$$

$$\xrightarrow{M_3(-1/2)} \left(\begin{array}{ccc|ccc} 1 & 0 & 1 & 3 & 0 & -2 \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & \frac{5}{2} & -\frac{1}{2} & -\frac{3}{2} \end{array}\right) \xrightarrow{A_{31}(-1)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & 0 & -1 & 0 & 1 \\ 0 & 0 & 1 & \frac{5}{2} & -\frac{1}{2} & -\frac{3}{2} \end{array}\right)$$

We conclude that A^{-1} exists, and is equal to

$$A^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 1 & -1 \\ -2 & 0 & 2 \\ 5 & -1 & -3 \end{pmatrix}$$

We can easily verify our result

$$AA^{-1} = \frac{1}{2} \begin{pmatrix} 1 & 2 & 1 \\ 2 & 1 & 0 \\ 1 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 & -1 \\ -2 & 0 & 2 \\ 5 & -1 & -3 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1-4+5 & 1+0-1 & -1+4-3 \\ 2-2+0 & 2+0+0 & -2+2+0 \\ 1-6+5 & 1+0-1 & -1+6-3 \end{pmatrix} = I_3$$

Example 2.72

Let

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 4 & 0 & 4 \\ 0 & 2 & 0 \end{pmatrix}$$

Calculate A^{-1}, if it exists.

$$(A \mid I_3) = \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 4 & 0 & 4 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{A_{12}(-4)} \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & -4 & -4 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{S_{23}} \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & -4 & -4 & 1 & 0 \end{array}\right) \xrightarrow{M_2(1/2), M_3(-1/4)} \left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 1 & -\frac{1}{4} & 0 \end{array}\right)$$

$$\xrightarrow{A_{31}(-2)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -1 & \frac{1}{2} & 0 \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 & 1 & -\frac{1}{4} & 0 \end{array}\right)$$

We conclude that A is invertible, with inverse

$$A^{-1} = \frac{1}{4} \begin{pmatrix} -4 & 2 & 0 \\ 0 & 0 & 2 \\ 4 & -1 & 0 \end{pmatrix}$$

We shall verify our calculation:

$$AA^{-1} = \frac{1}{4} \begin{pmatrix} 1 & 0 & 2 \\ 4 & 0 & 4 \\ 0 & 2 & 0 \end{pmatrix} \begin{pmatrix} -4 & 2 & 0 \\ 0 & 0 & 2 \\ 4 & -1 & 0 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} -4+0+8 & 2+0-2 & 0+0+0 \\ -16+0+16 & 8+0-4 & 0+0+0 \\ 0+0+0 & 0+0+0 & 0+4+0 \end{pmatrix} = I_3$$

Example 2.73

Let

$$A := \begin{pmatrix} 1 & 0 & 4 \\ 0 & 1 & 2 \\ -1 & 2 & 0 \end{pmatrix}$$

Calculate A^{-1}, or prove that A is singular.

$$(A \mid I_3) = \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ -1 & 2 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{A_{13}(1)} \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 2 & 4 & 1 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{A_{23}(-2)} \left(\begin{array}{ccc|ccc} 1 & 0 & 4 & 1 & 0 & 0 \\ 0 & 1 & 2 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & 1 \end{array}\right)$$

We stop at this step. Since the third row is zero, we conclude that A is singular and no inverse exists.

Example 2.74

Let

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

Calculate A^{-1}, if it exists.

$$(A \mid I_4) = \left(\begin{array}{cccc|cccc} 1 & 3 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & 0 & 1 & 0 & 0 \\ 3 & 1 & 2 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{A_{13}(-3)} \left(\begin{array}{cccc|cccc} 1 & 3 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & -8 & 5 & 1 & -3 & 0 & 1 & 0 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{S_{24}} \left(\begin{array}{cccc|cccc} 1 & 3 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \\ 0 & -8 & 5 & 1 & -3 & 0 & 1 & 0 \\ 0 & 2 & 1 & 1 & 0 & 1 & 0 & 0 \end{array}\right)$$

$$\xrightarrow{A_{21}(-3),A_{23}(8),A_{24}(-2)} \left(\begin{array}{cccc|cccc} 1 & 0 & -16 & -9 & 1 & 0 & 0 & -3 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 45 & 25 & -3 & 0 & 1 & 8 \\ 0 & 0 & -9 & -5 & 0 & 1 & 0 & -2 \end{array}\right)$$

$$\xrightarrow{M_3(1/45)} \left(\begin{array}{cccc|cccc} 1 & 0 & -16 & -9 & 1 & 0 & 0 & -3 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{5}{9} & -\frac{1}{15} & 0 & \frac{1}{45} & \frac{8}{45} \\ 0 & 0 & -9 & -5 & 0 & 1 & 0 & -2 \end{array}\right)$$

$$\xrightarrow{A_{34}(9)} \left(\begin{array}{cccc|cccc} 1 & 0 & -16 & -9 & 1 & 0 & 0 & -3 \\ 0 & 1 & 5 & 3 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & \frac{5}{9} & -\frac{1}{15} & 0 & \frac{1}{45} & \frac{8}{45} \\ 0 & 0 & 0 & 0 & -\frac{3}{5} & 1 & \frac{1}{5} & -\frac{2}{5} \end{array}\right)$$

We stop here, noticing that the final row is zero. Hence A is singular.

Example 2.75

Let

$$A = \begin{pmatrix} 1 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 2 & 1 & 1 & 0 \\ 2 & 0 & 2 & 0 \end{pmatrix}$$

Calculate A^{-1}, if it exists.

$$(A \mid I_4) = \left(\begin{array}{cccc|cccc} 1 & 1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 2 & 0 & 2 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{A_{12}(-2), A_{13}(-2)} \left(\begin{array}{cccc|cccc} 1 & 1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & -1 & -3 & 0 & -2 & 0 & 1 & 0 \\ 0 & -2 & -2 & 0 & -2 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow{S_{23}} \left(\begin{array}{cccc|cccc} 1 & 1 & 2 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & -3 & 0 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & -2 & -2 & 0 & -2 & 0 & 0 & 1 \end{array}\right) \xrightarrow{A_{21}(1), A_{24}(-2)} \left(\begin{array}{cccc|cccc} 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & -3 & 0 & -2 & 0 & 1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 2 & 0 & -2 & 1 \end{array}\right)$$

$$\xrightarrow{M_2(-1)} \left(\begin{array}{cccc|cccc} 1 & 0 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & 3 & 0 & 2 & 0 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 4 & 0 & 2 & 0 & -2 & 1 \end{array}\right)$$

$$\xrightarrow{A_{31}(1), A_{32}(-3), A_{34}(-4)} \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 2 & -1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -6 & 2 & -3 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -8 & 2 & -4 & -2 & 1 \end{array}\right)$$

$$\xrightarrow{M_4(-1/8)} \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 2 & -1 & 1 & 1 & 0 \\ 0 & 1 & 0 & -6 & 2 & -3 & -1 & 0 \\ 0 & 0 & 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & -\frac{1}{8} \end{array}\right)$$

$$\xrightarrow{A_{41}(-2), A_{42}(6), A_{43}(-2)} \left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & -\frac{1}{2} & 0 & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & -\frac{3}{4} \\ 0 & 0 & 1 & 0 & \frac{1}{2} & 0 & -\frac{1}{2} & \frac{1}{4} \\ 0 & 0 & 0 & 1 & -\frac{1}{4} & \frac{1}{2} & \frac{1}{4} & -\frac{1}{8} \end{array}\right)$$

We conclude that A is invertible, with

$$A^{-1} = \frac{1}{8} \begin{pmatrix} -4 & 0 & 4 & 2 \\ 4 & 0 & 4 & -6 \\ 4 & 0 & -4 & 2 \\ -2 & 4 & 2 & -1 \end{pmatrix}$$

Python 2.76

The following code verifies the result of Example 2.75.

```python
A = np.array([[1,1,2,0],[0,0,1,2],[2,1,1,0],[2,0,2,0]])
print(np.linalg.inv(A))
```

Chapter 3

Subspaces and bases of $\mathbb{R}^n$

3.1 Span and subspace

Suppose $\underline{v}, \underline{w} \in \mathbb{R}^3$ and consider the set of all possible linear combinations that we could make using them: we call this set the **span** of $\{\underline{v}, \underline{w}\}$ and we denote it synonymously by either $\langle \underline{v}, \underline{w} \rangle$ or $\text{span}\{\underline{v}, \underline{w}\}$; that is

$$\langle \underline{v}, \underline{w} \rangle = \text{span}\{\underline{v}, \underline{w}\} = \{\lambda \underline{v} + \mu \underline{w} \mid \lambda, \mu \in \mathbb{R}\}$$

Note that clearly $\langle \underline{v}, \underline{w} \rangle \subset \mathbb{R}^3$. More importantly, if you pick any two elements of $\langle \underline{v}, \underline{w} \rangle$, then their sum is also an element of it; indeed, so is any linear combination of them. After all, if we are given two vectors $\lambda_1 \underline{v} + \mu_1 \underline{w}$ and $\lambda_2 \underline{v} + \mu_2 \underline{w}$ in the span, then their sum is

$$(\lambda_1 \underline{v} + \mu_1 \underline{w}) + (\lambda_2 \underline{v} + \mu_2 \underline{w}) = (\lambda_1 + \lambda_2)\underline{v} + (\mu_1 + \mu_2)\underline{w}$$

which is visibly also in the span. We say that $\langle \underline{v}, \underline{w} \rangle$ is a subspace of $\mathbb{R}^3$, as in the following definition.

Definition 3.1

A **subspace** of $\mathbb{R}^n$ is a nonempty $W \subset \mathbb{R}^n$ with the property that for any $\underline{v}, \underline{w} \in W$ and any $\lambda \in \mathbb{R}$, we also have $\underline{v} + \underline{w} \in W$ and $\lambda \underline{v} \in W$.

There are two particular subspaces that we refer to as **trivial** subspaces: $\{\underline{0}\} \subset \mathbb{R}^n$ and $\mathbb{R}^n \subset \mathbb{R}^n$. Thus to say a subspace $W \subset \mathbb{R}^n$ is **nontrivial** is to say $W \neq \{\underline{0}\}$ and $W \neq \mathbb{R}^n$.

Example

In the situation $\langle \underline{v}, \underline{w} \rangle \subset \mathbb{R}^3$ above, there are three different types of behaviour that may happen.

- If $\underline{v}$ and $\underline{w}$ are not collinear, then the span $\langle \underline{v}, \underline{w} \rangle$ is a 2-dimensional plane inside $\mathbb{R}^3$ (that passes through the origin): it is exactly the same as the plane through $\underline{v}$, $\underline{w}$ and $\underline{0}$. For example, if $\underline{v} = \underline{e}_1$ and $\underline{w} = \underline{e}_2$ then

$$\langle \underline{v}, \underline{w} \rangle = \left\{ \begin{pmatrix} \lambda \\ \mu \\ 0 \end{pmatrix} \;\middle|\; \lambda, \mu \in \mathbb{R} \right\}$$

  is the $z = 0$ coordinate plane. As another example, if $\underline{v} = (-2, 1, 5)^T$ and $\underline{w} = (1, 1, 2)^T$ then by calculating as in §1.3, we can describe the span either parametrically (which is essentially its definition) or implicitly by an equation

$$\langle \underline{v}, \underline{w} \rangle = \{\lambda \underline{v} + \mu \underline{w} \mid \lambda, \mu \in \mathbb{R}\} = (x - 3y + z = 0) \subset \mathbb{R}^3$$

- If $\underline{v}$ and $\underline{w}$ are collinear but not both $\underline{0}$, then the span $\langle \underline{v}, \underline{w} \rangle$ is a 1-dimensional line inside $\mathbb{R}^3$ (that passes through the origin). For example, if $\underline{v} = \underline{e}_1$ and $\underline{w} = -3\underline{e}_1$ (or any other multiple of $\underline{v}$, including the zero multiple $\underline{w} = \underline{0}$) then

$$\langle \underline{v}, \underline{w} \rangle = \left\{ \begin{pmatrix} \lambda \\ 0 \\ 0 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\} = (y = z = 0) \subset \mathbb{R}^3$$

  is the x-axis (described first parametrically and then implicitly by equations).

- If $\underline{v} = \underline{w} = \underline{0}$, then the span $\langle \underline{v}, \underline{w} \rangle = \{\underline{0}\}$ is simply the zero vector.

It is true (as we shall see later, or you can persuade yourself now) that the only subspaces of $\mathbb{R}^3$, apart from the trivial ones, are lines through the origin and planes through the origin. That's why we went to all the fuss of thinking about them in §1.3.

In $\mathbb{R}^2$, the only nontrivial subspaces are lines through the origin.

More generally, we may construct the span of any number of vectors, and the span is a subspace.

Definition 3.2

Let $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$. Then their **span** is the set

$$\langle \underline{v}_1, \ldots, \underline{v}_s \rangle = \left\{ \sum_{i=1}^{s} \lambda_i \underline{v}_i \middle| \lambda_i \in \mathbb{R} \text{ for } i = 1, \ldots, s \right\}$$

consisting of all possible linear combinations of $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$.

It is convenient to refer to the vectors $\underline{v}_1, \ldots, \underline{v}_s$ as the **given generators** of $\langle \underline{v}_1, \ldots, \underline{v}_s \rangle$, but note that this is not standard usage, and most subspaces have many alternative generators.

Remark

In fact, for any subset $S \subset \mathbb{R}^n$ you may define the span $\langle S \rangle \subset \mathbb{R}^n$ in the same way, but be clear that when taking linear combinations you are only permitted **finite** sums: there are no infinite series here.

Proposition 3.3

For any $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$, the span $\langle \underline{v}_1, \ldots, \underline{v}_s \rangle \subset \mathbb{R}^n$ is a subspace of $\mathbb{R}^n$.

The proof is clear and routine: just check the rules of Definition 3.1: better to do this yourself, rather than read it, and certainly don't invest any effort to remember it.

Proof. Denote the span by $W = \langle \underline{v}_1, \ldots, \underline{v}_s \rangle$. Suppose $\underline{v}, \underline{w} \in W$. That is, there are scalars $\lambda_i, \mu_j \in \mathbb{R}$ for which

$$\underline{v} = \lambda_1 \underline{v}_1 + \ldots + \lambda_s \underline{v}_s \quad \text{and} \quad \underline{w} = \mu_1 \underline{v}_1 + \ldots + \mu_s \underline{v}_s.$$

Therefore, collecting coefficients together

$$\underline{v} + \underline{w} = (\lambda_1 + \mu_1)\underline{v}_1 + \ldots + (\lambda_s + \mu_s)\underline{v}_s$$

which is a linear combination of $\underline{v}_1, \ldots, \underline{v}_s$ and so lies in the span.

Similarly, if $\alpha \in \mathbb{R}$ is any scalar, then

$$\alpha\underline{v} = (\alpha\lambda_1)\underline{v}_1 + \ldots + (\alpha\lambda_s)\underline{v}_s$$

which is a linear combination of $\underline{v}_1, \ldots, \underline{v}_s$ and so also lies in the span, as required. Thus the span is a subspace of $\mathbb{R}^n$. □

Any matrix $A \in \mathrm{Mat}_{mn}$ determines a particularly important subspace of $\mathbb{R}^m$.

Definition 3.4

Let $A = (a_{ij}) \in \mathrm{Mat}_{mn}$. The **column span** $\mathrm{Colspan}(A)$ **of** A is the span of the columns of A: that is,

$$\mathrm{Colspan}(A) = \langle \underline{v}_1, \ldots, \underline{v}_n \rangle \subset \mathbb{R}^m$$

where $\underline{v}_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{mj} \end{pmatrix}$ is the jth column of A, for $j = 1, \ldots n$.

Example

Let

$$A = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 2 & 4 \\ -1 & 0 & -1 \end{pmatrix}$$

Then the column span $\mathrm{Colspan}(A)$ of A is

$$\mathrm{Colspan}(A) = \left\langle \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ -1 \end{pmatrix} \right\rangle$$

Since the third column $\underline{v}_3$ is equal a linear combination $\underline{v}_3 = \underline{v}_1 + 2\underline{v}_2$ of the first two columns, it can't contribute anything to the span that those first two columns don't already: if you ever see $\underline{v}_3$ in some linear expression, you can get rid of it by replacing it by $\underline{v}_1 + 2\underline{v}_2$. Therefore the third column can safely be omitted from the list of given generators. Doing so shows that $\mathrm{Colspan}(A)$ is equal to the column span of a smaller matrix:

$$\mathrm{Colspan}(A) = \left\langle \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix} \right\rangle = \mathrm{Colspan}(B) \quad \text{where} \quad B = \begin{pmatrix} 2 & -1 \\ 0 & 2 \\ -1 & 0 \end{pmatrix}$$

The idea we used in the example to remove the third column from the vectors generating the span in the example above is called **sifting**, and it can be used to optimise the collection of given generators in a span fairly generally. The key is the notion of linear (in)dependence, which we come to next.

We say that a subset $S \subset \mathbb{R}^n$ spans if and only if every vector $\underline{v} \in \mathbb{R}^n$ is a linear combination of (a finite collection of) vectors of S.

Definition 3.5

A subset $S \subset \mathbb{R}^n$ **spans** $\mathbb{R}^n$ if and only if $\langle S \rangle = \mathbb{R}^n$. (It is also common to express this as S **is a spanning set of** $\mathbb{R}^n$.)

Remark

It is ridiculous, but true, to say that $S = \mathbb{R}^n$ is a spanning set for $\mathbb{R}^n$. It is more interesting to try to find small spanning sets – we may think of them as a more efficient way of describing elements of $\mathbb{R}^n$ (without spelling out what *efficient* might mean). The question then becomes: is there a lower bound on the size of a spanning set for $\mathbb{R}^n$? Oh yes there is!

Proposition 3.6

Suppose $S \subset \mathbb{R}^n$ spans $\mathbb{R}^n$. Then S contains at least n elements.

Proof. Suppose that S has $m < n$ elements, say $\underline{v}_1, \ldots, \underline{v}_m$.

Consider the standard basis $\underline{e}_1, \ldots, \underline{e}_n \in \mathbb{R}^n$. Since S spans $\mathbb{R}^n$, each $\underline{e}_i$ is a linear combination of elements of S; that is, there are scalars $a_{ij} \in \mathbb{R}$ so that for each $i = 1, \ldots, n$

$$\underline{e}_i = \sum_{j=1}^{m} a_{ij} \underline{v}_j \tag{3.1}$$

Assemble the coefficients into an $n \times m$ matrix $A = (a_{ij})$ and denote its rows by $\underline{r}_i$ for $i = 1, \ldots, n$. The RREF of A is a product EA where $E \in \mathrm{Mat}_{nn}$ is an invertible matrix (in fact, E is a product of elementary matrices). For example, to set up notation, the final row of EA is of the form

$$k_1 \underline{r}_1 + \ldots + k_n \underline{r}_n \tag{3.2}$$

where $\underline{k} = (k_1, \ldots, k_n)$ is the final row of E. If $\underline{k} = \underline{0}$, then $E\underline{v} = \underline{0}$ for $\underline{v} = (0, \ldots, 0, 1)^T \in \mathbb{R}^n$; but this immediately gives a contradiction: $\underline{v} = E^{-1}E\underline{v} = E^{-1}\underline{0} = \underline{0}$. Therefore $\underline{k} \neq \underline{0}$.

Since $n > m$, the final row of EA must be zero $\underline{0} \in \mathbb{R}^m_{\text{row}}$. Therefore by (3.1) we also have that

$$\begin{aligned} k_1 \underline{e}_1 + \ldots + k_n \underline{e}_n &= k_1 \left(\sum_j a_{1j} \underline{v}_j \right) + \ldots + k_n \left(\sum_j a_{nj} \underline{v}_j \right) \\ &= \left(\sum_{i=1}^{n} k_i a_{i1} \right) \underline{v}_1 + \ldots + \left(\sum_{i=1}^{n} k_i a_{im} \right) \underline{v}_m \\ &= 0\underline{v}_1 + \ldots + 0\underline{v}_m = \underline{0} \end{aligned}$$

since the coefficient $\sum_i k_i a_{ij}$ of $\underline{v}_j$ is the jth entry of the last row of the product EA. But the left-hand side of this expression is simply $\underline{k}^T$, so this says that $\underline{k} = \underline{0}$, which is a contradiction. □

Remark

Were you able to see in your mind's eye the equations (3.1) as an array of the form

$$\begin{aligned} \underline{e}_1 &= a_{11}\underline{v}_1 + a_{12}\underline{v}_2 + \ldots + a_{1m}\underline{v}_m \\ \underline{e}_2 &= a_{21}\underline{v}_1 + a_{22}\underline{v}_2 + \ldots + a_{2m}\underline{v}_m \\ &\vdots \\ \underline{e}_n &= a_{n1}\underline{v}_1 + a_{n2}\underline{v}_2 + \ldots + a_{nm}\underline{v}_m \end{aligned}$$

The proof is saying that performing the row reduction of A is the same as what happens to the coefficients when you form a certain collection of linear combinations of the right-hand sides of these equations. After reducing, the right-hand sides of the final row (indeed any row after the mth row) is necessarily zero. But the left-hand side of that row cannot be zero: the entries of the different $\underline{e}_i$ cannot cancel, since they lie in different components (though the proof uses the power of the elementary matrices to demonstrate that point without having to imagine it).

3.2 Linear independence

The intuitive idea of linear independence of vectors is simple: vectors are called linearly independent if they do not point in the same direction, or, more precisely, if none of them is a linear combination of the others. Nevertheless, understanding the definition and its power properly needs a little care.

Definition 3.7

Vectors $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$ are called **linearly independent** if whenever

$$\lambda_1\underline{v}_1 + \ldots + \lambda_s\underline{v}_s = \underline{0} \text{ for scalars } \lambda_1, \ldots, \lambda_s \in \mathbb{R}$$

we necessarily have that

$$\lambda_1 = \ldots = \lambda_s = 0$$

In other words, $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$ are linearly independent if and only if the only linear combination of them that is $\underline{0}$ is the trivial one.

This definition is sometimes phrased slightly informally as: $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$ are linearly independent if and only if

$$\lambda_1\underline{v}_1 + \ldots + \lambda_s\underline{v}_s = \underline{0} \qquad \Longrightarrow \qquad \lambda_1 = \ldots = \lambda_s = 0$$

Since this is so important, it is worth defining what it means *not* to be linearly independent separately.

Definition 3.8

Vectors $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$ are called **linearly dependent** if and only if there is an equation

$$\lambda_1\underline{v}_1 + \ldots + \lambda_s\underline{v}_s = \underline{0} \tag{3.3}$$

for scalars $\lambda_1, \ldots, \lambda_s \in \mathbb{R}$ which are not all zero.

The equation (3.3) is referred to as a **linear dependence relation** for the vectors $\underline{v}_1, \ldots, \underline{v}_s$.

Clearly a set of vectors is either linearly independent or linearly dependent, and cannot be both.

Let's take this slowly in $\mathbb{R}^2$.

Example 3.9

The vectors $\underline{v} = (1, 2)^T$ and $\underline{w} = (-2, 3)^T \in \mathbb{R}^2$ are linearly independent. Indeed, suppose that $\lambda\underline{v} + \mu\underline{w} = \underline{0}$. Then

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \mu \begin{pmatrix} -2 \\ 3 \end{pmatrix} = \begin{pmatrix} \lambda - 2\mu \\ 2\lambda + 3\mu \end{pmatrix}$$

Considering the components separately gives two simultaneous linear equations:

$$\begin{aligned} \lambda - 2\mu &= 0 \\ 2\lambda + 3\mu &= 0 \end{aligned}$$

Solving these (for example by subtracting twice the top one from the bottom one) shows that $\lambda = \mu = 0$. That's exactly what linear independence is asking for: the only linear combination of $\underline{v}$ and $\underline{w}$ that equals the zero vector is the trivial linear combination.

In contrast, the vectors $\underline{u}_1 = (2, -6)$ and $\underline{u}_2 = (-3, 9)$ are not linearly independent. You may see at once that $3\underline{u}_1 + 2\underline{u}_2 = \underline{0}$, which is a nontrivial linear combination giving the zero vector. Even if you don't spot this at once, we may apply the definition to find this: if we write

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} 2 \\ -6 \end{pmatrix} + \mu \begin{pmatrix} -3 \\ 9 \end{pmatrix} = \begin{pmatrix} 2\lambda - 3\mu \\ -6\lambda + 9\mu \end{pmatrix}$$

which you solve (either by considering components, since this is such a small example, or by row reduction, since it is such a formidable tool) to get $\mu = s$ and $\lambda = 3s/2$ for any $s \in \mathbb{R}$. In particular, you may choose a nonzero $s \in \mathbb{R}$ to find a linear dependence relation: for example $s = 2$ gives the relation $3\underline{u}_1 + 2\underline{u}_2 = \underline{0}$ that we spotted above.

The linearly dependent vectors $\underline{u}_1, \underline{u}_2$ in the example were collinear. (Recall from Definition 1.5 that two vectors are collinear if and only if one of them is a multiple of the other.) This is a general fact.

Lemma 3.10

Consider $\underline{v}_1, \underline{v}_2 \in \mathbb{R}^n$. Then $\underline{v}_1, \underline{v}_2$ are linearly dependent if and only if they are collinear.

This lemma is a criterion for two vectors being linearly *dependent*, not independent. Negating it gives: $\underline{v}_1, \underline{v}_2$ are linearly independent if and only if neither is a multiple of the other – which is perfectly true, but is so convoluted to pronounce that it's not much help!

Proof. We check this using the definition. First, suppose there are $\lambda_1, \lambda_2 \in \mathbb{R}$ for which

$$\lambda_1\underline{v}_1 + \lambda_2\underline{v}_2 = \underline{0}$$

If $\lambda_1 \neq 0$, then $\underline{v}_1 = -(\lambda_2/\lambda_1)\underline{v}_2$ is a multiple of $\underline{v}_2$, while if $\lambda_2 \neq 0$, then $\underline{v}_2 = -(\lambda_1/\lambda_2)\underline{v}_1$ is a multiple of $\underline{v}_1$. So far, that says: if $\underline{v}_1, \underline{v}_2$ are linearly dependent then one is a multiple of the other, that is they are collinear. The converse is quicker: if $\underline{v}_1 = \mu\underline{v}_2$ then $\underline{v}_1 + (-\mu)\underline{v}_2 = \underline{0}$ is a linear dependence relation, since the coefficient 1 of $\underline{v}_1$ is nonzero, so it doesn't matter what the value of μ is. Similarly if $\underline{v}_2$ is a multiple of $\underline{v}_1$. □

Remark

Definitions 3.7 of linear independent and 3.8 of linearly dependent apply perfectly well to any subset $S \subset \mathbb{R}^n$, even if, for example, S is infinite: the only point is that linearly combinations can only have finitely many nonzero coefficients.

From here on, we will happily use these definitions for any subset S, though in practice in almost all situations S is finite and the definitions as stated above are ideal.

Example

Let $S = \{k\underline{e}_1 \mid k \in \mathbb{R}\} \subset \mathbb{R}^n$. Then S is linearly dependent. There are many linear dependence relations, but we only need to exhibit one: for example $3 \times (2\underline{e}_1) + 2 \times (-3\underline{e}_1) = \underline{0}$.

Example

At the other extreme, Definition 3.7 applies even if we are dealing with only $s = 1$ vector. If $\underline{v}_1 \in \mathbb{R}^n$, then it is linearly independent if and only if it is not the zero vector: if $\lambda_1\underline{v}_1 = \underline{0}$, then either $\lambda_1 = 0$ or $\underline{v}_1 = \underline{0}$. (If you find that confusing, just ignore it – you're merely confused by its triviality.)[a]

[a]This is a suitable point to remind ourselves of our favourite joke (The Puffin Joke Book, 1974). What's the difference between a duck? One of its legs are both the same! hahahahaha. . . [Aside for multi-dimensional readers: jokes are always funniest when you have to explain them. Here, the joke is that it has the syntax of a joke but not the semantics – the subject matter of the 'joke' is a piece of misdirection. Now it's funny, right?]

Remark

It is natural to ask at this point, how big can a linearly independent subset of $\mathbb{R}^n$ be? Prima facie we could imagine an infinite set of linearly independent vectors in $\mathbb{R}^n$, but in fact the upper limit is n. This fact is essentially a small piece of what we know about the solutions of systems of linear equations, and perhaps therefore it feels intuitively correct to you: from the point of view of linear equations, you already know the following important result. The key to its proof is that we already know a spanning set, namely the standard basis, which has n elements.

Proposition 3.11

Suppose a subset $S \subset \mathbb{R}^n$ is linearly independent. Then in fact S is a finite set and $\#S \leq n$.

Proof. Suppose S has strictly more than n elements (it could even be infinite, for example). Choose $n+1$ distinct elements $\underline{v}_1, \dots, \underline{v}_{n+1} \in S$; these are certainly linearly independent.

Consider the vectors $\underline{e}_1, \dots, \underline{e}_n$. Since they span $\mathbb{R}^n$, there are scalars $a_{ij} \in \mathbb{R}$ so that

$$\underline{v}_j = \sum_{i=1}^{n} a_{ij}\underline{e}_i$$

for each $j = 1, \dots, n+1$. (Equivalently, if you prefer, $\underline{v}_j = (a_{1j}, a_{2j}, \dots, a_{nj})^T$.)

Assemble the scalars a_{ij} into an $n \times (n+1)$ matrix $A = (a_{ij})$: the vectors $\underline{v}_j$ are the columns of A. Since A has fewer rows than columns, Corollary 2.69 says there is a nonzero vector $\underline{k} = (k_1, \ldots, k_{n+1})$ such that $A\underline{k} = \underline{0}$. Since the $\underline{v}_j$ are the columns of A, this is exactly saying that

$$k_1\underline{v}_1 + \ldots + k_{n+1}\underline{v}_{n+1} = \underline{0}$$

Since $\underline{k} \neq \underline{0}$, says that $\underline{v}_1, \ldots, \underline{v}_{n+1}$ are linearly dependent, which is a contradiction. □

3.3 Bases of $\mathbb{R}^n$

Definition 3.12

A sequence of vectors $\underline{v}_1, \ldots, \underline{v}_s \in \mathbb{R}^n$ is a **basis of** $\mathbb{R}^n$ if and only if it is linearly independent and spans $\mathbb{R}^n$.

The first example is of course the standard basis $\underline{e}_1, \ldots, \underline{e}_n \in \mathbb{R}^n$.

Propositions 3.6 and 3.11 prove the following important result at once.

Theorem 3.13

Let $S \subset \mathbb{R}^n$ be a basis. Then S is a finite set and $\#S = n$.

Language 3.14

A tiny point: it will prove fantastically useful to us that any basis we consider is a collection of vectors in some given fixed order – that's why the definition referred to a *sequence* of vectors. Of course the definitions of linear independence and spanning did not rely on the order: they referred merely to a *set* of vectors. Let's not fall out over this. In this module, let's just agree that whenever we have a (finite) basis it comes in some fixed order, and whether we call it a sequence or a set is irrelevant.

The second important result about bases is a kind of existence and uniqueness statement.

Proposition 3.15

Let $\underline{f}_1, \ldots, \underline{f}_n$ be a basis of $\mathbb{R}^n$. Then for any $\underline{w} \in \mathbb{R}^n$, there are unique scalars $\mu_1, \ldots, \mu_n \in \mathbb{R}$ so that

$$\underline{w} = \mu_1\underline{f}_1 + \ldots + \mu_n\underline{f}_n$$

Of course we know the scalars μ_i in the proposition exist: since the vectors $\underline{f}_1, \ldots, \underline{f}_n$ span $\mathbb{R}^n$, any vector may be expressed as a linear combination of them. The point is that since $\underline{f}_1, \ldots, \underline{f}_n$ are also linearly independent, there is exactly one way to do this for each vector $\underline{w}$. Try it yourself – imagine you had two expressions and then follow your nose – before comparing with the following proof.

Proof. As remarked above, since $\underline{f}_1, \ldots, \underline{f}_n$ spans $\mathbb{R}^n$, there certainly is at least one expression

$$\underline{w} = \mu_1\underline{f}_1 + \ldots + \mu_n\underline{f}_n$$

for some scalars $\mu_i \in \mathbb{R}$. Suppose we had two such expressions: that is, we also have

$$\underline{w} = \nu_1 \underline{f}_1 + \ldots + \nu_n \underline{f}_n$$

for some (possibly different) scalars $\nu_i \in \mathbb{R}$. Then subtracting these two expressions gives

$$\underline{0} = (\mu_1 - \nu_1)\underline{f}_1 + \ldots + (\mu_n - \nu_n)\underline{f}_n$$

But since the vectors $\underline{f}_1, \ldots, \underline{f}_n$ are linearly independent, it follows by Definition 3.7 that

$$\mu_1 - \nu_1 = \ldots = \mu_n - \nu_n = 0$$

that is, that $\mu_i = \nu_i$ for each $i = 1, \ldots, n$, as required. □

Remark

You probably noticed that Definition 3.12 only defined the notion of basis for a finite set of vectors. Later we will allow infinite bases, and we will see lots of interesting examples, but in our current context of $\mathbb{R}^n$ we know from Proposition 3.11 that any linearly independent set is finite, so it is enough to define the notion for finite sets.

Example 3.16

Let $\underline{f}_1 = (1, 1)^T$ and $\underline{f}_2 = (2, 1)^T \in \mathbb{R}^2$. Then $\underline{f}_1, \underline{f}_2$ is a basis of $\mathbb{R}^2$: you can easily check that it is linearly independent (by checking that the equation

$$\lambda_1 \underline{f}_1 + \lambda_2 \underline{f}_2 = \underline{0}$$

only has the solution $\lambda_1 = \lambda_2 = 0$) and that it spans (by checking that the equation

$$\lambda_1 \underline{f}_1 + \lambda_2 \underline{f}_2 = \underline{b}$$

has a solution for any $\underline{b} \in \mathbb{R}^2$). The notion of basis is entirely bound up with the existence and uniqueness of solutions of systems of linear equations, as Propositon 3.15 says precisely.

Here's another way to show that $\underline{f}_1, \underline{f}_2$ spans. We already know that the standard basis $\underline{e}_1, \underline{e}_2$ spans, so if we can write those two vectors as a linear combination of $\underline{f}_1, \underline{f}_2$ then we can write any vector, as required for spanning. But that's easy to see:

$$\underline{e}_1 = \underline{f}_2 - \underline{f}_1 \quad \text{and} \quad \underline{e}_2 = 2\underline{f}_1 - \underline{f}_2 \tag{3.4}$$

We will see later in complete generality that if a vector space has a basis of size n, then any linearly independent set of size n is a basis, and any spanning set of size n is a basis. Armed with that, the equations (3.4) constitute a complete proof that $\underline{f}_1, \underline{f}_2$ is a basis of $\mathbb{R}^2$.

Example 3.17

Continuing with Example 3.16, the following curious question arises, which turns out to be important. Consider the vector $\underline{w} = (-3, 5)^T \in \mathbb{R}^2$. It is extremely useful to regard that coordinate expression of $\underline{w}$ as the linear relation

$$\underline{w} = -3\underline{e}_1 + 5\underline{e}_2$$

That is, the components of $\underline{w}$ are really the coefficients of $\underline{w}$ when we write it with respect to the basis $\underline{e}_1, \underline{e}_2$ of $\mathbb{R}^2$.

By Example 3.16, we know another basis of $\mathbb{R}^2$, namely $\underline{f}_1, \underline{f}_2$. Since this is a basis, we may write $\underline{w}$ as a linear combination of these. What is that? Well, you could solve the equations

$$\underline{w} = \mu_1 \underline{f}_1 + \mu_2 \underline{f}_2 \tag{3.5}$$

You would find that $\underline{w} = 13\underline{f}_1 - 8\underline{f}_2$ but there is a much quicker way.

Consider the matrix

$$Q = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix}$$

whose columns are the vectors $\underline{f}_1, \underline{f}_2$ (in that order). Then form the inverse matrix

$$Q^{-1} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix}$$

(and quickly multiply them together in your head to be sure it is the inverse!) and observe that

$$Q^{-1}\underline{w} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} -3 \\ 5 \end{pmatrix} = \begin{pmatrix} 13 \\ -8 \end{pmatrix}$$

which, ta-dah!, are the coeffiicents of the expression of $\underline{w}$ with respect to the basis $\underline{f}_1, \underline{f}_2$.

Perhaps that doesn't feel quicker, since you had to compute the inverse matrix Q^{-1}. But if you had a hundred vectors $\underline{w}_1, \ldots, \underline{w}_{100}$ and needed to do this calculation for them all, you still only have to compute Q^{-1} once and then simply multiply out $Q^{-1}\underline{w}_i$ for each of them, whereas if you prefer to solve the system (3.5) for each one, then in practice you are repeating the calculation over and over.

Remark

The calculation of Example 3.17 is no fluke: it always works in exactly this way, and is one of the most powerful ideas of the whole theory. You should think of it as changing coordinates, and it is important to regard it as a simple thing, even though we all find it fiddly at first.

For comparison, you know from elementary integration that many calculations seem impossible in one set of coordinates but work easily if you change to some set of coordinates. When integrating, you are brilliant at this: you write $x = f(\vartheta)$ and so $dx = (df/d\vartheta)d\vartheta$, and, if you chose the change of coordinates f wisely, you then proceed with the calculation.

The point here is that if we get very good at expressing the same point $\underline{w}$ with respect to different bases, then we can choose a basis to make a given problem simpler to solve. This idea is referred to as 'change of basis'. It may seem tricky the first time you see it, but it becomes natural with practice. You may even realise that you've been doing it all along in all sorts of contexts (including Chapter 2!), but for now let's get to grips with how it works in $\mathbb{R}^n$ and why.

3.4 Change of basis is simplicity itself

Let's state this a little bit formally. Suppose $\underline{e}_1, \ldots, \underline{e}_n$ is the standard basis and $\underline{f}_1, \ldots, \underline{f}_n$ is any basis of $\mathbb{R}^n$.

Any vector $\underline{w} \in \mathbb{R}^n$ has a unique expression with respect to either basis: that is, there are unique scalars λ_i and μ_i for which

$$\begin{aligned} \underline{w} &= \lambda_1 \underline{e}_1 + \ldots + \lambda_n \underline{e}_n \qquad \text{and} \\ \underline{w} &= \mu_1 \underline{f}_1 + \ldots + \mu_n \underline{f}_n \end{aligned}$$

Of course, the coefficients with respect to the basis $\{\underline{e}_i\}$ are the components of $\underline{w} = (\lambda_1, \ldots, \lambda_n)^T$ as a column vector in $\mathbb{R}^n$. The question is, if you know the scalars $\lambda_1, \ldots, \lambda_n$, can you find the scalars $\mu_1, \ldots, \mu_n$ with minimal work? Oh yes you can!

We have to do some preparatory work as a one-off investment: for each $\underline{f}_i$, find scalars a_{ij} for which

$$\underline{f}_i = a_{1i} \underline{e}_1 + \ldots + a_{ni} \underline{e}_n$$

But that's no work at all: the a_{ij} are the components of the column vector $\underline{f}_i \in \mathbb{R}^n$. (You have to be careful about the i, j indexing of the coefficients a_{ij}.)

Now write the matrix

$$Q = (a_{ij})$$

that is, the matrix whose columns are the vectors $\underline{f}_i$ expressed in the standard $\underline{e}_j$ coordinates. Compute Q^{-1}, and finally

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = Q^{-1} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \tag{3.6}$$

Let's not think about why this works for a moment, but try it out.

Example

Consider the basis (you should check that it is indeed a basis)

$$\underline{f}_1 = 0 \times \underline{e}_1 + 1 \times \underline{e}_2 + 2 \times \underline{e}_3 \quad = \quad \begin{pmatrix} 0 \\ 1 \\ 2 \end{pmatrix}$$

$$\underline{f}_2 = 1 \times \underline{e}_1 + 2 \times \underline{e}_2 + 3 \times \underline{e}_3 \quad = \quad \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$$

$$\underline{f}_3 = 2 \times \underline{e}_1 + 3 \times \underline{e}_2 + 0 \times \underline{e}_3 \quad = \quad \begin{pmatrix} 2 \\ 3 \\ 0 \end{pmatrix}$$

and some other vector

$$\underline{w} = 5 \times \underline{e}_1 - 7 \times \underline{e}_2 + 11 \times \underline{e}_3 \quad = \quad \begin{pmatrix} 5 \\ -7 \\ 11 \end{pmatrix}$$

with all of them expressed in coordinates with respect to the standard basis $\underline{e}_1, \underline{e}_2, \underline{e}_3$.

The task is to find $\mu_1, \mu_2, \mu_3 \in \mathbb{R}$ for which

$$\underline{w} = \mu_1 \underline{f}_1 + \mu_2 \underline{f}_2 + \mu_3 \underline{f}_3$$

Since $\underline{f}_1, \underline{f}_2, \underline{f}_3$ is a basis, we know from Proposition 3.15 that the μ_i exist and are unique. In banal terms, we may say that the task is precisely to solve the system of equations

$$\begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 0 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = \begin{pmatrix} 5 \\ -7 \\ 11 \end{pmatrix}$$

and that's true, but let's follow the method above.

We write the matrix

$$Q = \begin{pmatrix} 0 & 1 & 2 \\ 1 & 2 & 3 \\ 2 & 3 & 0 \end{pmatrix}$$

and compute its inverse

$$\left(\begin{array}{ccc|ccc} 0 & 1 & 2 & 1 & 0 & 0 \\ 1 & 2 & 3 & 0 & 1 & 0 \\ 2 & 3 & 0 & 0 & 0 & 1 \end{array}\right) \text{ which row reduces to } \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{9}{4} & \frac{3}{2} & -\frac{1}{4} \\ 0 & 1 & 0 & \frac{3}{2} & -1 & \frac{1}{2} \\ 0 & 0 & 1 & -\frac{1}{4} & \frac{1}{2} & -\frac{1}{4} \end{array}\right)$$

so that

$$Q^{-1} = \frac{1}{4}\begin{pmatrix} -9 & 6 & -1 \\ 6 & -4 & 2 \\ -1 & 2 & -1 \end{pmatrix}$$

and then

$$\begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} = Q^{-1}\begin{pmatrix} 5 \\ -7 \\ 11 \end{pmatrix} = \frac{1}{2}\begin{pmatrix} -49 \\ 40 \\ -15 \end{pmatrix}$$

You can check at once that, yes indeed,

$$\underline{w} = -\frac{49}{2}\underline{f}_1 + 20\underline{f}_2 - \frac{15}{2}\underline{f}_3$$

The reason it works is simple. Write the vectors $\underline{f}_i$ as linear combinations of the basis $\underline{e}_1, \ldots, \underline{e}_n$:

$$\begin{aligned} \underline{f}_1 &= a_{11}\underline{e}_1 + a_{21}\underline{e}_2 + \ldots + a_{n1}\underline{e}_n \\ \underline{f}_2 &= a_{12}\underline{e}_1 + a_{22}\underline{e}_2 + \ldots + a_{n2}\underline{e}_n \\ \vdots &\quad \vdots \\ \underline{f}_n &= a_{1n}\underline{e}_1 + a_{2n}\underline{e}_2 + \ldots + a_{nn}\underline{e}_n \end{aligned} \tag{3.7}$$

for scalars $a_{ij} \in \mathbb{R}$. Once more, notice the i, j subscripts on the coefficients a_{ij}: they ordered carefully so that the coefficients are naturally column vectors.

Language 3.18

When writing the equations (3.7) above, we say that we are **expressing $\underline{f}_i$ with respect to the basis $\underline{e}_1, \ldots, \underline{e}_n$**, or simply **expressing $\underline{f}_i$ in the basis $\underline{e}_1, \ldots, \underline{e}_n$**. We may also say, when writing the coefficients as a column vector, that

$$\underline{f}_i = \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}$$

is a representation of $\underline{f}_i$ **in coordinates** with respect to the basis $\underline{e}_1, \ldots, \underline{e}_n$.

With the (square) matrix $Q = (a_{ij})$, we may express this formally as

$$\begin{pmatrix} \underline{f}_1 \\ \vdots \\ \underline{f}_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \vdots & & & \vdots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{pmatrix} \begin{pmatrix} \underline{e}_1 \\ \vdots \\ \underline{e}_n \end{pmatrix} = Q^T \begin{pmatrix} \underline{e}_1 \\ \vdots \\ \underline{e}_n \end{pmatrix}$$

where you note that we had to transpose the matrix Q to match the coefficients of (3.7). (By 'formally' we just mean that we are using matrix multiplication as a convenient short-hand notation for the equations (3.7), and we are not trying to image a column vector whose entries are other column vectors. You may continue to write in equation form as (3.7) if you prefer.)

On the other hand, $\underline{f}_1, \dots, \underline{f}_n$ is a basis too, so we may write the vectors $\underline{e}_j$ as linear combinations of them:

$$\begin{aligned} \underline{e}_1 &= b_{11}\underline{f}_1 + b_{21}\underline{f}_2 + \dots + b_{n1}\underline{f}_n \\ \underline{e}_2 &= b_{12}\underline{f}_1 + b_{22}\underline{f}_2 + \dots + b_{n2}\underline{f}_n \\ \vdots \quad & \quad \vdots \\ \underline{e}_n &= b_{1n}\underline{f}_1 + b_{2n}\underline{f}_2 + \dots + b_{nn}\underline{f}_n \end{aligned} \tag{3.8}$$

for scalars $b_{ij} \in \mathbb{R}$. Assemble the coefficients as a matrix $P = (b_{ij})$, so we may write this as

$$\begin{pmatrix} \underline{e}_1 \\ \vdots \\ \underline{e}_n \end{pmatrix} = P^T \begin{pmatrix} \underline{f}_1 \\ \vdots \\ \underline{f}_n \end{pmatrix}$$

Of course, substituting the expressions (3.7) for $\underline{f}_i$ into (3.8) must result in equations $\underline{e}_i = \underline{e}_i$ for $i = 1, \dots, n$, since it amounts to expressing the $\underline{e}_i$ in terms of the basis $\underline{e}_1, \dots, \underline{e}_n$, and there is only one such expression by Proposition 3.15. In other words,

$$P^T Q^T = I_n$$

or after transposing both sides using $(P^T Q^T)^T = (Q^T)^T (P^T)^T = QP$

$$QP = I_n$$

That is, Q is invertible and $P = Q^{-1}$.

Now, since the ith column of Q is the vector $\underline{f}_i$ in (the usual) coordinates, for any vector $\underline{w} \in \mathbb{R}^n$, expressed with respect to each of the two bases as

$$\underline{w} = \lambda_1 \underline{e}_1 + \dots + \lambda_n \underline{e}_n = \mu_1 \underline{f}_1 + \dots + \mu_n \underline{f}_n$$

it follows at once that

$$Q \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \mu_1 \underline{f}_1 + \dots + \mu_n \underline{f}_n = \underline{w} = \lambda_1 \underline{e}_1 + \dots + \lambda_n \underline{e}_n = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

which, after multiplying by Q^{-1}, is exactly the claim in (3.6).

This showed how to find the coefficients of a vector $\underline{w} \in \mathbb{R}^n$ that you know as a column vector (in other words, expressed in the standard basis $\underline{e}_1 \dots, \underline{e}_n$) with respect to another basis $\underline{f}_1, \dots, \underline{f}_n$. It is almost as simple to translate between any two bases, not just from the standard basis.

Example

Consider two bases $\underline{f}_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \underline{f}_2 = \begin{pmatrix} 1 \\ -2 \end{pmatrix}$ and $\underline{g}_1 = \begin{pmatrix} -1 \\ 3 \end{pmatrix}, \underline{g}_2 = \begin{pmatrix} 2 \\ -5 \end{pmatrix}$ of $\mathbb{R}^2$. Suppose

$$\underline{w} = -3\underline{f}_1 + 7\underline{f}_2$$

What are the coefficients with respect to the basis $\underline{g}_1, \underline{g}_2$? That is, find μ_1, μ_2 so that

$$\underline{w} = \mu_1 \underline{g}_1 + \mu_2 \underline{g}_2$$

The column vector expressions of $\underline{f}_i$ and $\underline{g}_i$ are their coefficients with respect to the standard basis, so using these as the columns of two matrices gives

$$Q_f = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \quad \text{and} \quad Q_g = \begin{pmatrix} -1 & 2 \\ 3 & -5 \end{pmatrix}$$

which (using the respective Q^{-1}) translate the column vector $\underline{w}$ into its coefficients with respect to each basis respectively. Therefore

$$Q_g^{-1} Q_f \begin{pmatrix} -3 \\ 7 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} -3 \\ 7 \end{pmatrix} = \begin{pmatrix} 5 & 2 \\ 3 & 1 \end{pmatrix} \begin{pmatrix} 4 \\ -17 \end{pmatrix} = \begin{pmatrix} -14 \\ -5 \end{pmatrix}$$

first translates the coefficients of $\underline{w}$ with respect to the $\underline{f}_i$ into its coefficients with respect to the standard basis (that is, its normal representation as a column vector in $\mathbb{R}^n$) and then translates again to the coefficients with respect to the $\underline{g}_i$. The conclusion is that

$$\underline{w} = -14\underline{g}_1 - 5\underline{g}_2$$

Of course we can check this in the usual standard coordinates, since

$$\underline{w} = \begin{pmatrix} 4 \\ -17 \end{pmatrix} = -14 \begin{pmatrix} -1 \\ 3 \end{pmatrix} - 5 \begin{pmatrix} 2 \\ -5 \end{pmatrix}$$

3.5 What about column operations?

Up to this point, given an $m \times n$ matrix $A \in \mathrm{Mat}_{mn}$, we have used row operations to put A into reduced row echelon form. We regard this a simplifying A, and in the context of augmented matrices $(A \mid \underline{b})$ it makes it almost trivial to read off everything we want to know about the solutions of the corresponding system of linear equations. We implement row operations as premultiplying A by a sequence of carefully chosen $m \times m$ elementary matrices, S_{ij}, $M_i(\lambda)$ (for $\lambda \neq 0$) and $A_{ij}(\mu)$.

So what about column operations? Well, exactly the same ideas apply, but with a different goal. Thus we may switch two columns, multiply a column by a nonzero scalar, or add any multiple of one column to a different column. There are analogous elementary matrices that we denote respectively by S^{ij}, $M^i(\lambda)$ and $A^{ij}(\mu)$. These are simply the transposes of the corresponding 'row' elementary matrices, but note that now they are invertible $n \times n$ matrices, and that we postmultiply by them. (The use of superscripts is barely necessary, but helps to remind us that they are being understood as column operations.)

Example

Consider the matrix A which is in RREF:

$$A = \begin{pmatrix} 1 & -3 & 0 & 5 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

By swapping columns around, we can move all the pivot columns to the front (left) of the matrix so we see an identity matrix (of suitable size) at the front: simply swap columns 2 and 3:

$$AS^{23} = \begin{pmatrix} 1 & -3 & 0 & 5 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & -3 & 5 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now we may add (or subtract) suitable multiples of column 1 to cancel the -3 and 5 in columns 3 and 4, and also use column 2 to cancel the 7 in column 4:

$$AS^{23}A^{13}(3)A^{14}(-5)A^{24}(-7) =$$

$$\begin{pmatrix} 1 & 0 & -3 & 5 \\ 0 & 1 & 0 & 7 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 3 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & -5 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & -7 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Matrices in this form are said to be in **Smith normal form**, which we discuss below.

The whole theory works seamlessly: we may even define a reduced column echelon form. There is nothing to prove: for $A \in \mathrm{Mat}_{mn}$, we generate a reduced column echelon form

$$AF_1 \cdots F_\ell \quad \text{for elementary matrices } F_i \in \mathrm{Mat}_{nn}$$

where we choose (row) elementary matrices E_i so that

$$E_\ell \cdots E_1 A^T$$

is in RREF, and F_i is defined to be $F_i = E_i^T$, giving

$$AF_1 \cdots F_\ell = (E_\ell \cdots E_1 A^T)^T$$

One value of column operations is that they find a basis of the column span $\mathrm{Colspan}(A)$ of a matrix A; recall Definition 3.4.

Proposition 3.19

Let $A \in \mathrm{Mat}_{mn}$. Then the nonzero columns of the reduced column echelon form of A form a basis of the column span $\mathrm{Colspan}(A)$ of A.

Proof. Let $\underset{\sim}{c}_1, \dots, \underset{\sim}{c}_n \in \mathbb{R}^m$ be the columns of A, so that

$$\mathrm{Colspan}(A) = \langle \underset{\sim}{c}_1, \dots, \underset{\sim}{c}_n \rangle \subset \mathbb{R}^m$$

Let $\underset{\sim}{d}_1, \dots, \underset{\sim}{d}_\ell$ be the nonzero columns of the reduced column echelon form of A, where $\ell \leq n$. By construction

$$\underline{d}_i \in \mathrm{Colspan}(A)$$

for each $i = 1, \dots, \ell$ since they are linear combinations of the columns of A. In fact

$$\langle \underline{d}_1, \dots, \underline{d}_\ell \rangle = \langle \underline{c}_1, \dots, \underline{c}_n \rangle$$

since by reversing the column operations we recover the $\underline{c}_i$ as linear combinations of $\underline{d}_1, \dots, \underline{d}_\ell$ (together with some zero columns, but of course those do not contribute anything to the linear combination). Thus $\underline{d}_1, \dots, \underline{d}_\ell$ span $\mathrm{Colspan}(A)$.

Finally $\underline{d}_1, \dots, \underline{d}_\ell$ are linearly independent since each of them has a pivot entry 1 in some row where all the others have a zero entry. □

Using row and column operations in combination, we may transform any matrix into a particularly simple form.

Definition 3.20

A matrix $A \in \mathrm{Mat}_{mn}$ is in **Smith normal form** if and only if it is in the form

$$\begin{aligned} A &= \begin{pmatrix} I_r & 0_{r,n-r} \\ 0_{m-r,r} & 0_{m-r,n-r} \end{pmatrix} \\ &= \left(\begin{array}{ccccc|ccc} 1 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & \cdots & 0 \\ \hline 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{array}\right) \end{aligned}$$

The **rank** of this Smith normal form is defined to be the integer $r \geq 0$.

Theorem 3.21

Let $A \in \mathrm{Mat}_{mn}$. Then there are elementary matrices $E_i \in \mathrm{Mat}_{mm}$ and elementary matrices $F_i \in \mathrm{Mat}_{nn}$ so that

$$E_k \cdots E_1 A F_1 \cdots F_\ell$$

is in Smith normal form of some rank $r \geq 0$.

We give a rather quick and dirty proof, that really relies on you drawing a sketch of the matrix after the first row reduction so you can understand the care that the proof requires in the second row reduction it employs; if you just say "row reduce A to B, and then row reduce B^T and finally transpose back" then you need to explain why the second (transposed) reduction didn't destroy the first. It is fine, but it is much more delicate than it pretends.

Proof. Choose elementary matrices $E_i \in \mathrm{Mat}_{mm}$ so that the matrix $B = E_k \cdots E_1 A$ is in RREF. Note the positions of the pivot entries $a_{ij} = 1$, for specific i, j. Now choose elementary matrices $D_i \in \mathrm{Mat}_{nn}$ so that $D_\ell \cdots D_1 B^T$ is in RREF, being careful to use the (previous) pivot entries

$a_{ji} = 1$ (now in their transposed positions) as the leading 1s for those columns. (This happens automatically if you always choose the top nonzero entry of any nonzero column as the pivot, which in practice you probably do.) Then setting $F_i = D_i^T$ gives the result. □

Remark

At this stage you could imagine choosing different E_i and F_j so that the resulting Smith normal form has a different rank $\neq r$. In fact, that is not possible, whichever way you choose the E_i and F_j, but it is not particularly clear yet why. The rank $r \geq 0$ of the Smith normal form is a crucial invariant of the matrix A, and we discuss it later in the context of linear maps.

Chapter 4

Vector spaces

Finally we come to the general theory of vector spaces (over $\mathbb{R}$). The main difference you will notice is that we no longer underline elements v, w and so on, unless they happen to be column vectors in $\mathbb{R}^n$.

4.1 Formal definition and examples

Definition 4.1

A **vector space** V is an abelian group under an operation $+$ that also admits **scalar multiplication** by elements of $\mathbb{R}$: for any $v \in V$ and any $\lambda \in \mathbb{R}$, there is another element $\lambda v \in V$, and this multiplication satisfies the rules:

(i) $\lambda(v + w) = \lambda v + \lambda w$

(ii) $(\lambda + \mu)v = \lambda v + \mu v$

(iii) $\lambda(\mu v) = (\lambda\mu)v$

(iv) $1v = v$

for any $\lambda, \mu \in \mathbb{R}$ and any $v, w \in V$. We usually denote the additive identity by 0_V and for any $v \in V$ its additive inverse by $-v$.

Remark

You may prefer to spell out all the axioms, rather than use the abelian group to carry some of the load. In that case, you define a vector space to be a set V that has, for any $u, v, w \in V$,

(i) a binary operation (**addition**) $V \times V \to V$ denoted $(v, w) \mapsto v + w$ which satisfies

- (i) $v + w = w + v$ (addition is commutative)
- (ii) $u + (v + w) = (u + v) + w$ (addition is associative)
- (iii) there is an (additive) identity $0_V \in V$ so that $0_V + v = v$
- (iv) there is an (additive) inverse denoted $-v$ so that $v + (-v) = 0_V$, and

(ii) an operation (**scalar multiplication**) $\mathbb{R} \times V \to V$ denoted $(\lambda, v) \mapsto \lambda v$ which satisfies

- (i) $\lambda(v + w) = \lambda v + \lambda w$

(ii) $(\lambda + \mu)v = \lambda v + \mu v$

(iii) $\lambda(\mu v) = (\lambda\mu)v$

(iv) $1v = v$

for any $\lambda, \mu \in \mathbb{R}$ and any $v, w \in V$.

Example

Of course, $V = \mathbb{R}^n$ is an example of a vector space for any fixed $n \geq 0$. We checked the points that the definition requires in Lemma 1.3.

Example

We denote the ring of all polynomials (in a variable x and with real coefficients) by

$$\mathbb{R}[x] = \left\{a_0 + a_1 x + a_2 x^2 + \ldots + a_s x^s \;\middle|\; s \in \mathbb{N} \text{ and all } a_i \in \mathbb{R}\right\}$$

You are probably very familiar with polynomials (see the following remark, if not, but otherwise ignore it). You can add them together by collecting the coefficients of each monomial x^i together, and you can multiply them by scalars simply by multiplying every coefficient, and it is easy to check that all the axioms of a vector space hold.

Remark

You need to be a tiny bit careful about the definition of the ring of polynomials if you read it, though you will surely be using it correctly already and it is simpler to work with this than to worry about the formulation (you may compare with Bourbaki if you really care).

To state the obvious, for any $f \in \mathbb{R}[x]$, there is some integer $s \geq 0$ and $a_i \in \mathbb{R}$ so that $f = a_0 + a_1 x + a_2 x^2 + \ldots + a_s x^s$. We refer to the a_i as the **coefficients** of f, and it is convenient for any power x^j not written (for example when $j > s$) to treat a_j as $a_j = 0 \in \mathbb{R}$. We may write a polynomial briefly as $\sum a_i x^i$, where it is understood that the sum is taken over integers $i \geq 0$ and, crucially, that only finitely many of the a_i are nonzero: in some jargons one may say **almost all** $a_i = 0$. With that in mind, two polynomials $f = \sum a_i x^i$ and $g = \sum b_i x^i$ are **equal** if and only if all their coefficients are identical: $a_i = b_i$ for all $i \geq 0$. The **zero polynomial** is by definition the polynomial with all coefficients $a_i = 0$.

For example, $f = 2 + 3x - 5x^3 + 0x^7$ is a polynomial with given coefficients $a_0 = 2$, $a_1 = 3$, $a_3 = -5$ and $a_7 = 0$, and implicitly all other $a_j = 0$. You see at once that the integer $s \geq 0$ is a bit of a red herring, since as written f has $s = 7$, but of course $g = 2 + 3x - 5x^3$ is equal to f. To mollify this trivial irritation, we define the **degree** of a polynomial $f = \sum a_i x^i$ to be the largest i for which $a_i \neq 0$: it is denoted

$$\deg f = \begin{cases} -1 & \text{if } f \text{ is the zero polynomial} \\ \max\{i \geq 0 \mid a_i \neq 0\} & \text{otherwise} \end{cases}$$

where you notice the slight care taken to make a special convention for the zero polynomial.

We may add polynomials and multiply by scalars: if $f = \sum a_i x^i$ and $g = \sum b_i x^i$ then

$$f + g = \sum (a_i + b_i) x^i \quad \text{and} \quad \lambda f = \sum (\lambda a_i) x^i$$

These operations make $\mathbb{R}[x]$ into a vector space; one must check all the details of Definition 4.1.

You know a great deal more about $\mathbb{R}[x]$, and being a vector space is perhaps the least interesting thing about it. For example, we may multiply polynomials together in the usual way. (This makes $\mathbb{R}[x]$ into an $\mathbb{R}$-algebra; we don't discuss those in this module.) Also you naturally think of polynomials $f \in \mathbb{R}[x]$ as real functions $f\colon \mathbb{R} \to \mathbb{R}$ by evaluating the variable x at any number $b \in \mathbb{R}$: the map is $b \mapsto f(b)$. This point of view is extremely useful, even though prima facie it had nothing to do with the definition (but beware there are contexts where analogous thinking can mislead). With this in mind, we may ask about roots of polynomials (solutions of the equation $f(x) = 0$), local maxima and minima (perhaps employing derivatives), pictorial representations as graphs, areas under graphs (perhaps using integrals), etc., etc. The only one of these we will see again in this module is differentiation, even though the tools of Linear Algebra are absolutely central to them all.

Example

A small variation on the previous example is, for any fixed $d \in \mathbb{N}$, the vector space

$$\mathbb{R}[x]_{\leq d} = \{f \in \mathbb{R}[x] \mid \deg(f) \leq d\} = \left\{ a_0 + a_1 x + \ldots + a_d x^d \;\middle|\; a_1, \ldots, a_d \in \mathbb{R} \right\}$$

You can check that this is a vector space too under the same addition and scalar multiplication.

Example 4.2

The set of functions

$$V = \left\{ f\colon \mathbb{R} \to \mathbb{R} \;\middle|\; f \text{ is twice continuously differentiable and } \frac{d^2 f}{dx^2} + 9f = 0 \right\}$$

is a vector space: the zero function is certainly a solution of this differential equation, and by the rules of differentiation if $f, g \in V$ and $\lambda, \mu \in \mathbb{R}$ then $\lambda f + \mu g$ is also in V.

This vector space V is defined in a rather subtle way, and we were able to check that it is a vector space in abstract terms, applying the linearity of differentiation rather than any knowledge of its elements. But in fact you probably already have a clear idea of the elements of V from your knowledge of differential equations: they are precisely all the functions $f\colon \mathbb{R} \to \mathbb{R}$ of the form

$$f = \lambda \sin(3x) + \mu \cos(3x)$$

for any $\lambda, \mu \in \mathbb{R}$.

Remark

There are unknown infinities of vector spaces ready to be our friends. We have seen one tiny corner of the tip of the iceberg. Just below that, but still a long way from getting our feet wet, all of the following are perfectly respectable examples.

There are vector spaces of

(i) bounded sequences: abbreviating the infinite sequence $(a_1, a_2, \dots)$ by $(a_n)_{n\in\mathbb{N}}$

$$\ell^\infty = \{(a_n)_{n\in\mathbb{N}} \mid \text{there is some } N \in \mathbb{R} \text{ such that } |a_i| < N \text{ for all } i\}$$

For example $(1/n)_{n\in\mathbb{N}}$ lies in ℓ^1 but $(n^2)_{n\in\mathbb{N}}$ does not.

(ii) sequences that give absolutely convergent series:

$$\ell^1 = \left\{ (a_n)_{n\in\mathbb{N}} \;\middle|\; \sum_{i=1}^{\infty} |a_i| < \infty \right\}$$

For example $(1/n^2)_{n\in\mathbb{N}}$ lies in ℓ^∞ but $(1/n)_{n\in\mathbb{N}}$ does not.

and any number of variations such as square-summable sequences which have $\sum a_n^2 < \infty$, and so on. You can quickly check (using Analysis) that each of these is a vector space.

There are also vector spaces of

(iii) Continuous functions:

$$\mathcal{C}(\mathbb{R}) = \{f\colon \mathbb{R} \to \mathbb{R} \mid f \text{ is continuous}\}$$

For example $x \mapsto |x|$ lies in this space.

(iv) Smooth functions:

$$\mathcal{C}^\infty(\mathbb{R}) = \{f\colon \mathbb{R} \to \mathbb{R} \mid f \text{ may be differentiated as many times as you like}\}$$

For example $x \mapsto \sin(x)$ lies in this space.

(v) Formal power series:

$$\mathbb{R}[\![x]\!] = \left\{a_0 + a_1 x + a_2 x^2 + \dots \;\middle|\; a_i \in \mathbb{R} \text{ for all } i \geq 0\right\}.$$

For example, $1 + 2x + 4x^2 + \dots + 2^n x^n + \dots$ lies in this space; it has radius of convergence zero, and so you cannot evaluate x at any nonzero value to produce a number: it is not a function on any ε-neighbourhood of the origin, however small $\epsilon > 0$ you choose.

and any number of variations: power series with radius of convergence at least 1 (such as the Taylor expansion of $1/(1-x)$), or power series with radius of convergence ∞ (such as the Taylor expansion of $\exp(x)$), and so on and so on. You can quickly check (using Analysis) that each of these is a vector space, and that $\mathbb{R}[x]$ is a subset of all of them.

All of these vector spaces are different from our most familiar collection of examples, $\mathbb{R}^n$ for $n \in \mathbb{N}$. Vector spaces are ubiquitous in mathematics, and while $\mathbb{R}^n$ is vitally important, and certainly is the central example in this module, we must never think that column vectors of some fixed length n is the only example of a vector space.

Essentially you can forget all these examples for the rest of the module – I will only mention them occasionally to illustrate particular points, and then most likely only to say that these examples are too subtle for me to understand.

There are a whole bunch of things not listed in the axioms of a vector space that nevertheless hold completely generally; compare Lemma 1.3. Their proof is usual routine fooling around, so you should do it and ignore my feeble efforts. (Recall from group theory that the additive identity 0_V is unique: after all, if you had another, $0'_V$ say, then $0_V = 0_V + 0'_V = 0'_V$, so it was the same all along.)

Lemma 4.3

Let V be a vector space. For any $\lambda \in \mathbb{R}$ and $v \in V$ we have

(i) $\lambda 0_V = 0_V$ and $0v = 0_V$.

(ii) $(-1)v = -v$ and more generally $(-\lambda)v = -(\lambda v) = \lambda(-v)$.

If you do accidentally find yourself reading the following drivel, please turn it into a useful exercise by indicating at each equals sign which axiom(s) is being used, and why it's right.

Proof. (i) The second claim first: $0v + 0v = (0 + 0)v = 0v$, so adding $-(0v)$ to both sides shows that $0v = 0_V$. So we only need to check the first claim for $\lambda \neq 0$: for any $v \in V$, $v + \lambda 0_V = \lambda(\frac{1}{\lambda}v + 0_V) = \lambda(\frac{1}{\lambda}v) = v$, so adding $-v$ to both sides shows that $\lambda 0_V = 0_V$.

(ii) The additive inverse $-v$ is characterised by what happens when it is added to v, so we compute $v + (-1)v = 1v + (-1)v = (1 + (-1))v = 0v = 0_V$ so $-v = (-1)v$. The general case is the same: $\lambda v + (-\lambda)v = (\lambda + (-\lambda))v = 0v = 0_V$, so $-(\lambda v) = (-\lambda)v$. And similarly: $\lambda v + \lambda(-v) = \lambda(v - v) = 0_V$, so $-(\lambda v) = \lambda(-v)$. Please make it stop.

□

4.2 Spans and subspaces

We define the span of a subset of a vector space V exactly as for $\mathbb{R}^n$.

Definition 4.4

Let $S \subset V$ be a non-empty subset of a vector space V. The **span of** S, denoted $\langle S \rangle$, is

$$\langle S \rangle = \{\lambda_1 v_1 + \ldots + \lambda_s v_s \mid v_1, \ldots, v_s \in S \text{ and } \lambda_1, \ldots, \lambda_s \in \mathbb{R}\}$$

By convention, we define the span of the empty subset $S = \varnothing \subset V$ to be $\langle \varnothing \rangle = \{0_V\}$.

We define subspace of a vector space; compare with subspaces of $\mathbb{R}^n$ in Definition 3.1 – it is identical.

Definition 4.5

Let V be a vector space. A **subspace** of V is a nonempty $W \subset V$ with the property that for any $v, w \in W$ and any $\lambda \in \mathbb{R}$, we also have $v + w \in W$ and $\lambda v \in W$.

There are two particular subspaces that we refer to as **trivial** subspaces: $\{0_V\} \subset V$ and $V \subset V$. Thus to say a subspace $W \subset V$ is **nontrivial** is to say $W \neq \{0_V\}$ and $W \neq V$.

Example

For any $d \in \mathbb{N}$, $W = \mathbb{R}[x]_{\leq d}$ is a subspace of $V = \mathbb{R}[x]$.

One detail to be sure about is that the two operations in W (addition and scalar multiplication) are exactly the same as those in V. In Definition 4.5 the plus sign $+$ was addition in V, but we probably came into this example thinking that W had an addition defined all of its own.

Remark

To prove some $W \subset V$ is a subspace of V, you just check the conditions of Definition 4.5. It is common to amalgamate them into a single equivalent condition: $W \subset V$ is a subspace if and only if

$$\lambda v + \mu w \in W$$

for every $v, w \in W$ and all $\lambda, \mu \in \mathbb{R}$.

Proposition 4.6

Let V be a vector space.

(i) If $S \subset V$ is any subset, then its span $\langle S \rangle$ is a subspace.

(ii) If $W_1, W_2 \subset V$ are subspaces, then so is $W_1 \cap W_2$.

Another routine proof: best to do yourself.

Proof. (i) If $v = \sum \alpha_i v_i$ and $w = \sum \beta_j w_j$ for $v_i, w_j \in S$ and $\alpha_i, \beta_j \in \mathbb{R}$, then for any $\lambda, \mu \in \mathbb{R}$

$$\lambda v + \mu w = \sum (\lambda \alpha_i) v_i + \sum (\mu \beta_j) w_j$$

is a (finite) linear combination of elements of S, and so lies in $\langle S \rangle$

(ii) Suppose $v, w \in W_1 \cap W_2$. We must prove that for any $\lambda, \mu \in \mathbb{R}$ also $\lambda v + \mu w \in W_1 \cap W_2$. But this is instant: $v, w \in W_1$ so $\lambda v + \mu w \in W_1$ since W_1 is a subspace, and similarly for W_2. □

Definition 4.7

If $W_1, W_2 \subset V$ are subspaces of a vector space V, then we define their **sum** to be exactly the same as their combined span:

$$W_1 + W_2 = \langle W_1 \cup W_2 \rangle$$

That is, for any $v \in V$ we have:

$$v \in W_1 + W_2 \iff \text{there are } w_1 \in W_1 \text{ and } w_2 \in W_2 \text{ so that } v = w_1 + w_2.$$

Example

Let $V = \mathbb{R}^3$ and consider

$$W_1 = \left\{ \begin{pmatrix} a \\ b \\ 0 \end{pmatrix} \middle| \; a, b \in \mathbb{R} \right\} \text{ and } W_2 = \left\{ \begin{pmatrix} 0 \\ c \\ d \end{pmatrix} \middle| \; c, d \in \mathbb{R} \right\}$$

Then

$$W_1 \cap W_2 = \left\{ \begin{pmatrix} 0 \\ b \\ 0 \end{pmatrix} \middle| \; b \in \mathbb{R} \right\} \text{ and } W_1 + W_2 = \mathbb{R}^3$$

Note that $W_1 + W_2 \neq W_1 \cup W_2$ in this example. In fact, you can easily prove (so do!) that

$$W_1 + W_2 = W_1 \cup W_2 \iff \text{either } W_1 \subset W_2 \text{ or } W_2 \subset W_1.$$

The next definition is the key one that identifies the main subject area of this module. It is very simple, but it is worth absorbing why it works in the context of the rest of the results of this section.

Definition 4.8

A vector space V is **finite dimensional** if and only if there is a finite subset $\{v_1, \ldots, v_s\} \subset V$ that spans V: that is $\langle v_1, \ldots, v_s \rangle = V$.

Example

We know lots of examples of finite-dimensional vector spaces:

(i) $\mathbb{R}^n$ for any $n \in \mathbb{N}$. The standard basis $\underline{e}_1, \ldots, \underline{e}_n$ provides a finite spanning set.

(ii) $\mathbb{R}[x]_{\leq d}$ for any $d \in \mathbb{N}$. The monomials $1, x, x^2, \ldots, x^d$ provide a finite spanning set.

(iii) the space of solutions of the differential equation in Example 4.2. The two functions $\sin(3x)$ and $\cos(3x)$ provide a finite spanning set.

By contrast, the ring of polynomials $\mathbb{R}[x]$ is not a finite-dimensional vector space. That needs proof. Suppose $f_1, \ldots, f_s \in \mathbb{R}[x]$ is a spanning set. Denote the degrees of these polynomials by $d_i = \deg f_i$. Since the set $\{d_i \mid i = 1, \ldots, d_s\}$ is a finite set of integers, it has a largest element: without loss of generality $d_s \geq d_i$ for all $i = 1, \ldots, s$. But then x^{d_s+1} cannot be written as a linear combination of $f_1, \ldots, f_s$, and so they cannot have formed a spanning set.

Remark

The last example $V = \mathbb{R}[x]$ is a salutary: although it is not finite dimensional, it is a perfectly lovely space, and we work in it without concern. The point to take from this is that, while our module will focus on finite-dimensional vector spaces and will prove theorems in that context, we should expect to work with infinite-(i.e. not finite-)dimensional vector spaces too.

(If you own a more philosophically constructivist view, you might object that nobody ever truly works with the whole of $\mathbb{R}[x]$, but always with some approximation such as $\mathbb{R}[x]_{<d}$. Fair enough, but in that case you're probably even more upset with $\mathbb{R}$, so let's just roll with it for now.)

4.3 Bases of vector spaces

We define linear (in)dependence for a set of elements of a vector space just as we did in $\mathbb{R}^n$; compare Definitions 3.7 and 3.8.

Definition 4.9

A subset $\mathcal{L} \subset V$ of a vector space V is **linearly independent** if and only if whenever

$$\lambda_1 v_1 + \ldots + \lambda_s v_s = 0_V \text{ for } v_1, \ldots, v_s \in \mathcal{L} \text{ and } \lambda_1, \ldots, \lambda_s \in \mathbb{R}$$

we necessarily have that

$$\lambda_1 = \ldots = \lambda_s = 0$$

A subset $\mathcal{L} \subset V$ is **linearly dependent** if and only if it is not linearly independent. That is, $\mathcal{L}$ is linearly dependent if and only if there is a **linear dependence relation** among (finitely many of) its elements, namely a relation of the form (for some $s \geq 1$)

$$\lambda_1 v_1 + \ldots + \lambda_s v_s = 0_V \tag{4.1}$$

for $v_1, \ldots, v_s \in \mathcal{L}$ and $\lambda_1, \ldots, \lambda_s \in \mathbb{R}$ which are not all zero.

Linearly independent subsets of finite-dimensional vector spaces are necessarily finite; compare the statement and proof of the following proposition with Proposition 3.11.

Proposition 4.10

Suppose V is a finite-dimensional vector space and that $\mathcal{L} \subset V$ is linearly independent. Then $\mathcal{L}$ is a finite set.

Proof. Since V is finite dimensional, there is a spanning set $w_1, \ldots, w_n \in V$. Suppose there are $n+1$ distinct elements $v_1, \ldots, v_{n+1} \in \mathcal{L}$. Since $w_1, \ldots, w_n$ span, there are $a_{ij} \in \mathbb{R}$ so that

$$v_j = \sum_{i=1}^{n} a_{ij} w_i \quad \text{for each } j = 1, \ldots, n+1.$$

Consider the $n \times n+1$ matrix $A = (a_{ij})$. Since A has fewer rows than columns, by Corollary 2.69 there is some $\underline{k} = (k_1, \ldots, k_{n+1})^T \neq \underline{0}$ such that $A\underline{k} = \underline{0}$; note that the ith row of this equation is

$$a_{i1}k_1 + \ldots + a_{i,n+1}k_{n+1} = 0 \tag{4.2}$$

Thus

$$\begin{aligned} k_1 v_1 + \ldots + k_{n+1} v_{n+1} &= k_1 \sum_{i=1}^{n} a_{i1} w_i + \ldots + k_{n+1} \sum_{i=1}^{n} a_{i,n+1} w_i \\ &= \sum_{i=1}^{n} (k_1 a_{i1} + \ldots + k_{n+1} a_{i,n+1}) \, w_i \\ &= 0_V \qquad \text{by (4.2)} \end{aligned}$$

This is a linear dependence relation among $v_1, \ldots, v_{n+1}$, since $\underline{k} \neq \underline{0}$, and so $\mathcal{L}$ is not linearly independent, which is a contradiction. $\square$

The next very simple lemma matches our intuition and proves useful several times.

Lemma 4.11

Let V be a vector space with $\mathcal{B} \subset V$ a linearly independent subset and $v \in V$ an element. Then exactly one of the following two cases occurs (and not both):

(i) $v \in \langle \mathcal{B} \rangle$, or

(ii) $\mathcal{B} \cup \{v\}$ is linearly independent.

The proof hinges on the triviality that if you have a linear expression involving v and elements of $\mathcal{B}$, then the coefficient of v is either zero or it isn't. It's a great first exercise, before reading on.

Proof. If $v \in \langle \mathcal{B} \rangle$, then $v = \lambda_1 w_1 + \ldots + \lambda_s w_s$ for some $w_1, \ldots, w_s \in \mathcal{B}$ and $\lambda_1, \ldots, \lambda_s \in \mathbb{R}$. This is a nontrivial linear dependence relation among $\mathcal{B} \cup \{v\}$, since the coefficient of v is not zero, so $\mathcal{B} \cup \{v\}$ is not linearly independent.

If $v \notin \langle \mathcal{B} \rangle$, then we must prove that $\mathcal{B} \cup \{v\}$ is linearly independent. Consider a linear combination

$$\lambda v + \mu_1 w_1 + \ldots + \mu_s w_s = 0_V \tag{4.3}$$

for some $w_1, \ldots, w_s \in \mathcal{B}$ and $\lambda, \mu_1, \ldots, \mu_s \in \mathbb{R}$. If $\lambda \neq 0$, then rearranging (4.3) gives

$$v_j = -\left(\frac{\mu_1}{\lambda}\right) w_1 - \ldots - \left(\frac{\mu_s}{\lambda}\right) w_s \in \langle \mathcal{B} \rangle$$

But $v_j \notin \langle \mathcal{B} \rangle$, so we must have $\lambda = 0$. Therefore (4.3) reads $\sum \mu_i w_i = 0_V$. Since $\mathcal{B}$ is linearly independent, it follows that all $\mu_i = 0$, and so all coefficients of (4.3) are zero as required. $\square$

Corollary 4.12

Let V be a vector space. If V is not finite dimensional, then V contains an infinite set $v_1, v_2, \ldots$ of linearly independent elements.

Note that there is no claim that the v_i span V.

Proof. Certainly $V \neq \{0_V\}$ so choose any nonzero $v_1 \in V$ and set $W = \langle v_1 \rangle$. Since V is not finite dimensional, $W \neq V$, so we may choose some element $v_2 \notin W$. The set $\{v_1, v_2\}$ is linearly independent by Lemma 4.11.

We proceed inductively: after m steps, $v_1, \ldots, v_m$ is a linearly independent set. Since V is not finite dimensional, $W = \langle v_1, \ldots, v_m \rangle \neq V$, so we may choose some element $v_{m+1} \notin W$. The set $\{v_1, \ldots, v_{m+1}\}$ is linearly independent by Lemma 4.11.

Thus we may continue the sequence $v_1, v_2, \ldots$ indefinitely so that any finite portion $v_1, \ldots, v_m$ of the sequence is linearly independent. But that means that the whole sequence is linearly independent, since any dependence relation would only involve finitely many of the v_i. $\square$

Example

For example $V = \mathbb{R}[x]$ is a vector space that is not finite dimensional, and the set of monomials $1, x, x^2, \ldots, x^n, \ldots$ is linearly independent: the zero polynomial is the additive identity 0_V, and by definition it is the polynomial whose coefficients are all zero.

The definition of basis of a vector space also mirrors that we used for $\mathbb{R}^n$; compare Definition 3.12.

Definition 4.13

Let V be a vector space. A subset $\mathcal{B} \subset V$ is a **basis of** V if and only if $\mathcal{B}$ is linearly independent and spans V.

We immediately state an equivalent criterion for being a basis, just as we did in the case of $\mathbb{R}^n$ in Proposition 3.15.

Proposition 4.14

A subset $\mathcal{B} \subset V$ is a basis of V if and only if for any $w \in V$, there is an expression

$$w = \mu_1 v_1 + \ldots + \mu_s v_s$$

with distinct elements $v_1, \ldots, v_s \in \mathcal{B}$ and scalars $\mu_1, \ldots, \mu_s \in \mathbb{R}$, and furthermore the scalars $\mu_1, \ldots, \mu_s$ in this expression are uniquely determined.

Compare with the statement and proof of Proposition 3.15: this version is a little more precise (it is 'if and only if') but the proof is essentially identical. The faintly bizarre wording is to avoid saying that the choice if $v_1, \ldots, v_s$ is unique: of course there may be some other $v_{s+1} \in \mathcal{B}$, and then one could write

$$w = \mu_1 v_1 + \ldots + \mu_s v_s + \mu_{s+1} v_{s+1} \quad \text{with } \mu_{s+1} = 0$$

and the wording is to stop that being regarded as a different way of expressing w as a linear combination of elements of $\mathcal{B}$. (This is ridiculous! Please ignore and correctly treat the expression as unique.)

Proof. Suppose $\mathcal{B}$ is a basis. Then for $w \in V$ there certainly is such an expression, since $\mathcal{B}$ spans V. If there were two such expressions, then we may assume they involve the same finite collection $v_1, \ldots, v_s \in \mathcal{B}$ (by setting $\mu_i = 0$ or $\lambda_i = 0$ as required), so that they are

$$\begin{aligned} w &= \mu_1 v_1 + \ldots + \mu_s v_s \quad \text{and} \\ w &= \lambda_1 v_1 + \ldots + \lambda_s v_s \end{aligned}$$

for $\lambda_i, \mu_i \in \mathbb{R}$. Subtracting these two equations shows that

$$(\mu_1 - \lambda_1) v_1 + \ldots (\mu_s - \lambda_s) v_s = 0_V$$

so each $\mu_i - \lambda_i = 0$ by linear independence of $\mathcal{B}$, and so indeed the coefficients μ_i are uniquely determined.

Conversely, the expression for w shows at once that $\mathcal{B}$ spans V, while if

$$\lambda_1 v_1 + \ldots + \lambda_s v_s = 0_V$$

for $\lambda_i \in \mathbb{R}$ and $v_i \in \mathcal{B}$, then each $\lambda_i = 0$, since the trivial linear combination is one expression for 0_V, and so by uniqueness must be the only one. □

Example

It follows immediately that $\mathcal{B} = \{x^i \mid i \geq 0\} = \{1, x, x^2, x^3, \dots\}$ is a basis for $\mathbb{R}[x]$. The uniqueness is simply the definition of equality for polynomials: two are equal if and only if all their coefficients are equal.

Proposition 4.15 (Sifting Lemma)

Let V be a (finite-dimensional) vector space, and suppose $V \neq \{O_V\}$. Suppose $\mathcal{S} = \{v_1, \dots, v_s\}$ is a spanning set of V. Then there is a subset $\mathcal{B} \subset \mathcal{S}$ that is a basis of V.

Moreover, if $\mathcal{L} \subset \mathcal{S}$ is any linearly independent set, then we may choose $\mathcal{B}$ to contain $\mathcal{L}$.

We give two proofs. The first is stylish but doesn't help directly with finding a basis. We prove the final statement, since this includes the first claim by setting $\mathcal{L} = \varnothing$. Note that 'maximal' here means maximal with respect to inclusion: there is no claim that there is a unique 'biggest' subset.

Proof. Let $\mathcal{B} \subset S$ be a maximal linearly independent subset that contains $\mathcal{L}$. It certainly exists, since $\mathcal{L}$ is linearly independent. The claim is that $\mathcal{B}$ is a basis.

We need only prove that $\mathcal{B}$ spans, as it is linearly independent by specification. Suppose not. Then there must be some v_j that is not in the span $\langle \mathcal{B} \rangle$ of $\mathcal{B}$. The set $\mathcal{B} \cup \{v_j\}$ is linearly independent by Lemma 4.11, but it strictly contains $\mathcal{B}$, which contradicts the maximality of $\mathcal{B}$. □

The second proof is by **sifting**; considers each of the elements of S in turn, and 'sift out' those that are not needed for a basis. It is this proof that gives Proposition 4.15 its name.

Proof. Without loss of generality $\mathcal{L} = \{v_1, \dots, v_r\}$ for some $r \geq 0$. We consider the elements $v_{r+1}, \dots, v_s$ in turn. Set $\mathcal{B} = \mathcal{L}$; we will adjust it as we go.

Suppose we are considering v_j for $r+1 \leq j \leq s$. If v_j lies in the span of $\mathcal{B}$, then discard it. On the other hand, if $v_j \notin \langle \mathcal{B} \rangle$ then replace $\mathcal{B}$ by $\mathcal{B} \cup \{v_j\}$, which is linearly independent by Lemma 4.11. If $j < s$, continue with this new $\mathcal{B}$ and consider v_{j+1}; if $j = s$ then stop. Clearly this process stops after finitely many steps, once we have considered v_s.

The resulting $\mathcal{B}$ is linearly independent by construction, and it spans trivially: each v_j that we discarded was a linear combination of other $v_i \in \mathcal{B}$ (in fact, with $i < j$), so in any linear combination of the elements of S, we may substitute for v_j by an expression involving only elements of $\mathcal{B}$. □

So every finite-dimensional vector space has a basis. Proposition 4.15 implies even more.

Corollary 4.16 ("You can extend a linearly independent set to a basis" Lemma)

Let V be a finite-dimensional vector space and $v_1, \dots, v_s \in V$ a linearly independent subset. Then there exist $v_{s+1}, \dots, v_n \in V$ so that $v_1, \dots, v_n$ is a basis of V.

Proof. Since V is finite dimensional, there is a subset $\mathcal{M} = \{w_1, \ldots, w_r\}$ that spans V. Apply Proposition 4.15 with $\mathcal{L} = \{v_1, \ldots, v_s\}$ and $S = \mathcal{L} \cup \mathcal{M}$. ☐

Example

Suppose $\underline{v}_1 = (1, -2, 4)^T$ and $\underline{v}_2 = (0, 1, -2) \in \mathbb{R}^3$. It is easy to check that $\mathcal{L} = \{\underline{v}_1, \underline{v}_2\}$ is linearly independent, but it does not span: for example, $\underline{e}_2 \notin \langle \underline{v}_1, \underline{v}_2 \rangle$: if $\underline{e}_2 = \alpha\underline{v}_1 + \beta\underline{v}_2$, then we must have $\alpha = 0$ to get the first component right, and then $\beta = 0$ for the third component, but that does not give $\underline{e}_2$. If we set

$$S = \{\underline{v}_1, \underline{v}_2, \underline{e}_1, \underline{e}_2, \underline{e}_3\}$$

then Proposition 4.15 guarantees some subset of S is a basis that contains $\mathcal{L}$. For example, follow the second proof and sift S. We start with $\mathcal{B} = \mathcal{L}$. Note that $\underline{e}_1 = \underline{v}_1 + 2\underline{v}_2$, so we discard it. We already observed that $\underline{e}_2 \notin \langle \mathcal{B} \rangle$, so we include $\underline{e}_2$ and consider $\mathcal{B} = \{\underline{v}_1, \underline{v}_2, \underline{e}_2\}$. Then $\underline{e}_3 = -\frac{1}{2}\underline{v}_2 + \frac{1}{2}\underline{e}_2 \in \langle \mathcal{B} \rangle$, and so we discard $\underline{e}_3$ and the basis is $\underline{v}_1, \underline{v}_2, \underline{e}_2$.

4.4 Dimension theory

The main result is the following.

Theorem 4.17

If V is a finite-dimensional vector space, then any two bases of V are finite and have the same number of elements.

We give a famous proof: one by one, swap an element of one basis for an element of the other, and think about when this process can stop. The lemma is stated in a slightly complicated way, that essentially includes its proof. That is unusual, but the precise details are needed later.

Lemma 4.18 (Exchange Lemma)

Let V be a vector space and $\mathcal{B} = \{v_1, \ldots, v_s\}$ be a basis of V. Suppose $w \in V$ with $w \neq 0_V$. Since $\mathcal{B}$ is a basis, there are (unique) scalars $\lambda_i \in \mathbb{R}$ so that

$$w = \lambda_1 v_1 + \ldots + \lambda_s v_s$$

If $\lambda_j \neq 0$, then

$$\{v_1, \ldots, v_{j-1}, v_{j+1}, \ldots, v_s\} \cup \{w\}$$

is also a basis of V, where the element v_j has been removed from the first factor of the union (though of course $w = v_j$ is perfectly possible in the second, in which case the union is $\mathcal{B}$).

In other words, if v_j appears in a nontrivial way in the expression for w, then you can remove v_j from the basis and replace it by w, and the result is still a basis. Let's consider an example before the proof.

Example

Let $\mathcal{B} = \{\underline{e}_1, \underline{e}_2, \underline{e}_3\}$ be the standard basis of $\mathbb{R}^3$ and consider

$$\underline{w} = (2, -3, 0)^T = 2\underline{e}_1 + (-3)\underline{e}_2$$

The Exchange Lemma says that $\mathcal{B}_1 = \{\underline{w}, \underline{e}_2, \underline{e}_3\}$ and $\mathcal{B}_2 = \{\underline{e}_1, \underline{w}, \underline{e}_3\}$ are both bases of $\mathbb{R}^3$. Consider the first of these, $\mathcal{B}_1$.

Clearly $\mathcal{B}_1$ spans: any element of $\underline{v} \in \mathbb{R}^3$ is of the form $\underline{v} = \sum \mu_i \underline{e}_i$, and so substituting for $\underline{e}_1$ using the relation $\underline{e}_1 = \frac{1}{2}\underline{w} + \frac{3}{2}\underline{e}_2$, we can rewrite this as

$$\underline{v} = \mu_1(\tfrac{1}{2}\underline{w} + \tfrac{3}{2}\underline{e}_2) + \mu_2\underline{e}_2 + \mu_3\underline{e}_3$$

which is a linear combination of $\mathcal{B}_1$ (we could collect the terms together if we wished, but there is no need).

Equally clearly $\mathcal{B}_1$ is linearly independent: if $\lambda\underline{w} + \lambda_2\underline{e}_2 + \lambda_3\underline{e}_3 = \underline{0}$, we must show that $\lambda = \lambda_2 = \lambda_3 = 0$. Substituting for $\underline{w}$ gives

$$\underline{0} = \lambda(2\underline{e}_1 + (-3)\underline{e}_2) + \lambda_2\underline{e}_2 + \lambda_3\underline{e}_3 = 2\lambda\underline{e}_1 + (-3\lambda + \lambda_2)\underline{e}_2 + \lambda_3\underline{e}_3$$

But the original basis $\mathcal{B}$ is linearly independent, so we know that

$$2\lambda = -3\lambda + \lambda_2 = \lambda_3 = 0$$

which implies what we want.

We may prove the basis $\mathcal{B}_2$ in a similar way. The Lemma makes no claim about $\{\underline{e}_1, \underline{e}_2, \underline{w}\}$, where we have substituted $\underline{w}$ for $\underline{e}_3$, but clearly this is not a basis: every vector has zero as its third component, so we can never express $\underline{e}_3$ as a linear combination.

The proof is nothing more than this example in general notation.

Proof. Suppose without loss of generality that $j = 1$, so that

$$w = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_s v_s \quad \text{with } \lambda_1 \neq 0. \tag{4.4}$$

Write $\mathcal{B}' = \{w, v_2, \ldots, v_s\}$. We must prove that $\mathcal{B}'$ is a basis.

Note that w is not in the span of $v_2, \ldots, v_s$, since the coefficients λ_i are unique by Proposition 4.14. (To spell that out: any expression $w = 0v_1 + \mu_2 v_2 + \ldots + \mu_s v_s$ clearly has a different coefficient of v_1 than the one in (4.4), but that is impossible.) Therefore $\mathcal{B}'$ is linearly independent by Lemma 4.11.

And $\mathcal{B}'$ spans: any linear combination of $\mathcal{B}$ is equal to a linear combination of $\mathcal{B}'$ by substituting for v_1 after rearranging (4.4). □

It remains to prove Theorem 4.17. It is just a form of book-keeping now.

Proof. Since V is finite dimensional, by Proposition 4.15 or Corollary 4.16, it certainly has at least one basis that is finite. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be such a basis of smallest size and let $\mathcal{B}' \subset V$ be any other basis; in particular, $\mathcal{B}'$ has at least n elements.

Pick any element of $\mathcal{B}'$; call it w_1. By the Exchange Lemma 4.18, there is some $v_j \in \mathcal{B}$ that we may remove and replace by w_1 so that the collection remains a basis. Without loss of generality (by renumbering the v_i if necessary) $j = 1$ and

$$\mathcal{B}_1 = \{w_1, v_2, \ldots, v_n\}$$

is a basis of V.

Next pick any element of $\mathcal{B}' \setminus \{w_1\}$; call it w_2. By the Exchange Lemma there is some $v_j \in \mathcal{B}_1$ that we may remove and replace by w_2. (Note that we may insist, as we did, that the element we remove from $\mathcal{B}_1$ is not w_1: w_1, w_2 are linearly independent so when we write $w_2 = \lambda_1 w_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$, it could not be that $\lambda_k = 0$ for all $k \geq 2$. Even though λ_1 may be nonzero, we have deilberately chosen not to replace w_1, and we have just observed that there is indeed some other vector that we may replace instead.) Without loss of generality $j = 2$ and

$$\mathcal{B}_2 = \{w_1, w_2, v_3, \ldots, v_n\}$$

is a basis of V.

Proceeding inductively, after n such steps we reach the point where we have constructed a basis

$$\mathcal{B}_n = \{w_1, w_2, \ldots, w_n\}$$

Now if $\mathcal{B}'$ has strictly more than n elements, we may continue and pick an element $w_{n+1} \in \mathcal{B}' \setminus \mathcal{B}_n$. But since $\mathcal{B}_n$ is a basis, there are scalars $\mu_i \in \mathbb{R}$ so that

$$w_{n+1} = \mu_1 w_1 + \ldots + \mu_n w_n$$

But this is a contradiction: that equation, $w_{n+1} - \sum \mu_i w_i = 0_V$ is a nontrivial dependence relation among some elements of $\mathcal{B}'$ (the coefficient of w_{n+1} is not zero) but $\mathcal{B}'$ is linearly independent. Therefore $\mathcal{B}'$ cannot have any more elements, and so $\#\mathcal{B}' = n = \#\mathcal{B}$, as claimed. □

The conclusion of all that work is that we may formulate a general definition of dimension, at least in the finite-dimensional case.

Definition 4.19

Let V be a finite-dimensional vector space. The **dimension of** V is the number of elements of any basis of V. It is denoted $\dim V$, or $\dim_{\mathbb{R}} V$ when it is useful to emphasise that the scalars are $\mathbb{R}$. By definition $\dim V \in \mathbb{N}$.

We now have a theory of dimension for vector spaces – or for finite-dimensional ones at least. As a first test of the flexibility of the theory we see that it accords with our idea that subspaces of vector spaces have lower dimension.

Proposition 4.20

Let V be a finite-dimensional vector space and $W \subset V$ a subspace. Then W is finite dimensional and $\dim W \leq \dim V$, with equality if and only if $W = V$.

Proof. We first prove that W is finite dimensional. Suppose not. Then there is an infinite set $v_1, v_2, \ldots \in W$ of linearly independent elements of W by Corollary 4.12. But they are linearly independent when considered as elements of V, which contradicts Proposition 4.10.

Let $v_1, \dots, v_s \in W$ be a basis of W. Considered as elements of V, they are linearly independent. Therefore, by Corollary 4.16 there are elements $v_{s+1}, \dots, v_n \in V$ so that $v_1, \dots, v_n$ is a basis of V. Thus

$$\dim W = s \leq n = \dim V$$

with equality if and only if $s = n$, which is to say that $v_1, \dots, v_s$ is a basis of V and so $W = V$. □

This proposition is frequently useful in the following curious formulation: the point is that $\dim W \geq s$ by Corollary 4.16.

Corollary 4.21

Let $W \subset V$ be a subspace of a finite-dimensional vector space. If $w_1, \dots, w_s \in W$ are linearly independent with $s = \dim V$, then $W = V$.

The final corollary packages the main points into two simple criteria.

Corollary 4.22

Let V be a vector space with $\dim V = n$.

(i) If $v_1, \dots, v_n \in V$ are linearly independent, then they form a basis of V.

(ii) If $v_1, \dots, v_n \in V$ span V, then they form a basis of V.

The first part follows at once from the previous corollary, while the second follows at once from the Sifting Lemma 4.15.

4.5 Direct sum of vector spaces

One of the tropes of beginning algebra is the idea of constructing new spaces from old ones. Having done that, you usually find that it moves quickly to the converse idea of breaking big spaces into smaller pieces, but let's not run before we can walk. This section merely provides yet another class of examples of vector spaces.

Definition 4.23

Let V and W be vector spaces. The **direct sum of V and W** is the vector space

$$V \oplus W = \{(v, w) \mid v \in V, w \in W\}$$

with vector space operations

$$\begin{aligned} (v_1, w_1) + (v_2, w_2) &= (v_1 + v_2, w_1 + w_2) \\ \lambda(v, w) &= (\lambda v, \lambda w) \end{aligned}$$

where the left-hand side of each line is the operation in $V \oplus W$, and it is defined by the right-hand side, which involves only operations in V and W.

The additive identity is $0_{V \oplus W} = (0_V, 0_W)$, and the additive inverse is $-(v, w) = (-v, -w)$.

Of course one must check that the operations defined above on $V \oplus W$ above really do make it into a vector space: they do, but please check.

Example

Let $V = \mathbb{R}^2$ and $W = \mathbb{R}^1$. Then

$$V \oplus W = \left\{ \left(\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, (\lambda_3) \right) \middle| \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R} \right\}$$

If we slightly cheekily treat the element $((\lambda_1, \lambda_2)^T, (\lambda_3))$ as a single column vector $(\lambda_1, \lambda_2, \lambda_3)^T$, then we see that the operations defined in Definition 4.23 agree with the operations in $\mathbb{R}^3$.

In any case, we see a natural basis of $V \oplus W$:

$$v_1 = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, (0) \right), \quad v_2 = \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, (0) \right), \quad v_3 = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, (1) \right)$$

It is easy to check that v_1, v_2, v_3 are linearly independent and span. From this point of view,

$$V \oplus W = \{\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 \mid \lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}\}$$

which makes our little charade with $\mathbb{R}^3$ above completely watertight (since the operations defined on $V \oplus W$ are exactly componentwise sum of such linear expressions).

We can be even more precise about this later, once we have linear maps and isomorphisms, but even now it is clear that we may, for example, build up $\mathbb{R}^n$ by direct sum of n copies of $\mathbb{R}$, and that larger direct sums $\mathbb{R}^n \oplus \mathbb{R}^m$ behave exactly like $\mathbb{R}^{n+m}$.

Proposition 4.24

Suppose $v_1, \ldots, v_s$ is a basis of V and $w_1, \ldots, w_t$ is a basis of W. Then

$$\mathcal{B} = \{(v_1, 0_W), \ldots, (v_s, 0_W)\} \cup \{(0_V, w_1), \ldots, (0_V, w_t)\}$$

is a basis of $V \oplus W$.

This is another routine proof for you to try: if you'd like to compare with my solution, here it is.

Proof. If $(v, w) \in V \oplus W$ then since using the given basis there are scalars $\lambda_i, \mu_j \in \mathbb{R}$ for which

$$v = \lambda_1 v_1 + \ldots + \lambda_s v_s \quad \text{and} \quad w = \mu_1 w_1 + \ldots + \mu_t w_t$$

and therefore $\mathcal{B}$ spans since

$$(v, w) = \lambda_1(v_1, 0_W) + \ldots + \lambda_s(v_s, 0_W) + \mu_1(0_V, w_1) + \ldots + \mu_t(0_V, w_t)$$

If for $\lambda_i, \mu_j \in \mathbb{R}$

$$\lambda_1(v_1, 0_W) + \ldots \lambda_s(v_s, 0_W) + \mu_1(0_V, w_1) + \ldots \mu_t(0_V, w_t) = 0_{V \oplus W}$$

then

$$\lambda_1 v_1 + \ldots + \lambda_s v_s = 0_V \quad \text{and} \quad \mu_1 w_1 + \ldots + \mu_t w_t = 0_W$$

and so all $\lambda_i = \mu_j = 0$ since the given bases of V and W are linearly independent. □

Counting the size of the basis $\mathcal{B}$ computes the dimension of $V \oplus W$.

Corollary 4.25

If V and W are finite-dimensional vector spaces, then so is $V \oplus W$ and moreover

$$\dim V \oplus W = \dim V + \dim W$$

Example 4.26

Let $V = \mathbb{R}^2$ and $W = \mathbb{R}[x]_{\leq 2}$. Then

$$V \oplus W = \left\{ \left(\begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}, \lambda_3 + \lambda_4 x + \lambda_5 x^2 \right) \middle| \lambda_1, \ldots, \lambda_5 \in \mathbb{R} \right\}$$

Choosing bases of V and W provides a basis for $V \oplus W$: for example

$$v_1 = \left(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, 0 \right), \; v_2 = \left(\begin{pmatrix} 0 \\ 1 \end{pmatrix}, 0 \right), \; v_3 = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, 1 \right), \; v_4 = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, x \right), \; v_5 = \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, x^2 \right)$$

so that

$$V \oplus W = \{\lambda_1 v_1 + \ldots + \lambda_5 v_5 \mid \lambda_1, \ldots, \lambda_5 \in \mathbb{R}\}$$

With this view, the key data that determines an element of $V \oplus W$ is the vector of coefficients $(\lambda_1, \ldots, \lambda_5)^T \in \mathbb{R}^5$. As far as the vector space operations go, we may treat the vector space $V \oplus W$ essentially as the same as $\mathbb{R}^5$. Of course we must remember that if we ever need to use properties of its elements then in fact it is very different from $\mathbb{R}^5$.

4.6 What does a finite basis do for me?

Example 4.26 illustrated how a finite basis boils down the elements of an n-dimensional vector space V to a simple vector of coefficients that we may regard as an element of $\mathbb{R}^n$. Put differently, when we work with a finite-dimensional vector space **together with a fixed choice of basis**, then we may often do some or all our work in $\mathbb{R}^n$ where we know exactly how to solve any linear problem.

Proposition 4.27

Let V be a vector space and $\mathcal{B} = \{v_1, \ldots, v_n\}$ a basis. Then there is a bijection

$$\begin{aligned} \chi_{\mathcal{B}} \colon V &\longrightarrow \mathbb{R}^n \\ v &\mapsto \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \quad \text{where } v = \lambda_1 v_1 + \ldots + \lambda_n v_n \end{aligned}$$

which respects the vector space operations: that is,

$$\chi_{\mathcal{B}}(v) + \chi_{\mathcal{B}}(w) = \chi_{\mathcal{B}}(v + w) \quad \text{and} \quad \chi_{\mathcal{B}}(\lambda v) = \lambda \chi_{\mathcal{B}}(v)$$

for all $v, w \in V$ and $\lambda \in \mathbb{R}$.

Notice that it is crucial that we regard the basis as being $v_1, \ldots, v_n$ **in that order**, despite the set-theoretic notation.

Proof. There is almost nothing to prove. The map $\chi_{\mathcal{B}}$ both exists and is surjective because $\mathcal{B}$ spans so every $v \in V$ has an expression as indicated, and every such expression gives an element of V. The map $\chi_{\mathcal{B}}$ is both well defined and injective because $\mathcal{B}$ is linearly independent so the indicated expression for v is unique by Proposition 4.14.

The map $\chi_{\mathcal{B}}$ respects the vector space operations because those operations are carried out componentwise in both the domain and the codomain. In detail, suppose $v = \sum \lambda_i v_i$ and $w = \sum \mu_i v_i$ so that $v + w = \sum (\lambda_i + \mu_i) v_i$ and $\lambda v = \sum (\lambda \lambda_i) v$. Then the ith component of the column vector $\chi_{\mathcal{B}}(v + w)$ is $\lambda_i + \mu_i$, which equals the ith component of $\chi_{\mathcal{B}}(v) + \chi_{\mathcal{B}}(w)$, and similarly for λv. □

Definition 4.28

When V is a (finite-dimensional) vector space with a fixed choice of basis $\mathcal{B} = \{v_1, \ldots, v_n\}$, we refer to the bijection $\chi_{\mathcal{B}} \colon V \to \mathbb{R}^n$ as the **coordinate map with respect to the basis** $\mathcal{B}$.

Remark

This is the great trick, or the great con, of (finite-dimensional) linear algebra: faced with a complicated and abstract situation V, we may simply choose a basis and do our calculations in some $\mathbb{R}^n$, and then use the basis to translate back to V.

This is almost exactly the same as being a forensic pathologist: we take samples v and w from the messy and complicated real life crime scene V back to the clean lab $\mathbb{R}^n$ where we have all our tools for cutting them into pieces and solving whatever linear equation mystery they may be involved in.

We will see this pay off next when we finally consider linear maps between vector spaces explicitly (and see that we have been using them implicitly all along).

That sounds great, but on the quiet down here, we can be honest about the problem: what happens if we choose a different basis? how do our conclusions gleaned from one choice of basis compare to the other? That is the change of basis question we saw in §3.4, and life is simplest if we keep thinking of it as the simple matter it is (once we have discussed it slowly and precisely).

Chapter 5

Linear Maps

Informally, a linear map $\varphi\colon V \to W$ between vector spaces V and W is a map of sets that respects the two vector space operations: this simply means the following.

Definition 5.1

Let V and W be vector spaces. Then a map $\varphi\colon V \to W$ is a **linear map** if and only if

$$\varphi(v+w) = \varphi(v) + \varphi(w) \quad \text{and} \quad \varphi(\lambda v) = \lambda\varphi(v)$$

for all $v, w \in V$ and $\lambda \in \mathbb{R}$. If you prefer an equivalent more concise version:

$$\varphi \text{ is a linear map} \iff \varphi(\lambda v + \mu w) = \lambda\varphi(v) + \mu\varphi(w) \quad \text{for all } v, w \in V \text{ and } \lambda, \mu \in \mathbb{R}.$$

The linear map φ that is also a bijection is called an **isomorphism** (or an **isomorphism of vector spaces** if the context is not clear). In this case we say V and W are **isomorphic** and we write $V \cong W$. (It is easy to see that in this case the inverse bijection $\varphi^{-1}\colon W \to V$ is also a linear map; see Lemma 5.6.)

Linear maps are the natural maps to consider between vector spaces, just as group homomorphisms (which respect the group operations) are the natural maps between groups, and ring homomorphisms (which respect the ring operation) are the natural maps between rings, and differentiable maps (which respect the differentiability of functions) are the natural maps between (differentiable) manifolds.

You are extremely familiar with the most famous linear map of them all: recall that $\mathbb{C}^\infty(\mathbb{R})$ is the vector space of (infinitely) differentiable functions $f\colon \mathbb{R} \to \mathbb{R}$. Differentiation

$$\begin{array}{rcl} \mathbb{C}^\infty(\mathbb{R}) & \to & \mathbb{C}^\infty(\mathbb{R}) \\ f & \mapsto & \dfrac{df}{dx} \end{array}$$

is a linear map: you have known forever that whenever f and g are differentiable and $\lambda \in \mathbb{R}$

$$\frac{d}{dx}(f+g) = \frac{df}{dx} + \frac{dg}{dx} \quad \text{and} \quad \frac{d(\lambda f)}{dx} = \lambda\frac{df}{dx}$$

At first, though, we consider more mundane examples.

Example

The map $\varphi\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2x - 5y \\ -x + 3y \end{pmatrix}$$

is a linear map; you can check that it satisfies the conditions of Definition 5.1. In fact it is a bijection, so it is an isomorphism. The inverse map $\varphi^{-1}\colon \mathbb{R}^2 \to \mathbb{R}^2$ is

$$\begin{pmatrix} u \\ v \end{pmatrix} \mapsto \begin{pmatrix} 3x + 5y \\ x + 2y \end{pmatrix}$$

which again is a linear map. You see at once the fingerprints of the matrix

$$A = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix} \quad \text{and its inverse} \quad A^{-1} = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$$

all over this crime scene.

We will see that this is no coincidence: this is how a typical linear map appears (see §5.3), and in fact the matrices are the good guys not the villains, so don't read them their rights just yet.

Whenever you see maps defined by **homogeneous linear expressions** in some parameters, as the map φ above is defined by the expressions $2x - 5y$ and $-x + 3y$, you should expect there is a vector space kicking around and that this is a linear map. If you see nonlinear terms such as $3x + y^2$ or $1 - 2x + y$, then the map is surely not a linear map, nice as it may be. These are not rules, but they are a good first approximation when you're out in the wild.

Language 5.2

Whenever someone says that $\varphi\colon V \to W$ is a linear map, it is understood that V and W are vector spaces in some way that they expect to be clear to you. Since in these notes we use the same notation for addition and scalar multiplication in every vector space we consider, we may omit the explicit mention that V and W are vector spaces. We will never deploy a linear map without the domain and codomain being vector spaces in some way that is unambiguous in the context.

This chapter studies the image and kernel of linear maps (mostly between finite-dimensional vector spacess) and applies this to questions of invertibility and isomorphism. These are the key definitions, so we give them next, even before looking at more examples.

5.1 Routine trivialities

Definition 5.3

Let $\varphi\colon V \to W$ be a linear map. The **image of** φ, denoted $\mathrm{Im}(\varphi)$, is the subset of W

$$\mathrm{Im}(\varphi) = \{w \in W \mid w = \varphi(v) \text{ for some } v \in V\}$$

This is precisely the same as the usual set-theoretic image, and is also often denoted by $\varphi(V)$.

The **kernel of** φ, denoted $\ker(\varphi)$, is the subset of V

$$\ker(\varphi) = \{v \in V \mid \varphi(v) = 0_W\}$$

This is precisely the same as the usual group-theoretic kernel if we consider V and W as abelian groups under their respective addition operations.

Remark

Since any linear map $\varphi\colon V \to W$ is, in particular, a homomorphism of abelian groups, with respective additive identities 0_V and 0_W, we already know the following lemma. I include a proof in case you have forgotten it, and I use scalar multiplication whether I need to or not.

Lemma 5.4

Let $\varphi\colon V \to W$ be a linear map.

(i) $\varphi(0_V) = 0_W$.

(ii) φ is injective if and only if $\ker\varphi = \{0_V\}$.

Proof. (i) Cashing in one axiom at a time we have $\varphi(0_V) + \varphi(0_V) = \varphi(0_V + 0_V) = \varphi(0_V)$, so adding $-\varphi(0_V)$ to both sides (without caring in the slightest what it actually is) gives the result.

(ii) Suppose that $\ker\varphi = \{0_V\}$ and we have $v_1, v_2 \in V$ with $\varphi(v_1) = \varphi(v_2)$. We must prove that $v_1 = v_2$. By linearity of φ,

$$\varphi(v_1 - v_2) = \varphi(v_1) + \varphi((-1)v_2) = \varphi(v_1) - \varphi(v_2) = 0_W$$

so that $v_1 - v_2 \in \ker\varphi$. Therefore $v_1 - v_2 = 0_V$, or in other words $v_1 = v_2$, as required.

The converse is quicker: if φ is injective, then at most one element may map to 0_W, and since $\varphi(0_V) = 0_W$ by (i), we have $\ker\varphi = \{0_V\}$. □

Remark

We should get on looking at examples, but there are a couple more useful yet routine results to note in passing. They are slightly harder than the previous lemma, but similar in spirit, in that they depend only on how the rules work at a fairly superficial level, so live in the mulch just one inch above the axioms. They are good to try as an exercise in shunting the symbols around.

Proposition 5.5

Let $\varphi\colon V \to W$ be a linear map.Then

(i) $\ker\varphi \subset V$ is a subspace of V.

(ii) $\operatorname{Im}\varphi \subset W$ is a subspace of W.

Proof. (i) We know $0_V \in \ker\varphi$ by Lemma 5.4(i). Suppose $v_1, v_2 \in \ker\varphi$ and $\lambda_1, \lambda_2 \in \mathbb{R}$. Then by linearity of φ

$$\varphi(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 \varphi(v_1) + \lambda_2 \varphi(v_2) = \lambda_1 0_V + \lambda_2 0_V = 0_V$$

and so $\lambda_1 v_1 + \lambda_2 v_2 \in \ker \varphi$, as required.

(ii) We know $0_W \in \ker \varphi$ by Lemma 5.4(i). Suppose $w_1, w_2 \in \operatorname{Im} \varphi$ and $\mu_1, \mu_2 \in \mathbb{R}$. Then by the definition of the image, there exist $v_1, v_2 \in V$ with $w_1 = \varphi(v_1)$ and $w_2 = \varphi(v_2)$. So by linearity of φ

$$\mu_1 w_1 + \mu_2 w_2 = \mu_1 \varphi(v_1) + \mu_2 \varphi(v_2) = \varphi(\mu_1 v_1 + \mu_2 v_2)$$

is also in $\operatorname{Im} \varphi$, as required. □

Lemma 5.6

If $\varphi\colon V \to W$ is a linear map that is a bijection, then its set-theoretic inverse map $\varphi^{-1}\colon W \to V$ is also a linear map.

Proof. Oh blimey, how do you do these ones? We'd better start with some $w_1, w_2 \in W$ and $\mu_1, \mu_2 \in \mathbb{R}$, and then we need somehow to deal with $\varphi^{-1}(\mu_1 w_1 + \mu_2 w_2)$. It pays to say out loud "I do not yet know that φ^{-1} is linear" until you find you have proved that it is. Now try it.

The trick with this sort of thing is usually to say: since φ is a bijection, there are $v_1, v_2 \in V$ so that $w_1 = \varphi(v_1)$, or equivalently $v_1 = \varphi^{-1}(w_1)$, and similarly w_2. Then since we know that at least φ is linear, we have

$$\begin{aligned} \varphi(\mu_1 v_1 + \mu_2 v_2) &= \mu_1 \varphi(v_1) + \mu_2 \varphi(v_2) \\ &= \mu_1 \varphi(\varphi^{-1}(w_1)) + \mu_2 \varphi(\varphi^{-1}(w_2)) \\ &= \mu_1 w_1 + \mu_2 w_2 \end{aligned}$$

and now applying φ^{-1} to both sides spells out exactly what we want. Miserable, but there it is. □

Proposition 5.7

Let $\varphi\colon V \to W$ be an isomorphism. If $\mathcal{B} \subset V$ is a basis of V, then $\varphi(\mathcal{B}) = \{\varphi(v) \mid v \in \mathcal{B}\}$ is a basis of W.

Remark

During the proof we see a typical move when working with linear maps. If φ is linear, then by applying the addition and scalar multiplication properties from Definition 5.1 one at a time, we know for example that

$$\begin{aligned} \varphi(\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3) &= \varphi(\lambda_1 v_1) + \varphi(\lambda_2 v_2 + \lambda_3 v_3) \\ &= \varphi(\lambda_1 v_1) + \varphi(\lambda_2 v_2) + \varphi(\lambda_3 v_3) \\ &= \lambda_1 \varphi(v_1) + \lambda_2 \varphi(v_2) + \lambda_3 \varphi(v_3) \end{aligned}$$

and more generally, by induction, that

$$\varphi(\lambda_1 v_1 + \ldots + \lambda_s v_s) = \lambda_1 \varphi(v_1) + \ldots + \lambda_s \varphi(v_s)$$

That is, when mapping any linear combination of vectors in V by φ, we may map all the individual vectors by φ first, and then rebuild the linear combination in W. It is probably easier simpler to do it than to say it in prose. You will see this manoeuvre many many times.

Good grief! here we go again. Note first that we are secretly given an inverse linear map $\varphi^{-1}\colon W \to V$, so we can translate any question in W back to V to solve it. Right, deep breath ...

Proof. We first show that $\varphi(\mathcal{B})$ spans. Let $w \in W$. Consider $v = \varphi^{-1}(w) \in V$. Since $\mathcal{B}$ is a basis of V, there are elements $v_1, \ldots, v_s \in \mathcal{B}$ and scalars $\lambda_1, \ldots, \lambda_s \in \mathbb{R}$ so that

$$v = \lambda_1 v_1 + \ldots + \lambda_s v_s$$

(Problem solved in V, the wrong place, so hit that with φ to get into W.) So applying φ to both sides we have

$$\begin{aligned} w = \varphi(v) &= \varphi(\lambda_1 v_1 + \ldots + \lambda_s v_s) \\ &= \lambda_1 \varphi(v_1) + \ldots + \lambda_s \varphi(v_s) \end{aligned}$$

which does indeed express w as a linear combination of elements of $\varphi(\mathcal{B})$.

To finish, we show that $\varphi(\mathcal{B})$ is linearly independent. Suppose

$$0_W = \lambda_1 \varphi(v_1) + \ldots + \lambda_s \varphi(v_s)$$

for elements $\varphi(v_i) \in \varphi(\mathcal{B})$ and scalars $\lambda_i \in \mathbb{R}$. We must show that all $\lambda_i = 0$. (Strategy: hit this with φ^{-1} to get into V and hope to solve it there.) So applying φ^{-1} to both sides we have

$$\begin{aligned} 0_V = \varphi^{-1}(0_W) &= \varphi^{-1}(\lambda_1 \varphi(v_1) + \ldots + \lambda_s \varphi(v_s)) \\ &= \lambda_1 \varphi^{-1}(\varphi(v_1)) + \ldots + \lambda_s \varphi^{-1}(\varphi(v_s)) \\ &= \lambda_1 v_1 + \ldots + \lambda_s v_s \end{aligned}$$

But $v_1, \ldots, v_s \in \mathcal{B}$, and $\mathcal{B}$ is linearly independent, so we conclude at once that $\lambda_1 = \ldots = \lambda_s = 0$, which is what we were required to do. Game over. □

Remark

You probably noticed that Proposition 5.7 did not assume that $\mathcal{B}$ is a finite set: the result holds for bases of any size, and therefore for vector spaces of any dimension. The next result follows at once in the case that V is finite dimensional.

Corollary 5.8

If $V \cong W$ and V is finite dimensional, then W is finite dimensional too and $\dim V = \dim W$. In particular, $\mathbb{R}^n \cong \mathbb{R}^m$ if and only if $n = m$.

5.2 Writing elements in coordinates is a linear map

We can express Proposition 4.27 in the language of isomorphisms.

Proposition 5.9

Let V be a vector space with a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ (regarded as being in that fixed order). Then the coordinate map $\chi_{\mathcal{B}}$ of Proposition 4.27

$$\chi_{\mathcal{B}} \colon V \longrightarrow \mathbb{R}^n$$

which takes $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$ to the column vector $(\lambda_1, \ldots, \lambda_n)^T \in \mathbb{R}^n$ is an isomorphism.

This has an immediate consequence that is slightly shocking at first sight.

Corollary 5.10

If V is a finite-dimensional vector space, then $V \cong \mathbb{R}^n$ for some unique $n \in \mathbb{N}$.

Proof. By Proposition 4.15 there is a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ of V. The coordinate map $\chi_{\mathcal{B}}$ then gives an isomorphism to $\mathbb{R}^n$. The uniqueness of n is by Corollary 5.8. □

Remark

So why on earth do we bother with a fully tooled-up abstract theory of vector spaces if up to isomorphism they are all just one of the $\mathbb{R}^n$? The first thing to say is that of course those are only the finite-dimensional ones. As we've seen, there are plenty of other vector spaces that we use every day that are not finite dimensional.

Another thing to say, and perhaps more important, is that the isomorphism $V \cong \mathbb{R}^n$ involved a choice of basis, and so it is more data that the vector space V alone. We will get a lot of profit from working with vector spaces without having to make that choice: as they say in the jargon, we frequently work 'coordinate free'.

Another thing worth saying is that, when vector spaces arise, they usually have personalities all of their own, involving many other ideas that just add, subtract and multiply by scalars. Have a look at the next example, and decide whether you like your old friend left as it is or would rather think of it as $\mathbb{R}^2$.

Example

The complex numbers $\mathbb{C}$ is a vector space: indeed you can add and subtract complex numbers, and you can certainly multiply them by real numbers (and if you review all the axioms and properties of $\mathbb{C}$, you will find that the vector space axioms do all hold). Of course you can do a whole lot more, but as a vector space you are not asked about that.

Fine, so what is its dimension? You probably already have a favourite basis: I'm guessing it is $\mathcal{B} = \{1, i\}$, where $i = \sqrt{-1}$ (whatever that means), and that you write

$$\mathbb{C} = \{a + ib \mid a, b \in \mathbb{R}\}$$

(where we usually quietly omit the 1: we usually write $a + ib$ rather than $a1 + bi$). Maybe that is even your definition of $\mathbb{C}$. In any case, it certainly satisfies all the axioms of a vector space, and since we have a basis we know that $\dim_{\mathbb{R}} \mathbb{C} = 2$ (where this seems like a wise time to emphasise that we are using $\mathbb{R}$ as scalars, so we write $\dim_{\mathbb{R}}$ rather than simply $\dim$ to say so).

In this case we may write the coordinate map $\chi_{\mathcal{B}}$ as

$$\begin{array}{rcl} \chi_{\mathcal{B}} \colon \mathbb{C} & \longrightarrow & \mathbb{R}^2 \\ a + ib & \mapsto & \begin{pmatrix} a \\ b \end{pmatrix} \end{array}$$

and this map is an isomorphism. As a vector space we never have to think about multiplying two complex numbers together, so the horrid mess we would see in coordinates in $\mathbb{R}^2$ if we did is not an issue.

That's all true, but it is pretty forgetful, ungrateful and uncouth. I'll stick with $\mathbb{C}$, thanks.

Incidentally, since $\mathbb{C} \cong \mathbb{R}^2$ as a vector space, it should be possible to draw $\mathbb{C}$ as a plane. That is exactly what the Argand diagram is: you probably label the point $(1,0)^T \in \mathbb{R}^2$ as the complex number 1 and $(0,1)^T \in \mathbb{R}^2$ as the complex number i, which is what $\chi_{\mathcal{B}}^{-1}$ tells you to do.

5.3 Linear maps $\mathbb{R}^n \to \mathbb{R}^m$

Any matrix $A \in \mathrm{Mat}_{mn}$ determines a linear map by left multiplication by A:

$$\begin{array}{rcl} L_A \colon \mathbb{R}^n & \to & \mathbb{R}^m \\ \underline{v} & \mapsto & A\underline{v} \end{array} \tag{5.1}$$

The linearity of L_A follows immediately from the properties of matrix multiplication, Proposition 2.42. Linear maps of this form are essential examples to understand.

We already understand the image and kernel of L_A in concrete terms.

Lemma 5.11

Let $A \in \mathrm{Mat}_{mn}$ and let $L_A \colon \mathbb{R}^n \to \mathbb{R}^m$ be the associated left multiplication linear map.

(i) The kernel of L_A is equal to the set of solutions of the equations $A\underline{v} = \underline{0}$:

$$\ker(L_A) = \{\underline{v} \in \mathbb{R}^n \mid A\underline{v} = \underline{0}\}.$$

(ii) The image of L_A is equal to the column span of A: $\mathrm{Im}(L_A) = \mathrm{Colspan}\, A$.

Proof. (i) This is immediate: $\underline{v} \in \ker L_A$ means exactly that $A\underline{v} = \underline{0}$.

(ii) Let $\underline{e}_1, \dots, \underline{e}_n$ be the standard basis of $\mathbb{R}^n$. Note that $L_A(\underline{e}_i) = \underline{c}_i$ is the ith column $\underline{c}_i$ of A treated as a vector in $\mathbb{R}^m$, so we have

$$\mathrm{Colspan}\, A = \langle \underline{c}_1, \dots, \underline{c}_n \rangle = \langle L_A(\underline{e}_1), \dots, L_A(\underline{e}_n) \rangle$$

For any $\underline{v} \in \mathbb{R}^m$ there are unique scalars $\lambda_i \in \mathbb{R}$ so that $\underline{v} = \lambda_1 \underline{e}_1 + \dots + \lambda_n \underline{e}_n$. Since L_A is linear

$$\begin{array}{rcl} L_A(\underline{v}) & = & L_A(\lambda_1 \underline{e}_1 + \dots + \lambda_m \underline{e}_n) \\ & = & \lambda_1 L_A(\underline{e}_1) + \dots + \lambda_m L_A(\underline{e}_n) \end{array}$$

so that $L_A(\underline{v})$ lies in $\mathrm{Colspan}\, A$. Thus $\mathrm{Im}(L_A) \subset \mathrm{Colspan}\, A$.

Conversely, if $w \in \operatorname{Colspan}(A)$ then there are scalars $\mu_i \in \mathbb{R}$ so that

$$\begin{aligned} w &= \mu_1 \underline{c}_1 + \ldots + \mu_n \underline{c}_n \\ &= \mu_1 L_A(\underline{e}_1) + \ldots + \mu_n L_A(\underline{e}_n) \\ &= L_A(\mu_1 \underline{e}_1 + \ldots + \mu_n \underline{e}_n) \end{aligned}$$

so that $w = L_A(v)$ where $v = \mu_1 \underline{e}_1 + \ldots + \mu_n \underline{e}_n$ and so $w \in \operatorname{Im}(A)$. Thus $\operatorname{Colspan} A \subset \operatorname{Im}(A)$, and so they are equal. □

Language 5.12

This connection between maps and matrices is so automatic that we frequently refer to the map as A rather than L_A. We will try to stick to L_A, but even then everybody uses the shortcut

$$\ker A = \{\underline{v} \in \mathbb{R}^n \mid A\underline{v} = \underline{0}\}$$

The only teeny catch is that some people prefer to multiply by matrices on the right, and that determines a linear map $\mathbb{R}^m_{\text{row}} \to \mathbb{R}^n_{\text{row}}$, but we will be explicit about it whenever that happens.

Just to be sure, recall that for maps $\varphi\colon X \to Y$ and $\psi\colon Y \to Z$, their composition is defined by

$$\begin{aligned} \psi \circ \varphi\colon X &\longrightarrow Z \\ x &\mapsto \psi(\varphi(x)) \end{aligned}$$

Lemma 5.13

Let $A \in \operatorname{Mat}_{mn}$ and $B \in \operatorname{Mat}_{\ell m}$. Then $L_{BA} = L_B \circ L_A\colon \mathbb{R}^n \to \mathbb{R}^\ell$. If $m = n$ then

(i) L_{I_n} is the identity map $\mathbb{R}^n \to \mathbb{R}^n$.

(ii) If A is an invertible matrix, then L_A is an invertible map and $(L_A)^{-1} = L_{A^{-1}}$.

Proof. Let $\underline{v} \in \mathbb{R}^n$. Then by the associativity of matrix multiplication

$$L_B(L_A(\underline{v})) = B(A\underline{v}) = BA(\underline{v}) = L_{BA}(\underline{v})$$

which is the first claim. The remaining claims hold because $A^{-1}A = AA^{-1} = I_n$ multiplies $\underline{v}$ to itself, and so L_A and $L_{A^{-1}}$ are mutual inverses, and so are inverse bijections. □

This framework provides a nice clean environment in which to understand the technology of elementary matrices and row and column operations.

Proposition 5.14

Let $A \in \operatorname{Mat}_{mn}$. If $E = E_k \cdots E_1$ is a product of elementary matrices $E_i \in \operatorname{Mat}_{mm}$ and $F = F_1 \cdots F_\ell$ is a product of elementary matrices $F_j \in \operatorname{Mat}_{nn}$, then

(i) L_{EA} is the composition $\mathbb{R}^n \xrightarrow{L_A} \mathbb{R}^m \xrightarrow{L_E} \mathbb{R}^m$ and

$$\ker L_{EA} = \ker L_A \quad \text{and} \quad \dim \operatorname{Im} L_{EA} = \dim \operatorname{Im} L_A.$$

(ii) L_{AF} is the composition $\mathbb{R}^n \xrightarrow{L_F} \mathbb{R}^n \xrightarrow{L_A} \mathbb{R}^m$ and

$$\dim \ker L_{AF} = \dim \ker L_A \quad \text{and} \quad \operatorname{Im} L_{AF} = \operatorname{Im} L_A.$$

(iii) L_{EAF} is the composition $\mathbb{R}^n \xrightarrow{L_F} \mathbb{R}^n \xrightarrow{L_A} \mathbb{R}^m \xrightarrow{L_E} \mathbb{R}^m$ and

$$\dim \ker L_{EAF} = \dim \ker L_A \quad \text{and} \quad \dim \operatorname{Im} L_{EAF} = \dim \operatorname{Im} L_A.$$

You will notice the difference in the statements above: sometimes we show that two spaces are the same, while sometimes we show only that they have the same dimension.

If you read it carefully, you will see that the spaces are the same only when they are both subsets of the same vector space in the relevant diagram $\mathbb{R}^a \to \mathbb{R}^b \to \mathbb{R}^c$. When instead the two spaces lie in the domain and codomain respectively of some map L, we only discuss their dimension: we do not know what L does, so it is not reasonable to expect them to be identical, but we may nevertheless compare them using L and the following completely general statement.

Lemma 5.15

Suppose $\varphi\colon V \to W$ is an isomorphism and $U \subset V$ is a subspace. Then φ gives an isomorphism between U and the subspace $\varphi(U) \subset W$.

In particular, if U is finite dimensional then $\dim U = \dim \varphi(U)$.

Proof. Regarding φ as a linear map $U \to W$, it is still injective, and it is of course surjective onto its image $\varphi(U)$, therefore it is a bijection $U \to \varphi(U)$, and so is an isomorphism as claimed. The final line follows from Corollary 5.8. □

With that, we may proceed with the proof of Proposition 5.14.

Proof. The only point we use about the elementary matrices is that E and F are both invertible matrices, so that in particular L_E and L_F are isomorphisms. The compositions of maps L_E, L_A and L_F all follow at once from Lemma 5.13 – we just need to check the equality claims.

(i) Since L_E is injective, $L_E(L_A(\underline{v})) = \underline{0}$ if and only if $L_A(\underline{v}) = \underline{0}$, which is the first claim. Now setting $U = \operatorname{Im} L_A \subset \mathbb{R}^m$, we have $\operatorname{Im} L_{EA} = L_E(U)$, and these have equal dimension by Lemma 5.15.

(ii) Since L_F is surjective, $L_A(L_F(\mathbb{R}^n)) = L_A(\mathbb{R}^n)$, which is the second claim. Now setting $U = \ker L_A \subset \mathbb{R}^n$, we have $\ker L_{AF} = L_F^{-1}(U)$, and these have equal dimension by Lemma 5.15.

(iii) This follows from the previous two. For example, setting $B = EA$, we know $\ker L_B = \ker L_A$ by (i) and $\dim \ker L_{BF} = \dim \ker L_B$ by (ii) so

$$\dim \ker L_{EAF} = \dim \ker L_{BF} \overset{\text{(ii)}}{=} \dim \ker L_B \overset{\text{(i)}}{=} \dim \ker L_A$$

as claimed. Similarly the image. □

This contains all the results we know when solving systems of linear equations, and then some that we did not yet know.

Corollary 5.16

Let $A \in \operatorname{Mat}_{mn}$. We regard A as the matrix of coefficients of a system of m linear equations in n unknowns.

(i) If EA is the RREF of A, then the solutions of $EA\underline{x} = \underline{0}$ are identical to the solutions of $A\underline{x} = \underline{0}$.

(ii) If AF is the RCEF of A, then $\operatorname{Colspan} AF = \operatorname{Colspan} A$.

(iii) The Smith normal form EAF of A is unique; in particular, its rank is well defined.

Proof. Each part follows from the corresponding part of Proposition 5.14. In particular, the solution set of $A\underline{x} = \underline{0}$ is $\ker L_A$ and the column span $\operatorname{Colspan} A$ is $\operatorname{Im} L_A$, so the proposition gives the claimed equalities.

Finally, the rank of the Smith normal form EAF of A is equal to the number of nonzero columns, which is the same as $\dim \operatorname{Im} \varphi_{EAF}$, which equals $\dim \operatorname{Im} L_A$ by the proposition, and so is independent of which elementary products E and F you used to compute it. □

Geometry and maps

We consider the vector space $\mathbb{R}^2$ together with the dot product, so that we may speak of lengths and angles, as in §1.2. You probably know that

$$A = \begin{pmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{pmatrix}$$

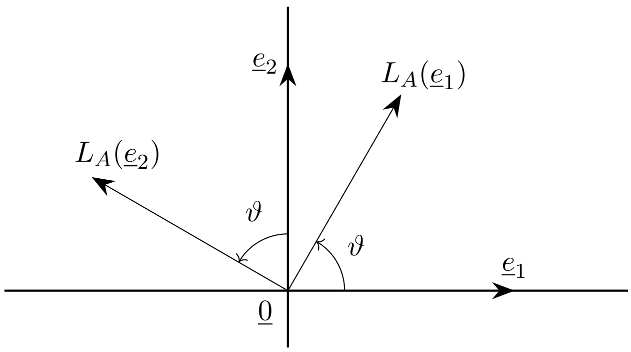
is a **rotation matrix**, but in any case we discuss it precisely now. The corresponding linear map $L_A \colon \mathbb{R}^2 \to \mathbb{R}^2$ takes the standard bases to

$$L_A(\underline{e}_1) = A\underline{e}_1 = \begin{pmatrix} \cos\vartheta \\ \sin\vartheta \end{pmatrix} \quad \text{and} \quad L_A(\underline{e}_2) = A\underline{e}_2 = \begin{pmatrix} -\sin\vartheta \\ \cos\vartheta \end{pmatrix}$$

We calculate the angle between each $\underline{e}_i$ and its image $\underline{v}_i = L_A(\underline{e}_i)$ using Definition 1.14. Suppose that $\vartheta \in [0, \pi]$. Then

$$\angle \underline{e}_1\underline{v}_1 = \cos^{-1}(\underline{e}_1 \cdot \underline{v}_1) = \cos^{-1}(\cos\vartheta) = \vartheta \quad \text{and} \quad \angle \underline{e}_2\underline{v}_2 = \cos^{-1}(\cos\vartheta) = \vartheta$$

and so we draw the familiar meaningful picture.



The great thing about linear maps is that what we see with our eyes actually happens: we see the standard coordinate vectors rotate by ϑ, and indeed every other vector does the same, as we naturally imagine. Indeed, given $\underline{v} = \begin{pmatrix} x \\ y \end{pmatrix}$, the angle to $\underline{w} = L_A(\underline{v}) = \begin{pmatrix} x\cos\vartheta + y\sin\vartheta \\ -x\sin\vartheta + y\cos\vartheta \end{pmatrix}$ is

$$\angle \underline{v}\underline{w} = \cos^{-1}(\hat{\underline{v}} \cdot \hat{\underline{w}}) = \cos^{-1}\left(\frac{x^2\cos\vartheta + y^2\cos\vartheta}{x^2+y^2}\right) = \vartheta$$

(We see in Proposition 5.17 below that it is always enough to understand what a linear map does on a basis, as we did here.)

In a similar spirit, the linear map $L_A\colon \mathbb{R}^2 \to \mathbb{R}^2$ corresponding to the **reflection matrix**

$$A = \begin{pmatrix} \cos\vartheta & \sin\vartheta \\ \sin\vartheta & -\cos\vartheta \end{pmatrix}$$

is the reflection in a line at angle $\vartheta/2$ above the x-axis. Again, you see the geometry by drawing a picture of how the corresponding linear map L_A takes the standard basis to the two columns of A. You can also check that $A^2 = I_2$, so that L_A is an involution of the plane (which, by definition, simply means that if you do it twice you get back to where you started).

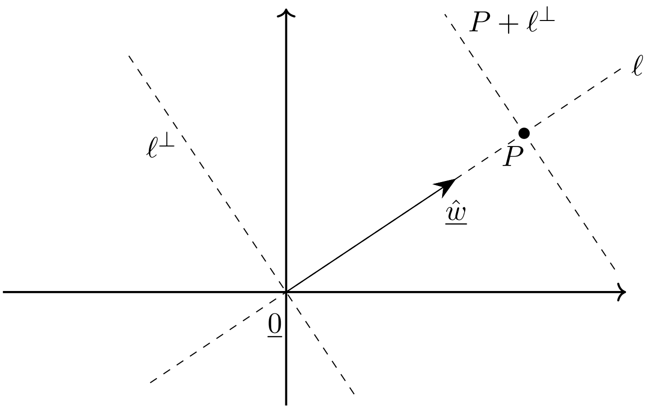
The first two matrices were invertible. For a non-invertible example, consider the **orthogonal projection** from vectors in $\mathbb{R}^2$ onto the line $\ell = (2x - 3y = 0) \subset \mathbb{R}^2$. We apply the formula for orthogonal projection from Definition 1.20: let $\hat{\underline{w}}$ be a unit vector along the line ℓ and then map $\underline{v} \mapsto (\underline{v} \cdot \hat{\underline{w}})\hat{\underline{w}}$. The vector $\underline{w} = (3,2)^T$ lies on ℓ and the unit vector in that direction is

$$\hat{\underline{w}} = \frac{1}{\sqrt{13}}\begin{pmatrix} 3 \\ 2 \end{pmatrix}$$

so by Definition 1.20 the linear map is

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \left(\begin{pmatrix} x \\ y \end{pmatrix} \cdot \hat{\underline{w}}\right)\hat{\underline{w}} = \frac{1}{13}(3x+2y)\begin{pmatrix} 3 \\ 2 \end{pmatrix} = A\begin{pmatrix} x \\ y \end{pmatrix} \quad \text{where} \quad A = \frac{1}{13}\begin{pmatrix} 9 & 6 \\ 6 & 4 \end{pmatrix}$$

In other words, the orthogonal projection to the line ℓ is the linear map $L_A\colon \mathbb{R}^2 \to \mathbb{R}^2$ for this matrix A. As ever, we have the power to check what we have calculated. You probably now see that $L_A(\underline{v}) \in \ell$ for any $\underline{v} \in \mathbb{R}^2$, simply because twice the first row of A equals 3 times the second row, and furthermore $A\hat{\underline{w}} = \hat{\underline{w}}$. In other words, $\operatorname{Im} L_A = \ell$. Which $\underline{v}$ map to the origin? By definition, any $\underline{v} \in \ker A$, and $\ker A$ is the line $\ell^\perp = (3x + 2y = 0) \subset \mathbb{R}^2$, which is indeed at right angles to ℓ. Furthermore, for any point $P = (a,b)^T \in \ell$, the set of points that maps to P is simply $P + \ell^\perp$.



5.4 Linear maps, bases and matrices

The essential point of this section is that a linear map is fully determined by how it maps a basis: linearity is precisely the condition that guarantees it. Once we have established that point, we show how to relate linear maps and matrices when you have chosen a basis in the domain and codomain.

Proposition 5.17

Let Let V be a vector space with basis $v_1, \ldots, v_n$ and W be any vector space (not necessarily finite dimensional).

(i) Suppose $\varphi_1 : V \to W$ and $\varphi_2 : V \to W$ are two linear maps, and suppose that they agree on the basis elements v_i: that is, $\varphi_1(v_i) = \varphi_2(v_i)$ for all $i = 1, \ldots, n$. Then $\varphi_1 = \varphi_2$.

(ii) For any choice of vectors $u_1, \ldots, u_n \in W$, there is a unique linear map $\varphi : V \to W$ with $\varphi(v_i) = u_i$ for $i = 1, \ldots, n$.

Proof. (i) Consider any $v \in V$. We must simply show that $\varphi_1(v) = \varphi_2(v)$. Expressing v in the given basis, there are scalars $\lambda_1, \ldots, \lambda_n$ for which $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$. Then, using the linearity of the two maps, and the fact they agree on the basis elements v_i,

$$\begin{aligned}
\varphi_1(v) &= \varphi_1(\lambda_1 v_1 + \ldots + \lambda_n v_n) \\
&= \lambda_1 \varphi_1(v_1) + \ldots + \lambda_n \varphi_1(v_n) \\
&= \lambda_1 \varphi_2(v_1) + \ldots + \lambda_n \varphi_2(v_n) \\
&= \varphi_2(\lambda_1 v_1 + \ldots + \lambda_n v_n) \\
&= \varphi_2(v)
\end{aligned}$$

as required.

(ii) Once again, consider any $v = \lambda_1 v_1 + \ldots + \lambda_n v_n \in V$. We must simply define a value for $\varphi(v)$ so that φ is linear and has the specified values on the basis elements v_i. Thus we *define*

$$\varphi(v) = \lambda_1 u_1 + \ldots + \lambda_n u_n \in W$$

It is immediate that $\varphi(v_i) = u_i$ for $i = 1, \ldots, n$, and φ is linear since if $w = \mu_1 v_1 + \ldots + \mu_n v_n$ and $\alpha, \beta \in \mathbb{R}$ are scalars, then by collecting coefficients of the v_i together we have

$$\begin{aligned}
\varphi(\alpha v + \beta w) &= \varphi(\alpha(\lambda_1 v_1 + \ldots + \lambda_n v_n) + \beta(\mu_1 v_1 + \ldots, \mu_n v_n)) \\
&= \varphi((\alpha\lambda_1 + \beta\mu_1)v_1 + \ldots + (\alpha\lambda_n + \beta\mu_n)v_n) \\
&= (\alpha\lambda_1 + \beta\mu_1)u_1 + \ldots + (\alpha\lambda_n + \beta\mu_n)u_n \\
&= \alpha(\lambda_1 u_1 + \ldots + \lambda_n u_n) + \beta(\mu_1 u_1 + \ldots + \mu_n u_n) \\
&= \alpha\varphi(v) + \beta\varphi(w)
\end{aligned}$$

as required. The uniqueness of φ follows from part (i). $\square$

So from now on, if you wish to define a linear map, you can simply choose a basis of the domain and specify any images in the codomain you would like those basis elements to have, and voilà there is a unique linear map that does just that.

Language 5.18

When we define a linear map φ by specifying the images of a basis $v_1, \ldots, v_n$ of V and applying Proposition 5.17(ii), in the jargon we say that φ is a **linear map defined on the basis** $v_1, \ldots, v_n$. To know which linear map we have just defined on the basis, we must also specify the image elements $u_1, \ldots, u_n \in W$.

Remark

In fact, there was no need to require that V is finite dimensional in Proposition 5.17: the issue is the existence of a basis B of V, not its size. In general, even if V has an infinite basis B, any elements $v, w \in V$ that you use involve only finitely many elements of B, and so the same proof works. We won't need this, but it is good to know the key point with no caveats: any choice of images of a basis uniquely determines a linear map.

If in addition to a fixed chosen basis $v_1, \ldots, v_n$ of V we fix a basis $w_1, \ldots, w_m$ of W, then the images $\varphi(v_i) \in W$ may be expressed in this basis of W. The key point is that we may assemble the coefficients of those expressions into a matrix, or conversely read them from a matrix. Thus we will see that matrices (of suitable size) and linear maps determine one another uniquely, as long as we have **fixed a basis of V and fixed a basis of W**.

Notice how a matrix $A \in \mathrm{Mat}_{mn}$ is being used in two slightly different ways. In §5.3 the matrix A determined a map $\mathbb{R}^n \to \mathbb{R}^m$ simply by multiplication (5.1). We use A next to define a map $V \to W$ between two vector spaces of dimensions n and m respectively. Essentially this is the same map, but we need to be careful to distinguish them as we do our first analysis. We only work with finite-dimensional vector spaces from now on, because the matrices we use have finite size.

Proposition 5.19

Let V be a vector space with basis $v_1, \ldots, v_n$ and W be a vector space with basis $w_1, \ldots, w_m$. If $A = (a_{ij}) \in \mathrm{Mat}_{mn}$ is a matrix, then there is a linear map φ_A defined on the basis of V by

$$\begin{aligned} \varphi_A \colon V &\longrightarrow W \\ v_i &\mapsto a_{1i}w_1 + \ldots + a_{mi}w_m \end{aligned}$$

where the coefficients of $\varphi_A(v_i)$ with respect to the basis of W are the ith column of A.

Proof. There is nothing to prove. We have stated where we want the basis elements to map to, and therefore the linear map exists by Proposition 5.17(ii). □

Language 5.20

Fixing a basis of one or more vector spaces is an essential part of all of the calculations we are discussing, and so not surprisingly there are many different ways of saying that the basis has been fixed. We refer synonymously to a **chosen basis** or a **given basis** or a **specified basis** to indicate that we are working with a **fixed basis**.

Definition 5.21

This map is called **the linear map of A with respect to the given bases of V and W**. It depends on the two given bases, $v_1, \ldots, v_n$ in the domain V and $w_1, \ldots, w_m$ in the codomain W, and it is not defined without explicit reference to them both.

The reason to be so insistent on saying *given* bases, is that if you choose different bases, either in V or in W or in both, then the map φ_A will most likely be a very different map. This point is so important that φ_A should probably have some complicated notation such as

$$\varphi_{A;v_1,\ldots,v_n;w_1,\ldots,w_m}$$

but you can see why I don't do that. We just have to be very very careful that we all agree which bases are in play at the time.

Example

If $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$, and if we choose the standard basis $\underline{e}_1, \ldots, \underline{e}_n \in V$ and the (unusually named) standard basis $\underline{e}'_1, \ldots, \underline{e}'_m \in W$, then $\varphi_A = L_A$ is just the map we studied in §5.3.

Indeed, by Proposition 5.17(i) it is enough to check that they agree on a basis, and

$$\varphi_A(\underline{e}_i) = a_{1i}\underline{e}'_1 + \ldots + a_{mi}\underline{e}'_m = \begin{pmatrix} a_{1i} \\ \vdots \\ a_{mi} \end{pmatrix} = A\underline{e}_i = L_A(\underline{e}_i)$$

are both simply the ith column of A.

Remark

It is worth taking time to pick apart the relationship between the matrix A and the map φ_A. In the notation of Proposition 5.19, if $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$, then

$$\varphi_A(v) = \mu_1 w_1 + \ldots + \mu_m w_m$$

where the coefficients μ_j of the image are defined by

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

To see this, it is helpful to write out A fully (and when you write out particular examples for yourself, you may find it clearer to consider a matrix that is not square, say 2×3):

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

Expanding out $\varphi_A(v) = \lambda_1\varphi_A(v_1) + \ldots + \lambda_n\varphi_A(v_n)$ using the definition of each $\varphi_A(v_i)$, we see, for example, that the coefficient of w_1 is

$$\mu_1 = a_{11}\lambda_1 + a_{12}\lambda_2 + \ldots + a_{1n}\lambda_n$$

where the coefficients a_{ij} of the λ_j are the entries of the 1st **row** of A.

With that established, the key is now to observe that the ith **column** of A comprises the coefficients of the image $\varphi(v_i)$ of the ith element v_i of the basis of V when expressed in the basis $w_1, \ldots, w_m$ of W. With that view, we could imagine writing formally

$$\varphi_A \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = A^T \begin{pmatrix} w_1 \\ \vdots \\ w_m \end{pmatrix} \quad \text{where} \quad A^T = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{nm} \end{pmatrix}$$

meaning just that

$$\varphi_A(v_1) = a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m \tag{5.2}$$

and similarly for $\varphi_A(v_2), \ldots, \varphi_A(v_n)$.

In a sense, whether you use A or A^T is a delicate matter, but of course it is not at all, since one of the vectors has length n while the other has length m, so only one expression makes sense. However, when you use this machine, it may happen that A is square, so this syntactic assistance is missing and you need to be in total control of the coefficients.

Now the key result. It says that every linear map is of the form φ_A for any given choice of basis in domain and codomain, and the proof is simply to observe that a formula of the shape of (5.2) must hold for basis elements, and that that determines both the map on any element and also the matrix A.

Theorem 5.22

Let V be a vector space with basis $v_1, \ldots, v_n$ and W be a vector space with basis $w_1, \ldots, w_m$. If $\varphi\colon V \to W$ is a linear map, then there is a matrix $A \in \mathrm{Mat}_{mn}$ so that $\varphi = \varphi_A$ is the linear map of A with respect to these specified bases.

Definition 5.23

The matrix A in Theorem 5.22 is called **the matrix of the linear map φ with respect to the given bases** of V and W, and we say that φ **is represented by the matrix A with respect to the given bases**.

The matrix A of φ depends on the two specified bases and it is not defined without explicit reference to them both. (Sounds repetitive? It is worth repeating.)

Proof. Define the entries a_{ij} of a matrix by expressing $\varphi(v_i)$ in the basis $w_1, \ldots, w_m$:

$$\begin{aligned} \varphi(v_1) &= a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m \\ \varphi(v_2) &= a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m \\ &\vdots \\ \varphi(v_n) &= a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_m \end{aligned}$$

If we set $A = (a_{ij}) \in \mathrm{Mat}_{mn}$, then $\varphi(v_i) = \varphi_A(v_i)$, using the defining formula of $\varphi_A(v_i)$ from Proposition 5.19. Therefore $\varphi = \varphi_A$, as claimed. □

Commutative squares

There is an extremely convenient technology that handles a lot of the bureaucracy of the kind of results we have been proving. We can package up the previous statement into a statement about equality of compositions of maps, as follows.

Definition 5.24

A **commutative square** is a collection of four vector spaces and four linear maps

$$\begin{array}{ccc} U_1 & \xrightarrow{\varphi_1} & U_2 \\ {\scriptstyle \chi_1}\downarrow & & \downarrow{\scriptstyle \chi_2} \\ V_1 & \xrightarrow{\varphi_2} & V_2 \end{array}$$

so that $\chi_2 \circ \varphi_1 = \varphi_2 \circ \chi_1$.

That is, a commutative square records two different compositions $U_1 \to U_2 \to V_2$ and $U_1 \to V_1 \to V_2$ that give the same linear map $U_1 \to V_2$.

We can package up all the results of this section in the following elegant corollary. It gives a nice point of view of the situation: the thing we are interested in is the map $\varphi\colon V \to W$, but to calculate anything we choose bases in V and W and present φ in coordinates as $L_A\colon \mathbb{R}^n \to \mathbb{R}^m$. In this view, L_A is a kind of computable model of φ, and the (invertible) coordinate maps $\chi_{\mathcal{B}}$ and $\chi_{\mathcal{B}'}$ translate between the two.

Corollary 5.25

Let V be a vector space with basis $\mathcal{B} = \{v_1, \ldots, v_n\}$ and W be a vector space with basis $\mathcal{B}' = \{w_1, \ldots, w_m\}$. If $\varphi\colon V \to W$ is a linear map, then $\varphi = \varphi_A$ for some matrix $A \in \mathrm{Mat}_{mn}$ and there is a commutative square of linear maps

$$\begin{array}{ccc} V & \xrightarrow{\varphi_A} & W \\ {\scriptstyle \chi_{\mathcal{B}}}\downarrow & & \downarrow{\scriptstyle \chi_{\mathcal{B}'}} \\ \mathbb{R}^n & \xrightarrow{L_A} & \mathbb{R}^m \end{array}$$

Proof. The maps are all defined: φ_A is in Proposition 5.19, L_A in (5.1) in §5.3, and $\chi_{\mathcal{B}}$ and $\chi_{\mathcal{B}'}$ in Proposition 4.27. The claim is simply that $\chi_{\mathcal{B}'} \circ \varphi_A = L_A \circ \chi_{\mathcal{B}}$.

Let $v = \lambda_1 v_1 + \ldots + \lambda_n v_n \in V$. On the one hand,

$$\chi_{\mathcal{B}}(v) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \quad \text{and so} \quad L_A(\chi_{\mathcal{B}}(v)) = A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

On the other hand,

$$\varphi_A(v) = \mu_1 w_1 + \ldots + \mu_m w_m$$

where the coefficients μ_j of the image are defined by

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \quad \text{and so} \quad \chi_{\mathcal{B}'}(\varphi_A(v)) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \end{pmatrix} = A \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

These two are equal, as claimed. □

The real benefit of this book-keeping tool is that it packages up the proofs of many statements that would be fiddly or repetitive to expand out in coordinates. For example, this association between a linear map and a matrix is natural with respect to composition.

Corollary 5.26

Let $\psi\colon U \to V$ and $\varphi\colon V \to W$ be linear maps between vector spaces. Let $\mathcal{B}$ be a basis of U, $\mathcal{B}'$ a basis of V, and $\mathcal{B}''$ a basis of W of sizes ℓ, n, m respectively.

Suppose ψ is represented by a matrix $A \in \mathrm{Mat}_{n\ell}$ with respect to these bases and φ is represented by $A' \in \mathrm{Mat}_{mn}$. Then there is a commutative square

$$\begin{array}{ccc} U & \xrightarrow{\varphi \circ \psi} & W \\ \downarrow{\scriptstyle \chi_{\mathcal{B}}} & & \downarrow{\scriptstyle \chi_{\mathcal{B}''}} \\ \mathbb{R}^\ell & \xrightarrow{L_{A'} \circ L_A = L_{A'A}} & \mathbb{R}^m \end{array}$$

That is, the composition $\varphi \circ \psi$ is represented by the product $A'A \in \mathrm{Mat}_{m\ell}$ with respect to the bases $\mathcal{B}$ of U and $\mathcal{B}''$ of W.

There is nothing to prove. The point is that you can glue commutative squares together whenever they have a map in common, and then you can walk around the resulting diagram following the arrows in any way you like and you always get the same answer. More precisely, it is immediate that if

$$\begin{array}{ccc} U_1 & \xrightarrow{\varphi_1} & U_2 \\ \downarrow{\scriptstyle \chi_1} & & \downarrow{\scriptstyle \chi_2} \\ V_1 & \xrightarrow{\varphi_2} & V_2 \end{array} \qquad \begin{array}{ccc} U_2 & \xrightarrow{\varphi_3} & U_3 \\ \downarrow{\scriptstyle \chi_2} & & \downarrow{\scriptstyle \chi_3} \\ V_2 & \xrightarrow{\varphi_4} & V_3 \end{array}$$

are two commutative squares (with the visible coincidence $\chi_2\colon U_2 \to V_2$) then

$$\begin{array}{ccc} U_1 & \xrightarrow{\varphi_3 \circ \varphi_1} & U_3 \\ {\scriptstyle \chi_1}\downarrow & & \downarrow{\scriptstyle \chi_3} \\ V_1 & \xrightarrow{\varphi_4 \circ \varphi_2} & V_3 \end{array}$$

is also a commutative square.

5.5 The Rank–Nullity Formula

There are two basic counting theorems of linear algebra: the Rank–Nullity Formula and the Dimension Formula. Their value throughout mathematics is impossible to overstate: they are used absolutely everywhere to count the number of parameters in a problem (that is, the dimensions of subspaces).

Definition 5.27

Let $\varphi\colon V \to W$ be a linear map. The **rank of** φ is $\operatorname{rk}\varphi = \dim \operatorname{Im}\varphi$. The **nullity of** φ is $\operatorname{nullity}\varphi = \dim\ker\varphi$.

Theorem 5.28 (Rank–Nullity formula)

Let $\varphi\colon V \to W$ be a linear map, and suppose that V is finite dimensional. Then

$$\dim \operatorname{Im}\varphi + \dim\ker\varphi = \dim V$$

The theorem gets its name, because this equation is $\operatorname{rk}\varphi + \operatorname{nullity}\varphi = \dim V$.

As this is so important, we give multiple proofs – though you might regard them as the same idea cast into the various different languages we have developed for speaking about the solutions of linear equations.

The first version is the vanilla one that only uses the technology of bases. You should think of it as easy: adjoin a basis of the kernel to a set that maps bijectively to a basis of the image, and prove that this is a basis of V. The work is all in the last clause, and it is easy once you are well practiced.

Proof. Both $\ker\varphi \subset V$ and $\operatorname{Im}\varphi \subset W$ are finite dimensional because V is. Pick bases $u_1, \ldots, u_s$ of $\ker\varphi$ and $w_1, \ldots, w_r$ of $\operatorname{Im}\varphi$. Since each w_i lies in the image, there are $v_1, \ldots, v_r \in V$ which satisfy $\varphi(v_i) = w_i$ for $i = 1, \ldots, r$.

We claim that $\mathcal{B} = \{u_1, \ldots, u_s, v_1, \ldots, v_r\}$ is a basis of V. This will complete the proof since $s = \dim\ker\varphi$ and $r = \dim\operatorname{Im}\varphi$. (It is now quicker to do this yourself than to read the rest.)

To show that $\mathcal{B}$ spans V, consider any $v \in V$. Set $w = \varphi(v) \in \operatorname{Im}\varphi$. Since $w_1, \ldots, w_r$ is a basis of $\operatorname{Im}\varphi$, there exist scalars $\mu_1, \ldots, \mu_r$ so that

$$w = \mu_1 w_1 + \ldots + \mu_r w_r$$

The key is to note that $v' = \mu_1 v_1 + \ldots + \mu_r v_r$ (using v_i in place of w_i) also maps to w, by the linearity of φ:

$$\begin{aligned}\varphi(v-v') &= \varphi(v) - \mu_1\varphi(v_1) - \ldots - \mu_r\varphi(v_r) \\ &= w - \mu_1 w_1 - \ldots - \mu_r w_r \\ &= 0_W\end{aligned}$$

so that $v - v' \in \ker\varphi$. Therefore there exist scalars $\lambda_1, \ldots, \lambda_s$ so that

$$v - v' = \lambda_1 u_1 + \ldots + \lambda_s u_s$$

Rearranging that gives $v = \lambda_1 u_1 + \ldots + \lambda_s u_s + \mu_1 v_1 + \ldots + \mu_r v_r$ so that $v \in \langle\mathcal{B}\rangle$, as required.

To show that $\mathcal{B}$ is linearly independent, suppose there are scalars λ_i, μ_j such that

$$\lambda_1 u_1 + \ldots + \lambda_s u_s + \mu_1 v_1 + \ldots + \mu_r v_r = 0_V \tag{5.3}$$

Applying φ, and noting that $\varphi(u_i) = 0_W$ and $\varphi(v_j) = w_j$, we have

$$0_W + \mu_1 w_1 + \ldots + \mu_r w_r = 0_W$$

Since $w_1, \ldots, w_r$ are linearly independent, being a basis of $\operatorname{Im}\varphi$, we have $\mu_1 = \ldots = \mu_r = 0$. Equation (5.3) now reads

$$\lambda_1 u_1 + \ldots + \lambda_s u_s = 0_V$$

Since $u_1, \ldots, u_r$ are linearly independent, being a basis of $\ker\varphi$, we have $\lambda_1 = \ldots = \lambda_s = 0$. So all coefficients of (5.3) are necessarily zero, and so $\mathcal{B}$ is linearly independent. □

Example

Consider $L_A \colon \mathbb{R}^3 \to \mathbb{R}^2$ where A is the matrix

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

Then

$$\operatorname{Im} L_A = \operatorname{Colspan} A = \left\langle \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\rangle \quad \text{has} \quad \operatorname{rk} L_A = \dim \operatorname{Im} L_A = 1$$

and

$$\ker L_A = \left\{ \begin{pmatrix} 3s-5t \\ 2s \\ 2t \end{pmatrix} \,\middle|\, s,t \in \mathbb{R} \right\} \quad \text{has} \quad \operatorname{nullity} L_A = \dim\ker L_A = 2$$

and indeed

$$\operatorname{rk} L_A + \operatorname{nullity} L_A = 1 + 2 = 3 = \dim \mathbb{R}^3$$

The map L_A squishes the whole of the 3-dimensional $\mathbb{R}^3$ down to a 1-dimensional line, and to do so it must pay the price of killing the 2-dimensional kernel.

The Smith normal form S of A, which is computed as $S = EAF$ for invertible matrices E and F, is simply

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

It is almost instant to calculate that

$$\dim \operatorname{Im} L_S = \dim \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle = 1 \quad \text{and} \quad \dim\ker L_S = \dim \left\langle \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\rangle = 2$$

and since the maps L_E and L_F provide isomorphisms between $\operatorname{Im} L_A$ and $\operatorname{Im} L_S$, and $\ker L_S$ and $\ker L_A$, respectively, we could have used this calculation in place of the harder one using A.

The second proof of the Rank–Nullity Theorem is more elementary, in that it translates the problem into what we know about matrices as hinted in the example above.

Proof. Let $n = \dim V$ and $m = \dim W$. Pick bases for V and W, and consider the matrix $A \in \mathrm{Mat}_{mn}$ that represents φ with respect to them. Compute the Smith normal form EAF of A, where $E \in \mathrm{Mat}_{mm}$ and $F \in \mathrm{Mat}_{nn}$ are invertible matrices. Let r be the rank of the Smith normal form. Then by Proposition 5.14(iii)

$$\operatorname{rk} \varphi = \dim \operatorname{Im} L_A = \dim \operatorname{Im} L_{EAF} = r$$

and

$$\operatorname{nullity} \varphi = \dim \ker L_A = \dim \ker L_{EAF} = n - r$$

and the result follows since $r + (n - r) = n$. □

Some more examples.

Example

Let $U = \{(x, y, z)^T \in \mathbb{R}^3 \mid 5x + 3y - 7z = 0\}$. What is $\dim U$? Until now, our method for tackling this would probably have been to find a basis for U. That's fine, but watch this instead.

We may use the equation that defines U to define a linear map:

$$\begin{aligned} \varphi \colon V &\to W \qquad\qquad \text{where } V = \mathbb{R}^3 \text{ and } W = \mathbb{R} \\ \begin{pmatrix} x \\ y \\ z \end{pmatrix} &\mapsto 5x + 3y - 7z \end{aligned}$$

that is, $\varphi = L_A$ for the matrix $A = \begin{pmatrix} 5 & 3 & -7 \end{pmatrix}$. You notice at once that $U = \ker \varphi$.

The key is that it is easier to calculate $\operatorname{Im} \varphi$ (or even just $\dim \operatorname{Im} \varphi$) than $\ker \varphi$ here. For example, $\operatorname{Colspan} A = \langle 1 \rangle = W$, using the RCEF $(1\ 0\ 0)$ of A, say. (Or, if you prefer, we could observe that $\varphi(\underline{e}_1) = 5$, and 5 is a basis of W, albeit an unusual one, so again we see $\operatorname{Im} \varphi = W$. Or you could say that $\operatorname{Im} \varphi$ is a subspace of W, and since $\dim W = 1$ it is either $\{0_W\}$ or W, and it's clearly not the former since you easily find a vector with nonzero image.)

In any case, a moment's thought tells you that $\dim \operatorname{Im} \varphi = 1$, and so by the rank–nullity formula

$$\dim U = \dim \ker \varphi = \dim V - \dim \operatorname{Im} \varphi = 3 - 1 = 2$$

Of course you could have used the RREF of A to observe that the solution set of $5x+3y-7z = 0$ has two parameters, $\lambda, \mu \in \mathbb{R}$, and that a general solution (clearing denominators) is $(x, y, z)^T = (-3\lambda + 7\mu, 5\lambda, 5\mu)^T$. With that we quickly obtain a linearly independent pair $(-3, 5, 0)^T$ and $(7, 0, 5)^T$, so that $\dim U \geq 2$. Of course $U \neq \mathbb{R}^3$ (just find any $(x, y, z)^T$ that does not satisfy the equation), so $\dim U < \dim \mathbb{R}^3 = 3$ by Proposition 4.20, and so we have proved that $\dim U = 2$, and moreover we have constructed a basis of U. But that feels like more work.

Example

What is the dimension of $U = \{f \in \mathbb{R}[x]_{\le 100} \mid f(3) = 0 \text{ and } df/dx(3) = 0\}$? Once again, we do not immediately know a basis of U, but we can understand U as the kernel of a linear map and apply the rank–nullity formula.

Let $V = \mathbb{R}[x]_{\le 100}$. Of course $\dim V = 101$, since we know a basis $1, x, \dots, x^{100}$. Consider the linear map

$$\begin{aligned} \varphi\colon V &\to W \qquad\qquad \text{where } W = \mathbb{R}^2 \\ f &\mapsto \begin{pmatrix} f(3) \\ df/dx(3) \end{pmatrix} \end{aligned}$$

This is rigged up so that $U = \ker \varphi$.

Clearly φ is surjective: for example, $\varphi(1) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\varphi(x) = \begin{pmatrix} 3 \\ 1 \end{pmatrix}$ form a basis of W. Therefore

$$\dim U = \dim \ker \varphi = \dim V - \dim \operatorname{Im} \varphi = 101 - 2 = 99$$

That's great: question answered. But what's more is that, now that we know this dimension, it is much easier to compute a basis of U if we need it: by the Sifting Lemma 4.15 or Corollary 4.22 we just need to find 99 linearly independent elements and do not need to check that they span. Clearly $f = (x-3)^2$ lies in U, so the linearly independent set $\mathcal{B} = \{x^i f \mid i = 0, \dots, 98\} \subset U$ is a basis of U. (Now try to imagine proving that $\mathcal{B}$ spans U without using rank–nullity.)

Once you appreciate this example, you see that more generally conditions on a vector space of (suitably differentiable) functions $f\colon \mathbb{R} \to \mathbb{R}$ such as

$$\frac{d^i}{dx^i} f(p) = 0 \qquad \text{for some } i \ge 0 \text{ and some (fixed!) } p \in \mathbb{R}$$

are linear, and so imposing them tends to reduce the dimension of the space of functions by 1. The only issue is that if you impose several such conditions you need to know that they are linearly independent: in our example, that is equivalent to the surjectivity of φ. You should add this idea to your set of ninja tools.

As a final thought, this gives you some intuition why (homogeneous) differential equations are so powerful and so hard to solve: they impose conditions like that at *every* point $p \in \mathbb{R}$: that is, they impose *uncountably many* linear conditions!

5.6 The Dimension Formula and Complements

We could have done this sooner: the moment we learned we could extend a linearly independent set to a basis, we owned this.

Theorem 5.29 (Dimension Formula)

Let $U_1, U_2 \subset V$ be two finite-dimensional subspaces of a vector space V. Then

$$\dim U_1 + \dim U_2 = \dim(U_1 + U_2) + \dim(U_1 \cap U_2).$$

Proof. Let $W = U_1 \cap U_2$, which is a finite-dimensional subspace of V. Pick a basis $\mathcal{B}_W = \{w_1, \dots w_t\}$ of W.

Since $W \subset U_1$, by Corollary 4.16 we may extend $\mathcal{B}_W$ to a basis $\{w_1, \dots w_t, u_1, \dots, u_s\}$ of U_1. Similarly, since $W \subset U_2$, we may extend $\mathcal{B}_W$ to a basis $\{w_1, \dots w_t, v_1, \dots, v_r\}$ of U_2.

We claim that $\mathcal{B} = \{u_1, \dots, u_s, v_1, \dots, v_r, w_1, \dots, w_t\}$ is a basis of the span $U_1 + U_2$. This will complete the proof since

$$s + t = \dim U_1 \qquad r + t = \dim U_2 \qquad t = \dim U_1 \cap U_2 \quad \text{and} \quad r + s + t = \dim U_1 + U_2$$

(It is now quicker to do this yourself than to read the rest.)

It is immediate that $\mathcal{B}$ spans: any element of $U_1 + U_2$ is of the form $p_1 + p_2$ with $p_i \in U_i$, and each of these p_i may be written as a linear combination of $\mathcal{B}$.

To show $\mathcal{B}$ is linearly independent, suppose there are scalars λ_i, μ_j, ν_k such that

$$\lambda_1 u_1 + \ldots + \lambda_s u_s + \mu_1 v_1 + \ldots + \mu_r v_r + \nu_1 w_1 + \ldots \nu_t w_t = 0_V \tag{5.4}$$

Rearranging this determines an element $p \in V$ defined by

$$p = \lambda_1 u_1 + \ldots + \lambda_s u_s + \nu_1 w_1 + \ldots + \nu_t w_t = -(\mu_1 v_1 + \ldots + \mu_r v_r)$$

which the two equal expressions show lies in both U_1 and U_2. Thus $p \in U_1 \cap U_2$, and so there are scalars τ_i so that

$$p = \tau_1 w_1 + \ldots + \tau_t w_t$$

Thus, subtracting two of the expressions for p, we have

$$\lambda_1 u_1 + \ldots + \lambda_s u_s + (\nu_1 - \tau_1) w_1 + \ldots + (\nu_t - \tau_t) w_t = 0_V$$

In particular, since the u_i, w_k form a basis of U_1 all the $\lambda_i = 0$. Now since the v_j, w_k form a basis of U_2, the equation (5.4) with all $\lambda_i = 0$ shows that all $\mu_j = 0$ and all $\nu_k = 0$, as required. ☐

Example

Let $V = \mathbb{R}^4$ and consider two subspaces $U_1 = \langle u_1, u_2, u_3 \rangle$, where

$$u_1 = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \quad u_2 = \begin{pmatrix} -2 \\ 0 \\ 5 \\ -3 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 6 \\ 1 \\ -4 \\ 3 \end{pmatrix}$$

which are clearly linearly independent, and $U_2 = \langle \underline{e}_1, \underline{e}_2 \rangle$, the span of the first two standard basis elements.

It is quickly clear that $U_1 + U_2 = V$: for example $\underline{e}_3$ is $u_2 + u_3$ minus suitable multiples of $\underline{e}_1$ and $\underline{e}_2$, and then it is easy to write $\underline{e}_4$ as a combination of $u_1, \underline{e}_2, \underline{e}_3$. Therefore by the Dimension Formula

$$\dim U_1 \cap U_2 = 3 + 2 - 4 = 1$$

which is not immediately clear from the given vectors.

The idea of a complementary subspace to a given subspace, or *complement*, is natural, and gives another point of view on the proof of the rank–nullity formula.

Definition 5.30

Let V be a vector space and $U \subset V$ a subspace. Then a subspace $U' \subset V$ is called a **complement to** U if and only if $V = U + U'$ and $U \cap U' = \{0_V\}$.

Example

For example, the y-z plane is a complement to the x-axis in $\mathbb{R}^3$.

The following dimension count is immediate from the Dimension Formula 5.29 since $\dim\{0_V\} = 0$.

Corollary 5.31

If U' is a complement to $U \subset V$, then $\dim V = \dim U + \dim U'$.

Remark

In this case, people sometimes write $V = U \oplus U'$ and refer to it as an **internal direct sum** to distinguish it from the direct sum of Definition 4.23. This is a slightly stronger notion: since both U and U' lie inside V, the vector space V is equal to this sum, not merely isomorphic to it, and any $v \in V$ may be written as $v = u + u'$ for unique $u \in U$ and $u' \in U'$. This is not worth any fuss. Of course the two notions of $U \oplus U'$ are isomorphic to each other in any case.

Lemma 5.32

If U' is a complement to $U \subset V$, then for every $v \in V$ there exist unique $u \in U$ and $u' \in U'$ so that $v = u + u'$.

Proof. There is such an expression since $V = U + U'$. If $v = u_2 + u_2'$ with $u_2 \in U$ and $u_2' \in U'$ is another, then subtracting the two expressions gives

$$U \ni u - u_2 = u_2' - u' \in U'$$

so $u - u_2 = u_2' - u' = 0_V$, as $U \cap U' = \{0_V\}$, and the two expressions are the same. □

Lemma 5.33

Let V be a finite-dimensional vector space and $U \subset V$ a subspace. Then there exists a complement $U' \subset V$ to U.

Proof. Let $u_1, \ldots, u_s$ be a basis of U. By Corollary 4.16, there are elements $v_1, \ldots, v_r$ of V so that $u_1, \ldots, u_s, v_1, \ldots, v_r$ is a basis of V. Let $U' = \langle v_1, \ldots, v_r\rangle$.

Clearly $U + U' = V$. If $w \in U \cap U'$, then

$$w = \lambda_1 u_1 + \ldots + \lambda_s u_s \quad \text{and} \quad w = \mu_1 v_1 + \ldots + \mu_r v_r$$

for scalars λ_i, μ_j. Subtracting these two expressions gives

$$\lambda_1 u_1 + \ldots + \lambda_s u_s - \mu_1 v_1 - \ldots - \mu_r v_r = 0_V$$

so by linear independence all the coefficients are zero, and so $w = 0_V$. Thus $U \cap U' = \{0_V\}$ and U' is a complement to U. □

The third proof of the Rank–Nullity Theorem uses complements.

Proof. Since V is finite dimensional, so is $U = \ker\varphi \subset V$ by Proposition 4.20 Let $U' \subset V$ be a complement to U. Define the (evidently linear) map

$$\begin{aligned} \psi\colon U' &\to \operatorname{Im}\varphi \\ u &\mapsto \varphi(u) \end{aligned}$$

which is simply the restriction of the map φ to U'.

We claim that ψ is an isomorphism $U' \to \operatorname{Im}\varphi$. This will complete the proof since then by Corollary 5.31 and Corollary 5.8

$$\dim V = \dim U + \dim U' = \dim\ker\varphi + \dim\operatorname{Im}\varphi$$

Firstly, ψ is injective since $\psi(u) = 0_W$ for $u \in U'$ only if also $u \in \ker\varphi$; but $U' \cap \ker\varphi = \{0_V\}$, since U' is a complement to $\ker\varphi$, and so we have: $\psi(u) = 0_W$ if and only if $u = 0_V$, as required.

It remains to show that ψ is surjective. Suppose $w \in \operatorname{Im}\varphi$. Then there is $v \in V$ so that $\varphi(v) = w$. Write $v = u + u'$ with $u \in U$ and $u' \in U'$. Then computing the image of $u' = v - u$ gives

$$\psi(u') = \varphi(v - u) = \varphi(v) - \varphi(u) = w - 0_V = w$$

so $w \in \operatorname{Im}\psi$, as required. □

5.7 Change of basis and equivalent matrices

Recall the change of basis calculation of §3.4. The situation is this. We have an n-dimensional vector space V. Any choice of basis $\mathcal{A} = \{v_1, \ldots, v_n\}$ of V expresses any element $v \in V$ as a unique linear combination

$$v = \lambda_1 v_1 + \ldots + \lambda_n v_n$$

and by presenting the coefficients λ_i as a column vector we describe an isomorphism

$$\begin{aligned} \chi_{\mathcal{A}}\colon V &\to \mathbb{R}^n \\ v = \lambda_1 v_1 + \ldots + \lambda_n v_n &\mapsto \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \end{aligned}$$

and which we call **writing v in coordinates with respect to the basis** $v_1, \ldots, v_n$.

Remark

We may even turn this around. If you have an isomorphism $\chi\colon V \to \mathbb{R}^n$, then you automatically know a basis of V, viz. the preimage of the standard basis, $\mathcal{A} = \left\{\chi^{-1}(\underline{e}_1), \ldots, \chi^{-1}(\underline{e}_n)\right\}$.

In other words, whether somebody gives you a basis of V or they give you an isomorphism of V with $\mathbb{R}^n$ are two sides of the same coin: they are exactly equivalent data.

> Note again that one of the features of $\mathbb{R}^n$ as a vector space is that is has a favoured basis, namely the standard basis $\underline{e}_1, \ldots, \underline{e}_n$. For other vector spaces, you may need to do a lot of work to find a basis at all, let alone one that is so natural that everyone would agree it's the best.

The issue described in §3.4 is that if $\mathcal{A}'$ is a different basis of V, then we have a different isomorphism $\chi_{\mathcal{A}'} \colon V \to \mathbb{R}^n$, and we may need to know how to translate between these two: if you know the column vector $\chi_{\mathcal{A}}(v)$, what is the column vector $\chi_{\mathcal{A}'}(v)$?

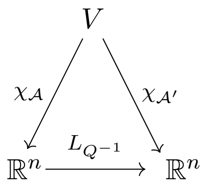
This is resolved by a straightforward procedure that you can simply learn: if

$$\chi_{\mathcal{A}}(v) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} \quad \text{and} \quad \chi_{\mathcal{A}'}(v) = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$$

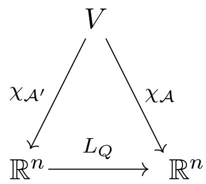
then

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = Q^{-1} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

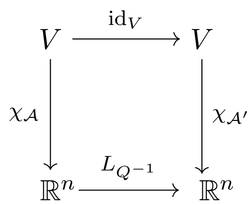
where Q is the (invertible) matrix whose columns are the coefficients of the elements of $\mathcal{A}'$ when expressed in the basis $\mathcal{A}$. We may express this idea using maps as $\chi_{\mathcal{A}'} = L_{Q^{-1}} \circ \chi_{\mathcal{A}}$ or, flipping L_Q to the other side, $L_Q \circ \chi_{\mathcal{A}'} = \chi_{\mathcal{A}}$, and we may visualise this in either of the following diagrams



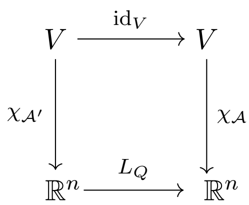
or flipping it over



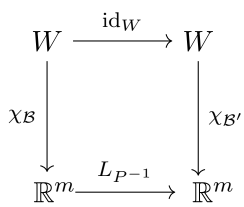
It may seem ridiculous, but using the identity map $\mathrm{id}_V \colon V \to V$, we may express this as a commutative square:



or flipping it over



Now suppose we have a map $\varphi \colon V \to W$ and also two bases $\mathcal{B}$ and $\mathcal{B}'$ of W: these two bases have a change of basis matrix P that translates between the two coordinate maps by



Suppose that $\varphi = \varphi_A$ for some $A \in \mathrm{Mat}_{mn}$ with respect to the bases $\mathcal{A}$ and $\mathcal{B}$. In other words, we have a commutative square

$$\begin{array}{ccc} V & \xrightarrow{\varphi_A} & W \\ \downarrow{\scriptstyle \chi_{\mathcal{A}}} & & \downarrow{\scriptstyle \chi_{\mathcal{B}}} \\ \mathbb{R}^n & \xrightarrow{L_A} & \mathbb{R}^m \end{array}$$

The question is, which matrix represents φ with respect to the bases $\mathcal{A}'$ and $\mathcal{B}'$? We may glue the commutative squares above into one big diagram

$$\begin{array}{ccccccc} V & \xrightarrow{\mathrm{id}_V} & V & \xrightarrow{\varphi_A} & W & \xrightarrow{\mathrm{id}_V} & W \\ \downarrow{\scriptstyle \chi_{\mathcal{A}'}} & & \downarrow{\scriptstyle \chi_{\mathcal{A}}} & & \downarrow{\scriptstyle \chi_{\mathcal{B}}} & & \downarrow{\scriptstyle \chi_{\mathcal{B}'}} \\ \mathbb{R}^n & \xrightarrow{L_Q} & \mathbb{R}^n & \xrightarrow{L_A} & \mathbb{R}^m & \xrightarrow{L_{P^{-1}}} & \mathbb{R}^m \end{array}$$

Reading around the edge we see a commutative square

$$\begin{array}{ccc} V & \xrightarrow{\varphi_A} & W \\ \downarrow{\scriptstyle \chi_{\mathcal{A}'}} & & \downarrow{\scriptstyle \chi_{\mathcal{B}'}} \\ \mathbb{R}^n & \xrightarrow{L_B} & \mathbb{R}^m \end{array}$$

where the bottom map L_B is represented by the matrix $B \in \mathrm{Mat}_{mn}$ that we seek. But of course by reading the bottom line of the 3 squares we see that $L_B = L_{P^{-1}} \circ L_A \circ L_Q = L_{P^{-1}AQ}$ so that so the matrix is $B = P^{-1}AQ \in \mathrm{Mat}_{mn}$.

The crucial thing to remember is the formula we have just derived, and which we summarise as follows.

Theorem 5.34 (Change of basis formula for linear maps)

Suppose $\varphi\colon V \to W$ is a linear map between finite-dimensional vector spaces. Suppose we have:

(i) bases $\mathcal{A} = \{v_1, \ldots, v_n\}$ of V and $\mathcal{B} = \{w_1, \ldots, w_m\}$ of W, so that with respect to these bases $\varphi = \varphi_A$ is represented by a matrix $A \in \mathrm{Mat}_{mn}$.

(ii) an alternative basis $\mathcal{A}' = \{v'_1, \ldots, v'_n\}$ of V with change of basis matrix Q for which $\chi_{\mathcal{A}} = L_Q \circ \chi_{\mathcal{A}'}$.

(iii) an alternative basis $\mathcal{B}' = \{w'_1, \ldots, w'_m\}$ of W with change of basis matrix P for which $\chi_{\mathcal{B}} = L_P \circ \chi_{\mathcal{B}'}$.

Then with respect to the bases $\mathcal{A}'$ of V and $\mathcal{B}'$ of W, the map φ is represented by the matrix $B = P^{-1}AQ$.

This motivates the following definition.

Definition 5.35

Two matrices $A, B \in \mathrm{Mat}_{mn}$ are called **equivalent** if and only if there exist invertible matrices $P \in \mathrm{Mat}_{mm}$ and $Q \in \mathrm{Mat}_{nn}$ so that $B = P^{-1}AQ$.

This is of course an equivalence relation on the set Mat_{mn} of all $m \times n$ matrices.

Example

Did you ever use the notation $\underline{i}, \underline{j}, \underline{k}$ for three elements of a basis $\mathcal{A}$ of $V = \mathbb{R}^3$? Let's do it. While we're at it, let's double up and use $\underline{i}, \underline{j}$ for the elements of a basis $\mathcal{B}$ of $W = \mathbb{R}^2$. That is, elements of V are quantities of the form $a\underline{i} + b\underline{j} + c\underline{k}$, and analogously for W.

Consider the linear map $\varphi\colon V \to W$ given by

$$a\underline{i} + b\underline{j} + c\underline{k} \mapsto (a + 2c)\underline{i} + (2a + 3b + c)\underline{j}$$

The coordinate maps with respect to these bases are

$$\chi_{\mathcal{A}}(a\underline{i} + b\underline{j} + c\underline{k}) = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \quad \text{and} \quad \chi_{\mathcal{B}}(d\underline{i} + e\underline{j}) = \begin{pmatrix} d \\ e \end{pmatrix}$$

and the map φ has coefficients from the matrix

$$A = \begin{pmatrix} 1 & 0 & 2 \\ 2 & 3 & 1 \end{pmatrix}$$

Thus all together we have the commutative square

$$\begin{array}{ccc} V & \xrightarrow{\varphi_A} & W \\ {\scriptstyle \chi_{\mathcal{A}}}\downarrow & & \downarrow{\scriptstyle \chi_{\mathcal{B}}} \\ \mathbb{R}^3 & \xrightarrow{L_B} & \mathbb{R}^2 \end{array}$$

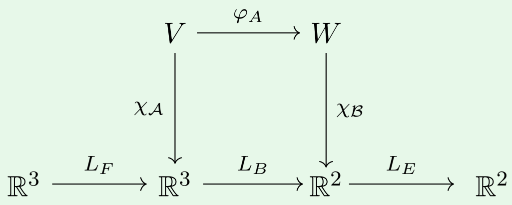
For a moment, just think about how you would calculate the Smith normal form of A. You might start by subtracting twice the top row from the bottom, and then scaling the bottom row by $1/3$: that is, premultiply A by $E = E_2E_1$

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

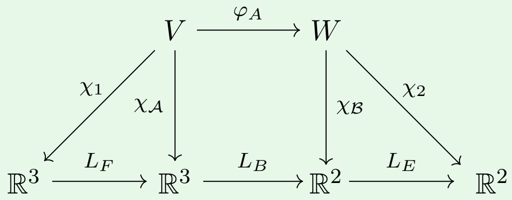
Then you might subtract twice the first column from the third and add the second to the third: that is, postmultiply the matrix above by $F = F_1F_2$

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

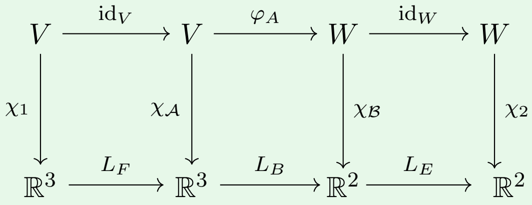
The result is the Smith normal form $S = EAF$. We can picture those compositions in coordinates on the bottom line of the picture above:



Now since $\chi_{\mathcal{A}}$ and L_F are isomorphisms (and similarly $\chi_{\mathcal{B}}$ and L_E) we may include another isomorphism $\chi_1 = L_{F^{-1}} \circ \chi_{\mathcal{A}}$ (and similarly $\chi_2 = L_E \circ \chi_{\mathcal{B}}$) in the picture



or if you prefer your triangles to be square



Of course χ_1 and χ_2 are the coordinate maps of some basis. Let's work those out.

Let $\underline{e}_1, \ldots, \underline{e}_3$ be the standard basis of $\mathbb{R}^3$. We compute its preimages under χ_1, that is

$$\begin{aligned} \chi_{\mathcal{A}}^{-1} \circ L_F(\underline{e}_1) &= \underline{i} \\ \chi_{\mathcal{A}}^{-1} \circ L_F(\underline{e}_2) &= \underline{j} \\ \chi_{\mathcal{A}}^{-1} \circ L_F(\underline{e}_3) &= -2\underline{i} + \underline{j} + \underline{k} \in V \end{aligned}$$

and similarly for the standard basis of $\mathbb{R}^2$, whose preimages under χ_2 are

$$\underline{i} + 2\underline{j} \quad \text{and} \quad 3\underline{j} \in W \quad \text{using } E^{-1} = \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$$

Now rewriting exactly the same map $\varphi = \mathrm{id}_W \circ \varphi \circ \mathrm{id}_V \colon V \to W$ with respect to these two bases gives

$$\begin{aligned} \lambda_1 \underline{i} + \lambda_2 \underline{j} + \lambda_3(-2\underline{i} + \underline{j} + \underline{k}) &= (\lambda_1 - 2\lambda_3)\underline{i} + (\lambda_2 + \lambda_3)\underline{j} + \lambda_3 \underline{k} \\ &\mapsto ((\lambda_1 - 2\lambda_3) + 2\lambda_3)\underline{i} + (2(\lambda_1 - 2\lambda_3) + 3(\lambda_2 + \lambda_3) + \lambda_3)\underline{k} \\ &= \lambda_1 \underline{i} + (2\lambda_1 + 3\lambda_2)\underline{j} \\ &= \lambda_1(\underline{i} + 2\underline{j}) + \lambda_2(3\underline{j}) \end{aligned}$$

which, reading in one go (with respect to the two new bases) as

$$\lambda_1 \underline{i} + \lambda_2 \underline{j} + \lambda_3(-2\underline{i} + \underline{j} + \underline{k}) \mapsto \lambda_1(\underline{i} + 2\underline{j}) + \lambda_2(3\underline{j})$$

is the map in coordinates

$$\begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} \mapsto \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix}$$

that corresponds to the matrix

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

Ta dah! The conceptually simple Smith normal form calculation that you are good at is in fact computing smarter bases in which to present the linear map as a matrix.

We can express some of what we know about equivalence of matrices and row-and-column reduction in a rather gratuitous way. Write $A \sim B$ for matrices $A, B \in \mathrm{Mat}_{mn}$ if and only if A is equivalent to B.

Proposition 5.36

Equivalence of matrices $A \sim B$ is an equivalence relation on Mat_{mn}. Moreover, in each equivalence class there is a unique matrix in Smith normal form.

Remark

It is absolutely brilliant that any linear map may be represented by a matrix. It is not a surprise that if we choose a different basis in the domain or codomain or both, then we get a different matrix. At first sight, we may regard this as a fiddly problem. But we derive more profit by thinking of it the other way round: this opens up the possibility that we may choose bases in domain and codomain so that the matrix is particularly useful or elegant or satisfies whatever property we like.

This is the fundamental question when representing maps by matrices: can you choose bases cleverly so that the matrix better suits your calculation. The Smith normal form is the first result of this type: for any matrix A we may choose invertible matrices E and F so that $S = EAF$ is in Smith normal form, and we regard $Q = F$ and $P = E^{-1}$ as the change of basis matrices.

5.8 The vector space $\mathrm{Hom}(U, V)$

We close this chapter by considering the set of all maps $U \to V$ for fixed finite-dimensional vector spaces U and V. There is essentially nothing new here, except the language. This set is denoted

$$\mathrm{Hom}(U, V) = \{\varphi \colon U \to V \mid \varphi \text{ is a linear map}\}$$

In fact, this set is a vector space in its own right, as follows. For $\varphi, \psi \in \mathrm{Hom}(U, V)$ and $\alpha \in \mathbb{R}$, we have linear maps $\varphi + \psi$ and $\alpha\varphi$ defined in the usual way: for any $v \in V$ we have

$$\begin{aligned} (\varphi + \psi)(u) &= \varphi(u) + \psi(u) \\ (\alpha\varphi)(u) &= \alpha\varphi(u) \end{aligned}$$

It is easy to check that these are both linear maps. The zero map $v \mapsto 0_V$ for all $u \in U$ is the additive identity, and so the additive inverse $-\varphi$ is defined as you would expect: $(-\varphi)(u) = -\varphi(u)$.

Theorem 5.37

If U and V are vector spaces, then $\mathrm{Hom}(U, V)$ is a vector space under the natural operations.

If furthermore $\dim U = n$ and $\dim V = m$, then for any choice of bases of U and V there is an isomorphism of vector spaces

$$\begin{array}{rcl} \mathrm{Hom}(U,V) & \to & \mathrm{Mat}_{mn} \\ \varphi & \mapsto & A \qquad \text{where } \varphi = \varphi_A \end{array}$$

that is, where A is the matrix that represents φ with respect to the two chosen bases.

In particular, $\dim \mathrm{Hom}(U, V) = mn$.

Proof. It is routine to check that $\mathrm{Hom}(U, V)$ is a vector space, and we omit it.

The map $\varphi \mapsto A$, where A Is the matrix that represents φ with respect to the given choice of bases, is certainly a well-defined bijection. We must check that it is a linear map. Suppose φ is represented by a matrix A and ψ is represented by B. Then writing $u \in U$ in coordinates as $\underline{\lambda}$

$$(\varphi + \psi)(u) = \varphi(u) + \psi(u) = A\underline{\lambda} + B\underline{\lambda} = (A + B)\underline{\lambda}$$

and

$$(\alpha\varphi)(u) = \alpha\varphi(u) = \alpha A\underline{\lambda} = (\alpha A)\underline{\lambda}$$

so the map is linear as required, and is therefore an isomorphism.

This isomorphism means that $\dim \mathrm{Hom}(U, V) = \dim \mathrm{Mat}_{mn}$. There is a further isomorphism

$$\begin{array}{rcl} \mathrm{Mat}_{mn} & \to & \mathbb{R}^{mn} \\ A = (a_{ij}) & \mapsto & \underline{v} \end{array}$$

where $\underline{v} = (a_{11}, a_{21}, \dots, a_{m1}, a_{12}, \dots, a_{m2}, a_{13}, \dots, a_{mn})$ is the column vector of all $m \times n$ entries of A. This map clearly respects addition and scalar multiplication, since these operations are done componentwise in both Mat_{mn} and $\mathbb{R}^{mn}$, and it is a bijection simply because the entries a_{ij} of the matrix may be any scalars. □

Remark

We essentially knew all of this apart from the language of Hom, and in fact we know a lot more. For example, given three vector spaces U, V and W, we may compose maps to give

$$\begin{array}{rcl} \mathrm{Hom}(U,V) \times \mathrm{Hom}(V,W) & \to & \mathrm{Hom}(U,W) \\ (\varphi, \psi) & \mapsto & \psi \circ \varphi \end{array}$$

We already checked that if $\varphi = \varphi_A$ and $\psi = \psi_B$ with respect to bases $\mathcal{B}, \mathcal{B}', \mathcal{B}''$ of $U \cong \mathbb{R}^\ell$, $V \cong \mathbb{R}^n$ and $W \cong \mathbb{R}^m$ respectively, then $\psi \circ \varphi$ is represented by the product matrix BA, and the map of Hom spaces agrees with

$$\begin{array}{rcl} \mathrm{Mat}_{n\ell} \times \mathrm{Mat}_{mn} & \to & \mathrm{Mat}_{m\ell} \\ (A, B) & \mapsto & BA \end{array}$$

after we represent each map by a matrix with respect to the given bases.

It is worth saying a word about what "agrees with" actually means. It is saying precisely that the diagram of maps

$$\begin{array}{ccc} \mathrm{Hom}(U,V)\times\mathrm{Hom}(V,W) & \xrightarrow[(\varphi_A,\varphi_B)\mapsto\varphi_B\circ\varphi_A]{\text{composition}} & \mathrm{Hom}(U,W) \\ \cong\Big\downarrow & & \Big\downarrow\cong \\ \mathrm{Mat}_{n\ell}\times\mathrm{Mat}_{mn} & \xrightarrow[(A,B)\mapsto BA]{\text{product}} & \mathrm{Mat}_{m\ell} \end{array}$$

is a commutative diagram, where the downward maps are the natural 'writing in coordinates' maps $(\varphi_A, \varphi_B) \mapsto (A, B)$ and $\varphi_C \mapsto C$ which take linear maps to the bases that represent them with respect to the chosen bases.

Chapter 6

Euclidean structures on vector spaces

In $\mathbb{R}^n$ we have the familiar **dot product** that we have been referring to as the **scalar product**. It is not part of the definition of vector space, but is an additional tool that is pretty much essential for most applications and that we treat as an integrated part of $\mathbb{R}^n$. Indeed we can't help thinking of the standard basis $\underline{e}_1, \ldots, \underline{e}_n$ as consisting of vectors of length 1 that are mutually orthogonal to one another (that is, at right angles to one another), whether we mentioned that we are using dot product to say all that or not. Such bases are called **orthonormal**, and there are lots of advantages to working with them rather than other bases.

The first thing we do is to prove that orthonormal bases exist. This is the famous **Gram–Schmidt orthogonalisation** process. As a corollary, it gives us a method of calculating orthogonal complements to subspaces.

A general n-dimensional vector space V has no natural given scalar product. We use the properties listed in Proposition 1.9 to define the kind of gadget it should be, and then go on to show that these so-called **inner products** or **Euclidean forms** exist on any V. They are intimately related to the usual scalar product when we represent elements of V as column vectors with respect to a basis.

Curiously, we gain a lot of insight into the general case already in the case $V = \mathbb{R}^n$. It turns out that there are many functions that could serve as a scalar product, and they give different notions of length and angle than our Euclidean geometry eyes are used to.

6.1 Gram–Schmidt orthogonalisation in $\mathbb{R}^n$

We start by working in the vector space $V = \mathbb{R}^n$ equipped with the usual dot product $\underline{v} \cdot \underline{w}$. The key new notion is *orthonormal basis*; compare Definition 1.18 and recall the Kronecker delta δ_{ij}.

Definition 6.1

A set of vectors $v_1, \ldots, v_s \in \mathbb{R}^n$ is **orthonormal** if and only $v_i \cdot v_j = \delta_{ij}$ for each $i, j = 1, \ldots, s$.

An **orthonormal basis** is a basis $v_1, \ldots, v_n$ of V that is orthonormal.

The key theoretical bonus of orthonormality is this: suppose $v_1, \ldots, v_s$ are orthonormal and consider

$$v = \lambda_1 v_1 + \ldots + \lambda_s v_s$$

Then since $v_1 \cdot v_1 = 1$ and $v_1 \cdot v_i = 0$ for $i \geq 2$, we have

$$v_1 \cdot v = v_1 \cdot (\lambda_1 v_1 + \ldots + \lambda_s v_s) = \lambda_1 v_1 \cdot v_1 + \lambda_2 v_1 \cdot v_2 + \ldots + \lambda_s v_1 \cdot v_s = \lambda_1$$

and similarly

$$\lambda_i = v_i \cdot v \text{ for all } i = 1, \ldots, s \tag{6.1}$$

Remark

I think you already use this all the time: it is also part of the machine in §1.2 we refer to as 'orthogonal projection', though is simpler than that, and you routinely use it to extract the coefficients of a vector. Even if it is not familiar, I hope you see it follows easily from the properties of scalar product listed in Proposition 1.9.

The reason to mention this here, is that I bet that when we change notation later and write $\langle v_i, v\rangle$ in place of $v_i \cdot v$, you will forget all about this and regard it as impossibly complicated. Nooo... you wouldn't do that, would you?

Let's warm up by stating a couple of things formally that you know inside out intuitively and proving them using that simple observation.

Proposition 6.2

(i) If $v_1, \ldots, v_s$ are orthonormal, then they are linearly independent.

(ii) If $v_1, \ldots, v_n$ is an orthonormal basis, then for any v we have

$$v = (v_1 \cdot v)v_1 + \ldots + (v_n \cdot v)v_n$$

Proof. (i) Suppose $\lambda_1 v_1 + \ldots + \lambda_s v_s = 0_V$. We must prove that all $\lambda_i = 0$. By (6.1) we have

$$\lambda_i = v_i \cdot (\lambda_1 v_1 + \ldots + \lambda_s v_s) = v_i \cdot 0_V = 0$$

as required.

(ii) Since $v_1, \ldots, v_n$ is a basis there are scalars $\lambda_1, \ldots, \lambda_n$ so that

$$v = \lambda_1 v_1 + \ldots + \lambda_n v_n$$

By (6.1) we have $\lambda_i = v_i \cdot v$, which completes the proof. □

Example

It is easy to see that the vectors

$$\underline{f}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\1\end{pmatrix}, \quad \underline{f}_2 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\end{pmatrix}$$

form an orthonormal basis of $\mathbb{R}^2$ (with respect to the usual dot product). Consider

$$\underline{v} = \begin{pmatrix}3\\-8\end{pmatrix} \in \mathbb{R}^2$$

What are the coefficients of $\underline{v}$ with respect to the given basis $\underline{f}_1, \underline{f}_2$? To answer that, we must calculate λ_1, λ_2 in the expression $\underline{v} = \lambda_1\underline{f}_1 + \lambda_2\underline{f}_2$. We do have methods for this already, but since the basis is orthnormal, there is a quick way using Proposition 6.2(ii):

$$\lambda_1 = \underline{f}_1 \cdot \underline{v} = \frac{-5}{\sqrt{2}}, \quad \lambda_2 = \underline{f}_2 \cdot \underline{v} = \frac{11}{\sqrt{2}}$$

It is easy now to check that $\lambda_1 \underline{f}_1 + \lambda_2 \underline{f}_2$ really does equal $\underline{v}$.

The Gram–Schmidt process that we discuss next is very famous: it takes any basis of $\mathbb{R}^n$ and returns an orthonormal basis. It seems a little involved, but once you try an example it is perfectly natural. The first example is very simple, the next a little more fiddly, but by then the algorithm is becoming clear, and stating and proving it formally turns out to be simpler than actually calculating an example.

Example

Consider

$$v_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

This is a basis of $\mathbb{R}^2$, but it is not an orthonormal basis. Taking the vectors in turn, the first problem is that v_1 is not a unit vector: $\|v_1\| = \sqrt{5}$. Well, that is easily fixed: define

$$w_1 = \frac{1}{\|v_1\|} v_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

and replace v_1 in the basis by w_1: the new set w_1, v_2 is of course still a basis.

The next problem is that w_1 and v_2 are not orthogonal: that is, $w_1 \cdot v_2 = 2/\sqrt{5} \neq 0$. Again, that is easily fixed. Since w_1 is a unit vector, the orthogonal projection of v_2 onto w_1 is, according to Definition 1.20, the vector

$$(v_2 \cdot w_1) w_1 = \frac{2}{\sqrt{5}} \times \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

If we subtract that from v_2, the result will necessarily be orthogonal to w_1: that is, we define

$$u_2 = v_2 - (v_2 \cdot w_1) w_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \frac{2}{5} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \frac{1}{5} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

You see at once that $u_2 \cdot w_1 = 0$.

The only remaining problem is that u_2 is not a unit vector, which again is easily fixed: define

$$w_2 = \frac{1}{\|u_2\|} u_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

Now we have $\|w_1\| = \|w_2\| = 1$ and $w_1 \cdot w_2 = 0$. In other words,

$$w_1 = \frac{1}{\sqrt{5}} \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \quad w_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} -2 \\ 1 \end{pmatrix}$$

is orthonormal, and in fact it is an orthonormal basis: it is a basis because $\langle w_1, w_2 \rangle = \langle v_1, v_2 \rangle$.

Remark

If all you care about is finding an orthonormal basis and you don't care about the order of the vectors you were given, then you might run the algorithm in the example above on the vectors in a different order. If we run the same sequence of ideas on the pair v_2, v_1, then the result is an orthogonal basis w_2', w_1'

$$w_2' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad w_1' = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

The calculations are easier, but the ideas are just the same. You might feel that this is a nicer basis, but in any case the algorithm worked flawlessly to produce an orthonormal basis.

Example 6.3

Consider

$$v_1 = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

This is a basis of $\mathbb{R}^3$, but it is not an orthonormal basis. However, we can use these vectors to find a different set of vectors w_1, w_2, w_3 that do form an orthonormal basis. The process is inductive: we define w_1 first, then w_2, then w_3, making sure that the set of w_i vectors we have made so far are orthonormal at each step.

Start with v_1. You see that $v_1 \cdot v_1 = 9$, that is $\|v_1\| = 3$. Consider the unit vector w_1 in the direction of v_1. That is,

$$w_1 = \frac{1}{\|v_1\|} v_1 = \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

Now consider v_2. Don't worry about its length for a moment, but instead find out whether it is at right angles to w_1: you can detect this by computing the component of v_2 in the direction of w_1 as usual by the scalar product $v_2 \cdot w_1$. If this component was zero, then setting w_2 to be v_2 divided by its length would give an orthonomal pair w_1, w_2.

But in general, even if as in this case $v_2 \cdot w_1 = \frac{1}{3}$ is not zero, we can make a vector at right angles to w_1 by subtracting from v_2 its orthogonal projection onto w_1. The formula from Definition 1.20 for orthogonal projection is simple: since w_1 is a unit vector, it is $(v_2 \cdot w_1)w_1$. So the formula for defining w_2 is in two steps: first subtract this projection, and then divide by the length of the resulting vector to normalise it to length 1: that is, set

$$u_2 = v_2 - (v_2 \cdot w_1)w_1 \quad \text{and then} \quad w_2 = \frac{1}{\|u_2\|} u_2$$

which in this example gives

$$u_2 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix} - \frac{1}{3} \times \frac{1}{3} \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} = \frac{4}{9} \begin{pmatrix} 2 \\ -5 \\ 4 \end{pmatrix} \quad \text{and then} \quad w_2 = \frac{1}{3\sqrt{5}} \begin{pmatrix} 2 \\ -5 \\ 4 \end{pmatrix}$$

(Of course, when going from u_2 to w_2 I simply ignored the $\frac{4}{9}$ and just calculated the unit vector in the direction of $(2, -5, 4)^T$ – it's easier and gives the same answer.)

At this point **you absolutely must** check that $w_1 \cdot w_2 = 0$ and that $\|w_1\| = \|w_2\| = 1$. If you are wrong now, then any following calculations will be wasted. In fact I think those checks work, and so the pair w_1, w_2 is orthonormal.

Now continue with v_3, but this time remove its orthogonal projection onto both w_1 and w_2. That is, first define

$$\begin{aligned} u_3 &= v_3 - (v_3 \cdot w_1)w_1 - (v_3 \cdot w_2)w_2 \\ &= v_3 - \frac{2}{3}w_1 - \frac{4}{3\sqrt{5}}w_2 \\ &= \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} - \frac{2}{9}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} - \frac{4}{45}\begin{pmatrix} 2 \\ -5 \\ 4 \end{pmatrix} \\ &= \frac{9}{45}\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix} \end{aligned}$$

(which by construction is at right angles to both w_1 and w_2) and then normalise it to be length 1, giving

$$w_3 = \frac{1}{\|u_3\|}u_3 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

Once more **you absolutely must** check that $w_1 \cdot w_3 = w_2 \cdot w_3 = 0$ and that $\|w_3\| = 1$. In fact those checks pass, and the result is an orthonormal basis w_1, w_2, w_3, which in all its glory is

$$\frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad \frac{1}{3\sqrt{5}}\begin{pmatrix} 2 \\ -5 \\ 4 \end{pmatrix}, \quad \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

Remark

Once again, we may try the algorithm on the same vectors in a different order. For example, if we put v_3 first, then the result is an orthogonal basis w_3', w_1', w_2'

$$w_3' = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad w_1' = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \quad w_2' = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \\ 0 \end{pmatrix}$$

The calculations are easier, but the ideas are just the same.

The examples above illustrate the general idea of Gram–Schmidt orthogonalisation follows.

Theorem 6.4 (Gram–Schmidt orthogonalisation)

Let $V = \mathbb{R}^n$, equipped with the usual scalar (dot) product. If $v_1, \ldots, v_n$ is a basis of V, then the following algorithm determines an orthonormal basis $w_1, \dots, w_n$ of V:

$$\begin{aligned} & & w_1 &= \frac{1}{\|v_1\|}v_1 \\ u_2 = v_2 - (v_2 \cdot w_1)w_1 & \quad \text{and} & w_2 &= \frac{1}{\|u_2\|}u_2 \\ & & &\vdots \\ u_j = v_j - \sum_{k=1}^{j-1}(v_j \cdot w_k)w_k & \quad \text{and} & w_j &= \frac{1}{\|u_j\|}u_j \\ & & &\vdots \\ u_n = v_n - \sum_{k=1}^{n-1}(v_n \cdot w_k)w_k & \quad \text{and} & w_n &= \frac{1}{\|u_n\|}u_n \end{aligned}$$

where at each step $w_1, \dots, w_i$ is an orthonormal basis of its span $\langle w_1, \dots, w_i \rangle = \langle v_1, \dots, v_i \rangle$.

Proof. This is an induction. Certainly $v_1 \neq 0_V$ so w_1 is defined, and $\langle w_1 \rangle = \langle v_1 \rangle$.

We proceed by induction. Suppose we have $w_1, \dots, w_{i-1}$ as claimed. Then u_i is defined and lies in $\langle w_1, \dots, w_{i-1}, v_i \rangle = \langle v_1, \dots, v_i \rangle$ (the equality by induction). Note that $u_i \neq 0_V$, otherwise v_i would be in the span of $\langle w_1, \dots, w_{i-1} \rangle = \langle v_1, \dots, v_{i-1} \rangle$, contradicting linear independence of $v_1, \dots, v_n$. So w_i is defined, it has $\|w_i\| = 1$, and by Lemma 4.11 it is linearly independent of $w_1, \dots, w_{i-1}$.

If $j < i$ then since by induction $w_1, \dots, w_{i-1}$ is orthonormal, then for the nonzero scalar $\alpha = \frac{1}{\|u_i\|}$,

$$\begin{aligned} w_i \cdot w_j &= \alpha \left(v_i - \sum_{k=1}^{i-1}(v_i \cdot w_k)w_k \right) \cdot w_j \\ &= \alpha \left(v_i \cdot w_j - (v_i \cdot w_j) w_j \cdot w_j \right) \\ &= 0 \end{aligned}$$

and so $w_1, \dots, w_i$ are orthonormal as required. □

Corollary 6.5 (Extend an orthonormal set to an orthonormal basis)

Let $V = \mathbb{R}^n$, equipped with the usual scalar (dot) product. If $v_1, \dots, v_s$ is an orthonormal set, then there are vectors $v_{s+1}, \dots, v_n \in V$ so that $v_1, \dots, v_n$ is an orthonormal basis of V.

Proof. Let $w_{s+1}, \dots, w_n$ be any vectors that extend $v_1, \dots, v_s$ to a basis $v_1, \dots, v_s, w_{s+1}, \dots, w_n$, and apply Gram–Schmidt process to this basis. Of course the initial elements $v_1, \dots, v_s$ will be unchanged by the process, since any orthogonal projection from one of these to the others is zero. □

Definition 6.6

Let $V = \mathbb{R}^n$ equipped with the usual scalar (dot) product, and let $U \subset V$ be a subspace. We define the **orthogonal complement** $U^\perp$ **of** U **in** V to be the set of elements of V that are orthogonal to every element of U:

$$U^\perp = \{v \in V \mid u \cdot v = 0 \text{ for all } u \in U\}$$

Theorem 6.7 (Orthogonal complements)

Let $V = \mathbb{R}^n$ equipped with the usual scalar (dot) product and let $U \subset V$ be a subspace. Then $U^\perp$ is a subspace of V, and moreover it is a complement to U in the sense of Definition 5.30: that is, $U \cap U^\perp = \{0_V\}$ and U and $U^\perp$ together span V.

In particular $V \cong U \oplus U^\perp$ and $\dim V = \dim U + \dim U^\perp$.

One could prove this directly, but the Gram-Schmidt process gives an elegant proof.

Proof. Choose any basis $\mathcal{B}$ of U and apply Gram–Schmidt. The result is an orthonormal basis $v_1, \dots, v_s$ of U: Theorem 6.4 guarantees both that it has the same span U as $\mathcal{B}$ and that it is a basis of that span (or you may say that $v_1, \dots, v_s$ is linearly independent by Proposition 6.2(i)).

By Corollary 6.5, extend this orthonormal set $v_1, \dots, v_s$ to an orthonormal basis $v_1, \dots, v_n$ of V. Then we claim that

$$U^\perp = \langle v_{s+1}, \dots, v_n \rangle$$

which completes the proof. Clearly each v_j for $j \geq s+1$ lies in $U^\perp$, while conversely if $v = \lambda_1 v_1 + \ldots + \lambda_n v_n \in U^\perp$, then for all $j = 1, \dots s$

$$0 = v_j \cdot v = v_j \cdot (\lambda_1 v_1 + \ldots + \lambda_n v_n) = \lambda_j$$

so in fact $v \in \langle v_{s+1}, \dots, v_n \rangle$ as claimed. □

Example

We calculate the orthogonal complement of

$$U = \left\langle \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix} \right\rangle \subset V = \mathbb{R}^3$$

Following the proof, we first find a orthonormal basis of U: for example

$$w_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

Then we extend it to a basis of $\mathbb{R}^3$: for example

$$w_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Then we apply Gram–Schmidt to this basis **in this order**. The result, by Example 6.3 is

$$w_1 = \frac{1}{3}\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}, \quad w_2 = \frac{1}{3\sqrt{5}}\begin{pmatrix} 2 \\ -5 \\ 4 \end{pmatrix}, \quad w_3 = \frac{1}{\sqrt{5}}\begin{pmatrix} -2 \\ 0 \\ 1 \end{pmatrix}$$

The proof then shows that

$$U^\perp = \left\langle \frac{1}{3\sqrt{5}}\begin{pmatrix}2\\-5\\4\end{pmatrix}, \frac{1}{\sqrt{5}}\begin{pmatrix}-2\\0\\1\end{pmatrix}\right\rangle$$

and the given vectors are an orthonormal basis of $U^\perp$.

Of course, for the purposes of describing $U^\perp$ we may not have needed to produce an orthonormal basis of it, in which case we could equally well say, for example, that

$$U^\perp = \left\langle \begin{pmatrix}2\\-5\\4\end{pmatrix}, \begin{pmatrix}-2\\0\\1\end{pmatrix}\right\rangle \quad \text{or} \quad U^\perp = \left\langle \begin{pmatrix}0\\-1\\1\end{pmatrix}, \begin{pmatrix}-2\\0\\1\end{pmatrix}\right\rangle$$

and so on, but Gram–Schmidt produced an orthonormal basis automatically.

6.2 Inner products, lengths and angles

Compare the following definition with Proposition 1.9.

Definition 6.8

An **inner product** on V associates a scalar, denoted $\langle v, w\rangle$, to any $v, w \in V$ subject to the following rules:

(i) $\langle v, w\rangle = \langle w, v\rangle$ for any $v, w \in V$.

(ii) $\langle(\lambda_1 v_1 + \lambda_2 v_2), w\rangle = \lambda_1 \langle v_1, w\rangle + \lambda_2 \langle v_2, w\rangle$ for any $v_1, v_2, w \in V$ and any $\lambda_1, \lambda_2 \in \mathbb{R}$.

(iii) For any $v \in V$, $\langle v, v\rangle \geq 0$, and furthermore $\langle v, v\rangle = 0$ if and only if $v = 0_V$.

If you prefer, an inner product is a function $V \times V \to \mathbb{R}$ denoted $(v, w) \mapsto \langle v, w\rangle$ that satisfies those rules.

Example

If $V = \mathbb{R}^n$ then $\langle \underline{v}, \underline{w}\rangle = \underline{v} \cdot \underline{w}$ is an inner product, but there are many others.

Let $V = \mathbb{R}^2$ and consider the matrix

$$A = \begin{pmatrix}2 & 1\\1 & 3\end{pmatrix}$$

This is a **symmetric matrix**, meaning just that it equals its own transpose: $A = A^T$. We may use A to define an inner product:

$$\langle \underline{v}, \underline{w}\rangle = \underline{v}^T A \underline{w}$$

or in coordinates $\underline{v} = \begin{pmatrix}x_1\\y_1\end{pmatrix}$ and $\underline{w} = \begin{pmatrix}x_2\\y_2\end{pmatrix}$

$$\left\langle \begin{pmatrix}x_1\\y_1\end{pmatrix}, \begin{pmatrix}x_2\\y_2\end{pmatrix}\right\rangle = \begin{pmatrix}x_1 & y_1\end{pmatrix}\begin{pmatrix}2 & 1\\1 & 3\end{pmatrix}\begin{pmatrix}x_2\\y_2\end{pmatrix} = 2x_1x_2 + x_1y_2 + y_1x_2 + 3y_1y_2$$

The first two properties of inner product follow from the fact that A is symmetric, and from the properties of matrix multiplication.

The final point needs more care but is elementary. Suppose $\underline{v} = (x, y)^T$ with $|x| \geq |y|$. Then

$$\langle \underline{v}, \underline{v} \rangle = 2x^2 + 2xy + 3y^2 \geq 3y^2 \geq 0$$

since the only issue is if either $x < 0$ or $y < 0$, and in that case $x^2 + xy \geq x^2 - |xy| \geq |x|(|x| - |y|) \geq 0$. Moreover, this shows that if $\langle \underline{v}, \underline{v} \rangle = 0$, then $y^2 = 0$, so $y = 0$ and then also $x = 0$. The case $|x| \leq |y|$ is similar, using $3y^2$ to absorb any negative contributions.

Example

Let $a \leq b$ and consider

$$V = \{f\colon [a, b] \to \mathbb{R} \mid f \text{ is continuous}\}$$

Then it is easy to check that

$$\langle f, g \rangle = \int_a^b fg$$

defines an inner product on V. (You need to use definition of continuity, the idea of integration as area under the curve, and the standard properties of integrals.)

For example, on $[a, b] = [-\pi, \pi]$ we have

$$\int_{-\pi}^{\pi} \sin(x)\cos(x)dx = 0$$

since $\sin$ is an odd function and $\cos$ is an even function. Therefore $\langle \sin(x), \cos(x) \rangle = 0$ and we see that $\sin(x)$ and $\cos(x)$ are orthogonal to one another. They are not quite orthonormal, since

$$\int_{-\pi}^{\pi} \sin^2(x)dx = \int_{-\pi}^{\pi} \cos^2(x)dx = \pi$$

but you can divide them by π if you need that, or start again and redefine the inner product with a factor $1/\pi$ if you prefer.

Remark

You may be thinking that we could run through all the definitions, theorems and calculations that we know for the usual dot product and try them with a general inner product $\langle \cdot, \cdot \rangle$.

You're right! We can use an inner product to define length and angle and orthogonal projection and all the rest, and it all works out nicely. Here goes . . . simply replacing the dot product $v \cdot w$ by the inner product $\langle v, w \rangle$ wherever we see it.

Definition 6.9

A **Euclidean space** is a finite-dimensional vector space V (over $\mathbb{R}$) equipped with a fixed choice of inner product. (The inner product is denoted by $\langle \cdot, \cdot \rangle$, even if it is not mentioned explicitly.)

First we define lengths; compare Definition 1.10.

Definition 6.10 (Length of a vector)

Let V be a Euclidean space. We define the **length** of a vector $v \in V$ to be

$$\|v\| = \sqrt{\langle v, v \rangle}$$

which is a non-negative real number. (Notice the double lines in the notation.)

Notice that the inner product is an essential part of the definition: it has to have been chosen and fixed in advance, and everyone has to know which inner product we are using. (In that respect it's a bit like choosing a basis: it is some extra information that everyone is meant to know about.)

Now for angles; compare Definition 1.14

Definition 6.11 (Angle between vectors)

Let V be a Euclidean space and let $v, w \in V \setminus \{0_V\}$ be nonzero vectors. We define the **angle between v and w**, denoted $\angle vw$, to be the real number

$$\angle vw = \cos^{-1}\left(\frac{\langle v, w \rangle}{\|v\|\|w\|}\right)$$

where we take the principal preimage of $\cos$, so that $\angle vw$ lies in the interval $[0, \pi]$.

As before, we often write $\vartheta = \angle vw$ so that the formula is

$$\langle v, w \rangle = \|v\|\|w\| \cos(\vartheta)$$

You know the issue: we need the following result so that $\cos$ has a preimage.

Proposition 6.12 (Cauchy–Schwartz inequality)

Let V be a Euclidean space. For any $v, w \in V$,

$$|\langle v, w \rangle| \leq \|v\|\|w\|$$

and furthermore equality is achieved only when $v = \lambda w$ for some $\lambda \in \mathbb{R}$ or $w = 0_V$.

We omit the proof: both the second and third proofs we gave in §1.2 can be adapted easily to this general situation.

We define orthonormal vectors as you expect.

Definition 6.13

Let V be a Euclidean space. A set of vectors $v_1, \ldots, v_s \in V$ is **orthonormal** if and only $\langle v_i, v_j \rangle = \delta_{ij}$ for each $i, j = 1, \ldots, s$.

An **orthonormal basis** is a basis $v_1, \ldots, v_n$ of V that is orthonormal.

Finally we consider orthogonal projection; compare Definition 1.20.

Definition 6.14

Let V be a Euclidean space. Let $v, w \in V$ with $w \neq 0_V$ and let $\hat{w} = w/\|w\|$ be the unit vector in the direction of w. Then the scalar quantity $\langle v, \hat{w} \rangle$ is called **component of** v **in the direction of** w, and the vector $\langle v, \hat{w} \rangle \, \hat{w}$ is the **orthogonal projection of** v **in the direction of** w or **onto** w.

Example

Let $V = \mathbb{R}^2$ with the usual dot product. Then

$$\underline{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \underline{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

is an orthonormal basis. It is certainly a basis, and it is easy to see at once that $\underline{v}_i \cdot \underline{v}_j = \delta_{ij}$.

But now consider a different inner product on V:

$$\langle \underline{v}, \underline{w} \rangle = \underline{v}^T A \underline{w} \quad \text{where} \quad A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$$

or in coordinates $\underline{v} = \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$ and $\underline{w} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$

$$\left\langle \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \right\rangle = \begin{pmatrix} x_1 & y_1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = 2x_1x_2 + x_1y_2 + y_1x_2 + 3y_1y_2$$

For example, collecting the $\sqrt{2}$ factors and multiplying the row vector and matrix first,

$$\langle \underline{v}_1, \underline{v}_2 \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 3 & 4 \end{pmatrix} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = -\frac{1}{2}$$

We can run the Gram–Schmidt process on the pair $\underline{v}_1, \underline{v}_2$ using this inner product to calculate lengths and projections. We have

$$\langle \underline{v}_1, \underline{v}_1 \rangle = \frac{7}{2} \quad \text{and} \quad \|\underline{v}_1\| = \sqrt{\frac{7}{2}}$$

so the unit vector in the direction of $\underline{v}_1$ is

$$\underline{w}_1 = \frac{\underline{v}_1}{\|\underline{v}_1\|} = \frac{1}{\sqrt{7}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

The orthogonal projection of $\underline{v}_2$ onto $\underline{w}_1$ is

$$\langle \underline{v}_2, \underline{w}_1 \rangle \, \underline{w}_1 = \frac{-1}{\sqrt{2}\sqrt{7}} \times \frac{1}{\sqrt{7}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = -\frac{1}{7\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

so, as usual, we define

$$\underline{u}_2 = \underline{v}_2 - \langle \underline{v}_2, \underline{w}_1 \rangle \, \underline{w}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \frac{1}{7\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \frac{\sqrt{2}}{7} \begin{pmatrix} 4 \\ -3 \end{pmatrix}$$

It is quick to see that $\langle \underline{w}_1, \underline{u}_2 \rangle = 0$, so it only remains to scale $\underline{u}_2$ to a unit vector and we have the orthonormal basis

$$\underline{w}_1 = \frac{1}{\sqrt{7}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \underline{w}_2 = \frac{1}{\sqrt{5}} \begin{pmatrix} 4 \\ -3 \end{pmatrix}$$

using the inner product $\langle \cdot, \cdot \rangle$.

6.3 Gram–Schmidt orthogonalisation in general

The Gram–Schmidt orthogonalisation process described in Theorem 6.4 works in exactly the same way if we replace the dot products $v \cdot w$ that arise by any other inner product $\langle v, w \rangle$. With that change, the statement and proof are exactly analogous, and the corollaries follow as before with the same change.

That is, given a basis $v_1, \ldots, v_n$ of any Euclidean space V, the Gram–Schmidt process works inductively, constructing an orthonormal set $w_1, \ldots, w_{i-1}$ (with the same span as $v_1, \ldots, v_{i-1}$) and then defines the (nonzero) vector

$$u_i = v_i - \sum_{k=1}^{i-1} \langle v_i, w_k \rangle \, w_k \quad \text{and sets} \quad w_i = \frac{1}{\|u_i\|} u_i$$

Theorem 6.15

Let V be a Euclidean space. Then V has an orthonormal basis.

Proof. Let $v_1, \ldots, v_n$ be any basis of V. The Gram–Schmidt orthogonalisation process is a well-defined terminating algorithm that produces an orthogonal basis. □

Corollary 6.16 (Extend an orthonormal set to an orthonormal basis)

Let V be a Euclidean space. If $v_1, \ldots, v_s$ is an orthonormal set, then there are vectors $v_{s+1}, \ldots, v_n \in V$ so that $v_1, \ldots, v_n$ is an orthonormal basis of V.

Corollary 6.17 (Orthogonal complements)

Let V be a Euclidean space and let $U \subset V$ be a subspace. Then $U^\perp$ is a subspace of V, and moreover it is a complement to U in the sense of Definition 5.30: that is, $U \cap U^\perp = \{0_V\}$ and U and $U^\perp$ together span V.

In particular $V \cong U \oplus U^\perp$ and $\dim V = \dim U + \dim U^\perp$.

6.4 What does an orthonormal basis do for me?

Any basis gives an isomorphism of V with $\mathbb{R}^n$. One way to think of an orthonormal basis is that this isomorphism also matches the given inner product on V with the standard dot product on $\mathbb{R}^n$.

Although we prove this precisely in a moment, I'm not sure it's such a useful idea for us. It's main value is perhaps just that it explains that it is not such a jungle out there, and inner products (on finite-dimensional vector spaces) are not such very bizarre beasts after all.

The important stuff is everything in the previous section: if you have an inner product and need to work with a basis, then please make it an orthonormal basis unless you have a clear reason not to.

Theorem 6.18 (Isomorphism with $\mathbb{R}^n$ with the usual dot product)

Let V be a Euclidean space and let $\mathcal{B}$ be an orthonormal basis. Then the coordinate isomorphism

$$\chi_{\mathcal{B}} \colon V \to \mathbb{R}^n$$

matches the inner product of V with the usual dot product of $\mathbb{R}^n$, in the sense that

$$\langle v, w \rangle = \chi_{\mathcal{B}}(v) \cdot \chi_{\mathcal{B}}(w)$$

for all $v, w \in V$.

Proof. Let $\mathcal{B} = \{w_1, \ldots, w_n\}$ so that by definition the coordinate isomorphism $\chi_{\mathcal{B}}$ is determined by $\chi_{\mathcal{B}}(w_i) = \underline{e}_i \in \mathbb{R}^n$. Observe first that therefore $\chi_{\mathcal{B}}$ has the required property on the elements of $\mathcal{B}$: for any $i, j \in \{1, \ldots, n\}$

$$\langle w_i, w_j \rangle = \delta_{ij} = \underline{e}_i \cdot \underline{e}_j$$

Now by linearity, writing any $v \in V$ with respect to $\mathcal{B}$ as $v = \sum_{i=1}^n \lambda_i w_i$, we have

$$\begin{aligned} \langle v, w_j \rangle = \left\langle \sum_{i=1}^n \lambda_i w_i, w_j \right\rangle &= \sum_{i=1}^n \lambda_i \langle w_i, w_j \rangle \\ &= \sum_{i=1}^n \lambda_i \left(\underline{e}_i \cdot \underline{e}_j \right) \\ &= \left(\sum_{i=1}^n \lambda_i \underline{e}_i \right) \cdot \underline{e}_j \\ &= \chi_{\mathcal{B}}(v) \cdot \chi_{\mathcal{B}}(w_j) \end{aligned}$$

and similarly in the right-hand factor of $\langle \cdot, \cdot \rangle$. $\square$

Chapter 7

Eigenvalues and eigenvectors

Throughout this chapter V is a finite-dimensional vector space. We consider linear maps $\varphi\colon V \to V$ from V to itself. These are sometimes referred to as linear operators to emphasise that the domain and codomain are the same.

Definition 7.1

Let V be a vector space. A **linear operator on** V is a linear map $V \to V$.

In particular, whenever we choose a basis of V, such a map φ is represented by a square matrix with respect to that basis in both the domain and the codomain. It is worth noting how the rank–nullity formula works in this context, both for linear operators and the square matrices that represent them.

Remark

Let $\varphi\colon V \to V$ be a linear operator. Therefore φ is surjective if and only if $\operatorname{rk}\varphi = \dim V$, and for any linear map φ is injective if and only if $\operatorname{nullity}\varphi = 0$. Therefore the rank–nullity formula

$$\operatorname{rk}\varphi + \operatorname{nullity}\varphi = \dim V$$

shows that

$$\varphi \text{ is injective} \quad\Longleftrightarrow\quad \varphi \text{ is surjective}$$

We may express this for a square matrix $A \in \mathrm{Mat}_{nn}$ as

$$\dim\ker L_A = 0 \quad\Longleftrightarrow\quad \ker L_A = \{\underline{0}\} \quad\Longleftrightarrow\quad \operatorname{rk} A = n \quad\Longleftrightarrow\quad \dim\operatorname{Im} L_A = n$$

We define eigenvalues and eigenvectors for both linear operators and square matrices, so that the definitions agree whenever we choose a basis.

Definition 7.2 (Eigenvalues and eigenvectors)

Consider a linear operator $\varphi\colon V \to V$. We say that $\lambda \in \mathbb{R}$ is an **eigenvalue** of φ if there exists some **nonzero** $v \in V$ (re-emphasise: $v \neq 0_V$) such that

$$\varphi(v) = \lambda v$$

We call v an **eigenvector** of φ corresponding to the eigenvalue λ.

Let $A \in \mathrm{Mat}_{nn}$. We say that $\lambda \in \mathbb{R}$ is an **eigenvalue** of A if there exists some **nonzero** $\underline{v} \in \mathbb{R}^n$ (re-emphasise: $\underline{v} \neq \underline{0}$) such that

$$A\underline{v} = \lambda \underline{v}$$

We call $\underline{v}$ an **eigenvector** of A corresponding to the eigenvalue λ.

In passing, we check that these definitions do indeed match. Let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for V and let $A = (a_{ij}) \in \mathrm{Mat}_{nn}$ be the matrix such that $\varphi = \varphi_A$. Then writing $v \in V$ in coordinates as $\underline{v} = (\alpha_1, \ldots, \alpha_n)^T \in \mathbb{R}^n$ we have

$$\varphi(v) = \lambda v \quad \text{if and only if} \quad A\underline{v} = \lambda \underline{v}$$

Therefore computing eigenvalues and eigenvectors of linear transformations is equivalent to computing eigenvalues and eigenvectors of matrices, and we often deliberately conflate the two.

Example

Let $V = \mathbb{R}^2$ and consider the linear map $L_A \colon V \to V$ for the matrix

$$A = \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}$$

It is not immediately clear what L_A does in geometric terms, but if you know to consider the following vectors, you see that

$$A\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = -\begin{pmatrix} 1 \\ -1 \end{pmatrix} \quad \text{and} \quad A\begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2\begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

Giving names to these vectors, we may write

$$Av_1 = \lambda_1 v_1 \text{ where } v_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \text{ and } \lambda_1 = -1$$

and similarly

$$Av_2 = \lambda_2 v_2 \text{ where } v_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } \lambda_2 = 2$$

That is, L_A is a reflection of the line $\ell_1 = \langle v_1 \rangle$ through v_1 onto itself, and it scales the line $\ell_2 = \langle v_2 \rangle$ through v_2 by a factor of 2. Intuitively, you may think of the map L_A as interpolating between these two behaviours in the rest of the picture. In particular, if you consider any other line through the origin in V, you will see that it is not mapped to itself at all. For example,

$$A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

so that the line $\ell = \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle$ (also known as the y-axis) is mapped to the different line $\left\langle \begin{pmatrix} 0 \\ 2 \end{pmatrix} \right\rangle$ (also known as the x-axis). You may easily check any other line through the origin similarly.

Remark

That example illustrates the key property of eigenvectors. If $v \in V$ is an eigenvector (for any eigenvalue λ), then the 1-dimensional subspace $\ell = \langle v \rangle \subset V$ is mapped to itself by φ. Indeed, any element $w \in \ell$ is simply a scalar multiple $w = \alpha v$ of v, so that $\varphi(w) = \alpha\varphi(v) \in \langle v \rangle$. That is, $\varphi(\ell) \subset \ell$ where $\ell = \langle v \rangle$ for any eigenvector v.

There is a minor nuance: if $\lambda = 0$ then $\varphi(\langle v \rangle) = \{0_V\}$, while if $\lambda \neq 0$ then $\varphi(\langle v \rangle) = \langle v \rangle$, and in the latter case φ is an isomorphism of this 1-dimensional subspace with itself.

This nuance is worth a moment's thought, since in prose we easily assume the wrong idea: if $U \subset V$ is a subspace of V, we say U **is mapped to itself** (by φ) to mean only that $\varphi(U) \subset U$. There is no claim that $\varphi(U) = U$: we are not claiming that the map φ restricted to U is surjective onto U.

Example 7.3

Alongside the previous remark, it is useful to have in mind various different types of geometric behaviour that a linear map $L_A \colon \mathbb{R}^2 \to \mathbb{R}^2$ determined by a matrix A may have, and to notice the effect on 1-dimensional subspaces (lines through the origin) in each case:

(i) The scalar matrix $A = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ stretches the whole plane by a factor of 2. In particular, every 1-dimensional linear subspace is mapped to itself.

(ii) The diagonal matrix $A = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right)$ stretches the x-axis by a factor of 2 and the y-axis by a factor of 3. These two axes are the only 1-dimensional linear subspaces that are mapped to themselves.

(iii) The shear matrix $A = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ fixes the x-axis and tilts the y-axis over by $\pi/4$. The x-axis is the only 1-dimensional linear subspace that is mapped to itself.

(iv) The rotation matrix $A = \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & -1 \\ 1 & 1 \end{smallmatrix}\right)$ that rotates the whole plane by an angle $\pi/4$ anti-clockwise about the origin does not map any 1-dimensional linear subspace to itself.

(v) The projection matrix $A = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$ that maps the whole plane onto the x-axis, with each vertical line squished to the point where it meets the x-axis, certainly fixes the x-axis pointwise. In disguise, there is exactly one other 1-dimensional subspace that is mapped to itself: the y-axis is mapped to the origin, which is indeed a point of the y-axis. So the y-axis is mapped to itself, although the map is not an isomorphism: it is the zero map.

(vi) OK then, let's do the daft one: the zero matrix maps all 1-dimensional subspaces to themselves, in the trivial sense that they all map to the origin. Daft it may be, but it is an important part of the whole picture.

You are right to regard each of the claims above about 1-dimensional subspaces as an exercise that you need to do.

The notions of eigenvalue and eigenvector work for linear operators on any vector space, not only finite-dimensional ones, and some cases are very familiar.

Example 7.4

Differentiation is a linear operator on the vector space $C^\infty(\mathbb{R})$ of infinitely differentiable functions:

$$\begin{aligned} C^\infty(\mathbb{R}) &\to C^\infty(\mathbb{R}) \\ f &\mapsto \frac{d}{dx} f \end{aligned}$$

Although $C^\infty(\mathbb{R})$ is not a finite-dimensional vector space, the ideas of eigenvalue and eigenvector still make sense, and you already know what they are. The exponential function $e^{\lambda x}$ is an eigenvector of d/dx with corresponding eigenvalue $\lambda \in \mathbb{R}$, since

$$\frac{d}{dx} e^{\lambda x} = \lambda e^{\lambda x}$$

and of course any scalar multiple of $e^{\lambda x}$ is also an eigenvector. Eigenvectors belonging to function spaces are often called **eigenfunctions**.

7.1 Eigenvectors in $\mathbb{R}^2$

We consider linear maps $\mathbb{R}^2 \to \mathbb{R}^2$, as in Example 7.3.

Example 7.5

Let

$$A = \begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix} \in \mathrm{Mat}_{22}$$

We will show that $\lambda = 2$ is an eigenvalue of A. To do this, we need to find

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathbb{R}^2$$

not equal to zero, such that

$$\begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} 2\alpha \\ 2\beta \end{pmatrix}$$

Equivalently, we need a non-trivial solution to the system of linear equations

$$\begin{aligned} -5\alpha + 2\beta &= 2\alpha \\ -7\alpha + 4\beta &= 2\beta \end{aligned}$$

Rearranging so that the right-hand side is zero we get two copies of the equation

$$-7\alpha + 2\beta = 0$$

Thus we see that there is a one-parameter family of possible eigenvectors given by

$$\begin{pmatrix} 2t \\ 7t \end{pmatrix} \in \mathbb{R}^2 \qquad \text{for any } t \in \mathbb{R}, t \neq 0$$

Example 7.6

Consider the map

$$\begin{aligned} \varphi\colon \mathbb{R}^2 &\to \mathbb{R}^2 \\ (x, y) &\mapsto (-5x + 2y, -7x + 4y) \end{aligned}$$

Then $\varphi = \varphi_A$, where A is the matrix in Example 7.5. Hence $\lambda = 2$ is an eigenvalue, and any vector $\underline{v} \in \{(2t, 7t) \mid t \in \mathbb{R}, t \neq 0\}$ an eigenvector for λ.

Consider a 2×2 matrix A with eigenvalue $\lambda \in \mathbb{R}$ and eigenvector $\underline{v} \in \mathbb{R}^2$. That simply means that $A\underline{v} = \lambda\underline{v}$. We may rearrange this equation to say instead that, equivalently,

$$(A - \lambda I_2)\underline{\mathbf{v}} = \underline{0}$$

Since by definition eigenvectors are not zero, the matrix $A - \lambda I_2$ cannot be invertible: indeed there is a non-zero vector $\underline{v} \in \ker(A - \lambda I_2)$, and so $\text{nullity}(A - \lambda I_2) = \dim\ker(A - \lambda I_2) \geq 1$.

Recall the notion of **determinant** of a 2×2 matrix, viz.

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

The crucial point is the necessary and sufficient condition for a 2×2 matrix to be invertible:

$$A \text{ is invertible} \quad \Longleftrightarrow \quad \det A \neq 0$$

or equivalently the criterion for A to be singular, that is for $\text{nullity}\, A \geq 1$:

$$A \text{ is singular} \quad \Longleftrightarrow \quad \det A = 0$$

We will come to this formally in the next section, but for now just keep those facts in mind.

Eigenvectors with eigenvalue λ are the nonzero elements of $\ker(A - \lambda I_2)$, and so given A as above we consider the equation

$$\det(A - \lambda I_2) = 0$$

In detail, writing

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \text{Mat}_{22}$$

we have

$$A - \lambda I_2 = \begin{pmatrix} a - \lambda & b \\ c & d - \lambda \end{pmatrix}$$

which has determinant

$$\begin{aligned} \det(A - \lambda I_2) &= (a - \lambda)(d - \lambda) - bc \\ &= \lambda^2 - \lambda(a + d) + ad - bc \end{aligned}$$

By the criterion for A to be singular, the eigenvalues of A are precisely the (real) roots of this polynomial. This is a quadratic in λ, so we can find the roots via the quadratic equation, viz.

$$\lambda = \frac{a + d \pm \sqrt{(a - d)^2 + 4bc}}{2} \tag{7.1}$$

Remark

Equation (7.1) raises an important point. The value $(a - d)^2 + 4bc$ can be negative, giving complex eigenvalues $\lambda \in \mathbb{C}$. We will address this point later. Strictly speaking, the definition of eigenvalue we have only considers those roots λ that lie in $\mathbb{R}$. That's fine and simple, but of course it would be uncouth simply to ignore what a complex eigenvalue is trying to tell you.

Example 7.7

Consider the matrix

$$A = \begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix}$$

We could simply use equation (7.1) to compute the eigenvalues

$$\lambda = \frac{-1 \pm \sqrt{81 - 56}}{2} = \frac{-1 \pm 5}{2} = 2 \text{ and } -3$$

Fine, but in fact it is usually just as easy to compute them from first principles. We need to compute the roots of

$$\begin{aligned} \det(A - \lambda I_2) &= \det \begin{pmatrix} -5 - \lambda & 2 \\ -7 & 4 - \lambda \end{pmatrix} \\ &= (\lambda + 5)(\lambda - 4) + 14 \\ &= \lambda^2 + \lambda - 6 \\ &= (\lambda + 3)(\lambda - 2) \end{aligned}$$

and once again we obtain the two eigenvalues $\lambda = 2$ and -3.

We computed the eigenvectors for the eigenvalue $\lambda = 2$ in Example 7.5. Let's do the case when $\lambda = -3$. We want

$$\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \in \mathbb{R}^2$$

such that

$$\begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} = \begin{pmatrix} -3\alpha \\ -3\beta \end{pmatrix}$$

Thus we need to solve the linear system of equations

$$\begin{aligned} -5\alpha + 2\beta &= -3\alpha \\ -7\alpha + 4\beta &= -3\beta \end{aligned}$$

Notice (after a little rearranging) that once again this reduces to a single equation

$$\alpha - \beta = 0$$

Hence there is a one-parameter family of eigenvectors:

$$\begin{pmatrix} t \\ t \end{pmatrix} \in \mathbb{R}^2 \qquad \text{for any } t \in \mathbb{R}, t \neq 0$$

There is a very simple calculation we can do with the eigenvectors we calculated so far, and it is the key to the main result at this stage of the theory (which in turn invites the next questions that are the basis of the next development in the theory next year). Watch closely: you're going to be impressed!

Example 7.8

Let

$$A = \begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix}$$

In Example 7.5 we saw that an eigenvalue when $\lambda = 2$ is

$$v_1 = \begin{pmatrix} 2 \\ 7 \end{pmatrix}$$

and in Example 7.7 we saw that an eigenvalue when $\lambda = -3$ is

$$v_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

Notice that $\{v_1, v_2\}$ is a linearly independent set, and hence a basis for $\mathbb{R}^2$. Write

$$P = \begin{pmatrix} 2 & 1 \\ 7 & 1 \end{pmatrix}$$

We can easily calculate the inverse matrix

$$P^{-1} = \frac{1}{5}\begin{pmatrix} -1 & 1 \\ 7 & -2 \end{pmatrix}$$

and

$$\begin{aligned} P^{-1}AP &= \frac{1}{5}\begin{pmatrix} -1 & 1 \\ 7 & -2 \end{pmatrix}\begin{pmatrix} -5 & 2 \\ -7 & 4 \end{pmatrix}\begin{pmatrix} 2 & 1 \\ 7 & 1 \end{pmatrix} \\ &= \frac{1}{5}\begin{pmatrix} -1 & 1 \\ 7 & -2 \end{pmatrix}\begin{pmatrix} 4 & -3 \\ 14 & -3 \end{pmatrix} \\ &= \begin{pmatrix} 2 & 0 \\ 0 & -3 \end{pmatrix} \end{aligned}$$

where the entries on the diagonal are the eigenvalues of A.

What makes the example work? Of course, we may regard P as a change of basis matrix, in which case the example is saying that if we choose the basis carefully, then the linear map is represented by a diagonal matrix with the eigenvalues down the diagonal. The key technical point is already evident: this works if you can use a basis of eigenvectors. We need to pronounce all this carefully and precisely, and in particular we need to determine whether we can find a basis of eigenvectors or not. The striking point is that this is not always possible, and this is already evident from the collection of different geometric behaviours listed at the beginning of the chapter. But first we need to address the determinant in the room.

7.2 Determinants of square matrices

Determinants are a very powerful tool but notoriously fiddly. The first rule of Determinant Club is: you probably already know all you need to know about determinants and you must not forget that.

The key point about determinants, the thing they **determine**, is that for $A \in \mathrm{Mat}_{nn}$

$$\det A = 0 \iff \text{there is a nontrivial solution to } A\underline{v} = \underline{0}$$

(which is Theorem 7.12 below) so that using the definition of kernel and the rank–nullity formula you may also say

$$\det A = 0 \iff \dim\ker A \geq 1 \iff \mathrm{rk}\, A < n$$

We work out the essential parts of the theory. We take a curious approach that you might not have seen before, but that is fairly routine in higher mathematics: we will simply list all the properties we would like determinant to have, and then prove all the results we need using those, and only then will we answer the question of whether the determinant exists.

Definition 7.9

For any (square) matrix $A \in \mathrm{Mat}_{nn}$, the **determinant of** A is scalar, denoted $\det A$, which, regarded as a function,

$$\begin{array}{rcl} \mathrm{Mat}_{nn} & \longrightarrow & \mathbb{R} \\ A & \mapsto & \det A \end{array}$$

has the following properties with respect to the elementary row operations

(i) $\det(S_{ij}A) = -\det A$ (swapping two rows changes the sign of $\det A$)

(ii) $\det(M_j(\lambda)A) = \lambda \det A$ (multiplying a row by λ multiples $\det A$ by λ, even for $\lambda = 0$)

(iii) $\det(A_{ij}(\lambda)A) = \det A$ (adding a multiple of one row to another does not change $\det A$)

(iv) $\det I_n = 1$

For now, we simply **assume that such a function exists** (and also that it is unique). With that, we can derive almost everything we ever use at once, including techniques for computing $\det A$, and in the process we get a clue about why it should exist at all. First a load of immediate observations from the properties of determinant above.

Proposition 7.10

The determinant $\det A$ of a matrix A satisfies the following properties:

(i) If a row of A is zero, then $\det A = 0$.

(ii) If two rows of A are identical, then $\det A = 0$.

(iii) If A is a diagonal matrix, that is $A_{ij} = 0$ whenever $i \neq j$, then $\det A = a_{11}a_{22}\cdots a_{nn}$.

(iv) If A is an upper triangular matrix, that is $A_{ij} = 0$ whenever $i > j$, then $\det A = a_{11}a_{22}\cdots a_{nn}$.

(v) If A is a lower triangular matrix, that is $A_{ij} = 0$ whenever $i < j$, then $\det A = a_{11}a_{22}\cdots a_{nn}$.

Proof. (i) Multiplying that row by, say, 2 does not change A but doubles the determinant, so $\det A = 0$. (ii) Subtract one row from the other (without changing the determinant) to make a zero row and apply (i) (or swap them to pick up a minus sign but without changing the determinant). (iii) A may be constructed by performing a series of row multiplications $M_i(a_{ii})$ for $i = 1, \ldots, n$ starting with the identity matrix I_n. The result follows since $\det I_n = 1$ and the ith row operation multiplies this by a_{ii}. (iv–v) Use row addition operations (which do not change the diagonal entries nor $\det A$) to replace A by a diagonal matrix, and apply (iii). □

Second, we prove two structural results about determinants: the key point mentioned above is one, and the multiplicativity of determinant is the other. Both rely on the following observation that was clear to us when we worked on RREF in Chapter 2.

Remark

The RREF of A is the matrix EA which arises by premultiplying A by a product E of (invertible) elementary matrices (with $\lambda \neq 0$ in $M_i(\lambda)$). Therefore, $\det EA$ and $\det A$ differ by a nonzero factor (a product of -1s and nonzero λs), and so $\det A = 0$ if and only if $\det EA = 0$.

Theorem 7.11

Let $A \in \mathrm{Mat}_{nn}$. Then $\det A = 0$ if and only if $\ker L_A \neq \{\underline{0}\}$.

Proof. Let the RREF of A be EA, where E is a product of invertible elementary matrices. Either $EA = I_n$ (and $\det EA = 1$) or EA has at least one zero row (and $\det EA = 0$). In the first case $\ker L_{EA} = \{\underline{0}\}$, while in the second case $\dim \ker L_{EA} \geq 1$. Therefore $\det EA = 0$ if and only if $\dim \ker L_{EA} \geq 1$. The proof is completed by recalling that $\ker L_{EA} = \ker L_A$, either by Proposition 5.14 or by recalling how row operations correspond to solving linear equations. □

Theorem 7.12

Let $A, B \in \mathrm{Mat}_{nn}$. Then $\det(AB) = \det(A)\det(B)$.

Proof. Notice first that by Definition 7.9 applied with A being an elementary matrix, we have

$$\det S_{ij} = -1, \qquad \det M_j(\lambda) = \lambda, \qquad \det A_{ij}(\lambda) = 1$$

(for example $A = S_{ij}$ gives $1 = \det I_n = \det(S_{ij}S_{ij}) = -\det S_{ij}$) so Definition 7.9 shows that

$$\det(FA) = \det(F)\det(A) \qquad \text{where } F \text{ is any of } S_{ij},\ M_j(\lambda) \text{ or } A_{ij}(\lambda)$$

That is the statement for elementary matrices; the proof is almost complete.

Now let EA be the RREF of A where $E = E_k \cdots E_1$ is a product of invertible elementary matrices. Suppose first that $\det A \neq 0$, so that $EA = I_n$. Rewriting that as $A = F_1 \cdots F_k$, where $F_i = E_i^{-1}$, and applying the previous result repeatedly gives

$$\begin{aligned}
\det(AB) &= \det(F_1 \cdots F_k B) \\
&= \det(F_1)\det(F_2 \cdots F_k B) \\
&= \det(F_1)\det(F_2)\det(F_3 \cdots F_k B) \\
&\ \vdots \\
&= \det(F_1)\det(F_2)\cdots\det(F_{k-1})\det(F_k)\det(B) \\
&= \det(F_1)\det(F_2)\cdots\det(F_{k-1}F_k)\det(B) \\
&\ \vdots \\
&= \det(F_1 \cdots F_k)\det(B) \\
&= \det(A)\det(B)
\end{aligned}$$

where you notice that first we break up the product $AB = F_1 \cdots F_k B$ and then we reassemble the product $A = F_1 \cdots F_k$.

Now suppose $\det A = 0$. Therefore $\operatorname{nullity} L_A > 0$ by Theorem 7.12, so $\operatorname{rk} L_A < n$ by rank–nullity, and so L_A is not surjective. But therefore $L_{AB} = L_A \circ L_B$ is not surjective either, and running the argument back again shows that $\operatorname{rk} L_{AB} < n$ so $\operatorname{nullity} L_{AB} > 0$ so $\det(AB) = 0$, as required. □

Remark

What we've seen so far makes it clear that if there really is a well-defined function $\det(A)$ of square matrices $A \in \mathrm{Mat}_{nn}$ that satisfies all the properties of Definition 7.9, then it is uniquely defined: indeed Definition 7.9 gives enough tools to calculate $\det(A)$ uniquely from its RREF.

So we turn now to the question of showing that there really is such a function at all. It's worth noting that very often the properties we've seen above are all that we need both to calculate $\det(A)$ and to use it theoretically in proofs, but still we need to show somehow that there really is a well-defined function there at all.

Proof of existence for $n = 2, 3$

Consider the 2×2 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

We **define**

$$\det A = a_{11}a_{22} - a_{12}a_{21} = \sum_{\sigma \in S_2} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \tag{7.2}$$

where the first expression is probably familiar, while the second curious expression needs a little explanation.

Recall that the symmetric (or permutation) group on 2 symbols is $S_2 = \{\mathrm{id}, (1,2)\}$, where $\sigma = (1,2)$ denotes the transposition $1 \leftrightarrow 2$, or more formally the map $\sigma(1) = 2$ and $\sigma(2) = 1$. The sign of a permutation $\operatorname{sgn}\sigma \in \{1, -1\}$ satisfies $\operatorname{sgn}\sigma = -1$ if σ is a product of an odd number of transpositions, and $\operatorname{sgn}\sigma = +1$ is σ is a product of an even number of transpositions. Thus we have

$$\operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} = +a_{11}a_{12} \qquad \text{for } \sigma = \mathrm{id}$$

and

$$\operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} = -a_{12}a_{21} \qquad \text{for } \sigma = (1,2)$$

and so the two expressions in (7.2) are the same.

We check that the four properties (i–iv) required by Definition 7.9 are satisfied.

Point (i) holds because swapping rows 1 and 2 amounts to replacing σ by the permutation $(1,2)\sigma$ in the curious formula, which multiplies the sign by -1. Notice that the formula is **linear** in the entries of the first row $\begin{pmatrix} a_{11} & a_{12} \end{pmatrix}$ of A and also in the entries of the second row $\begin{pmatrix} a_{21} & a_{22} \end{pmatrix}$ of A, so that (ii) is automatic by linearity.

Linearity also helps (iii). To be concrete, let's add λ times row 2 to row 1, that is, let's work out $A_{12}(\lambda)A$ – the other case $A_{21}(\lambda)$ works by an identical argument. The 1st row of $A_{12}(\lambda)A$ is $\begin{pmatrix} a_{11} + \lambda a_{21} & a_{12} + \lambda a_{22} \end{pmatrix}$, so by linearity of the formula in the rows of A

$$\det(A_{12}(\lambda)A) = \det A + \lambda \det A' \qquad \text{where} \qquad A' = \begin{pmatrix} a_{21} & a_{22} \\ a_{21} & a_{22} \end{pmatrix}$$

that is, A' has two equal rows. But then $\det A' = 0$ by Proposition 7.10, so $\det(A_{ij}(\lambda)A) = \det A$ as Definition 7.9 requires. Finally point (iv): $\det I_2 =$ since the formula has only one term (which in the curious formula of (7.2) corresponds to $\sigma = \mathrm{id}$).

This is good: the formula (7.2) defines a function that has the properties that Definition 7.9 demands.

Remark

The determinant is called a **multilinear** function of the rows of A, since it is separately linear in each row. It is extremely useful to be able to recognise this and use it, as we saw above (and more significantly in a moment).

Such multilinearity phenomena occur in lots of places: for example, an inner product is linear in each of its arguments, so is multilinear in 2 arguments, or bilinear, as we put it at the time.

The determinant is also called an **alternating** function in the rows of A, since swapping them introduces a factor -1, so that if two rows are equal then the result is zero.

People sometimes even say that determinant is **normalised as 1 on the identity** to say that $\det I_n = 1$. (We need this, since any scalar multiple of the determinant also satisfies conditions (i–iii) of Definition 7.9.)

With this jargon, the determinant function is the unique alternating multilinear function in the rows of a square matrix, normalised as 1 on the identity – well, it is once we show that it exists!

Let's do the same for a 3×3 matrix

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

We use the curious formula to define

$$\det A = \sum_{\sigma \in S_3} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} a_{3\sigma(3)} \tag{7.3}$$

and then work out what that means using a little knowledge of the permutation group on 3 symbols

$$S_3 = \{\mathrm{id}, (1,2), (1,3), (2,3), (1,2,3), (1,3,2)\}$$

(where of course $(1,2,3) = (1,2)(2,3)$ and $(1,3,2) = (1,3)(2,3)$ both have sign $+1$) so that working through S_3 in that order we have

$$\det A = a_{11}a_{22}a_{33} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32}$$

in which you notice the individual terms each contain exactly one entry from each row and on entry from each column, and there are exactly $6 = 3!$ ways to do this. Although the expression is complicated, you can easily parse it as the perhaps more familiar expression

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

The great thing about the crazy formula (7.3) is that the properties of Definition 7.9 follow formally just as they did in the 2×2 case, and so once again we have shown the existence of a determinant function: (i) and (iv) are immediate, while (ii) and (iii) follow from multilinearity.

Proof of existence in the general case (non-examinable)

Finally, for any $A \in \mathrm{Mat}_{nn}$, we define a scalar $\det A$ by the crazy formula

$$\det A = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\sigma(1)} a_{2\sigma(2)} \cdots a_{n\sigma(n)} \tag{7.4}$$

where S_n is the symmetric (permutation) group on n symbols, and $\operatorname{sgn}(\sigma) \in \{+1, -1\}$ is the usual permutation representation. This sum is a form of madness: it has $n!$ terms.

Proposition 7.13

The formula (7.4) defines a function $\det(A)$ that satisfies the properties of Definition 7.9 for a determinant function.

We omit further proof: the discussions in the cases $n = 2$ and 3 above contain all the ideas: we just check the four properties of Definiion 7.9: (i) and (iv) are immediate, while (ii) and (iii) follow from multilinearity.

One thing we have only discussed in the 2×2 cases so far is how to expand a determinant out along a row (or a column). We cover that next briefly, explaining the expansion along the first row; all other cases are similar (after taking care about the sign). This is most useful as another proof that a matrix with a zero row (or column) has determinant zero, or when calculating in small cases when a row has almost all entries zero.

Definition 7.14

Let $A \in \mathrm{Mat}_{nn}$ and $i, j \in \{1, \ldots, n\}$. Then the i, j **minor of** A is the matrix $A_{ij} \in \mathrm{Mat}_{n-1,n-1}$ which is constructed by removing the ith row and the jth column from A.

The i, j **cofactor of** A is the scalar $\det A_{ij}$.

Proposition 7.15

Let $A \in \mathrm{Mat}_{nn}$. Then

$$\det A = a_{11} \det A_{11} + \ldots + a_{1n} \det A_{1n}$$

Proof. Consider the crazy formula (7.4). Every term has exactly one factor a_{1j} for some j; indeed $j = \sigma(1)$ for some $\sigma \in S_n$ in the crazy formula (7.4). Consider all the terms that have a_{11} as a factor. As a summand of the right-hand side of (7.4) they are

$$a_{11} \times \sum_{\sigma \in S_n, \sigma(1)=1} \mathrm{sgn}(\sigma) a_{2\sigma(2)} \cdots a_{n\sigma(n)}$$

which is $a_{11} \det A_{11}$, since the sum is over all permutations of $2, \ldots, n$. The terms with factors $a_{12}, \ldots, a_{1n}$ are exactly analogous. □

Remark

Throughout this section we have worked with row operations. All analogous results work for column operations. One can either work through them repeating the arguments, or instead apply the following result.

Proposition 7.16

Let $A \in \mathrm{Mat}_{nn}$. Then $\det A^T = \det A$.

Proof. **(non-examinable)** Let $A = (a_{ij})$ and set $b_{ij} = a_{ji}$ so that $A^T = (b_{ij})$. Then

$$\det A^T = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) b_{1\sigma(1)} \cdots b_{n\sigma(n)} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{\sigma(1)1} \cdots a_{\sigma(n)n} = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) a_{1\tau(1)} \cdots a_{n\tau(n)}$$

where $\tau = \sigma^{-1}$. Note that $\operatorname{sgn}(\tau) = \operatorname{sgn}(\sigma)$ so that continuing the calculation gives

$$\det A^T = \sum_{\tau \in S_n} \operatorname{sgn}(\tau) a_{1\tau(1)} \cdots b_{n\tau(n)} = \det A$$

as required. □

7.3 The characteristic polynomial, eigenspaces and multiplicity

We work here with square matrices $A \in \mathrm{Mat}_{nn}$ and column vectors $\underline{v} \in \mathbb{R}^n$, rather than with linear operators $\varphi\colon V \to V$.

Recall that if $A \in \mathrm{Mat}_{nn}$ then an eigenvector $\underline{v} \neq \underline{0}$ corresponding to the eigenvalue λ satisfies

$$A\underline{v} = \lambda \underline{v}$$

Rearranging the equation, this is equivalent to saying

$$\underline{v} \in \ker(A - \lambda I_n) \qquad \text{and so in particular} \qquad \operatorname{nullity}(A - \lambda I_n) > 0$$

This motivates the following definition.

Definition 7.17

Let $A \in \mathrm{Mat}_{nn}$. The **characteristic polynomial of** A is the polynomial $c_A(x)$ defined by

$$c_A(x) = \det(A - xI_n)$$

By the discussion above, the eigenvalues of A are the roots of the characteristic polynomial c_A. These roots need not be distinct: the multiplicity of a root λ is called the **algebraic multiplicity** of the eigenvalue λ.

(Incidentally, you may see alternative notation for $c_A(x)$, such as $p_A(x)$ or $\chi_A(x)$. Do ask if it's ever not clear, since there are other polynomials related to matrices that crop up.)

Remark

For an $n \times n$ matrix $A \in \mathrm{Mat}_{nn}$, the characteristic polynomial $c_A(x)$ is a polynomial of degree n: indeed the coefficient of x^n is $(-1)^n$. Hence by the Fundamental Theorem of Algebra, A has exactly n complex eigenvalues, counted with multiplicity.

Example 7.18

Consider

$$A = \begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix} \in \mathrm{Mat}_{22}$$

The characteristic polynomial is (using the variable λ in place of x)

$$\begin{aligned} c_A(\lambda) &= \det(A - \lambda I_2) \\ &= \det \begin{pmatrix} 2-\lambda & 3 \\ 3 & -6-\lambda \end{pmatrix} \\ &= (\lambda-2)(\lambda+6) - 9 \\ &= \lambda^2 + 4\lambda - 21 \\ &= (\lambda+7)(\lambda-3) \end{aligned}$$

The eigenvalues are the roots of c_A: namely, $\lambda = 3$ and -7. Both eigenvalues have algebraic multiplicity one.

Example 7.19

Consider

$$A = \begin{pmatrix} 1 & 3 & 5 & 6 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 1 & 7 \\ 0 & 0 & 0 & 5 \end{pmatrix} \in \mathrm{Mat}_{44}$$

Then

$$A - \lambda I_4 = \begin{pmatrix} 1-\lambda & 3 & 5 & 6 \\ 0 & 2-\lambda & 1 & 0 \\ 0 & 0 & 1-\lambda & 7 \\ 0 & 0 & 0 & 5-\lambda \end{pmatrix}$$

By Proposition 7.10(iv), the determinant of an upper triangular matrix is simply the product of the diagonal entries, so

$$c_A(\lambda) = \det(A - \lambda I_4) = (1-\lambda)^2(2-\lambda)(5-\lambda)$$

and the eigenvalues are

$\lambda = 1$	with algebraic multiplicity 2
$\lambda = 2$	with algebraic multiplicity 1
$\lambda = 5$	with algebraic multiplicity 1

Recall that the kernel of any linear map, including those represented by matrices such as $A - \lambda I_n$, is a subspace of $\mathbb{R}^n$.

Definition 7.20

Let $A \in M_{nn}$ and let λ be an eigenvalue of A. The subspace

$$E_\lambda = \{\underline{v} \in \mathbb{R}^n \mid A\underline{v} = \lambda\underline{v}\}$$

is called the **eigenspace** of A corresponding to λ. It is simply the set of all eigenvectors associated to the eigenvalue λ together with the zero vector $\underline{0}$.

The dimension $\dim E_\lambda$ is called the **geometric multiplicity** of the eigenvalue.

Remark

Let $A \in \mathrm{Mat}_{nn}$ and λ be an eigenvalue of A. By definition an eigenvalue cannot be zero, and so the eigenspace E_λ will always contain a non-zero vector. In particular, $\dim E_\lambda \geq 1$.

In fact, there is also an upper bound on $\dim(E_\lambda)$: if λ has algebraic multiplicity s_λ, then

$$\dim E_\lambda \leq s_\lambda$$

That is, the geometric multiplicity of λ is less than or equal to the algebraic multiplicity of λ. We do not need that here, so we omit the proof, though you have all the tools you need if you wish to prove it yourself.

Example 7.21

Consider the matrix

$$A = \begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix} \in \mathrm{Mat}_{22}$$

From Example 7.18 we know this has eigenvalues 3 and -7. We will compute the eigenspace in each case.

$\lambda = 3$: We need to find the solutions to the system of linear equations

$$\begin{aligned} 2\alpha + 3\beta &= 3\alpha \\ 3\alpha - 6\beta &= 3\beta \end{aligned}$$

or, equivalently,

$$\alpha - 3\beta = 0$$

Hence

$$E_\lambda = \left\{ \begin{pmatrix} 3t \\ t \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\} = \left\langle \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\rangle$$

Therefore the geometric multiplicity of λ is $\dim E_\lambda = 1$.

$\lambda = -7$: Proceeding similarly, we have the system of linear equations

$$\begin{aligned} 2\alpha + 3\beta &= -7\alpha \\ 3\alpha - 6\beta &= -7\beta \end{aligned}$$

which reduces to the single equation

$$3\alpha + \beta = 0$$

Hence

$$E_\lambda = \left\{ \begin{pmatrix} t \\ -3t \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\} = \left\langle \begin{pmatrix} 1 \\ -3 \end{pmatrix} \right\rangle$$

Therefore the geometric multiplicity of λ is $\dim E_\lambda = 1$.

Notice that the two eigenvectors

$$\begin{pmatrix} 3 \\ 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

are linearly independent.

Proposition 7.22

Let $A \in \mathrm{Mat}_{nn}$ and let $\underline{v}_1, \ldots, \underline{v}_r \in \mathbb{R}^n$ be eigenvectors corresponding to distinct eigenvalues $\lambda_1, \ldots, \lambda_r$ of A. Then $\underline{v}_1, \ldots, \underline{v}_r$ are linearly independent.

Proof. Suppose for a contradiction that $\{\underline{v}_1, \ldots, \underline{v}_r\}$ is linearly dependent. Let k be the smallest positive integer such that

$$\underline{v}_k \in \langle \underline{v}_1, \ldots, \underline{v}_{k-1} \rangle \tag{7.5}$$

Then there exists $\mu_1, \ldots, \mu_{k-1} \in \mathbb{R}$ such that

$$\underline{v}_k = \mu_1 \underline{v}_1 + \cdots + \mu_{k-1} \underline{v}_{k-1} \tag{7.6}$$

Left-multiplying by A gives

$$\begin{aligned} A\underline{v}_k &= A(\mu_1 \underline{v}_1 + \cdots + \mu_{k-1} \underline{v}_{k-1}) \\ &= \mu_1 A\underline{v}_1 + \cdots + \mu_{k-1} A\underline{v}_{k-1} \end{aligned}$$

so that, since each $\underline{v}_i$ is an eigenvector of A corresponding to the eigenvalue λ_i, we have

$$\lambda_k \underline{v}_k = \mu_1 \lambda_1 \underline{v}_1 + \cdots + \mu_{k-1} \lambda_{k-1} \underline{v}_{k-1} \tag{7.7}$$

Multiplying both sides of (7.6) by λ_k and then subtracting (7.7) gives

$$\underline{0} = \mu_1(\lambda_k - \lambda_1)\underline{v}_1 + \cdots + \mu_{k-1}(\lambda_k - \lambda_{k-1})\underline{v}_{k-1}$$

Now $\underline{v}_1, \ldots, \underline{v}_{k-1}$ are linearly independent since k was chosen to be the smallest positive integer satisfying (7.5), so $\mu_1 = \cdots = \mu_{k-1} = 0$. Substituting these values for μ_i in (7.6) gives $\underline{v}_k = \underline{0}$. But this is a contradiction: $\underline{v}_k$ is an eigenvector, so $\underline{v}_i \neq \underline{0}$. □

7.4 Diagonalisation of square matrices

We now see one of the main theoretical uses of eigenvalues: transforming a matrix to a diagonal matrix.

Definition 7.23

We say that matrices $A, B \in \mathrm{Mat}_{nn}$ are **similar** if and only if

$$B = P^{-1}AP$$

where $P \in \mathrm{Mat}_{nn}$ is invertible. It is easy to prove that being similar is an equivalence relation on Mat_{nn}.

Remark

Compare this definition with that of equivalence of matrices, Definition 5.35. Thinking in terms of the linear map $L_A \colon \mathbb{R}^n \to \mathbb{R}^n$, the point is that the domain and codomain of L_A are equal, so rather than permitting different changes of basis in the domain and codomain, this definition requires that we use the same change of basis $Q = P$ in both.

Theorem 7.24

Similar matrixes have the same eigenvalues.

Proof. Let $A, B \in \mathrm{Mat}_{nn}$ be two similar matrices. Then there exists an invertible matrix $P \in \mathrm{Mat}_{nn}$ such that $B = P^{-1}AP$. Then $A = PBP^{-1}$. We have that

$$\begin{aligned}\det(A - \lambda I_n) &= \det(PBP^{-1} - \lambda PP^{-1}) \\ &= \det(P(B - \lambda I_n)P^{-1}) \\ &= \det(P)\det(B - \lambda I_n)\det(P^{-1}) \\ &= \det(P)\det(P^{-1})\det(B - \lambda I_n) \\ &= \det(B - \lambda I_n)\end{aligned}$$

□

Definition 7.25

A matrix $A \in \mathrm{Mat}_{nn}$ is said to be **diagonalisable** if it is similar to a diagonal matrix.

Example 7.26

Let

$$A = \begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix} \in \mathrm{Mat}_{22}$$

We saw in Example 7.21 that two linearly independent choices of eigenvectors for A are

$$\underline{v}_1 = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \qquad \text{and} \qquad \underline{v}_2 \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

Let

$$P = \begin{pmatrix} 3 & 1 \\ 1 & -3 \end{pmatrix} \in \mathrm{Mat}_{22}$$

be the matrix whose columns are given by $\underline{v}_1$ and $\underline{v}_2$. The matrix P is invertible since $\{\underline{v}_1, \underline{v}_2\}$ is a basis for $\mathbb{R}^2$ (a set of two linearly independent vectors in a two-dimensional vector space must form a basis). We can quickly calculate P^{-1}, which is

$$P^{-1} = \frac{1}{10}\begin{pmatrix} 3 & 1 \\ 1 & -3 \end{pmatrix}$$

Let us calculate

$$\begin{aligned}P^{-1}AP &= \frac{1}{10}\begin{pmatrix} 3 & 1 \\ 1 & -3 \end{pmatrix}\begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & -3 \end{pmatrix} \\ &= \frac{1}{10}\begin{pmatrix} 3 & 1 \\ 1 & -3 \end{pmatrix}\left(\begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix}\begin{pmatrix} 3 & 1 \\ 1 & -3 \end{pmatrix}\right) \\ &= \frac{1}{10}\begin{pmatrix} 3 & 1 \\ 1 & -3 \end{pmatrix}\begin{pmatrix} 9 & -7 \\ 3 & 21 \end{pmatrix} \\ &= \begin{pmatrix} 3 & 0 \\ 0 & -7 \end{pmatrix}\end{aligned}$$

Hence A is diagonalisable.

The main theorem is this.

Theorem 7.27 (The Diagonalisation Theorem)

A matrix $A \in \mathrm{Mat}_{nn}$ is diagonalisable if and only if A has n linearly independent eigenvectors, that is, if and only if there is a basis of eigenvectors of A.

More precisely, $P^{-1}AP = B$ where B is a diagonal matrix if and only if the columns of P are linearly independent eigenvectors of A. In this case, the diagonal entries of B are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

Proof. Suppose that $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n \in \mathbb{R}^n$ are n linearly independent eigenvectors of A corresponding to eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (which are not necessarily distinct). Let

$$P = \begin{pmatrix} \underline{v}_1 & \underline{v}_2 & \ldots & \underline{v}_n \end{pmatrix} \in \mathrm{Mat}_{nn}$$

be the matrix whose columns are equal to the $\underline{v}_i$. Then

$$\begin{aligned} AP &= \begin{pmatrix} A\underline{v}_1 & A\underline{v}_2 & \ldots & A\underline{v}_n \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1\underline{v}_1 & \lambda_2\underline{v}_2 & \ldots & \lambda_n\underline{v}_n \end{pmatrix} \\ &= \begin{pmatrix} \underline{v}_1 & \underline{v}_2 & \ldots & \underline{v}_n \end{pmatrix} \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} \\ &= PB \end{aligned}$$

where

$$B = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

Note that P is invertible, since $\mathrm{Colspan}(B) = \mathbb{R}^n$. Hence

$$AP = PB \qquad \Longrightarrow \qquad P^{-1}AP = B$$

as required.

Conversely, suppose there is an invertible matrix P so that $P^{-1}AP = B$, where B is a diagonal matrix, $B = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$.

Define $\underline{v}_i = P\underline{e}_i$, the ith column of P, for each $i = 1, \ldots, n$. Note that $\underline{v}_i \neq \underline{0}$, since P is invertible, so $\det P \neq 0$ and so its ith column is not $\underline{0}$. In fact, more strongly, $\underline{v}_1, \ldots, \underline{v}_n$ is a basis since $L_P \colon \mathbb{R}^n \to \mathbb{R}^n$ is an isomorphism, so it maps any basis to another basis, in this case the standard basis $\underline{e}_1, \ldots, \underline{e}_n$ to the claimed basis.

Then since $AP = PB$, we have

$$\begin{aligned} A\underline{v}_i &= AP\underline{e}_i \\ &= PB\underline{e}_i \\ &= P\lambda_i\underline{e}_i \\ &= \lambda_i\underline{v}_i \end{aligned}$$

and so $\underline{v}_i$ is an eigenvector of A for λ_i, and therefore $\underline{v}_1, \ldots, \underline{v}_n$ is a basis of eigenvectors. □

It is not true that every matrix has a basis of eigenvectors, but there is a simple situation where it does hold: namely if it has n distinct eigenvalues.

Corollary 7.28

If $A \in \mathrm{Mat}_{nn}$ has n distinct eigenvalues, then A is diagonalisable. More precisely, we can write $P^{-1}AP = B$ where the columns of P are eigenvectors of A, and B is a diagonal matrix whose diagonal entries are eigenvalues of A that correspond, respectively, to the eigenvectors in P.

Proof. Suppose that A has n distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, with corresponding choices of eigenvectors $\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_n \in \mathbb{R}^n$. By Proposition 7.22 we have that these eigenvectors are linearly independent, and hence form a basis for $\mathbb{R}^n$. The result follows from the theorem. $\square$

Example 7.29

Let

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix}$$

This matrix is lower-triangular, so we see immediately that the eigenvalues are

$$\begin{aligned} \lambda = 1 &\qquad \text{with algebraic multiplicity } 2 \\ \lambda = 2 &\qquad \text{with algebraic multiplicity } 1 \end{aligned}$$

Let us compute some eigenvectors.

$\lambda = 1$: We need to solve the system of linear equations

$$\begin{aligned} \alpha &= \alpha \\ \beta &= \beta \\ \beta + 2\gamma &= \gamma \end{aligned}$$

Hence the eigenvectors are of the form

$$\begin{pmatrix} s \\ t \\ -t \end{pmatrix} \qquad \text{where } s, t \in \mathbb{R},\ s \text{ and } t \text{ not both zero}$$

We can write down two linearly independent eigenvectors

$$\underline{v}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \qquad \text{and} \qquad \underline{v}_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$$

$\lambda = 2$: We need to solve the system of linear equations

$$\begin{aligned} \alpha &= 2\alpha \\ \beta &= 2\beta \\ \beta + 2\gamma &= 2\gamma \end{aligned}$$

We immediately see that the eigenvectors are of the form

$$\begin{pmatrix} 0 & 0 & t \end{pmatrix} \qquad \text{where } t \in \mathbb{R}, t \neq 0$$

Hence an eigenvector is

$$\underline{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

The set $\{\underline{v}_1, \underline{v}_2, \underline{v}_3\}$ is linearly independent, hence by Theorem 7.27 we have that A is diagonalisable. More specifically, let

$$P = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix}$$

Then

$$P^{-1}AP = \mathrm{diag}(1, 1, 2)$$

We shall verify this via direct calculation. We can quickly compute

$$P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

Then

$$\begin{aligned} P^{-1}AP &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1 & 2 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \end{aligned}$$

Example 7.30

We shall show that the matrix

$$A = \begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix} \in \mathrm{Mat}_{22}$$

is not diagonalisable. In this case we immediately see that there is only one eigenvalue, $\lambda = 2$, with algebraic multiplicity $s_\lambda = 2$. In order to compute the eigenspace E_λ we need to solve the system of linear equations

$$\begin{aligned} 2\alpha + \beta &= 2\alpha \\ 2\beta &= 2\beta \end{aligned}$$

We have that

$$E_\lambda = \left\{ \begin{pmatrix} t \\ 0 \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\} = \left\langle \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\rangle$$

Hence the geometric multiplicity $\dim E_\lambda = 1$ does not equal the algebraic multiplicity $s_\lambda = 2$. So by Theorem 7.27 we conclude that A is not diagonalisable.

Corollary 7.31

Let $A \in \mathrm{Mat}_{nn}$ be a diagonalisable matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, counted with multiplicity. Then

$$\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$$

Proof. We know by Theorem 7.27 that there exists an invertible matrix $P \in \mathrm{Mat}_{nn}$ such that

$$P^{-1}AP = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

Writing $B = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, we have that $A = PBP^{-1}$. Hence

$$\begin{aligned} \det(A) &= \det(PBP^{-1}) \\ &= \det(P)\det(B)\det(P^{-1}) \\ &= \det(P)\det(P^{-1})\det(B) \\ &= \det(B) \\ &= \lambda_1 \lambda_2 \cdots \lambda_n \end{aligned}$$

□

Chapter 8

Orthogonal and symmetric matrices

Chapter 7 considered matrices up to similarity. One of the main results, the Diagonalisation Theorem 7.27, proved that if a matrix A has a basis of eigenvectors then it is similar to a diagonal matrix $P^{-1}AP$. In other words, there is (possibly) new basis with respect to which the linear map L_A is represented by a diagonal matrix – indeed, the new basis is simply any basis of eigenvectors, and the matrix P simply has that basis as its columns.

If we now include dot product on $\mathbb{R}^n$, we may ask whether the new basis can be chosen to be orthonormal? This is a reasonable request: the initial standard basis is orthonormal, and we may regard it as a good idea to preserve that. We prove such orthogonal diagonalisability in the special case of symmetric matrices with distinct eigenvalues. **The proofs in this chapter are not examinable.**

8.1 Orthogonal matrices and orthonormal bases

We work in $\mathbb{R}^n$ with the usual scalar (dot) product as inner product.

Definition 8.1

A matrix $P \in \mathrm{Mat}_{nn}$ is **orthogonal** if and only if $P^T = P^{-1}$.

Remark

Observe that $\det P = 1$ or -1 for an orthogonal matrix:

$$1 = \det I_n = \det(P^{-1}P) = \det(P^TP) = \det(P^T)\det(P) = \det(P)^2$$

since $P^{-1} = P^T$ and $\det(P^T) = \det(P)$.

Remark

Orthogonal matrices are really about geometry: for example, rotation and reflection matrices are orthogonal. More precisely, the simple observation that for $\underline{v}, \underline{w} \in \mathbb{R}^n$

$$P\underline{v} \cdot P\underline{w} = \underline{v}^T P^T P \underline{w} = \underline{v}^T \underline{w} = \underline{v} \cdot \underline{w}$$

or, in words, that orthogonal matrices preserve dot product, implies at once that both lengths

$$\|P\underline{v}\| = \|\underline{v}\|$$

and the angle between vectors

$$\angle P\underline{v}P\underline{w} = \angle \underline{v}\underline{w}$$

are preserved by orthogonal P.

In terms of the associated linear map $L_P \colon \mathbb{R}^n \to \mathbb{R}^n$ preserves lengths and angles and so is a *Euclidean isometry*.

Proposition 8.2

A matrix $\in \mathrm{Mat}_{nn}$ is orthogonal if and only if its rows $\underline{r}_1, \ldots, \underline{r}_n$ form an orthonormal basis.

Equivalently, P is orthogonal if and only if its columns $\underline{c}_1, \ldots, \underline{c}_n$ form an orthonormal basis.

Proof. The equation $^{-1} = I_n$ with $^{-1} =^T$ is exactly the claim for rows, given that we are using the usual dot product. (For columns, $^{-1} = I_n$ proves the claim.) ☐

Remark

Note that since $(PQ)^T = Q^T P^T$, we have that orthogonal matrices form a group. the Orthogonal group:

$$O(n) = \{P \in \mathrm{Mat}_{nn} \mid P \text{ is orthogonal}\}$$

By the observation that $\det P$ is 1 or -1 we have that

$$O(n) \subset \mathrm{SL}(n) = \{P \in \mathrm{Mat}_{nn} \mid \det P = 1 \text{ or } -1\}$$

is a subgroup of the Special Linear group. They are evidently not equal, since there are shear matrices

$$\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} \in \mathrm{SL}(n) \setminus O(n)$$

8.2 Symmetric matrices

Definition 8.3

A matrix $A \in \mathrm{Mat}_{nn}$ is **symmetric** if and only if $A^T = A$.

Proposition 8.4

Let $A \in \mathrm{Mat}_{nn}$ be a symmetric matrix. If λ, μ are distinct eigenvalues of A and $\underline{v}, \underline{w}$ are corresponding eigenvectors, then $\underline{v} \cdot \underline{w} = 0$.

Proof. Since $A\underline{v} = \lambda\underline{v}$ and $A\underline{w} = \mu\underline{w}$, we have

$$\underline{v}^T A\underline{w} = \underline{v}^T(A\underline{w}) = \underline{v}^T(\mu\underline{w}) = \mu\underline{v}^T\underline{w}$$

Similarly we have $\underline{w}^T A\underline{v} = \lambda\underline{w}^T\underline{v}$, which after transposing and using $A = A^T$ gives

$$\underline{v}^T A\underline{w} = (\underline{w}^T A^T \underline{v})^T = (\underline{w}^T A\underline{v})^T = (\lambda\underline{w}^T\underline{v})^T = \lambda\underline{v}^T\underline{w}$$

Subtracting these two expressions gives

$$\underline{0} = (\mu - \lambda)\underline{v}^T \underline{w}$$

Since $\lambda \neq \mu$ and $\underline{v} \cdot \underline{w} = \underline{v}^T \underline{w}$, the result follows. □

Corollary 8.5

Let $A \in \mathrm{Mat}_{nn}$ be a symmetric matrix. If A has n distinct eigenvalues $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, then there is an orthonormal matrix P such that $P^{-1}AP$ is diagonal.

Proof. Let $v_1, \dots, v_n$ be eigenvectors for $\lambda_1, \dots, \lambda_n$ respectively. Without loss of generality, each v_i has length $\|v_i\| = 1$: indeed the unit vector in the direction v_i is still an eigenvector for λ_i. By the proposition, $v_1, \dots, v_n$ are mutually orthogonal, and so they form an orthonormal basis. Let P be the matrix which has $v_1, \dots, v_n$ as its columns. By Proposition 8.2, P is an orthonormal matrix. The product $P^{-1}AP$ is diagonal by (the proof of) the Diagonalisation Theorem 7.27. □

It is evidently not true that every symmetric matrix has a set of n distinct real eigenvalues (you can write down diagonal matrices that do not have distinct diagonal entries, after all) but we can at least show that all of its eigenvalues are real.

Example

The 2×2 case is already familiar. If

$$A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

is a symmetric matrix, then $c_A(x) = x^2 - (a + c)x - b^2$. This quadratic has discriminant $(a + c)^2 + 4b^2 \geq 0$, so all its roots are real.

Proposition 8.6

Let $A \in \mathrm{Mat}_{nn}$ be a symmetric matrix. Then all the (complex) eigenvalues A lie in $\mathbb{R}$.

The proof is a nice example of the value of working in vector spaces more generally. We work in the vector space $\mathbb{C}^n$ of complex column vectors with scalars $\mathbb{C}$ (rather than $\mathbb{R}$ as we have throughout the module). Note that RREF works for matrices with entries in $\mathbb{C}$ (or any other field) in exactly the same way as it does for matrices with real entries.

Proof. Let $\lambda \in \mathbb{C}$ be any root of the characteristic polynomial $c_A(x) = \det(A - xI_n)$ (which is the same polynomial whether we are thinking of $\mathbb{R}^n$ or $\mathbb{C}^n$). If all such λ are real, then the proof is complete, so suppose λ is not real.

As λ is not real, we would not usually look for an eigenvalue, since λ is not a scalar. The novelty is that if we consider vectors in $\mathbb{C}^n$, then there is a corresponding eigenvector $\underline{v} \in \mathbb{C}^n$: we simply find any nonzero $\underline{v} \in \ker(A - \lambda I_n)$, and we may calculate that kernel using RREF as ever. That is, we have an equation

$$A\underline{v} = \lambda \underline{v} \tag{8.1}$$

where the entries of A are real but λ and the entries of $\underline{v}$ may be complex. The result follows by comparing the transpose of equation (8.1) with its complex conjugate.

First, if we transpose both sides of equation (8.1) we get $\underline{v}^T A^T = \lambda \underline{v}^T$, which is the same as

$$\underline{v}^T A = \lambda \underline{v}^T$$

since $A = A^T$ is a symmetric matrix.

Second, if we take complex conjugate of every complex number in equation (8.1) (writing $\overline{\underline{v}}$ for the column vector of the complex conjugates of the entries of $\underline{v}$), we have another equation

$$A\overline{\underline{v}} = \overline{\lambda}\overline{\underline{v}}$$

where $\overline{A} = A$ since A has real entries.

Combining these two equations gives

$$\lambda \underline{v}^T \overline{\underline{v}} = (\underline{v}^T A)\overline{\underline{v}} = \underline{v}^T (A\overline{\underline{v}}) = \underline{v}^T (\overline{\lambda}\overline{\underline{v}}) = \overline{\lambda}\underline{v}^T \overline{\underline{v}}$$

Certainly $\underline{v}^T \overline{\underline{v}} \neq 0$. (To check: if $\underline{v} = (a_1, \ldots, a_n)^T \in \mathbb{C}^n$ then $\underline{v}^T \overline{\underline{v}} = a_1 \overline{a_1} + \ldots + a_n \overline{a_n} = |a_1|^2 + \ldots + |a_n|^2$ can only be zero if $\underline{v} = \underline{0}$, which it is not as it is an eigenvector.) Therefore cancelling $\underline{v}^T \underline{v}$ gives $\lambda = \overline{\lambda}$, as required. $\square$

Example

Let

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

which is a symmetric matrix. The eigenvalues $\lambda = 1$ and -1 are indeed real, and in fact are distinct. We choose length 1 eigenvectors

$$\frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

respectively. They are clearly orthonormal. Using these as the columns of the change of basis matrix P gives

$$P = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

which is orthogonal since $P^T P = I_2$, and we quickly check that

$$P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

as the theorem claims.

Example

Let

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix}$$

which is a symmetric matrix. The eigenvalues are $\lambda = 2$ (with multiplicity 2) and $\lambda = -1$ (with multiplicity 1). The key is that in this case there is a basis of eigenvectors. We can of course make them length 1, and those corresponding to different eigenvalues are orthogonal, but the key is that in the 2-dimensional eigenspace we may choose the two vectors orthonormal: in general terms this is what Gram–Schmidt does for us, but in a simple concrete situation like this we may work by bare hands as follows.

The eigenspaces are

$$\ker \begin{pmatrix} -2 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & -1 \end{pmatrix} = \left\langle \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\rangle$$

and

$$\ker \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} = \left\langle \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} \right\rangle$$

Using those three vectors as the columns of a matrix P (in that order), you see at once that

$$P^{-1}AP = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

as the proof of the theorem shows.

Remark

In fact, symmetric matrices are diagonalisable by orthogonal matrices, but we need a little more technology to prove this in complete generality.