Linear Algebra

Dr Julia Goedecke

Year 2020-21

Contents

Preamble

The Notes

These notes are *not* verbatim what I will write in the lectures, but the content is exactly the same. (Every year some people complain about this, but this is why they are called my "lecture notes". Many other students appreciate that it is this way.) The main difference is that they have more complete sentences, and they may have some comments that I only said in lectures. I might try to make such comments green, but it may not be consistent.
If you find any errors and typos in the notes, please do let me know (email address is on www.julia-goedecke.de), even if they look trivial.
The notes are partly based on the book *Elementary Linear Algebra* by Anton and Rorres.

How to use the notes

These notes represent the material covered in the corresponding course in Leicester. They were accompanied by video lectures and live lectures with discussions, which are not freely available. The following comments refer back to these, but can be adapted to be more general advice to students.
Possible uses:

- ⋄ Use them for revision or when doing Workbook questions.
- ⋄ Look back at some details you didn't quite understand in lectures.
- ⋄ For automatic searching of a term/concept.
- ⋄ Many more ways: it is up to you to find out how they are most useful for you.

What to do during lecture videos

- ⋄ In maths lectures, you are not expected to understand everything in real time. You will have to go over some bits again, do exercises, work on the material to understand it fully.
- ⋄ Uni maths lectures will have more mathematical language and symbols than you might be used to from school. You will slowly get used to this. If you have a question on some particular notation that is not answered in lectures, do ask someone.
- ⋄ Make a note if there is something you need to come back to later.
- ⋄ Try to visualise the mathematical concepts.
- ⋄ If you think of a link to a topic you've seen before, make a quick note and explore it more after.

Some comments on note-taking

- ⋄ Many students (including myself) find that taking down notes in lectures is actually the best way to concentrate and to learn something. If this applies to you, set these notes aside for a while and use them just to fill in gaps later.
- ⋄ If you do take notes, make sure you also listen to what I say during the lecture (video or live). With videos, you can pause them or play on slower speed if you find you can't keep up.
- ⋄ If you can't concentrate on listening and taking notes at the same time (like my brother), then just use these notes instead and concentrate on listening.

- ◇ DO NOT try to read the notes (for the first time) while I'm lecturing the same material: you will most likely not hear what I say and not fully take in either.

So my advice is: use one of these three methods:

- ◇ Take notes in lectures. (And listen as well.)
- ◇ Read ahead, or pause the video and read in between, and just listen in lectures knowing you've already seen it in the notes.
- ◇ Listen in lectures hoping/trusting that it will all be written in the notes.

What to do in live lectures

- ◇ Participate in all activities, e.g. true/false questions or multiple choice questions.
- ◇ Think actively about the material, as prompted by the lecture.
- ◇ Ask questions to help you understand the material.

What to do between/after lectures

To help your understanding of the material:

- ◇ Do the exercises suggested.
- ◇ Try out statements of propositions and their proofs on an example after the lecture (video): this can help you understand it.
- ◇ Try to make your own examples for concepts.
- ◇ If you've tried lecture notes, friends, and exercises and still need help, use the lecturer's office hours or email them: they are happy to help you as well.
- ◇ Suggestion from older students: make your own bank of (new for you) mathematical symbols with their description. You will learn them better if you make this yourself than use someone else's. You can add to it through your university maths journey.

Colours

Text in green is a more informal comment, which often is meant to aid your understanding.
Some comments are in purple. I may not manage to be terribly consistent between green and purple, but purple comments might not be so informal.
Text in red is very important.

Definitions and results are underlayed in yellow.

I have also put a red border round especially important results.

Exercises

Sometimes in the notes I will say "exercise". The main reason to have these is for you to be able to check your understanding by doing a fairly straight-forward exercise yourself. As maths is not a spectator sport, it is important you keep doing exercises and don't just read notes passively. It is helpful for you to try exercises even outside the Workbook Questions I set you to hand in, so you get enough practice. You will also find Numbas questions which you can use for practice: they will give you immediate feedback, and many of them are randomised, which means you can try the same question again with different numbers, to get as much practice as you personally need. Another source of such exercises are situations in a proof where it says "similarly". This means you can try out that bit yourself, and it will be very close to what was done before.

Introduction

What is Linear Algebra about?
The ingredients are:

- ◇ vectors, e.g. $\begin{pmatrix}1\\2\end{pmatrix}$ or $\begin{pmatrix}3\\5\\4\end{pmatrix}$ or $\begin{pmatrix}x_1\\\vdots\\x_n\end{pmatrix}$ or v, ...
- ◇ scalars, i.e. real numbers, e.g. 0.5, $\sqrt{2}$, λ, ...
- ◇ matrices, e.g. $\begin{pmatrix}1&0\\0&1\end{pmatrix}$ or $\begin{pmatrix}3&5&2\\6&1&2\end{pmatrix}$, ...
- ◇ some other concepts we will see later.

Some of the most important things we will do with these ingredients are:

- ◇ **linear combinations**:
  - · **vector addition** e.g. $\begin{pmatrix}3\\5\end{pmatrix}+\begin{pmatrix}2\\1\end{pmatrix}=\begin{pmatrix}5\\6\end{pmatrix}$
  - · **scalar multiplication** e.g. $3\cdot\begin{pmatrix}1\\3\\5\end{pmatrix}=\begin{pmatrix}3\\9\\15\end{pmatrix}$

  Combine the two to get a linear combination:

  $$3\cdot\begin{pmatrix}1\\3\\5\end{pmatrix}+5\cdot\begin{pmatrix}-1\\2\\6\end{pmatrix}-\sqrt{2}\cdot\begin{pmatrix}1\\0\\0\end{pmatrix}$$

  Linear combinations will play a very imortant role throughout the course.
- ◇ Matrices acting on vectors, e.g.

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

  We will learn how this works in the course.

What do vectors stand for/represent?

- ◇ An arrow in two or three dimensions.
- ◇ Experimental data: n different measurements.
- ◇ Logistics: e.g. give locations of trucks in different warehouses.
- ◇ Images: hugh, saturation and brightness as well as position of a pixel.
- ◇ lots more.

What do matrices mean or stand for/represent?
There are many options, depending on the situation.

- ◇ a function sending a vector to matrix times vector
- ◇ a system of linear equations
- ◇ data (lots of different options, see above)
- ◇ ...

What does “linear” mean?
Intuitively, it has to do with “lines”, so no quadratics, cubics, etc. Just a scalar times a variable, or a scalar times a vector.

More directly: linear means we can add vectors or variables, and multiply them by a scalar. But we cannot do any squaring, roots, sin, cos, log, exponentials, ...
We will see that even this restriction gives a rich area of maths to study.

Why is Linear Algebra useful/important?
It is used in many many areas of maths and applications.

- ◇ Approximating functions by lines (via tangents) gives a linear situation. This is more useful than one might think.
- ◇ Used in solving differential equations.
- ◇ Can be used in Probability/Statistics.
- ◇ Used in Machine Learning.
- ◇ Used in a variety of numerical methods.
- ◇ Used in coordinate geometry, etc.
- ◇ many many more

CHAPTER 1

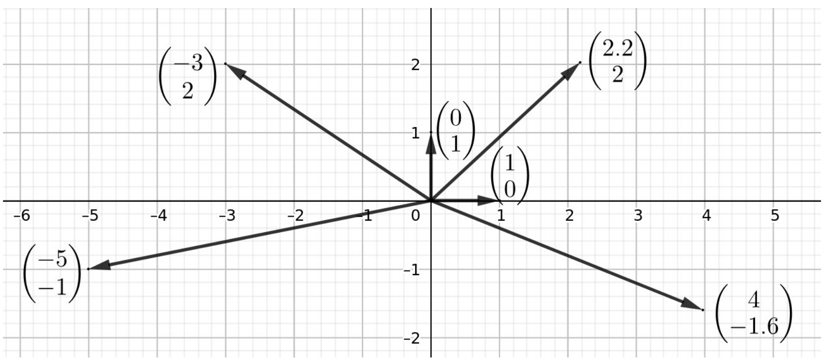
Vectors and Matrices

A. Vectors in two and three dimensions

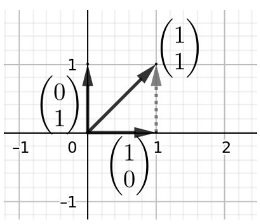
Since we said vectors are one of the main ingredients of Linear Algebra, let's start with vectors in the plane.

Definition 1.1: A **two-dimensional vector** has two entries written vertically, with both entries being real numbers. We also call this a **column vector**.

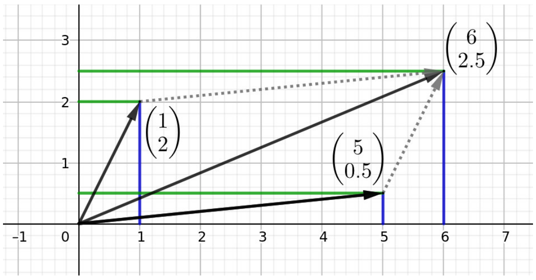
For example $\begin{pmatrix}1\\0\end{pmatrix}$, $\begin{pmatrix}0\\1\end{pmatrix}$, $\begin{pmatrix}-3\\2\end{pmatrix}$, $\begin{pmatrix}\sqrt{2}\\\frac{1}{2}\end{pmatrix}$.
You can think of this vector as describing a point in the plane, or a direction vector starting at the origin to that point in the plane.

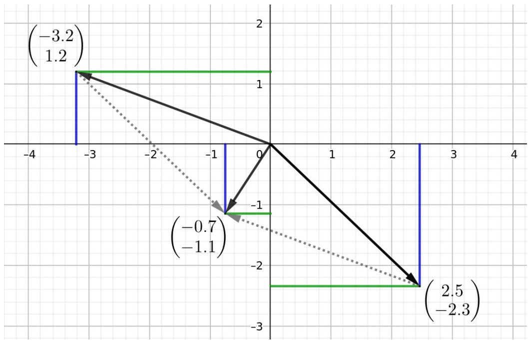


We write a general such vector as $\begin{pmatrix}x\\y\end{pmatrix}$ or as $\begin{pmatrix}x_1\\x_2\end{pmatrix}$. Varying over all possible combinations of real numbers x and y, we get the whole plane, which we write as $\mathbb{R}^2$ (denoting two entries of real numbers).
Given two such vectors, we can add them.



Geometrically, you add two vectors by putting one at the end of the other and then joining the origin to the second endpoint, as in the picture above.
But it is easy to see from the next two images that this corresponds exactly to adding the x-components and the y-components separately:





$$\begin{pmatrix}1\\2\end{pmatrix}+\begin{pmatrix}5\\0.5\end{pmatrix}=\begin{pmatrix}1+5\\2+0.5\end{pmatrix}=\begin{pmatrix}6\\2.5\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}-3.2\\1.2\end{pmatrix}+\begin{pmatrix}2.5\\-2.3\end{pmatrix}=\begin{pmatrix}-3.2+2.5\\1.2-2.3\end{pmatrix}=\begin{pmatrix}-0.7\\-1.1\end{pmatrix}$$

It is also clear from this algebraic way of adding them that the order does not matter:

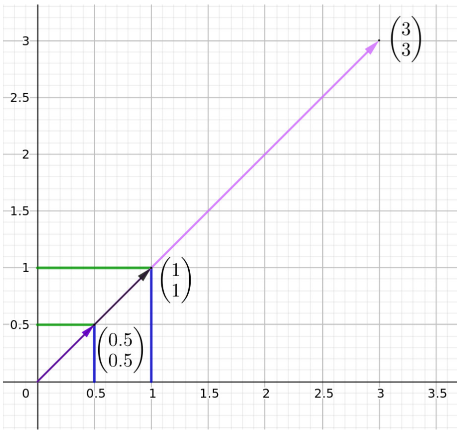
$$\begin{pmatrix}1\\2\end{pmatrix}+\begin{pmatrix}5\\0.5\end{pmatrix}=\begin{pmatrix}5\\0.5\end{pmatrix}+\begin{pmatrix}1\\2\end{pmatrix}$$

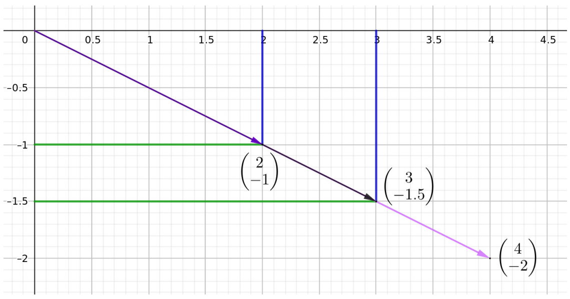
Definition 1.2: So in general, we define

$$\begin{pmatrix}x_1\\x_2\end{pmatrix}+\begin{pmatrix}y_1\\y_2\end{pmatrix}=\begin{pmatrix}x_1+y_1\\x_2+y_2\end{pmatrix}.$$

This is called **componentwise addition**, because we are adding each component or entry of the vector separately.

We can also scale a vector by a real number: leaving the direction the same and only changing the length of the vector.





$$\frac{1}{2}\cdot\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}0.5\\0.5\end{pmatrix} \quad \text{and} \quad 3\cdot\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}3\\3\end{pmatrix} \qquad \frac{2}{3}\cdot\begin{pmatrix}3\\-1.5\end{pmatrix}=\begin{pmatrix}2\\-1\end{pmatrix} \quad \text{and} \quad \frac{4}{3}\cdot\begin{pmatrix}3\\-1.5\end{pmatrix}=\begin{pmatrix}4\\-2\end{pmatrix}$$

Again we can calculate this algebraically:

Definition 1.3: We define **scalar multiplication** of a vector $\begin{pmatrix}x_1\\x_2\end{pmatrix}$ by a real number λ as

$$\lambda\begin{pmatrix}x_1\\x_2\end{pmatrix}=\begin{pmatrix}\lambda\cdot x_1\\\lambda\cdot x_2\end{pmatrix}.$$

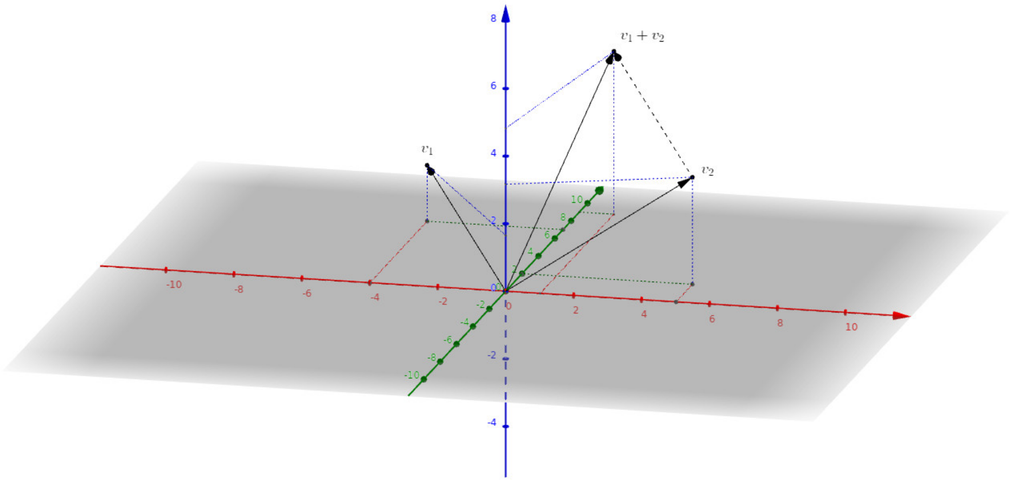
We take the scalar and multiply each entry of the vector by that scalar.
The symbol λ is called lambda, it is a greek letter we often use for scalars.

Exercise 1.4: Calculate the following: $\diamond\ 0 \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ $\diamond\ \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ $\diamond\ 3 \cdot \begin{pmatrix} \sqrt{3} \\ \pi \end{pmatrix} + 6 \cdot \begin{pmatrix} -\frac{\sqrt{3}}{2} \\ \pi \end{pmatrix}$

Now we can do the same thing for vectors in three-dimensional space.

Definition 1.5: A **three-dimensional vector** has three entries written vertically, with all entries being real numbers. The collection of all possible three-dimensional vectors is called $\mathbb{R}^3$, which is three-dimensional space.

For example $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 4 \\ -2 \\ 0 \end{pmatrix}$, $\begin{pmatrix} -\pi \\ 6.4 \\ -3 \end{pmatrix}$.
We can add such vectors just as we added two-dimensional vectors.



It is slightly harder to draw in three dimensions, so why don't you go to this GeoGebra picture and rotate it and play around with it to get a feel for it. `https://www.geogebra.org/3d/xnttqnrc`

Definition 1.6: For three-dimensional vectors, the componentwise addition is defined as

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \end{pmatrix}.$$

Scalar multiplication is defined as

$$\lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \lambda \cdot x_1 \\ \lambda \cdot x_2 \\ \lambda \cdot x_3 \end{pmatrix}.$$

So they are exactly the same rules as we had for two-dimensional vectors, doing everything componentwise. The only difference is that we now have three components.
You can see that we have a geometric interpretation and an algebraic way of calculating for these vectors in two and three dimensions. It is very useful to have both, but we will see soon that when we get into higher dimensions, it's harder to visualise geometrically. The topic is called Linear *Algebra* because we will focus on the algebraic side of doing things. Do keep the geometric intuition in two and three dimensions alongside this though. We will now and then fall back on it.

B. Lines and planes in $\mathbb{R}^2$ and $\mathbb{R}^3$

What can we do with this vector addition and scalar multiplication? Let's have a look at lines and planes.

Definition 1.7: A **line** in $\mathbb{R}^2$ is a set of points of the form

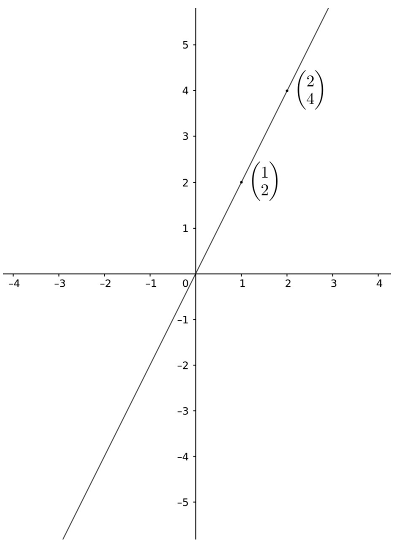
$$\left\{ \begin{pmatrix} a \\ b \end{pmatrix} + \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\}$$

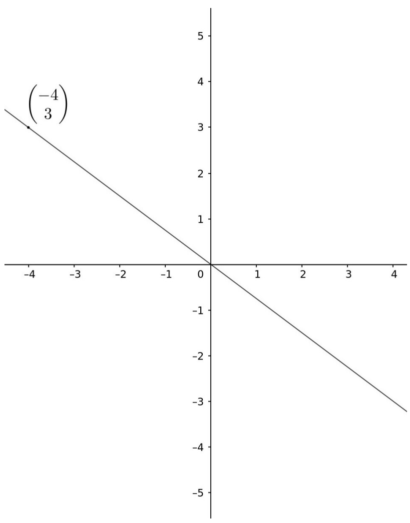
for fixed vectors $\begin{pmatrix} a \\ b \end{pmatrix}$ and $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. The scalar λ varies through all real numbers to give all points on the line. It is sometimes called a **parameter**. The vector $\begin{pmatrix} a \\ b \end{pmatrix}$ is some point on the line, and the vector $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ gives the direction of the line. This direction vector cannot be zero: $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix}$.

The curly brackets are called **set brackets**. You can think of a set as a collection of things. The symbol $\in$ means **is an element of**, and can also be read as "in". It means the thing on the left is an element of the set on the right, or is in the set on the right. The vertical line is read as "such that" or "with" or "where". So it is "the set of vectors of the form a, b plus lambda times x-one, x-two with lambda in R" (or where lambda is a real number).

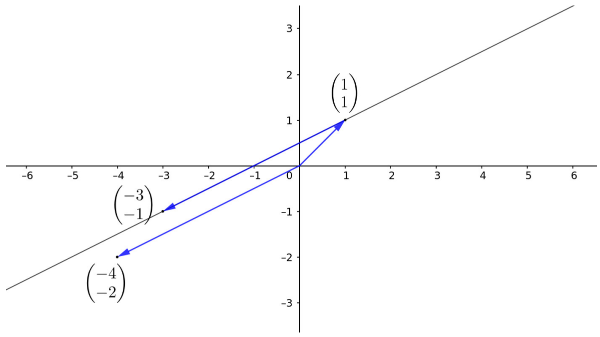
Examples 1.8:

- $\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 2 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\}$ is the line through the origin and the point $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$. If the line goes through the origin, we usually leave out the $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and just write $\left\{ \lambda \begin{pmatrix} 1 \\ 2 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\}$.
- Another line through the origin is $\left\{ \lambda \begin{pmatrix} -4 \\ 3 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\}$.





- ◇ The same way, with any vector, you get a line through the origin. Some of those are the same lines though: $\left\{\lambda\begin{pmatrix}1\\2\end{pmatrix}\middle|\lambda\in\mathbb{R}\right\}=\left\{\lambda\begin{pmatrix}2\\4\end{pmatrix}\middle|\lambda\in\mathbb{R}\right\}$
- ◇ We can also have lines that don't go through the origin, for example $\left\{\begin{pmatrix}1\\1\end{pmatrix}+\lambda\begin{pmatrix}-4\\-2\end{pmatrix}\middle|\lambda\in\mathbb{R}\right\}$. You can see from the image that the line does not go through $\begin{pmatrix}-4\\-2\end{pmatrix}$: this is the direction vector, so it is parallel to the line. When $\lambda=1$, we get the point $\begin{pmatrix}1\\1\end{pmatrix}+\begin{pmatrix}-4\\-2\end{pmatrix}=\begin{pmatrix}-3\\-1\end{pmatrix}$.



So here both vector addition and scalar multiplication play together.
It might be a little confusing at first that a vector can mean a point on the plane, as well as the direction vector from the origin to that point, or the same direction vector moved to somewhere else. However, if you do the algebra, you will always get the right answer. With some practice you will know when it is helpful to think of a vector as a point and when to think of it as a direction vector, and how to switch between the two. But if you're not sure, just do the algebra.

We know that lines are **one-dimensional**. Intuitively, this means we have only one direction we can move in (both positively and negatively). It can be helpful to think of it as having "one degree of freedom": there is one real number we can choose to determine a point on a line, namely the parameter λ.
There is a difference between lines that go through the origin and lines that don't go through the origin: if we take a point on a line and multiply it by some scalar, will it still be on the line? Or if we add two points on a line, will they still be on the line? The answer is yes if the line goes through the origin, and no if it doesn't.

Exercise 1.9:

- ◇ Verify that if $k\begin{pmatrix}x_1\\x_2\end{pmatrix}$ is any point on the line $\left\{\lambda\begin{pmatrix}x_1\\x_2\end{pmatrix}\middle|\lambda\in\mathbb{R}\right\}$, then any multiple $\mu\cdot k\begin{pmatrix}x_1\\x_2\end{pmatrix}$ (μ is read mu) still lies on the same line. And if $l\begin{pmatrix}x_1\\x_2\end{pmatrix}$ is a second point on the same line, verify that the sum of the two points is still on the line. Here $\begin{pmatrix}x_1\\x_2\end{pmatrix}\neq\begin{pmatrix}0\\0\end{pmatrix}$ is a fixed vector defining the line.

  If you find it tricky to get started, first try it in the case where you choose some actual numbers for x_1 and x_2, for example one of the lines through the origin we had earlier. You can also choose specific numbers for k and l, but then see if you can still do the calculations using the symbols k and l.
- ◇ Verify that for a line which does not go through the origin, adding two points on the line does not give a point on the line.

What would a line in three-dimensional space look like? We can use the same expression, just using vectors with three entries everywhere:

$$\left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \middle| \lambda \in \mathbb{R} \right\}$$

You can see it could be tedious to write everything out separately for two-dimensional vectors and three-dimensional vectors. So we use some convenient notation that lets us write both at once:

Definition 1.10: Writing u, v for vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, a **line** is a set of the form $\{u + \lambda \cdot v \mid \lambda \in \mathbb{R}\}$, where $v \neq 0$.

So u might be $\begin{pmatrix} a \\ b \end{pmatrix} \in \mathbb{R}^2$, or it might be $\begin{pmatrix} a \\ b \\ c \end{pmatrix} \in \mathbb{R}^3$, or as we'll see in the next section, even a vector with n entries. And similarly for v. This way we can write expressions that are shorter, and more general, so we don't have to write similar expressions for different situations which only vary in how many entries each vector has.
We tend to use letters u, v, w for vectors, and sometimes also x, y. We tend to use greek letters λ, μ, ν (lambda, mu, nu) for scalars, but you can also use k and l. For the real numbers which are the entries of a vector, we sometimes use x, y, z (then you'd have to know from context whether x is a real number or a vector, or say it in the sentence), and more often x_1, x_2, x_3 etc. Generally it needs to be made clear what a letter stands for, these are just guidelines.

Let's now look at planes.

Definition 1.11: A **plane** (in $\mathbb{R}^3$) is a set of the form

$$\{u + \lambda v + \mu w \mid \lambda, \mu \in \mathbb{R}\}$$

where u is a fixed vector on the plane, and v and w are two direction vectors which are not zero and do not go in the same direction.

What does "do not go in the same direction" mean here? Let's look at some examples.

Examples 1.12:

- Just as for lines, we can have planes that go through the origin, so where $u = 0$. For example, the x, y-plane in three-dimensional space is

$$\left\{ \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}.$$

  So every vector in this plane has the form $\begin{pmatrix} \lambda \\ \mu \\ 0 \end{pmatrix}$, so it is some point in the x, y-plane.

- Similarly we can have the x, z-plane

$$\left\{ \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$$

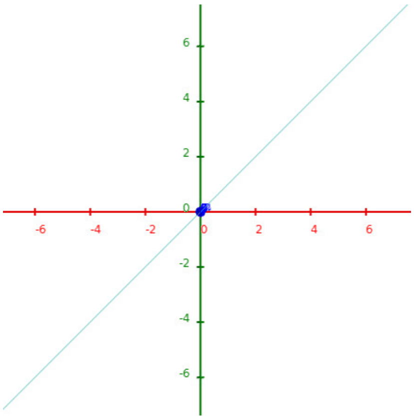
  and the y, z-plane

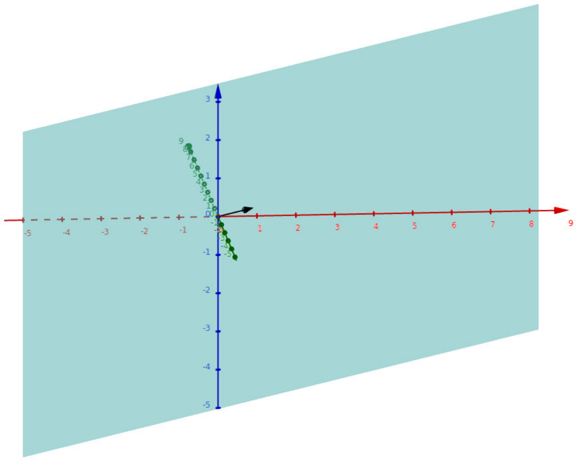
$$\left\{ \lambda \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}.$$

◇ Another plane through the origin is

$$\left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}.$$

If you look straight down onto the x, y-plane, you see it just as the line through $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, but if you rotate your viewpoint a bit, you can see that it is a vertical plane in that direction. You can rotate it in GeoGebra to get different viewpoints: `https://www.geogebra.org/3d/bkxrcnfk`





◇ Is this also a plane?

$$\left\{ \lambda \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$$

If we look at the expression algebraically, we can simplify it to $(\lambda + 3\mu) \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$. The two directions $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 3 \\ 3 \\ 0 \end{pmatrix}$ are not really different directions, they go in the same direction. This does not give us a plane, but only a line.

We can see from the last example that the condition "two direction vectors which do not go in the same direction" is crucial for getting a plane rather than a line. If one vector goes in the same direction (this includes with opposite sign) as the other, then it does not reach any other points than can be reached with just one vector.

Definition 1.13: We say two vectors u, v are **parallel** if one is a scalar multiple of the other: if $u = \lambda v$ for some $\lambda \in \mathbb{R}$ (or the other way round).

The "or the other way round" is to cover the situation where v might be the zero vector $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$. The zero vector is parallel to any other vector, because it is 0 times any other vector.

Two vectors which are not parallel are called **independent**. To have a plane, we need two independent directions. This is an example of an important concept called **linear independence** which we will study much more later on.

Exercise 1.14: Determine which of the following are planes. For each plane, also say whether it goes through the origin or not. (You can use the tick boxes given.) CAREFUL: while we said that if the plane (or line) goes through the origin, we don't use the fixed first vector (the one that does not have a parameter in front of it), the following planes may not be written in the most efficient way, so you have to check explicitely whether they go through the origin $\begin{pmatrix}0\\0\\0\end{pmatrix}$ or not.

a) $\left\{ \lambda \begin{pmatrix}1\\0\\4\end{pmatrix} + \mu \begin{pmatrix}8\\2\\0\end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$

b) $\left\{ \begin{pmatrix}1\\1\\1\end{pmatrix} + \lambda \begin{pmatrix}1\\2\\0\end{pmatrix} + \mu \begin{pmatrix}0\\4\\2\end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$

c) $\left\{ \begin{pmatrix}3\\1\\1\end{pmatrix} + \lambda \begin{pmatrix}1\\0\\3\end{pmatrix} + \mu \begin{pmatrix}-2\\0\\-6\end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$

d) $\left\{ \begin{pmatrix}1\\2\\1\end{pmatrix} + \lambda \begin{pmatrix}3\\6\\2\end{pmatrix} + \mu \begin{pmatrix}1\\2\\1\end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$

e) $\left\{ \begin{pmatrix}-1\\3\\\frac{1}{2}\end{pmatrix} + \lambda \begin{pmatrix}-2\\6\\2\end{pmatrix} + \mu \begin{pmatrix}1\\-3\\-\frac{3}{2}\end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$

Exercise 1.15: As for lines, show that if a plane goes through the origin, then adding two vectors on the plane gives another vector on the plane, and multiplying a vector on the plane by a scalar gives another vector on the plane.

A plane is **two-dimensional**: we can choose two different real numbers, the two parameters λ and μ. So we have "two degrees of freedom". Any choice of the two parameters gives a different point on the plane. Of course this is only true if the two direction vectors really are independent! If we choose only one parameter, we have not determined a point in the plane.

C. Vectorspace $\mathbb{R}^n$

We saw that we can write a line equation in a way that works for two-dimensional vectors and three-dimensional vectors at the same time. We also saw that while we can use our geometric intuition, we can also just use algebra to work out questions about vectors.
Now we are going to look at higher dimensions, where we can't easily visualise things geometrically any more. If we have more than three entries in a vector, we as humans can't easily imagine it in our three-dimensional space, but we can still do the algebra, and use some of the intuition from two and three dimensions to guide us.

Definition 1.16: A **vector in** $\mathbb{R}^n$ is a column vector with n entries, each of which is a real number:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$$

Vector addition and **scalar multiplication** work entrywise, as we saw in $\mathbb{R}^2$ and $\mathbb{R}^3$:

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n-1} \\ y_n \end{pmatrix} = \begin{pmatrix} x_1+y_1 \\ x_2+y_2 \\ \vdots \\ x_{n-1}+y_{n-1} \\ x_n+y_n \end{pmatrix} \quad \text{and} \quad \lambda \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_{n-1} \\ \lambda x_n \end{pmatrix}$$

Examples 1.17:
- ◇ When $n = 2$ or 3, these are just the examples we saw earlier.
- ◇ When $n = 1$, it is just a number: $\mathbb{R}^1 = \mathbb{R}$, and we add numbers and multiply numbers as we are used to.
- ◇ Here are some vectors in $\mathbb{R}^4$: $\begin{pmatrix} 4 \\ -2 \\ \sqrt{2} \\ 3 \end{pmatrix}, \begin{pmatrix} \frac{1}{2} \\ \frac{1}{100} \\ \pi \\ -4.2 \end{pmatrix}$
- ◇ Here are some vectors in $\mathbb{R}^5$: $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$
- ◇ If we want to say something about vectors of any length, we use dots to indicate the pattern of entries, as in $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$. If n happens to be 3, this would still be $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, even though the general vector gives four named entries with dots inbetween.
- ◇ We often use notation such as $v \in \mathbb{R}^n$ so we don't have to write out entries with dots. We can then still write entries when we need to.
- ◇ For any n, there is a vector with 0 in every entry, which we were calling "the origin" before. We also call this the **zero vector** and write it as 0.

Definition 1.18: A **linear combination** of vectors u, v is an expression of the form $\lambda u + \mu v$, with scalars $\lambda, \mu \in \mathbb{R}$.

This combines vector addition and scalar multiplication. The concept of linear combination will be very important throughout the course.

Examples 1.19:
- ◇ If we set $\lambda = \mu = 1$, then just the vector sum of two vectors is an example of a linear combination.
- ◇ Setting $\mu = 0$ gives λu as a possible linear combination.

◇ We can have linear combinations of more than two vectors:

$$\lambda \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \nu \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

is a linear combination of three vectors. (ν is pronounced nu)

Exercise 1.20: ◇ If v is any vector, what is $0 \cdot v$?

◇ Can you write any vector $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ as a linear combination of the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$?

◇ Can you write any vector $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \in \mathbb{R}^4$ as a linear combination of the vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$? Can you do it with just the first three of these vectors?

We see that we can add vectors of the same size, and multiply a vector by a scalar, which results again in a vector of the same size. Actually, these operations satisfy some rules:

Proposition 1.21: (Properties of vector addition and scalar mult)
Any vectors $u, v, w \in \mathbb{R}^n$ and any scalars $\lambda, \mu \in \mathbb{R}$ satisfy:

VA0	$u + v \in \mathbb{R}^n$	*(closure under vector addition)*
VA1	*The zero vector* 0 *satisfies* $v + 0 = v = 0 + v$.	*(zero vector)*
VA2	*There are negative vectors satisfying* $v + (-v) = 0 = (-v) + v$.	*(negative vectors)*
VA3	$(u + v) + w = u + (v + w)$	*(associativity of vector addition)*
VA4	$u + v = v + u$	*(commutativity of vector addition)*
SM0	$\lambda v \in \mathbb{R}^n$	*(closure under scalar multiplication)*
SM1	$1 \cdot v = v$	*(unit scalar)*
SM2	$\lambda \cdot (\mu v) = (\lambda \cdot \mu) v$	*(associativity of scalar mult)*
SM3	$(\lambda + \mu)\, v = \lambda v + \mu v$	*(distributivity of scalar mult over real addition)*
SM4	$\lambda(u + v) = \lambda u + \lambda v$	*(distribuitivity of scalar mult over vector addition)*

We call anything that satisfies these axioms a ***vector space****.*

PROOF. You can verify these easily using the definitions of vector addition and scalar multiplication. (If you're not sure, try it first for vectors in $\mathbb{R}^2$ and $\mathbb{R}^3$, and then see that it doesn't really change when you use vectors in $\mathbb{R}^n$.) □

We will come back to these axioms/rules later and see that not only vectors in $\mathbb{R}^n$ satisfy them, but also other sets of things. They will become our foundation of the theory of vector spaces, which is the foundation of linear algebra.

VA stands for vector addition, and SM stands for scalar multiplication. You might wonder why we start counting at 0: this is because the definition of vector addition and scalar multiplication already implicitely includes these two properties. But it's quite useful to state them anyway, as it reminds us to check them later when we want to show that something satisfies these rules.
Associativity is a property that allows us to "leave away brackets" when we are dealing with more than two vectors or more than one scalar. We say that vector addition and scalar multiplication are **associative**. If you think this is something unimportant and surely always true, think about the operation "to the power of". Is $2^{(3^2)} = (2^3)^2$? No! $2^{(3^2)} = 2^9$ and $(2^3)^2 = 2^6$.

Commutativity means that the order of addition doesn't matter. We will see some other kind of operations soon where the order does matter! You can also use "to the power of" again: 2^3 is not the same as 3^2. For vector addition, the order doesn't matter, because we're adding in each entry separately, and we know that when adding real numbers, the order doesn't matter. We say vector addition is **commutative**.
Distributivity tells us how to resolve brackets when we have addition (which could be vector addition, or the addition of real numbers in the scalars) and scalar multiplication. Again this property comes from the same property for real numbers.

We can define lines and planes in $\mathbb{R}^n$: while the intuition might not work any more of imagining an actual line in space, the algebraic way of defining these concepts still works:

Definition 1.22: A **line** in $\mathbb{R}^n$ is a set of the form $\{u + \lambda \cdot v \mid \lambda \in \mathbb{R}\}$, where $u, v \in \mathbb{R}^n$ are fixed vectors with $v \neq 0$.

Definition 1.23: Two vectors in $\mathbb{R}^n$ are called **parallel** if one is a scalar multiple of the other.

Definition 1.24: A **plane** in $\mathbb{R}^n$ is a set of the form $\{u + \lambda \cdot v + \mu w \mid \lambda, \mu \in \mathbb{R}\}$, where $u, v, w \in \mathbb{R}^n$ are fixed vectors with $v \neq 0 \neq w$, and v and w are not parallel.

We are particularly interested in lines and planes that go through the origin, because they themselves also satisfy the vector space axioms VA0-4 and SM0-4.

Example 1.25: Taking the x, y-plane $\left\{ \lambda \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \mu \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \middle| \lambda, \mu \in \mathbb{R} \right\}$ in $\mathbb{R}^3$, we see that it looks almost exactly like $\mathbb{R}^2$, except that we put a 0 as a third entry for each vector: given $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$, then $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ is in the x, y-plane of $\mathbb{R}^3$. As $\mathbb{R}^2$ satisfies the vector space conditions, so does the x, y-plane in $\mathbb{R}^3$.
Similarly the x, z-plane can be thought of as $\mathbb{R}^2$ with a zero inserted in the middle, and the y, z-plane as $\mathbb{R}^2$ with a zero added before the entries.

So we are now interested in such subsets that behave like (smaller) kinds of $\mathbb{R}^n$ themselves.

Definition 1.26: A **subspace** of $\mathbb{R}^n$ is a set of vectors in $\mathbb{R}^n$ that also satisfy all the vector space conditions VA0-4 and SM0-4.

Some of the properties stay true automatically when we look at a subset of $\mathbb{R}^n$: if $u + v = v + u$ for any vector $u, v \in \mathbb{R}^n$, then this property is automatically still true if we restrict u, v to a plane or a line or any other subset of vectors.

- VA3 associativity of vector addition
- VA4 commutativity of vector addition
- SM1 unit scalar
- SM2 associativity of scalar multiplication
- SM3 distributivity of scalar mult over vector addition
- SM4 distributivity of scalar mult over real addition

are all of the kind that stay true in any subset of vectors.
So if we have a given line or plane or similar kind of subset, what we have to check is:

VA0 If we add two vectors in the subset, is the sum still in the subset?
VA1 Is the zero vector in this subset?
VA2 If some vector is in the subset, is its negative also in the subset?
SM0 If we multiply a vector in the subset by some scalar, is it still in the subset?

If these are true, then we have a subset that is a subspace.

Notation 1.27: We write $S \subseteq \mathbb{R}^n$, a round symbol, for *subset*, and $S \leqslant \mathbb{R}^n$, a pointed symbol, for *subspace*.

Actually, we can reduce it even further:

Proposition 1.28: *A subset S of $\mathbb{R}^n$ is a subspace if and only if it satisfies:*

- $0 \in S$ *(zero vector is in the set)*
- *for any $u, v \in S$, $u + v \in S$* *(closed under vector addition)*
- *for any $v \in S$ and any $\lambda \in \mathbb{R}$, $\lambda v \in S$* *(closed under scalar mult)*

PROOF. If we check these three, then we automatically get negative vectors: $-v = (-1) \cdot v$. □

BE CAREFUL: you might think that we can get rid of checking whether 0 is in the set as well, because $0 \cdot v = 0$. But we do need to have *some* vector in the set: if there are no vectors at all, then the "closed under" conditions are vacuously true (because there are no cases to check). But if we don't have any vectors at all, then we also don't get 0, so it does not satisfy the conditions to be a vector space.

Examples 1.29:

- As we saw in an earlier exercise, a line through 0 (in $\mathbb{R}^2$, or $\mathbb{R}^3$ or $\mathbb{R}^n$) is a subspace. You can think of it as a copy of the real number line $\mathbb{R}$ inside $\mathbb{R}^n$: choosing the paramter λ, one real number, gives you a point on the line.
- A line which doesn't go through 0 is not a subspace: it fails all three of the conditions. These lines have their uses, but for much of the course we prefer lines that go through 0.
- A plane through the origin 0 is a subspace. A plane which does not go through 0 is not a subspace.
- You can have "bigger" subspaces as well:

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \end{pmatrix} \middle| x_1, x_2, x_3 \in \mathbb{R} \right\}$$

  is a subspace of $\mathbb{R}^4$ that looks a lot like $\mathbb{R}^3$. Similarly

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ 0 \\ x_4 \end{pmatrix} \middle| x_1, x_2, x_4 \in \mathbb{R} \right\} \quad \text{and} \quad \left\{ \begin{pmatrix} 0 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \middle| x_2, x_3, x_4 \in \mathbb{R} \right\} \quad \text{etc.}$$

- $\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1 - x_2 \\ 3x_3 \end{pmatrix} \middle| x_1, x_2, x_3 \in \mathbb{R} \right\}$ is a subspace of $\mathbb{R}^5$.
  - The zero vector is of that type: use $x_1 = x_2 = x_3 = 0$.

- Adding two vetors of that type gives another vector of that type:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1 - x_2 \\ 3x_3 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_1 - y_2 \\ 3y_3 \end{pmatrix} = \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \\ x_3 + y_3 \\ (x_1 + y_1) - (x_2 + y_2) \\ 3(x_3 + y_3) \end{pmatrix}$$

- A scalar multiple still stays of that type:

$$\lambda \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1 - x_2 \\ 3x_3 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \\ \lambda x_3 \\ (\lambda x_1) - (\lambda x_2) \\ 3(\lambda x_3) \end{pmatrix}$$

Exercise 1.30: For the first examples, show that they are a subspace by checking the three conditions. (This is meant to give you practice in remembering which conditions to check.)

Fact 1.31: These are all the possible subspaces of $\mathbb{R}^2$:

- ◇ Just the zero vector $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$.
- ◇ Any line through the origin.
- ◇ All of $\mathbb{R}^2$.

These are all possible subspaces of $\mathbb{R}^3$:

- ◇ Just the zero vector $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$.
- ◇ Any line through the origin.
- ◇ Any plane through the origin.
- ◇ All of $\mathbb{R}^3$.

We will not prove it now, but we will learn tools later in the course that prove this.

D. Matrices

Our second main ingredient for linear algebra is something called a matrix.

Definition 1.32: A **matrix** is a rectangular array of numbers or terms. A matrix has **rows** and **columns**. We refer to the **size of a matrix** as $m \times n$, where m is the number of rows and n is the number of columns. We write $\mathscr{M}_{m,n}$ for the set of all $m \times n$ matrices. If $m = n$, so a matrix has the same number of rows as columns, we call it a **square** matrix.

Read the size "m by n", not "m times n".
What does a matrix stand for or mean? The answer is: one of many things, depending on the context! A matrix can represent

- ◇ a collection of data,
- ◇ a function sending vectors of a certain size to other vectors,
- ◇ a system of linear equations,
- ◇ some other possibilities.

In the course we will mostly focus on a matrix representing a function, and a matrix representing a system of linear equations. To explain both of these, we will have to learn how to multiply a matrix times a vector. But let's just do some slightly easier things first.

Notation 1.33: We usually denote matrices by capital letters such as A, B, C etc, or perhaps M. We write a_{ij} or A_{ij} for the entries of A: the first index i gives the row and the second index j gives the column.

Examples 1.34:
- $\diamond$ $A = \begin{pmatrix} 1 & 0 \end{pmatrix}$ is a 1×2 matrix with entries $a_{11} = 1$ and $a_{12} = 0$.
- $\diamond$ $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ is a 2×2 matrix. This is a square matrix.
- $\diamond$ $B = \begin{pmatrix} \frac{3}{4} & -\frac{2}{5} & \frac{1}{2} \\ \frac{9}{102} & \pi & \sqrt{3} \end{pmatrix}$ is a 2×3 matrix. Some of the entries are $b_{12} = -\frac{2}{5}$, $b_{23} = \sqrt{3}$.
- $\diamond$ A vector in $\mathbb{R}^n$ can also be viewed as a $n \times 1$ matrix.
- $\diamond$ A 1×1 matrix is just a number: x. We usually leave out the brackets then.
- $\diamond$ $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}$ is a general 3×2 matrix.

Notice that many examples given in these notes and in other books use integers unproportionally often. This is just because we humans find it much easier to calculate with integers, so when we want to explain something new, we keep the surrounding things as simple as possible. In the "real world", you will not come across as many integers. But if we can let computers deal with the calculations, it only matters that we understand what is going on, and we can learn that just as well on integers.

Definition 1.35: Two matrices A and B are **equal** if they have the same size and all their entries agree.

Examples 1.36:
- $\diamond$ $\begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix}$ and $\begin{pmatrix} 3 & 2 & 0 \\ 1 & 3 & 0 \end{pmatrix}$ are not equal, as they don't have the same size.
- $\diamond$ $A = \begin{pmatrix} 3 & 2 \\ 1 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} x & 2 \\ 1 & y \end{pmatrix}$ are not equal, as not all their entries agree. Can you choose values for x and y such that $A = B$?
- $\diamond$ Let A and B be matrices with entries $A_{ij} = i^j$ and $B_{ij} = i^j$. Must $A = B$? NO: Suppose A is a 3×3 matrix but B is a 2×3 matrix. They are not the same, but we might not have noticed if we just check the formula for the entries.

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 4 & 8 \\ 3 & 9 & 27 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 4 & 8 \end{pmatrix}$$

Definition 1.37: In a square matrix, the entries A_{ii} are called the **diagonal entries**. The sum of the diagonal entries is called the **trace** of the matrix, written $\mathrm{tr}A$.

Example 1.38:

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \text{ has trace } \mathrm{tr}A = a_{11} + a_{22} + a_{33}.$$

Definition 1.39: A square matrix in which all diagonal entries are 1 and all other entries are 0 is called an **identity matrix** and written I, or I_n if it has size $n \times n$.

Examples 1.40:

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

Definition 1.41: Given two matrices A and B of the same size, their **matrix sum** $A + B$ is calculated entrywise: $(A + B)_{ij} = A_{ij} + B_{ij}$.
We can multiply a matrix by a scalar by multiplying each entry: $(\lambda A)_{ij} = \lambda A_{ij}$.

Examples 1.42: Matrices have to be the same size if we want to add them! We cannot add matrices of different sizes.

- ⋄ $\begin{pmatrix} 3 & 1 \\ 2 & 9 \end{pmatrix} + \begin{pmatrix} 7 & 2 \\ 8 & 3 \end{pmatrix} = \begin{pmatrix} 10 & 3 \\ 10 & 12 \end{pmatrix}$
- ⋄ $3 \cdot \begin{pmatrix} 1 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 9 \end{pmatrix}$
- ⋄ $4 \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 3 \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 3 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{pmatrix}$

  As for vectors, we call this combination of matrix sum and multiplying by scalars a **linear combination of matrices**.

Exercise 1.43: If it is possible, add the following matrices.

- ⋄ $\begin{pmatrix} 3 & 1 \\ 2 & 4 \end{pmatrix} + \begin{pmatrix} 2 & 1 \\ 9 & 3 \end{pmatrix}$
- ⋄ $\begin{pmatrix} 1 & 2 \\ 8 & 2 \end{pmatrix} + \begin{pmatrix} 9 & 1 & 0 \\ 3 & 8 & 4 \\ 9 & 2 & 2 \end{pmatrix}$
- ⋄ $\begin{pmatrix} 1 & 2 & 9 \end{pmatrix} + \begin{pmatrix} 2 & 3 & 2 \\ 9 & 8 & 1 \end{pmatrix}$
- ⋄ $4 \cdot \begin{pmatrix} 9 & 1 & 8 \\ 3 & 7 & 3 \\ 8 & 1 & 7 \end{pmatrix} + 2 \cdot \begin{pmatrix} 2 & 9 & 1 \\ 8 & 2 & 7 \\ 3 & 6 & 4 \end{pmatrix}$

Proposition 1.44: (Matrices form a vector space)
The set $\mathscr{M}_{m,n}$ of all matrices of a given size satisfy the conditions VA0-4 and SM0-4.

Proof. You can check this in the same way as for vectors: all the operations just happen in each entry. □

We can "turn matrices the other way around":

Definition 1.45: The **transpose** of an $m \times n$ matrix A is the $n \times m$ matrix A^T with entries $(A^T)_{ij} = A_{ji}$.

So the rows and columns are being swapped.

Examples 1.46: We can transpose square and non-square matrices.

$\diamond \begin{pmatrix} 1 & 3 & 2 \\ 9 & 2 & -1 \end{pmatrix}^T = \begin{pmatrix} 1 & 9 \\ 3 & 2 \\ 2 & -1 \end{pmatrix}$ $\quad \diamond \begin{pmatrix} 2 & 3 \\ 1 & 8 \end{pmatrix}^T = \begin{pmatrix} 2 & 1 \\ 3 & 8 \end{pmatrix}$ $\quad \diamond \begin{pmatrix} 3 & 2 & 1 \end{pmatrix}^T = \begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix}$

You can see that the diagonal entries stay the same in the transpose.

Proposition 1.47: (Properties of Transpose)
For any matrices A, B of the same size, and $\lambda \in \mathbb{R}$, we have
(i) $(A^T)^T = A$;
(ii) $(A+B)^T = A^T + B^T$;
(iii) $(\lambda A)^T = \lambda A^T$.

Proof. First check that the matrices on either side of each equation have the same size. Then:
(i) $((A^T)^T)_{ij} = (A^T)_{ji} = A_{ij}$
(ii) $((A+B)^T)_{ij} = (A+B)_{ji} = A_{ji} + B_{ji} = (A^T)_{ij} + (B^T)_{ij}$
(iii) $((\lambda A)^T)_{ij} = (\lambda A)_{ji} = \lambda A_{ji} = \lambda (A^T)_{ij}$ □

Example 1.48:

$$\left(\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T\right)^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

E. Matrix multiplication

So now we will learn the crucial way that a matrix acts on a vector.

Notation 1.49: Given a matrix A, we refer to the columns by $A_1, A_2, \dots, A_n$ or $a_1, a_2, \dots, a_n$:

$$\begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$$

Definition 1.50: (Matrix times vector) An $m \times n$ matrix A can act on a vector $v \in \mathbb{R}^n$ to give a vector $Av \in \mathbb{R}^m$, in the following way:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n-1} & a_{2n} \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ a_{m-11} & a_{m-12} & \cdots & a_{m-1n-1} & a_{m-1n} \\ a_{m1} & a_{m2} & \cdots & a_{mn-1} & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \\ a_{m-11}x_1 + a_{m-12}x_2 + \cdots + a_{m-1n}x_n \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n \end{pmatrix}$$

or

$$\begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} \uparrow \\ a_1 \\ \downarrow \end{pmatrix} + x_2 \begin{pmatrix} \uparrow \\ a_2 \\ \downarrow \end{pmatrix} + \cdots + x_{n-1} \begin{pmatrix} \uparrow \\ a_{n-1} \\ \downarrow \end{pmatrix} + x_n \begin{pmatrix} \uparrow \\ a_n \\ \downarrow \end{pmatrix}$$

Examples 1.51: ◇ If the matrix is just one row, you get a single number as the answer, and you might have seen something like this as a "dot product of vectors":

$$\begin{pmatrix} a & b & c & d & e \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = ax_1 + bx_2 + cx_3 + dx_4 + ex_5$$

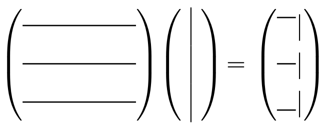
◇ The 2×2 case is perhaps easiest written as

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

◇ Here is the 3×3 case written out:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ a_{31}x_1 + a_{32}x_2 + a_{33}x_3 \end{pmatrix}$$

◇ Here is a visual picture that might help:



Think about "a row times a column": first entry in row times first entry in column, plus second entry in row times second entry in column, and so on. So it's just several times the very first example.

◇ Here's an example with numbers:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix} = \begin{pmatrix} 1 \cdot 7 + 2 \cdot 8 + 3 \cdot 9 \\ 4 \cdot 7 + 5 \cdot 8 + 6 \cdot 9 \end{pmatrix}$$

◇ Multiplying a vector by the identity matrix does not change it:

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

◇ The second way of viewing it can be helpful sometimes: A matrix times a vector is the linear combination of the columns of the matrix, with the entries of the vector as the scalars for the linear combination. Using the second expression of the definition in the 3×3 case, it looks like this:

$$\begin{pmatrix} \uparrow & \uparrow & \uparrow \\ a_1 & a_2 & a_3 \\ \downarrow & \downarrow & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} \uparrow \\ a_1 \\ \downarrow \end{pmatrix} + x_2 \begin{pmatrix} \uparrow \\ a_2 \\ \downarrow \end{pmatrix} + x_3 \begin{pmatrix} \uparrow \\ a_3 \\ \downarrow \end{pmatrix}$$

This viewpoint tends to be more useful in understanding of certain properties and relationships than for explicit calculation, though you can use it if you find it helpful.

Exercise 1.52: Calculate the following:

◇ $\begin{pmatrix} 3 & 2 \\ 5 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ ◇ $\begin{pmatrix} 1 & 3 & 2 \\ 9 & 2 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$ ◇ $\begin{pmatrix} 1 & 9 & 0 \\ 8 & 0 & 1 \\ 0 & 2 & 8 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ ◇ $\begin{pmatrix} 1 & 9 \\ 3 & 2 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} -1 \\ 2 \end{pmatrix}$

This now allows us to view a matrix as a function:

Definition 1.53: An $m \times n$ matrix A can be viewed as a function $T_A \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$: it takes a vector $v \in \mathbb{R}^n$ and sends it to $T_A(v) = Av$, the matrix times the vector. We call this a **matrix transformation**.

Examples 1.54:

- $\diamond$ $A = \begin{pmatrix} 3 & 2 \\ 5 & 1 \end{pmatrix}$ sends a vector $\begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2$ to the vector $\begin{pmatrix} 3x + 2y \\ 5x + y \end{pmatrix} \in \mathbb{R}^2$, so it represents a matrix transformation $T_A \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$.
- $\diamond$ $B = \begin{pmatrix} 1 & 3 & 2 \\ 9 & 2 & -1 \end{pmatrix}$ gives a matrix transformation $T_B \colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$.
- $\diamond$ $C = \begin{pmatrix} 1 & 9 \\ 3 & 2 \\ 2 & -1 \end{pmatrix}$ gives a matrix transformation $T_C \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3$, the other way round to B.
- $\diamond$ An identity matrix I_n corresponds to the **identity function** $\mathrm{id} \colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$: each vector gets sent to itself. $I_n v = v$, so $T_{I_n}(v) = v$.

As well as multiplying a matrix times a vector, we can multiply matrices of matching sizes.

Definition 1.55: (Matrix multiplication) If A is a $m \times r$ matrix and B is a $r \times n$ matrix, then we can form the **matrix product** AB with entries $(AB)_{ij} = \sum\limits_{k=1}^{r} A_{ik} B_{kj}$. This is a $m \times n$ matrix.
We can also write down the matrix product using the columns of the second matrix: if

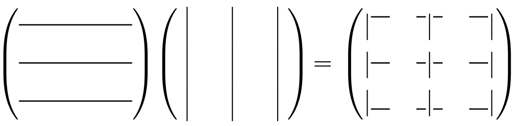
$$B = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ b_1 & \cdots & b_k & \cdots & b_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix},$$

then

$$AB = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ Ab_1 & \cdots & Ab_k & \cdots & Ab_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix},$$

using the matrix times vector operation we have already defined.

You can visualise it:



The entry in row i, column j of the product AB is calculated by taking row i of A times column j of B, as calculated in the first example of matrix times vector.
Important! You can only multiply matrices if their sizes match. Each single entry is calculated by a row of the first matrix times a column of the second matrix, and this is only possible if those rows have the same size as the columns. So to be able to multiply two matrices, the middle numbers of their sizes have to match:

$$\begin{array}{ccccc} A & \cdot & B & = & AB \\ m \times r & & r \times n & & m \times n \end{array}$$

And this middle matching size "drops out" to give you the size of the product.
You can see that a similar pattern of middle matching index happens in the formula for an entry of the matrix product.

Examples 1.56:

- $\diamond$ Our previous examples of matrix times vector still are examples for this, as a vector can be seen as a matrix with just one column. So in particular, if

$A = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \end{pmatrix}$ is a 1×5 matrix and $B = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix}$ is a 5×1 matrix, the product AB will be a 1×1 matrix, i.e. a number.

$$\begin{pmatrix} a_1 & a_2 & a_3 & a_4 & a_5 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} = a_1b_1 + a_2b_2 + a_3b_3 + a_4b_4 + a_5b_5 = \sum_{k=1}^{5} a_kb_k = \sum_k A_{1k}B_{k1}$$

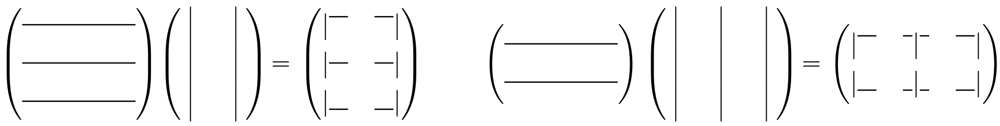
So for any other example, per entry we are doing exactly this calculation with the appropriate row from the first matrix and the appropriate column from the second matrix.

◇ If A and B are both 2×2 matrices, we get

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{pmatrix}$$

If you look carefully, you can see that for the i, jth entry, we have $a_{i1}b_{1j} + a_{i2}b_{2j}$. This is the sum in the definition written out for this case.

◇ Here are visualisations for non-square examples:



3×3 3×2 3×2 2×3 3×3 2×3

◇ If A is a 3×3 matrix and B is a 3×2 matrix, as in the first visual example above, then for example the entry in the first row and second column of the product is

$$(AB)_{12} = a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} = \sum_k a_{1k}b_{k2}$$

◇ Here are some examples with actual numbers:

$$\underset{3\times 3}{\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}} \underset{3\times 2}{\begin{pmatrix} 10 & 11 \\ 12 & 13 \\ 14 & 15 \end{pmatrix}} = \underset{3\times 3}{\begin{pmatrix} 1\cdot 10 + 2\cdot 12 + 3\cdot 14 & 1\cdot 11 + 2\cdot 13 + 3\cdot 15 \\ 4\cdot 10 + 5\cdot 12 + 6\cdot 14 & 4\cdot 11 + 5\cdot 13 + 6\cdot 15 \\ 7\cdot 10 + 8\cdot 12 + 9\cdot 14 & 7\cdot 11 + 8\cdot 13 + 9\cdot 15 \end{pmatrix}}$$

$$\underset{3\times 2}{\begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{pmatrix}} \underset{2\times 3}{\begin{pmatrix} 10 & 11 & 12 \\ 13 & 14 & 15 \end{pmatrix}} = \underset{3\times 3}{\begin{pmatrix} 1\cdot 10 + 2\cdot 13 & 1\cdot 11 + 2\cdot 14 & 1\cdot 12 + 2\cdot 15 \\ 3\cdot 10 + 4\cdot 13 & 3\cdot 11 + 4\cdot 14 & 4\cdot 12 + 4\cdot 15 \\ 5\cdot 10 + 6\cdot 13 & 5\cdot 11 + 6\cdot 14 & 5\cdot 12 + 6\cdot 15 \end{pmatrix}}$$

CAREFUL! In general, $AB \neq BA$. Indeed, for most size of matrices, these products are not both defined. But even for square matrices, it is not the case:

Example 1.57:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 4 \\ 3 & 8 \end{pmatrix}$$

but

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 6 & 8 \end{pmatrix}.$$

There is one matrix which is very easy to multiply by:

Proposition 1.58: (Product with identity matrix)
If A is an $m \times n$ matrix, then $AI_n = A$ and $I_m A = A$: multiplying by an identity matrix (of the correct size) does not change a matrix.

You can think of this as the matrix equivalent of multiplying by 1.

PROOF. By definition of matrix multiplication, $(AI_n)_{ij} = \sum_{k=1}^{n} A_{ik}(I_n)_{kj}$. But $(I_n)_{kj} = 0$ when $k \neq j$, and $(I_n)_{jj} = 1$. So this sum reduces to $(AI_n)_{ij} = \sum_{k=1}^{n} A_{ik}(I_n)_{kj} = A_{ij}(I_n)_{jj} = A_{ij}$. So A and AI_n have exactly the same entries (and the same size), so $A = AI_n$.
Similarly for $I_m A$. □

When it says "Similarly" in a proof, this is a "hidden exercise": you can test whether you really understand what is going on by trying to do that part yourself. First try to do it without looking at the previous part of the proof, and only use that as a hint if you get stuck.
When we multiply more than two matrices, it does not matter how we set brackets:

Proposition 1.59: (Associativity of matrix multiplication)
Matrix multiplication is associative: $(AB)C = A(BC)$ for any matrices whose sizes make this multiplication possible.

PROOF. First we can work out that the sizes of $(AB)C$ and $A(BC)$ must be the same. (Exercise.)
Let's work out an entry of the expressions on both sides.

$$((AB)C)_{ij} = \sum_k (AB)_{ik} C_{kj} = \sum_k \left(\sum_s A_{is} B_{sk} \right) C_{kj} = \sum_k \sum_s A_{is} B_{sk} C_{kj}$$

where the last step is multiplying out the bracket.

$$(A(BC))_{ij} = \sum_s A_{is} (BC)_{sj} = \sum_s A_{is} \left(\sum_k B_{sk} C_{kj} \right) = \sum_s \sum_k A_{is} B_{sk} C_{kj}$$

We can see that both expressions are the same, so $A(BC) = (AB)C$. □

We can think of this result as saying: When multiplying several matrices, we don't need brackets.

Exercise 1.60: Let A be $m \times r$, B be $r \times l$ and C be $l \times n$. Write out the above proof with expanded summations. E.g.

$$((AB)C)_{ij} = (AB)_{i1}C_{1j} + (AB)_{i2}C_{2j} + (AB)_{i3}C_{3j} + \cdots + (AB)_{il}C_{lj}$$

and then the same for each $(AB)_{ik}$ entry and so on.
Purpose of the exercise: see how convoluted matrix multiplication can get; this is not as obvious a result as $(A + B) + C = A + (B + C)$! Also see the advantage of the summation notation.

Given a square matrix, we can multiply it with itself again and again:

Definition 1.61: If A is a square matrix, then the k**th power** of A is

$$A^k = AA \cdots A,$$

the product of k copies of A. We define $A^0 = I$.

Proposition 1.62: (Transpose of product)
$(AB)^T = B^T A^T$ for all matrices A and B for which the product AB is defined.

PROOF. First we have to verify that $(AB)^T$ and B^TA^T have the same size. (Exercise.) Then we look at an entry.

$$((AB)^T)_{ij} = (AB)_{ji} = \sum_k A_{jk}B_{ki}$$

and

$$(B^TA^T)_{ij} = \sum_k (B^T)_{ik}(A^T)_{kj} = \sum_k B_{ki}A_{jk}.$$

As both expressions are the same, we have $(AB)^T = B^TA^T$. $\square$

Example 1.63:

$$\left(\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\right)^T = \begin{pmatrix} 1 & 4 \\ 1 & -2 \end{pmatrix}^T = \begin{pmatrix} 1 & 1 \\ 4 & -2 \end{pmatrix}$$

but

$$\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}^T \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^T = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}.$$

On the other hand

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^T \begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}^T = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 4 & -2 \end{pmatrix}.$$

If you ever need to work out what $(AB)^T$ might be and can't quite remember, then taking non-square matrices and working out the relevant sizes can help you work out what it must be.

Example 1.64:

$$\underset{2\times 3}{\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}} \underset{3\times 4}{\begin{pmatrix} 10 & 11 & 12 & 13 \\ 14 & 15 & 16 & 17 \\ 18 & 19 & 20 & 21 \end{pmatrix}} = AB$$

can be multiplied.

$$\underset{4\times 3}{\begin{pmatrix} 10 & 14 & 18 \\ 11 & 15 & 19 \\ 12 & 16 & 20 \\ 13 & 17 & 21 \end{pmatrix}} \underset{3\times 2}{\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}} = B^TA^T$$

can also be multiplied, but

$$\underset{3\times 2}{\begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}} \underset{4\times 3}{\begin{pmatrix} 10 & 14 & 18 \\ 11 & 15 & 19 \\ 12 & 16 & 20 \\ 13 & 17 & 21 \end{pmatrix}} = A^TB^T$$

cannot be multiplied: the sizes don't fit.

We saw that a matrix can act as a function, so what happens when we multiply two matrices?

Definition 1.65: Given any functions $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$, the function $g\circ f\colon X \longrightarrow Z$ defined by $(g\circ f)(x) = g(f(x))$ is the **composite** of g and f.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "g\circ f"'] & Y \arrow[d, "g"] \\ & Z \end{tikzcd}$$

This composite is just doing first f and then g. Notice that when we write $g \circ f$, we do f first: it is easy to remember if you imagine applying it to an element x, as the f is next to the x.

Proposition 1.66: (Composite matrix transformation)
Given two matrices $A \in \mathscr{M}_{m,r}$ and $B \in \mathscr{M}_{r,n}$, the composite of their matrix transformations $T_A \colon \mathbb{R}^r \longrightarrow \mathbb{R}^m$ and $T_B \colon \mathbb{R}^n \longrightarrow \mathbb{R}^r$ is the matrix transformation of the product:

$$T_{AB} = T_A \circ T_B \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m.$$

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "T_B"] \arrow[dr, "T_{AB}"'] & \mathbb{R}^r \arrow[d, "T_A"] \\ & \mathbb{R}^m \end{tikzcd}$$

PROOF. For any $v \in \mathbb{R}^n$, we have $T_{AB}(v) = ABv$, and also $(T_A \circ T_B)(v) = T_A(T_B(v)) = ABv$. so $T_{AB} = T_A \circ T_B$. □

Finally, we've already mentioned that linear combinations are one of the most important concepts in Linear Algebra, and matrix multiplication preserves linear combinations.

Proposition 1.67: (Matrix multiplication is linear)
Matrix multiplication preserves linear combinations. That is, if A, A' are $m \times r$ matrices and B, B' are $r \times n$ matrices, and $\lambda, \mu \in \mathbb{R}$, then

$$A(\lambda B + \mu B') = \lambda AB + \mu AB' \qquad \text{and} \qquad (\lambda A + \mu A')B = \lambda AB + \mu A'B.$$

PROOF. To show the two matrices on each side of the equation are equal, we have to check that they have the same size, and that all entries agree.
A is an $m \times r$ matrix, and $\lambda B + \mu B'$ is an $r \times n$ matrix, since both B and B' are. So $A(\lambda B + \mu B')$ is an $m \times n$ matrix. So too are AB and AB', so $\lambda AB + \mu AB'$ is also an $m \times n$ matrix, so the sizes are the same. We now look at an entry of each:

$$\begin{aligned} (A(\lambda B + \mu B'))_{ij} &= \sum_{k=1}^{r} A_{ik}(\lambda B + \mu B')_{kj} && \text{def of matrix mult} \\ &= \sum_{k=1}^{r} A_{ik}(\lambda B_{kj} + \mu B'_{kj}) && \text{def of matrix sum} \\ &= \sum_{k=1}^{r} \lambda A_{ik} B_{kj} + \mu A_{ik} B'_{kj} && \text{calculating in } \mathbb{R} \\ &= (\lambda AB + \mu AB')_{ij} && \text{def of matrix mult} \end{aligned}$$

Similarly for the second equation. It is a good exercise for you to do the proof of the other equation: try it without looking at this proof, and only use it as hints. This will help you see if you understand, and find the points where you might need to ask some more questions or go over something again.
You could also try writing out the summations in long-hand. Maybe try it for 2×2 matrices. □

Corollary 1.68: *In particular, matrix transformations are linear. That is,*

$$T_A(\lambda u + \mu v) = \lambda T_A(u) + \mu T_A(v).$$

PROOF. This is because matrix transformations are calculated with matrix multiplication: $T_A(v) = Av$. □

We won't use this very much now, but it is one of the most important concepts of Linear Algebra, so seeing it several times will hopefully help you to become friends with it.

F. Vectors and Matrices: Study guide

At the end of each Chapter, you will see a study guide like this, with concept review and skills.
The concept review lists concepts you need to understand: can you write down the definition or statement of this? Can you give some examples and counterexamples of this concept?

The skills are things you need to be able to do, so you need to practise them.
You can use these lists to help you check if you have covered the important things of the section, while you're learning it and also when you come to revision.

Concept review.

- ◇ Vectors in $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^n$.
- ◇ Linear combination.
- ◇ Lines and planes in $\mathbb{R}^2$, $\mathbb{R}^3$ and $\mathbb{R}^n$.
- ◇ (Informally) degree of freedom, e.g. of lines, planes.
- ◇ Parallel vectors.
- ◇ Vector space axioms for $\mathbb{R}^n$.
- ◇ Subspaces of $\mathbb{R}^n$.
- ◇ Matrices, size of matrix, square matrix.
- ◇ Matrix as function: matrix transformation.
- ◇ Transpose matrix.
- ◇ Linearity of matrix multiplication.

Skills.

- ◇ Add vectors, multiply a vector by a scalar, form linear combinations of vectors.
- ◇ Determine whether a given set is a line or a plane, and whether it goes through 0 or not.
- ◇ Determine whether two vectors are parallel.
- ◇ Determine whether a given set is a subspace of $\mathbb{R}^n$.
- ◇ Add matrices, multiply a matrix by a scalar, form linear combinations of matrices.
- ◇ Form the transpose of a matrix.
- ◇ Multiply a matrix and a vector.
- ◇ Multiplly two matrices of matching size.
- ◇ Determine whether two matrices can be multiplied or not.

CHAPTER 2

Linear Systems

Another way to view matrices is that they represent a linear system. We will see in this chapter what a linear system is, how we can solve it using matrices, and what form the solutions take. One reason for doing this is that it is a very efficient way of solving linear systems. Also, when we learn more linear algebra theory later in the course, we will come across the need to solve such linear systems again and again in different contexts. So this chapter will give you an essential tool for the rest of the course, and beyond.

A. Linear equations

Definition 2.1: A **linear equation in n unknowns (or variables)** is an equation which only involves scalar multiples of the unknowns on one side of the equation, and a constant on the other side:

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

with $x_1, \cdots, x_n$ the unknowns, and $a_1, \cdots, a_n, b \in \mathbb{R}$.
A linear equation is called **homogeneous** if $b = 0$.

Examples 2.2:
- ⋄ $3x-4y = 0$ and $3x-4y = 8$ are linear equations with two unknowns/variables. The first is homogeneous, the second is not: it is inhomogeneous.
- ⋄ $\pi x_1 + 3.5x_2 - 4x_3 + x_4 = \sqrt{3}$ is a linear equation with four unknowns.

Examples 2.3: (Counterexamples) The following are *not* linear:
- ⋄ $x^2 + 2y = 0$, $xy = 1$: products of variables are not allowed.
- ⋄ $2x - \sqrt{y} = 1$: no roots of variables are allowed.
- ⋄ $\sin(x) - \cos(y) = 1$: functions like sine, cos, log etc are not allowed in linear equations.
- ⋄ $e^x = 2$: unknowns can't appear in exponents.

We are interested in the solutions to such linear equations. Let's see what these solutions have to do with lines.

Example 2.4: The linear equation $3x - 4y = 0$ has many solutions: we can pick y to be what we want, and then x is determined. For example, $\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 4 \\ 3 \end{pmatrix}$ is a solution. Any other solution is a multiple of this: $\begin{pmatrix} 4t \\ 3t \end{pmatrix}$, or $t\begin{pmatrix} 4 \\ 3 \end{pmatrix}$. So the set of all solutions is a line $\left\{ t\begin{pmatrix} 4 \\ 3 \end{pmatrix} \middle| t \in \mathbb{R} \right\}$.
If we now take the same equation but make it inhomogeneous, we will get something similar: $3x - 4y = 2$ has solutions $\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} + t\begin{pmatrix} 4 \\ 3 \end{pmatrix} \middle| t \in \mathbb{R} \right\}$. This is a line which does not go through zero, but is parallel to the line that solved the homogeneous equation. $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$ is a **particular solution** of the inhomogeneous equation, but since $3 \cdot 4 - 4 \cdot 3 = 0$, adding a multiple of $\begin{pmatrix} 4 \\ 3 \end{pmatrix}$ to this particular solution still gives a solution: $3(2 + t \cdot 4) - 4(1 + t \cdot 3) = (3 \cdot 2 - 4 \cdot 1) + t(3 \cdot 4 - 4 \cdot 3) = 2 + t \cdot 0$.

As we have two variables and one equation, we have "one degree of freedom": we can choose one of the two variables, and then the other is determined by the equation. "One degree of freedom" corresponds to a line.

Example 2.5: The solutions to a linear equation with three variables form a plane rather than a line: now we have "two degrees of freedom."

$2x_1 - 3x_2 + 6x_3 = 0$ has solutions $\left\{ \begin{pmatrix} \frac{3}{2}s - 3t \\ s \\ t \end{pmatrix} \middle| s, t \in \mathbb{R} \right\}$: we can choose two of the variables, $x_2 = s$ and $x_3 = t$, and then x_1 is determined by the equation.

If we write this as $\left\{ s \cdot \begin{pmatrix} \frac{3}{2} \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \middle| s, t \in \mathbb{R} \right\}$, we can see that this is indeed a plane. We can choose s and t independently, which gives us two possible solution vectors which are not parallel.

What happens if we have several such equations? Then a solution to the whole system is the intersection of all the solutions to the individual equations.

Definition 2.6: A **system of linear equations** is a collection of one or more linear equations. Such a system is called **homogeneous** if all the equations in it are homogeneous. A **solution of a system of linear equations** is an n-tuple $x_1, x_2, \ldots, x_n$ which solves all equations simultaneously.

Notice that one solution of the system needs to provide a value for each unknown: all these together form one solution. We can think of a solution as a vector $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$.

Examples 2.7:

- ◇ Consider the linear system

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

These two equations represent two lines:

$$\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} + t \begin{pmatrix} 4 \\ 3 \end{pmatrix} \right\} \quad \text{and} \quad \left\{ s \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$$

A solution to the system is an intersection of the two lines. You know how to solve such a system from school: the second equation gives $y = 2x$, and using this in the first, we get $3x - 8x = 2$, or $x = -\frac{2}{5}$. Then $y = -\frac{4}{5}$. So $\begin{pmatrix} -\frac{2}{5} \\ -\frac{4}{5} \end{pmatrix}$ is the solution to this sytem. There is only one solution.

- ◇ This system has no solutions:

$$\begin{aligned} 3x - 4y &= 0 \\ 6x - 8y &= 4 \end{aligned}$$

The two lines are parallel, so they don't intersect.

- ◇ This system

$$\begin{aligned} 3x - 4y &= 2 \\ 6x - 8y &= 4 \end{aligned}$$

has a whole line of solutions: the second equation is just twice the first equation, so both equations represent the line

$$\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} + t \begin{pmatrix} 4 \\ 3 \end{pmatrix} \right\},$$

which gives all solutions to this system of equations.

Later we will investigate further when a linear system has a unique solution, when it has no solution and when it has infinitely many solutions, like the line in the third example.

Definition 2.8: A linear system is **consistent** if it has some solution. A linear system is called **inconsistent** if it has no solution.
If a linear system has infinitely many solutions, then a parametric expression from which all solutions can be obtained by choosing values for the parameters is called a **general solution**.

Example 2.9: The general solution to $\begin{aligned} 3x - 4y &= 2 \\ 6x - 8y &= 4 \end{aligned}$ is $\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} + t \begin{pmatrix} 4 \\ 3 \end{pmatrix} \right\}$.

In the previous example, the first and last linear system are consistent, and the middle one is inconsistent.

The aim of this chapter is to learn a systematic way of solving such systems of equations, using matrices.

Notation 2.10: A system of linear equations

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

can be represented in matrix form as

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

Here all the coefficients of the variables are collected in a matrix. If A is this matrix, and we write x for the vector of variables and b for the vector on the right hand side, the system becomes a matrix equation

$$Ax = b.$$

Remember how matrix multiplication works: if you multiply the left hand side out, you literally get the left hand sides of all the equations.

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

You can see that the matrix has as many columns as there are variables or unknowns, and as many rows as there are equations. Each row of the matrix represents one equation (or at least the left hand side of it).

So if you have a homogeneous system, you know all you need to know about it just from the matrix

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n-1} & a_{2n} \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn-1} & a_{mn} \end{pmatrix}.$$

So we can use this matrix as a short hand notation for the corresponding homogeneous system of equations. If we want to put in the bs on the right hand side as well, we can do it:

Definition 2.11: An **augmented matrix** is a matrix of the form

$$\left(\begin{array}{ccccc|c} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n-1} & a_{2n} & b_2 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn-1} & a_{mn} & b_m \end{array}\right).$$

So we just add the column of bs on the right hand side, and we use the vertical line to remember that they are on the other side of the equation.

Notation 2.12: A homogeneous system of linear equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n-1}x_{n-1} + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n-1}x_{n-1} + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn-1}x_{n-1} + a_{mn}x_n &= 0 \end{aligned}$$

can be represented by a matrix

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

An inhomogeneous system

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

can be represented by an augmented matrix

$$\left(\begin{array}{ccccc|c} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n-1} & a_{2n} & b_2 \\ \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn-1} & a_{mn} & b_m \end{array}\right).$$

Each row of the matrix represents an equation.

You see it is short hand because we're not writing out the variables $x_1, \ldots, x_n$: their presence is implied by the number of columns in the left hand part of the matrix.

Examples 2.13: Here are some linear systems and their corresponding (augmented) matrices.

$$\begin{aligned} 3x_1 - x_2 &= 0 \\ 2x_1 - 4x_2 &= 0 \\ -5x_1 + 6x_2 &= 0 \end{aligned} \qquad \begin{pmatrix} 3 & -1 \\ 2 & -4 \\ -5 & 6 \end{pmatrix}$$

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 2 \\ -x_1 - x_2 + x_3 &= 1 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ -1 & -1 & 1 & 1 \end{array}\right)$$

$$\begin{aligned} 3x_1 - x_2 + 4x_3 \qquad\quad &= 2 \\ -x_1 - 3x_2 \qquad + \sqrt{2}x_4 &= 0 \\ \pi\, x_2 - 3x_3 \qquad\quad &= -5 \end{aligned} \qquad \left(\begin{array}{cccc|c} 3 & -1 & 4 & 0 & 2 \\ -1 & -3 & 0 & \sqrt{2} & 0 \\ 0 & \pi & -3 & 0 & -5 \end{array}\right)$$

What is the easiest kind of linear system to solve? One which already gives all the solutions:

$$\begin{array}{cccccc} x_1 & & & & = & b_1 \\ & x_2 & & & = & b_2 \\ & & \ddots & & \vdots & \vdots \\ & & & x_n & = & b_n \end{array} \qquad \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \cdots & 0 & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & b_n \end{array}\right)$$

We can see that in this case we have the same number of equations as we have unknowns, so we get a square matrix (if we leave out the bs). But only the diagonal entries of this square matrix are non-zero.

Definition 2.14: A **diagonal** matrix is a square matrix in which only the diagonal entries are non-zero:

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 & 0 \\ 0 & a_{22} & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1n-1} & 0 \\ 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

Clearly we can solve it just as easily if the diagonal entries are not 1: we just have to divide the b_i by a_{ii} (as long as $a_{ii} \neq 0$).

The next best system is one like this:

$$\begin{array}{ccccccccc} x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ & & x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & \ddots & & \vdots & & \vdots \\ & & & & & & x_n & = & b_n \end{array} \qquad \left(\begin{array}{cccc|c} 1 & a_{12} & \cdots & a_{1n} & b_1 \\ 0 & 1 & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & b_n \end{array}\right)$$

We can substitute the last equation into the one before to get x_{n-1}, then substitute that into the previous one, and so on upwards until we have all the solutions. Again, we see it doesn't matter so much if the diagonal entries are not 1.

Definition 2.15: An **upper triangular** matrix is a square matrix in which all entries below the diagonal are zero:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n-1} & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n-1} & a_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & a_{n-1n-1} & a_{n-1n} \\ 0 & 0 & \cdots & 0 & a_{nn} \end{pmatrix}$$

A **lower triangular** matrix is a square matrix in which all entries above the diagonal are zero.

So our aim is to transform any linear system into one of these upper triangular ones, or a diagonal one, and then the solution can be easily determined or read off. Or at least as close to this as possible: not all systems can be transformed quite into this nice form. We will see what the closest possible is a little later.

B. Elementary row operations

There are three very simple types of operations we can perform on a system of equations, which are enough to transform the system into a diagonal form which gives the solutions. Any operation performed on the rows of the augmented matrix is an operation performed on the equations of the corresponding linear system.

Definition 2.16: The three types of **elementary row operations** are

1. Multiply one row by a non-zero scalar.
2. Swap two rows.
3. Add a multiple of one row to another row.

None of these operations change the solutions to a linear system.

Example 2.17: Applying the elemetary row operations to the augmented matrix corresponds to manipulating the equations of the linear system.

$$\begin{aligned} 7x_1 + 3x_2 &= 3 \\ x_1 + x_2 &= 1 \end{aligned} \qquad \left(\begin{array}{cc|c} 7 & 3 & 3 \\ 1 & 1 & 1 \end{array}\right)$$

Swap first and second row:

$$\begin{aligned} x_1 + x_2 &= 1 \\ 7x_1 + 3x_2 &= 3 \end{aligned} \qquad \left(\begin{array}{cc|c} 1 & 1 & 1 \\ 7 & 3 & 3 \end{array}\right)$$

Add -7 times first row to second row:

$$\begin{aligned} x_1 + x_2 &= 1 \\ 0 - 4x_2 &= -4 \end{aligned} \qquad \left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & -4 & -4 \end{array}\right)$$

Multiply second row by $-\frac{1}{4}$ (or divide by -4):

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_2 &= 1 \end{aligned} \qquad \left(\begin{array}{cc|c} 1 & 1 & 1 \\ 0 & 1 & 1 \end{array}\right)$$

Add -1 times second row from first row:

$$\begin{aligned} x_1 &= 0 \\ x_2 &= 1 \end{aligned} \qquad \left(\begin{array}{cc|c} 1 & 0 & 0 \\ 0 & 1 & 1 \end{array}\right)$$

So this system has a unique solution, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

You can see that using the augmented matrices rather than the equations is more efficient and visibly much clearer.

In Linear Algebra, you can often check your own solutions. For example, take the solution we have just arrived at and check it in the original equations:

$$\begin{aligned} 7 \cdot 0 + 3 \cdot 1 &= 3 \\ 0 + 1 &= 1 \end{aligned}$$

all checks out.

You can also see that all these elementary row operations can be reversed, so we never lose any information.
Careful! Make sure you never multiply a row by 0! This deletes all information from that row.

Definition 2.18: If matrix B can be obtained from matrix A by elementary row operations, we say that A and B are **row equivalent**.

As every elementary row operation can be done "in the other direction", this is a symmetric relationship: if B can be obtained from A via elementary row operations, then A can also be obtained from B via elementary row operations.

C. Gauss algorithm

We will now learn an algorithm that tells us which elementary row operations to do to transform a linear system into a form where a solution can easily be read off.
We saw that the nicest would be a diagonal form, but this is not always possible.

Example 2.19: The system

$$\begin{aligned} x_1 + 2x_2 + x_3 &= 2 \\ -x_1 - x_2 + x_3 &= 1 \end{aligned} \qquad \left(\begin{array}{ccc|c} 1 & 2 & 1 & 2 \\ -1 & -1 & 1 & 1 \end{array}\right)$$

cannot be made diagonal or upper triangular: it does not have enough equations.
The next best thing is

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

corresponding to

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

We can then choose x_3 to be whatever we like (say 1, or a parameter t), and then read off x_1 and x_2 from the above.
Notice that we can keep track of the elementary row operations in each step: here the roman numerals represent the rows.

Definition 2.20: A matrix is in **(row) echelon form** if the following hold:

- ⋄ If a row does not consist entirely of zeros, then the first non-zero entry in the row is a 1. We call this a **leading** 1.
- ⋄ Any rows consisting entirely of zeros are grouped together at the bottom of the matrix.
- ⋄ In any two successive non-zero rows, the leading 1 in the lower row occurs further to the right than the leading 1 in the higher row.

If in addition each column that contains a leading 1 has zeros in every other entry, then the matrix is in **reduced (row) echelon form**.

A matrix in echelon form looks like this:

$$\begin{pmatrix} 1 & * & * & \cdots & * & * & * & \cdots \\ 0 & 1 & * & \cdots & * & * & * & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \end{pmatrix}$$

A $*$ represents an arbitrary entry. A matrix in *reduced* row echelon forms has zeros above all the 1s:

$$\begin{pmatrix} 1 & {\color{red}0} & * & \cdots & * & {\color{red}0} & * & \cdots \\ 0 & 1 & * & \cdots & * & {\color{red}0} & * & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \end{pmatrix}$$

For an augmented matrix, we just look at the part on the left hand side of the line to determine whether it is in echelon form or not.

Examples 2.21: The following matrices are in echelon form, but not reduced.

◇ $\begin{pmatrix} 1 & 8 & -2 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$. ◇ $\left(\begin{array}{ccc|c} 1 & 3 & 2 & 1 \\ 0 & 0 & 1 & 1 \end{array}\right)$ ◇ $\begin{pmatrix} 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ ◇ $\left(\begin{array}{cccc|c} 1 & -1 & 2 & 3 & 1 \\ 0 & 1 & 1 & 2 & -1 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right)$

Exercise 2.22: Bring the above matrices into reduced row echelon form.

Examples 2.23: The following matrices are in reduced echelon form.

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

Exercise 2.24: Determine whether the systems above are consistent (whether they have a solution), and if yes, give the general form of the solution.

Gauss Algorithm: The Gauss algorithm transforms a matrix into row echelon form by the folllowing steps:

- ◇ Locate the left-most column that does not consist entirely of zeros.
- ◇ Swap rows to bring a non-zero entry to the top of the column you identified in the first step. We call this non-zero top entry the **pivot element**.
- ◇ Divide the first row by the pivot element to obtain a 1 at the top of the column.
- ◇ Add suitable multiples of the top row to the rows below so that the entries below the leading 1 become zeros.
- ◇ Now cover the top row of the matrix and begin again from the start, now applying the steps to the submatrix that remains.
- ◇ Continue this way until the matrix is in row echelon form.

Informally, what the Gauss algorithm does is to simplify the equations, get rid of any superfluous ones (which will transform into zero rows), and get the rest into a form where a solution can be read off more easily.
This is best seen on examples:

Examples 2.25: ◇ We use the Gauss algorithm on the following matrix. There is already a 1 in the top of the first column, so we clear the entries below it:

$$\left(\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ -4 & 5 & 9 & -9 \end{array}\right) \xrightarrow{\text{III}+4\text{I}} \left(\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & -3 & 13 & -9 \end{array}\right)$$

Now we cover the first row and continue with the rest of the matrix. There is a 2 at the (new) top of the second column, so we divide the second row by 2. Then we clear the entry below that in the third row.

$$\left(\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 2 & -8 & 8 \\ 0 & -3 & 13 & -9 \end{array}\right) \xrightarrow{\frac{1}{2}\text{II}} \left(\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & -3 & 13 & -9 \end{array}\right) \xrightarrow{\text{III}+3\text{II}} \left(\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

Now the matrix is in echelon form.

◇ This time we find that the top entry of the first row is 0, so we have to switch another row into the top, and then divide that row so we get the leading 1. Then we clear the entries below it.

$$\left(\begin{array}{ccc|c} 0 & 1 & -4 & 8 \\ 2 & -3 & 2 & 1 \\ 5 & -8 & 7 & 1 \end{array}\right) \xrightarrow{\text{I}\leftrightarrow\text{II}} \left(\begin{array}{ccc|c} 2 & -3 & 2 & 1 \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{array}\right) \xrightarrow{\frac{1}{2}\text{I}} \left(\begin{array}{ccc|c} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 5 & -8 & 7 & 1 \end{array}\right) \xrightarrow{\text{III}-5\text{I}} \left(\begin{array}{ccc|c} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 0 & -\frac{1}{2} & 2 & -\frac{3}{2} \end{array}\right)$$

Now we repeat for the last two rows: there is already a leading 1 in the second column, so we clear the entry below it:

$$\left(\begin{array}{ccc|c} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 0 & -\frac{1}{2} & 2 & -\frac{3}{2} \end{array}\right) \xrightarrow{\text{III}+\frac{1}{2}\text{II}} \left(\begin{array}{ccc|c} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & \frac{5}{2} \end{array}\right)$$

Now the matrix (up to the vertical line) is in row echelon form. We see that this system is inconsistent: the last line says $0 \cdot x_1 + 0 \cdot x_2 + 0 \cdot x_3 = \frac{5}{2}$, which is of course not possible.

Once a matrix is in echelon form, we can work out the solution. Many computer systems that solve systems of equations for huge matrices use this method.
However, for us as humans working usually on reasonably small (up to 5×5 or so) matrices, it is extremely helpful to keep going until the matrix is in *reduced* row echelon form, as we can then just read off the solutions.

Gauss-Jordan Algorithm: The Gauss-Jordan algorithm transforms a matrix into reduced row echelon form by the following steps:

- ◇ Use the Gauss algorithm to bring the matrix into echelon form.
- ◇ Working upwards from the lowest non-zero row, clear the entries above each leading 1.

Examples 2.26: Finishing off the two examples above:

◇ We clear first the third column and then the second column.

$$\left(\begin{array}{ccc|c} 1 & -2 & 1 & 0 \\ 0 & 1 & -4 & 4 \\ 0 & 0 & 1 & 3 \end{array}\right) \xrightarrow{\text{II}+4\text{III},\text{I}-\text{III}} \left(\begin{array}{ccc|c} 1 & -2 & 0 & -3 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right) \xrightarrow{\text{I}+2\text{II}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & 29 \\ 0 & 1 & 0 & 16 \\ 0 & 0 & 1 & 3 \end{array}\right)$$

So the unique solution to this system is $\begin{pmatrix} 29 \\ 16 \\ 3 \end{pmatrix}$.

- Since the second system we had was inconsistent, it makes no sense to keep going, as it doesn't have a solution. So let's change it slightly so that the echelon form is

$$\left(\begin{array}{ccc|c} 1 & -\frac{3}{2} & 1 & \frac{1}{2} \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{\mathrm{I}+\frac{3}{2}\mathrm{II}} \left(\begin{array}{ccc|c} 1 & 0 & -5 & \frac{25}{2} \\ 0 & 1 & -4 & 8 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

This is now in reduced row echelon form, and there is one column that has no leading 1s, highlighted in green. Such a column means that the corresponding variable, here x_3, can be anything, so we can make it a parameter, $x_3 = t$. We can choose that variable freely. This is why I informally call this column a **freedom column**, or a **stuff column**, because it has stuff in it that we can't clear.

So how do we read off our solution: let's first transform it back into equations so we can explain what's going on.

$$\begin{aligned} x_1 - 5x_3 &= \frac{25}{2} \\ x_2 - 4x_3 &= 8 \end{aligned}$$

Choosing $x_3 = t$ and putting it on the other side, we get

$$\begin{aligned} x_1 &= \frac{25}{2} + 5t \\ x_2 &= 8 + 4t \end{aligned}$$

So the solution set to this system is

$$\left\{ \begin{pmatrix} \frac{25}{2} \\ 8 \\ 0 \end{pmatrix} + t \begin{pmatrix} 5 \\ 4 \\ 1 \end{pmatrix} \middle| \; t \in \mathbb{R} \right\}.$$

- Let's do one more example where we get two freedom columns. For simplicity we'll make it a homogeneous system.

$$\begin{pmatrix} 1 & 2 & 6 & 8 \\ 2 & -1 & 2 & -4 \\ 3 & 1 & 8 & 4 \\ 1 & -3 & -4 & -12 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 6 & 8 \\ 0 & -5 & -10 & -20 \\ 0 & -5 & -10 & -20 \\ 0 & -5 & -10 & -20 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 6 & 8 \\ 0 & -5 & -10 & -20 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 2 & 6 & 8 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 2 & 0 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now we have two "stuff" columns, one green (third column), one purple (fourth column). So we can choose $x_3 = s$ and $x_4 = t$. So our general solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} -2s \\ -2s - 4t \\ s \\ t \end{pmatrix}$$

or we can write it as a solution set:

$$\left\{ s \begin{pmatrix} -2 \\ -2 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} 0 \\ -4 \\ 0 \\ 1 \end{pmatrix} \middle| \; s, t \in \mathbb{R} \right\}.$$

(See videos on Blackboard for more steps.)

Example 2.27: (Worked example: Linear system with parameter) For which a does the following linear system have 0, 1 or infinitely many solutions? For each consistent case, find all solutions.

$$\begin{aligned} x_1 + 3x_2 - 2x_3 &= 7 \\ 2x_2 - 4x_3 &= 8 \\ x_1 + 5x_2 + (a^2-7)x_3 &= a + 14 \end{aligned}$$

As usual, we first write it as an augmented matrix, and start doing the Gauss algorithm.

$$\left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 2 & -4 & 8 \\ 1 & 5 & a^2-7 & a+14 \end{array}\right) \xrightarrow{\text{III}-\text{I}} \left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 2 & -4 & 8 \\ 0 & 2 & a^2-5 & a+7 \end{array}\right) \xrightarrow{\frac{1}{2}\text{II}} \left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 1 & -2 & 4 \\ 0 & 2 & a^2-5 & a+7 \end{array}\right)$$

$$\xrightarrow{\text{III}-2\text{II}} \left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 1 & -2 & 4 \\ 0 & 0 & a^2-1 & a-1 \end{array}\right) = \left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 1 & -2 & 4 \\ 0 & 0 & (a-1)(a+1) & a-1 \end{array}\right)$$

At this stage we have to consider separate cases: do we get a leading 1 in the third row or not?

- ◇ Case $a = -1$: In this case we have

$$\left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 1 & -2 & 4 \\ 0 & 0 & 0 & -2 \end{array}\right)$$

which is an inconsistent system: there are no solutions.

- ◇ Case $a = 1$: In this case the third row becomes a zero row:

$$\left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 1 & -2 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right) \xrightarrow{\text{I}-3\text{II}} \left(\begin{array}{ccc|c} 1 & 0 & 4 & -5 \\ 0 & 1 & -2 & 4 \\ 0 & 0 & 0 & 0 \end{array}\right)$$

So we can finish the Gauss-Jordan algorithm. So solutions here are

$$\left\{ \begin{pmatrix} -5 \\ 4 \\ 0 \end{pmatrix} + t \begin{pmatrix} -4 \\ 2 \\ 1 \end{pmatrix} \;\middle|\; t \in \mathbb{R} \right\}.$$

- ◇ Case $a \neq \pm 1$: In this case we can divide by $a^2 - 1$:

$$\left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 1 & -2 & 4 \\ 0 & 0 & a^2-1 & a-1 \end{array}\right) \xrightarrow{\frac{1}{a^2-1}\text{III}} \left(\begin{array}{ccc|c} 1 & 3 & -2 & 7 \\ 0 & 1 & -2 & 4 \\ 0 & 0 & 1 & \frac{1}{a+1} \end{array}\right) \xrightarrow{\text{II}+2\text{III},\text{I}+2\text{III}} \left(\begin{array}{ccc|c} 1 & 3 & 0 & 7+\frac{2}{a+1} \\ 0 & 1 & 0 & 4+\frac{2}{a+1} \\ 0 & 0 & 1 & \frac{1}{a+1} \end{array}\right)$$

$$\xrightarrow{\text{I}-3\text{II}} \left(\begin{array}{ccc|c} 1 & 0 & 0 & \frac{7a+9}{a+1} - 3\frac{4a+6}{a+1} \\ 0 & 1 & 0 & \frac{4a+6}{a+1} \\ 0 & 0 & 1 & \frac{1}{a+1} \end{array}\right)$$

So the unique solution is

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \frac{-5a-9}{a+1} \\ \frac{4a+6}{a+1} \\ \frac{1}{a+1} \end{pmatrix} = \frac{1}{a+1} \begin{pmatrix} -(5a+9) \\ 4a+6 \\ 1 \end{pmatrix}.$$

D. Numbers of solutions

Now that we know how to solve a linear system using the Gauss-Jordan algorithm, let's think about how many solutions such a linear system can have.

Proposition 2.28: (Solutions of homogeneous system)
A homogeneous linear system always has at least one solution: the zero vector.
If a homogeneous linear system has more than one solution, then it has infinitely many solutions.

Proof. Let's say our homogenous linear system has n variables, $x_1, \ldots, x_n$. Let $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.
Then we can write the homogeneous linear system as $Ax = 0$ for some matrix A with n columns and as many rows as our system has equations.

Setting $x = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} = 0$, the zero vector, is definitely a solution to $Ax = 0$.

It could be that this is the *only* solution: this corresponds to the reduced row echelon form of A having exactly n non-zero rows, each with it's own leading 1, and then possibly some zero rows.

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

Any other reduced row echelon form of a system with n variables can only have fewer columns with leading 1s, so it will have at least one column which does not have a leading 1. Any such column corresponds to a variable which can be assigned any value, which means the general solution has a free parameter. So there are infinitely many solutions.

$$\begin{pmatrix} 1 & 0 & * & \cdots & * & 0 & * & \cdots \\ 0 & 1 & * & \cdots & * & 0 & * & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \end{pmatrix}$$

□

Proposition 2.29: *The solutions of a homogeneous linear system with n variables form a subspace of $\mathbb{R}^n$.*

Proof. Let $Ax = 0$ be the linear system, so $x \in \mathbb{R}^n$. We verify the three conditions for a subspace, from Proposition 1.28.

- ◇ The zero vector is a solution of $Ax = 0$ (see Prop. 2.28).
- ◇ If $u, v \in \mathbb{R}^n$ are solutions, then $Au = 0$ and $Av = 0$. Then also $A(u + v) = Au + Av = 0 + 0 = 0$ because matrix multiplication is linear (Prop. 1.67).
- ◇ If $u \in \mathbb{R}^n$ is a solution and $\lambda \in \mathbb{R}$, then $A(\lambda u) = \lambda Au = \lambda \cdot 0 = 0$ as matrix multiplication is linear.

So the solutions form a subspace of $\mathbb{R}^n$. □

Proposition 2.30: (Homogeneous and inhomogeneous solutions)
Given a particular solution to an inhomogeneous linear system, then adding any solution of the corresponding homogeneous linear system gives another solution.

Proof. Let the inhomogeneous linear system be $Ax = b$, with $b \neq 0$. Suppose a is a particular solution of this system, i.e. $Aa = b$, and that v is a solution to the corresponding homogeneous system $Ax = 0$. Then

$$A(a+v) = Aa + Av = b + 0 = b$$

so $a + v$ is also a solution to the inhomogeneous linear system. □

Proposition 2.31: (Solutions of inhomogeneous system)
An inhomogeneous linear system can have zero, one or infinitely many solutions.

Proof. If the inhomogeneous system is inconsistent, then it has zero solutions. This corresponds to the row echelon form having a zero row with a non-zero entry in the augmented part.

$$\left(\begin{array}{cccc|c} 1 & * & \cdots & * & b_1 \\ 0 & 1 & \cdots & * & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & b_m \neq 0 \end{array}\right)$$

If the inhomogeneous system does have some solution, then we know that the corresponding homogeneous system has 1 or infinitely many solutions (Prop. 2.28), and that adding any of these to the given one gives a new solution for the inhomogeneous system (Prop. 2.30). So the inhomogeneous system has either one solution or infinitely many. These options correspond to the same reduced row echelon forms as were given for the homogeneous systems, with the addition that any zero row also has a zero in the augmented part.

$$\left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & b_1 \\ 0 & 1 & \cdots & 0 & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & b_n \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{array}\right) \qquad \left(\begin{array}{cccccccc|c} 1 & 0 & * & \cdots & * & 0 & * & \cdots & b_1 \\ 0 & 1 & * & \cdots & * & 0 & * & \cdots & b_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 & * & \cdots & b_m \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{array}\right)$$

□

Examples 2.32: We have already seen examples of each of these three cases. Look back at them now.

The above results allow us to answer the following questions:

Exercise 2.33:
Determine whether the statement is true or false.

- ◇ A homogeneous linear system in n unknowns whose corresponding augmented matrix has a reduced row echelon form with r leading 1's has $n - r$ free parameters in the general solution.
- ◇ All leading 1's in a matrix in row echelon form must occur in different columns.
- ◇ If a homogeneous linear system of n equations in n unknowns has a corresponding augmented matrix with a reduced row echelon form containing n leading 1's, then the linear system has only the trivial solution.

- ◇ If the reduced row echelon form of the augmented matrix for a linear system has a row of zeros, then the system must have infinitely many solutions.
- ◇ If a linear system has more unknowns than equations, then it must have infinitely many solutions.

We will discuss this in lectures and write down explanations and/or examples and counterexamples.

The last one is important enough that we will want to refer to it, so let's write it as a proposition.

Proposition 2.34: *If a homogeneous linear system has more unknowns than equations, then it has infinitely many solutions.*

Proof. Note that a homogeneous linear system cannot be inconsistent, so it has at least one solution.
The system $Ax = 0$ has matrix of size $m \times n$ with $m < n$. So its row echelon form can have at most m leading 1s, as each row can only have one leading 1. As $m < n$, this leaves some columns in which there is no leading 1, which give us free parameters in the general solution. □

E. Linear Systems: Study guide

Concept review.

- ◇ Linear equation, linear system.
- ◇ Homogeneous and inhomogeneous linear systems.
- ◇ Solution to linear system.
- ◇ Augmented matrix.
- ◇ Matrix representation of linear system.
- ◇ Diagonal, upper and lower triangular matrices.
- ◇ Row echelon form and reduced row echelon form.
- ◇ Elementary row operations.
- ◇ Row equivalent matrices.
- ◇ Gauss and Gauss-Jordan algorithms.
- ◇ Numbers of solutions of linear systems.

Skills.

- ◇ Determine whether an equation is linear.
- ◇ Write a linear system into matrix form.
- ◇ Determine whether a matrix is in row echelon form, or reduced row echelon form.
- ◇ Perform elementary row operations on a matrix.
- ◇ Perform the Gauss algorithm to bring a matrix into row echelon form.
- ◇ Perform the Gauss-Jordan algorithm to bring a matrix into reduced row echelon form.
- ◇ Write down the general solution for a linear system which is in reduced row echelon form.
- ◇ Determine whether an inhomogeneous system is consistent or not.
- ◇ Find solutions to linear systems using the above methods.

CHAPTER 3

Inverse Matrices and Determinants

If we have a single equation $ax = b$, with a, b real numbers, then we can solve it by dividing the equation by a, *as long as* $a \neq 0$. We will now see what the equivalent is for matrices.

A. Inverse matrices

Definition 3.1: An **inverse** of a square matrix A is a matrix B of the same size such that $BA = I = AB$, where I is the identity matrix of the same size.

You can think of of this as being the equivalent of $\frac{1}{a} \cdot a = 1 = a \cdot \frac{1}{a}$ for a real number a. However, for matrices, $\frac{1}{A}$ does not make any sense at all, so don't write it!.
Notice that we need the matrix to be square so that we can multiply it with its inverse both ways round to give the *same* size identity matrix each time.
Also notice that the condition is symmetrical in A and B, so we also say that A and B are inverses of each other.

Exercise 3.2: Show that $A = \begin{pmatrix} 2 & -5 \\ -1 & 3 \end{pmatrix}$ and $B = \begin{pmatrix} 3 & 5 \\ 1 & 2 \end{pmatrix}$ are inverses of each other.

Examples 3.3:

- ◇ The inverse of an identity matrix is itself: $II = I$.
- ◇ The inverse of a diagonal matrix with no zero entries on the diagonal is the diagonal matrix with the reciprocal diagonal entries:
$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} \begin{pmatrix} \frac{1}{\lambda_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$
and the other way around. (Here $\lambda_i \neq 0$.)
- ◇ Not all matrices have an inverse: for example, a diagonal matrix with a zero row is not invertible: there is no way we can get 0·something= 1.
- ◇

A 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has inverse $A^{-1} = \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad-bc & -ab+ab \\ cd-cd & -cb+ad \end{pmatrix}$$

so

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

and the same the other way around. Obviously this needs $ad - bc \neq 0$.

Definition 3.4: A matrix is called **invertible** if it has an inverse. A matrix which is not invertible is called **singular**. An invertible matrix is also called **non-singular**.

Note that only a square matrix can be invertible: if a matrix is not square, it is definitely not invertible, as we defined an inverse only for square matrices.

Example 3.5: A 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$.

For example,

- $\begin{pmatrix} 1 & 3 \\ 2 & 6 \end{pmatrix}$ is not invertible: $ad - bc = 6 - 6 = 0$.
- $\begin{pmatrix} 9 & 2 \\ 4 & 7 \end{pmatrix}$ is invertible: $ad - bc = 63 - 8 \neq 0$.

Proposition 3.6: (Uniqueness of inverses)
If a matrix has an inverse, then that inverse is unique.

Proof. Suppose a matrix A has two inverses, B and C. Then

$$\begin{aligned} B &= BI && \text{mult by identity matrix does not change it} \\ &= B(AC) && \text{as } C \text{ is an inverse} \\ &= (BA)C && \text{matrix mult is associative} \\ &= IC && \text{as } B \text{ is an inverse} \\ &= C \end{aligned}$$

□

Because of this result, we say *the* inverse of A is A^{-1}. So

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

This allows us to show

Proposition 3.7: (Properties of inverse)

(i) Every invertible matrix is the inverse of its inverse:
$$(A^{-1})^{-1} = A$$
(ii) If A is invertible, then so is A^T, and
$$(A^T)^{-1} = (A^{-1})^T.$$
*(iii) **("Socks and shoes")** If two $n \times n$ matrices A and B are both invertible, then so is their product AB, with inverse*
$$(AB)^{-1} = B^{-1}A^{-1}.$$

Proof. (i) By uniqueness of inverses and $AA^{-1} = I = A^{-1}A$.
(ii) Taking transpose of these two equations gives
$$(AA^{-1})^T = I^T, \qquad \text{so} \qquad (A^{-1})^T A^T = I,$$
and
$$(A^{-1}A)^T = I^T, \qquad \text{so} \qquad A^T (A^{-1})^T = I.$$
So $(A^{-1})^T$ satisfies the definition for the inverse of A^T, so by uniqueness of inverses, $(A^T)^{-1} = (A^{-1})^T$.

(iii) We check whether $B^{-1}A^{-1}$ satisfies the definition of an inverse:

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I$$

and

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

So by uniqueness of inverses, $(AB)^{-1} = B^{-1}A^{-1}$. □

Notice that for the inverse of a product, we have to change the order, as we did when forming the transpose of a product. You can remember it this way:
If you first put your socks on and then your shoes, then to undo it you have to take your shoes off first, and then your socks.
Let's also record it as

The product of invertible matrices is invertible.

Example 3.8: ◇ Inverse of inverse: $\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}^{-1} = -\frac{1}{3}\begin{pmatrix} -1 & -2 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}$, and

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

◇ Inverse of transpose: $\left(\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}^T\right)^{-1} = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{2}{3} & -\frac{1}{3} \end{pmatrix} = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & -\frac{1}{3} \end{pmatrix}^T$

◇ Inverse of product: $\left(\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}\right)^{-1} = \begin{pmatrix} 1 & 4 \\ 1 & -2 \end{pmatrix}^{-1} = -\frac{1}{6}\begin{pmatrix} -2 & -4 \\ -1 & 1 \end{pmatrix}$

but

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

On the other hand

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}^{-1}\begin{pmatrix} 1 & 2 \\ 1 & -1 \end{pmatrix}^{-1} = \frac{1}{2}\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}\cdot\left(-\frac{1}{3}\right)\begin{pmatrix} -1 & -2 \\ -1 & 1 \end{pmatrix} = -\frac{1}{6}\begin{pmatrix} -2 & -4 \\ -1 & 1 \end{pmatrix}.$$

Exercise 3.9: ◇ Use "socks and shoes" to show that if A is invertible, then for any natural number k, A^k is invertible with $(A^k)^{-1} = (A^{-1})^k$.

This allows us to form **negative powers** of an invertible matrix:

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

◇ Prove that if A is invertible, then for any $\lambda \neq 0$, so is λA, with $(\lambda A)^{-1} = \frac{1}{\lambda}A^{-1}$.

Some matrices are easily seen to be not invertible:

Proposition 3.10: *If a matrix has a zero column, then it is not invertible.*

Proof. Let $A = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$ be a square matrix with $a_k = 0$, a zero column.
Then for any matrix B of the same size, the product

$$BA = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ Ba_1 & \cdots & Ba_k & \cdots & Ba_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ Ba_1 & \cdots & 0 & \cdots & Ba_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$$

also has a zero column, so it can never be the identity matrix. □

Corollary 3.11: *If a matrix has a zero row, then it is not invertible.*

PROOF. Say A has a zero row. Then A^T has a zero column, so A^T is not invertible. But if A were invertible, then A^T would be too. But as A^T isn't, A can't be invertible either. $\square$

Let's think about inverses interact with the two "meanings" of matrices we have seen.

Definition 3.12: Given a function $f: X \longrightarrow Y$, the **inverse** of f is a function $f^{-1}: Y \longrightarrow X$ such that $f^{-1}(f(x)) = x$ for all $x \in X$ and $f(f^{-1}(y)) = y$ for all $y \in Y$. We also write this as $f^{-1} \circ f = \mathrm{id}_X$ and $f \circ f^{-1} = \mathrm{id}_Y$, where id_X and id_Y are the **identity functions** on X and Y.

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[dr, "\mathrm{id}_X"'] & Y \arrow[d, "f^{-1}"'] \arrow[dr, "\mathrm{id}_Y"] & \\ & X \arrow[r, "f"'] & Y \end{tikzcd}$$

You can think of the inverse function as a function that "undoes" the original function. It has to work both ways round though.
Not every function has an inverse.

Proposition 3.13: (Inverse matrix transformation)
If A is an invertible matrix, then the matrix transformation of A^{-1} is the inverse of the matrix transformation of A: $T_A^{-1} = T_{A^{-1}}$.

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "T_A"] \arrow[dr, "\mathrm{id}"'] & \mathbb{R}^n \arrow[d, "T_{A^{-1}}"'] \arrow[dr, "\mathrm{id}"] & \\ & \mathbb{R}^n \arrow[r, "T_A"'] & \mathbb{R}^n \end{tikzcd}$$

PROOF. Suppose A is an $n \times n$ matrix. Then $T_{A^{-1}}(T_A(v)) = A^{-1}Av = Iv = v$ for any $v \in \mathbb{R}^n$, and $T_A(T_{A^{-1}}(v)) = AA^{-1}v = Iv = v$ for any $v \in \mathbb{R}^n$. $\square$

How do invertible matrices interact with linear systems?

Proposition 3.14: *A linear system $Ax = b$ with an invertible matrix A has a unique solution.*

PROOF. As A is invertible, we can multiply the matrix equation $Ax = b$ by A^{-1} on the left:

$$\begin{aligned} & & Ax &= b \\ &\Leftrightarrow & A^{-1}Ax &= A^{-1}b \\ &\Leftrightarrow & x &= A^{-1}b \end{aligned}$$

So this is the unique solution. $\square$

Notice that this implies that if A is invertible, then $Ax = b$ is consistent *for any* b.

Important!!! When you are dealing with matrix equations (including vectors), you have to specify whether you multiply by a matrix on the left (of both sides of the equation) or on the right (of both sides of the equation)! This is because matrix multiplication is not commutative: the order matters.

Example 3.15: Given that A, B are invertible, and all matrices are $n \times n$ matrices, resolve for X:

$$\begin{aligned} & & ABXA^{-1}B &= C & & \\ &\Leftrightarrow & A^{-1}ABXA^{-1}B &= A^{-1}C & &\text{multiply by } A^{-1} \text{ on the left} \\ &\Leftrightarrow & XA^{-1}B &= B^{-1}A^{-1}C & &\text{multiply by } B^{-1} \text{ on the left} \\ &\Leftrightarrow & XA^{-1} &= B^{-1}A^{-1}CB^{-1} & &\text{multiply by } B^{-1} \text{ on the right} \\ &\Leftrightarrow & X &= B^{-1}A^{-1}CB^{-1}A & &\text{multiply by } A \text{ on the right} \end{aligned}$$

This result tells us something about the reduced row echelon form that an invertible matrix has: it must have n leading 1s (if A is an $n \times n$ matrix), which makes it the identity matrix.
In fact, we can use this to calculate the inverse of a matrix.

B. Inverse algorithm

We can calculate inverse matrices with the following algorithm.

The **inverse algorithm** applied to a square matrix A has the following steps:

- ◇ Write the matrix A and the identity matrix of the same size next to each other.
- ◇ Apply the Gauss-Jordan algorithm to A, and apply the same steps in the same order to the identity matrix next to it.
- ◇ If the reduced row echelon form of A is the identity matrix, then the matrix next to it is A^{-1}.

We will have to prove that it really gives us the inverse, but let's see it on some examples first.

Examples 3.16: We will calculate the inverse of $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{pmatrix}$.

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{pmatrix} \left| \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right.$$

$$\text{II} - 2\text{I}, \text{III} - \text{I} : \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -3 \\ 0 & -2 & 5 \end{pmatrix} \left| \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix} \right.$$

$$\text{III} + 2\text{II} : \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & -1 \end{pmatrix} \left| \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ -5 & 2 & 1 \end{pmatrix} \right.$$

$$(-1) \cdot \text{III} : \quad \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & -3 \\ 0 & 0 & 1 \end{pmatrix} \left| \begin{pmatrix} 1 & 0 & 0 \\ -2 & 1 & 0 \\ 5 & -2 & -1 \end{pmatrix} \right.$$

$$\text{II} + 3\text{III}, \text{I} - 3\text{III} : \quad \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left| \begin{pmatrix} -14 & 6 & 3 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{pmatrix} \right.$$

$$\text{I} - 2\text{II} : \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \left| \begin{pmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{pmatrix} \right.$$

So $A^{-1} = \begin{pmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{pmatrix}$.

We should always check our answer: it's very easy to make small numerical mistakes, and checking will flag this up. Multiplying AA^{-1} should give us I:

$$\begin{pmatrix} 1 & 2 & 3 \\ 2 & 5 & 3 \\ 1 & 0 & 8 \end{pmatrix} \begin{pmatrix} -40 & 16 & 9 \\ 13 & -5 & -3 \\ 5 & -2 & -1 \end{pmatrix} = \begin{pmatrix} -40 + 26 + 15 & 16 - 10 - 6 & 9 - 6 - 3 \\ -80 + 65 + 15 & 32 - 25 - 6 & 18 - 15 - 3 \\ -40 + 40 & 16 - 16 & 9 - 8 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

So our answer is correct.

Exercise 3.17: Use the inverse algorithm to compute the inverse of $A = \begin{pmatrix} 1 & 6 & 4 \\ 2 & 4 & 0 \\ -1 & 2 & 6 \end{pmatrix}$.

We have seen from the viewpoint of linear systems that if A is invertible, doing elementary row operations in the Gauss-Jordan algorithm gives us the identity matrix. But why is the matrix on the right the inverse?
To prove this, we need to link the elementary row operations to elementary matrices.

Definition 3.18: There are three types of **elementary matrices**, corresponding to elementary row operations.

- A diagonal matrix with 1s on the diagonal except for a $\lambda \neq 0$ in the ith row,
$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$
is obtained by multiplying the ith row of the identity matrix by λ.
- A matrix with 1s on the diagonal and zeros everywhere else, except for the 1 in the ith row being in column j and the 1 in the jth row being in column i,
$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$
is obtained by swapping rows i and j of the identity matrix.
- A matrix with 1s on the diagonal and also a λ in position i, j,
$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & \lambda & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix},$$
corresponds to adding λ times row j to row i in the identity matrix.

Proposition 3.19: *Multiplying a matrix on the left by an elementary matrix has the effect of performing the corresponding elementary row operation to the matrix.*

Proof. Omitted. You can just try it out for 2×2 or 3×3 matrices. □

Exercise 3.20: Calculate the following:

- $\begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 9 & 3 \end{pmatrix}$
- $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$
- $\begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$

Example 3.21: Let $E_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$, $E_2 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 4 & 0 & 1 \end{pmatrix}$, $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$.
$E_1,\ E_2,\ E_3$ are elementary matrices, and

$$E_1A = \begin{pmatrix} a & b & c \\ d & e & f \\ 2g & 2h & 2i \end{pmatrix}, \quad E_2A = \begin{pmatrix} d & e & f \\ a & b & c \\ g & h & i \end{pmatrix}, \quad E_3A = \begin{pmatrix} a & b & c \\ d & e & f \\ g+4a & h+4b & i+4c \end{pmatrix}.$$

So when we do the elementary row operations in the inverse algorithm, we are multiplying A and I successively with more and more elementary matrices on the left.

$$\begin{aligned} A &\mid I \\ E_1A &\mid E_1I \\ E_2E_1A &\mid E_2E_1I \\ \vdots\ &\mid\ \vdots \\ E_k\cdots E_2E_1A &\mid E_k\cdots E_2E_1I \end{aligned}$$

When we are finished, we have $(E_k\cdots E_2E_1)A = I$ on the left of the line, so the product $E_k\cdots E_2E_1 = B$ is a very good candidate for A^{-1}. And we have this same matrix on the right of the line: $E_k\cdots E_2E_1I = B$.
To prove that this product of elementary matrices really is the inverse of A, we need either some way of checking the product the other way round: is AB also $= I$? Or we need to argue in some other way that we can be sure that B is the inverse of A. We can do this by proving that

Lemma 3.22: *All elementary matrices are invertible.*

Proof. The inverse of an elementary matrix is the matrix corresponding to the inverse elementary row operation:

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\lambda} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \qquad \text{divide row i by } \lambda\ (\lambda \neq 0)$$

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \qquad \text{self-inverse: swap rows again}$$

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & \lambda & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & -\lambda & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} \qquad \text{subtract } \lambda \text{ times row } j \text{ from row } i$$

□

So we can show

Proposition 3.23: (Inverse algorithm works)
The inverse algorithm really does produce the inverse of a matrix A, provided that the reduced row echelon form of A is I.

Proof. The inverse algorithm

$$\begin{aligned} A &\mid I \\ E_1 A &\mid E_1 I \\ E_2 E_1 A &\mid E_2 E_1 I \\ \vdots &\mid \vdots \\ E_k \cdots E_2 E_1 A &\mid E_k \cdots E_2 E_1 I \end{aligned}$$

produces a matrix $B = E_k \cdots E_2 E_1$ with $BA = I$. Because every elementary matrix is invertible (Lemma 3.22) and the product of invertible matrices is invertible (Prop. 3.7, "socks and shoes"), the matrix B is invertible with $B^{-1} = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$, and so

$$\begin{aligned} & & BA &= I \\ \Rightarrow & & B^{-1}BA &= B^{-1}I \\ \Rightarrow & & A &= B^{-1} \\ \Rightarrow & & A^{-1} &= B & & \text{as } B \text{ is the inverse of its inverse.} \end{aligned}$$

So indeed $A^{-1} = E_k \cdots E_2 E_1$, which is the matrix we obtain on the right hand side of the line at the end of the algorithm. □

We now have several conditions that tell us whether A is invertible.

Theorem 3.24: (Invertibility conditions)
For a square matrix A, the following are equivalent:
(i) A is invertible.
(ii) $Ax = 0$ has only the trivial solution $x = 0$.
(iii) The reduced row echelon form of A is the identity I.
(iv) A is the product of elementary matrices.

Proof. We will prove the implications in a circle: i⇒ii⇒iii ⇒iv⇒i. Then we automatically also get the other implications such as iv⇒iii and iii⇒i, by going two or three steps in the circle.

- ◇ i ⇒ii: We proved this already (Prop. 3.14): if A is invertible, then
$$Ax = 0 \quad \Rightarrow \quad A^{-1}Ax = A^{-1}0 \quad \Rightarrow \quad x = 0,$$
because any matrix times the zero vector gives the zero vector.
- ◇ ii ⇒iii: If $Ax = 0$ has only the trivial solution, then the proof of Prop. 2.28 (solutions of homogeneous system) shows that the reduced row echelon form of A has n leading 1s, where n is the number of variables. As A is a square matrix, this means the reduced row echelon form of A is I_n.
- ◇ iii ⇒iv: The Gauss-Jordan algorithm (or the inverse algorithm) produces $E_k \cdots E_2 E_1 A$ as the reduced row echelon form of A. If this is I, then $E_k \cdots E_2 E_1 A = I$, and so $A = E_1^{-1} E_2^{-1} \cdots E_k^{-1}$, as all elementary matrices are invertible (Lemma 3.22). The inverse of an elementary matrix is also an elementary matrix, so A is the product of elementary matrices.
- ◇ iv⇒i: If A is the product of elementary matrices, then it is invertible: the product of invertible matrices is invertible (Prop. 3.7, "socks and shoes"). □

Exercise 3.25: Is $A = \begin{pmatrix} 1 & 6 & 4 \\ 2 & 4 & -1 \\ -1 & 2 & 5 \end{pmatrix}$ invertible? Calculate it's reduced row echelon form to check this.

The second statement, that $Ax = 0$ only has the trivial solution, can also be interpreted in terms of the matrix transformation $T_A : \mathbb{R}^n \longrightarrow \mathbb{R}^n$: it tells us that only the zero vector $x = 0$ gets mapped to 0. If you had two different vectors, $v \neq w$ with $Av = 0 = Aw$, then there is no way to "undo" multiplication by A: if we tried to do $A^{-1}0$, we would need to get v *and* w, which is not possible! We will look more at this viewpoint in the second semester.

However, we can already use this theorem to show that actually, for a square matrix to be invertible, we only have to check one way round if $BA = I_n$: this will automatically give us $AB = I_n$ as well.

Proposition 3.26: (Check one get one free for invertible matrices)
If A and B are $n \times n$ matrices with $BA = I_n$, then A is invertible with $A^{-1} = B$.

PROOF. Suppose $Av = 0$ for some $v \in \mathbb{R}^n$. We want to show that then $v = 0$, so that A satisfies one of the invertibility conditions (Theorem 3.24).
We have $B(Av) = B0 = 0$, but also $(BA)v = I_n v = v$, so that $v = 0$. So $Ax = 0$ has only the trivial solution $x = 0$, so A is invertible.
Since A^{-1} exists, multiplying $BA = I_n$ by A^{-1} on the right gives $B = A^{-1}$. □

Corollary 3.27: *If A and B are $n \times n$ matrices with $AB = I_n$, then A is invertible with $A^{-1} = B$.*

PROOF. By "check one get on free for invertible matrices", we get that B is invertible with $B^{-1} = A$. But then of course also A is invertible with $A^{-1} = B$, see Prop. 3.7. □

So it doesn't matter which way round we check the multiplication in this case.

This result also gives us a partial converse to the result "a product of invertible matrices is invertible":

Proposition 3.28: (Invertible product)
If the product AB of two $n \times n$ matrices A and B is invertible, then both A and B are also invertible.

PROOF. Let $C = (AB)^{-1}$. Then we have $C(AB) = I_n = (AB)C$. If we set $D = BC$, then $AD = I_n$, so by "check one get one free for invertible matrices" (Prop. 3.26), A is invertible with inverse D.
Similarly, setting $E = CA$ gives $EB = I_n$, so B is invertible by "check one get one free". □

Corollary 3.29: *For $n \times n$ matrices A and B, we have*

$$AB \text{ is invertible} \quad \Longleftrightarrow \quad A \text{ and } B \text{ are invertible.}$$

Let's also record all implications an invertible matrix has on a linear system.

Theorem 3.30: (Invertible linear system)
Given a square matrix A, the following are equivalent:
(i) *A is invertible.*
(ii) *The homogeneous linear system $Ax = 0$ has only the trivial solution $x = 0$.*
(iii) *The inhomogeneous linear system $Ax = b$ is consistent for any $b \in \mathbb{R}^n$.*
(iv) *The inhomogeneous linear system $Ax = b$ has exactly one solution for any $b \in \mathbb{R}^n$.*

PROOF. We have already proved i $\Leftrightarrow$ii. We will now prove i $\Rightarrow$iv $\Rightarrow$iii $\Rightarrow$i.

- i $\Rightarrow$iv: This is Proposition 3.14:
$$Ax = b \quad \Rightarrow \quad A^{-1}Ax = A^{-1}b$$
giving the unique solution $x = A^{-1}b$.
- iv $\Rightarrow$iii: If the system $Ax = b$ has a unique solution, then it has some solution, so it is consistent.

◇ iii⇒i: We will construct an inverse for A from solutions to several different inhomogeneous systems $Ax = b$.

Let $e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, $e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$, …, $e_n = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$. Then we can find solutions $x_1,\ x_2, \ldots, x_n$ to the linear systems $Ax = e_1$, $Ax = e_2$, …, $Ax = e_n$. Let $B = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ x_1 & \cdots & x_k & \cdots & x_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$ be the matrix with these solutions as columns. Then

$$AB = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ Ax_1 & \cdots & Ax_k & \cdots & Ax_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ e_1 & \cdots & e_k & \cdots & e_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} = I_n.$$

So by "check one get one free for invertible matrices" (Prop. 3.26), A is invertible. □

C. Inverse matrices: Study guide

Concept review.

◇ Inverse of a matrix.
◇ Invertible, non-singular, singular matrices.
◇ Inverse of a 2×2 matrix.
◇ Uniqueness of inverse.
◇ Inverse of inverse, inverse of transpose, inverse of product.
◇ Negative integer powers of an invertible matrix.
◇ Inverse matrix transformation.
◇ Linear system with invertible matrix.
◇ Inverse algorithm.
◇ Elementary matrices.
◇ Invertibility conditions: equivalent conditions for a matrix to be invertible.
◇ Check one get one free for inverse matrices.

Skills.

◇ Find the inverse of a 2×2 matrix.
◇ Use the definition of an inverse (and uniqueness of inverses) to prove properties of inverses.
◇ Determine elementary matrices corresponding to a given elementary row operation.
◇ Use the inverse algorithm to determine whether a matrix is invertible, and to calculate it's inverse.
◇ Solve a linear system using inverse matrices.

D. Determinant of a 2×2 matrix

When introducing inverses, we saw that a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad-bc \neq 0$. This is a useful quantity, which also comes back in the formula for the inverse of a 2×2 matrix.

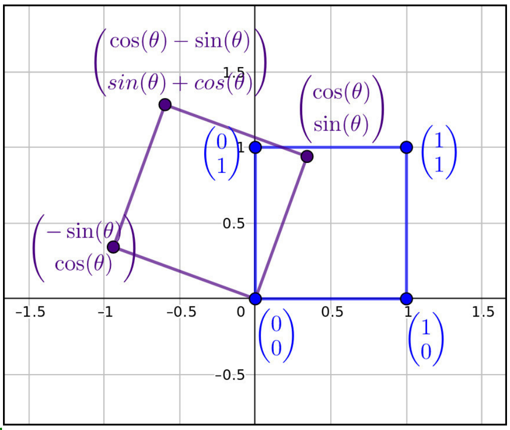
Definition 3.31: The **determinant** of a 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $\det A = ad - bc$.

Examples 3.32: ◇ $\det \begin{pmatrix} 1 & 4 \\ 8 & 3 \end{pmatrix} = 3 - 32 = -29$ ◇ $\det \begin{pmatrix} 7 & 0 \\ 0 & 3 \end{pmatrix} = 21$ ◇ $\det \begin{pmatrix} a & b \\ 2a & 2b \end{pmatrix} = 2ab - 2ab = 0$

This determinant has a geometric significance:

Fact 3.33: Consider the square with corners $\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$. This has area 1. If we transform this square using the matrix transformation T_A, then the resulting quadrilateral has area $\det A$. Here a negative determinant signifies that the orientation of the corners round the square has been reversed.

Examples 3.34:



$A = \begin{pmatrix}\cos(\theta) & -\sin(\theta)\\ \sin(\theta) & \cos(\theta)\end{pmatrix}$ has $\det A = \cos^2(\theta) + \sin^2(\theta) = 1$.
The vectors

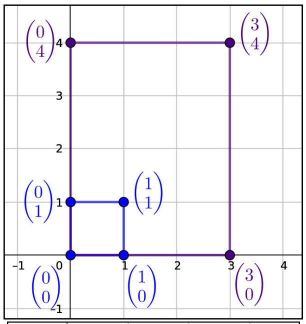
$$\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1\\0\end{pmatrix}, \begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}0\\1\end{pmatrix}$$

get mapped to

$$\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\cos(\theta)\\ \sin(\theta)\end{pmatrix}, \begin{pmatrix}\cos(\theta)-\sin(\theta)\\ \sin(\theta)+\cos(\theta)\end{pmatrix}, \begin{pmatrix}-\sin(\theta)\\ \cos(\theta)\end{pmatrix}.$$

This represents the same square rotated by θ.

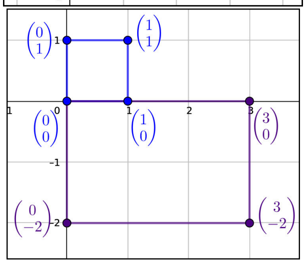
You can change the angle and see what happens in GeoGebra:
`https://www.geogebra.org/graphing/yt8v8cdq`



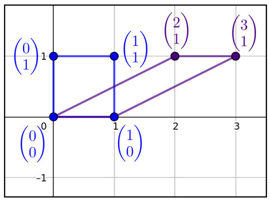
$A = \begin{pmatrix}3 & 0\\ 0 & 4\end{pmatrix}$ has $\det A = 12$. This map scales everything in the x-direction by 3 and everything in the y-direction by 4, so the unit square gets mapped to

$$\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}3\\0\end{pmatrix}, \begin{pmatrix}3\\4\end{pmatrix}, \begin{pmatrix}0\\4\end{pmatrix}$$

so we get a rectangle with area 12.



$A = \begin{pmatrix}3 & 0\\ 0 & -2\end{pmatrix}$ has negative determinant $\det A = -6$. This corresponds to the orientation of the square being reversed.



$A = \begin{pmatrix}1 & 2\\ 0 & 1\end{pmatrix}$ is a shear: $\det A = 1$, the area stays the same, the square just gets transformed into a parallelogram. You can imagine the square being made up out of lots of horizontal sheets of paper, and then you gently push the stack so that they slip sideways.

So we have already seen that

Proposition 3.35: *A 2×2 matrix A is invertible if and only if* $\det A \neq 0$.

Proof. Let $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. If $\det A = ad - bc \neq 0$, then

$$\frac{1}{\det A}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{\det A}\begin{pmatrix} da - bc & db - bd \\ -ca + ac & -cb + ad \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

So A has an inverse.
Conversely, if $\det A = ad - bc = 0$, then $ad = bc$, so

- ◇ either A has a zero row or zero column, and so it is not invertible (Prop. 3.10 and Corollary 3.11),
- ◇ or $\frac{d}{b} = \frac{c}{a}$, so multiplying the first row by this number gives the second row. This means that the row echelon form has a zero row, so the matrix is not invertible. □

This determinant satisfies the following properties:

Formula 3.36: (Properties of 2×2 determinants)

(1) $\det \begin{pmatrix} ra & b \\ rc & d \end{pmatrix} = rad - rcb = r \det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$,

(2) $\det \begin{pmatrix} a+x & b \\ c+z & d \end{pmatrix} = (a+x)d - b(c+z) = ad - bc + xd - bz = \det \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \det \begin{pmatrix} x & b \\ z & d \end{pmatrix}$

(3) $\det \begin{pmatrix} b & a \\ d & c \end{pmatrix} = bc - ad = -(ad - bc) = -\det \begin{pmatrix} a & b \\ c & d \end{pmatrix}$.

(4) $\det \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1$.

Our aim now is to define the determinant of arbitrary $n \times n$ matrices such that A is invertible if and only if the determinant is non–zero. We want this determinant of an $n \times n$ matrix to have the same properties as we've just written down for the 2×2 determinant.

E. Determinants by cofactor expansion

Definition 3.37: (Cofactor expansion across first row) Let $A = (a_{ij})$ be a $n \times n$–matrix, $n \geqslant 2$. The **determinant** of A is defined by

$$\det A = \sum_{j=1}^{n} a_{1j}(-1)^{1+j} \det(\widehat{A}_{1j}),$$

where $\widehat{A}_{ij}$ is the matrix one gets by deleting the i–th row and the j–th column of A.
The determinant of the matrix $\widehat{A}_{ij}$ is called the **minor** of entry a_{ij} and $(-1)^{i+j}\det(\widehat{A}_{ij})$ is called the **cofactor** of entry a_{ij}.

The hat is often used when we leave out something, so in $\widehat{A}_{ij}$ we leave out row i and column j.
Note that we introduce the determinant inductively: the determinant of a 3×3 matrix is by definition given by a linear combination of determinants of 2×2 matrices, which we have already defined. Similarly, the determinant of an $n \times n$ matrix can be computed if one knows the determinant of $n - 1 \times n - 1$ matrices.

Notation 3.38: When we write out a matrix with it's entries, we often use vertical lines to denote determinant:

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

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix}$$

$$\det\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix} = \begin{vmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{vmatrix}$$

We don't use this notation when giving the matrix a name! So we write $\det A$ not $|A|$.

Example 3.39: In the 3×3 case, we get

$$\begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} = a \cdot \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \cdot \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \cdot \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

Example 3.40: We will look at two 3×3 examples.

a) Let $A = \begin{pmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{pmatrix}$. Then

$$\widehat{A}_{11} = \begin{pmatrix} 4 & -1 \\ -2 & 0 \end{pmatrix}, \ \widehat{A}_{12} = \begin{pmatrix} 2 & -1 \\ 0 & 0 \end{pmatrix}, \ \widehat{A}_{13} = \begin{pmatrix} 2 & 4 \\ 0 & -2 \end{pmatrix},$$

and

$$\begin{aligned} \det A &= \sum_{j=1}^{3} (-1)^{1+j} a_{1j} \det \widehat{A}_{1j} \\ &= \underbrace{(-1)^{1+1}}_{=1} \underbrace{a_{11}}_{=1} \det \widehat{A}_{11} + \underbrace{(-1)^{1+2}}_{=-1} \underbrace{a_{12}}_{=5} \det \widehat{A}_{12} + \underbrace{(-1)^{1+3}}_{=1} \underbrace{a_{13}}_{=0} \det \widehat{A}_{13} \\ &= 1 \cdot 1 \cdot \begin{vmatrix} 4 & -1 \\ -2 & 0 \end{vmatrix} + (-1) \cdot 5 \cdot \begin{vmatrix} 2 & -1 \\ 0 & 0 \end{vmatrix} + 1 \cdot 0 \cdot \begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} \\ &= -2 - 5 \cdot 0 + 0 = -2. \end{aligned}$$

b) The determinant of

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

by definition is

$$\det A = 3 \cdot \begin{vmatrix} -4 & 3 \\ 4 & -2 \end{vmatrix} - 1 \cdot \begin{vmatrix} -2 & 3 \\ 5 & -2 \end{vmatrix} + 0 \cdot \begin{vmatrix} -2 & -4 \\ 5 & 4 \end{vmatrix} = -1\,.$$

The term determinant was first introduced by the German mathematician Carl Friedrich Gauss in 1801 (see Section 2.1 in the WileyPLUS book), who used them to "determine" properties of certain kinds of functions. The term minor is apparently due to the English

mathematician James Sylvester (see Section 2.1 in the WileyPLUS book), who wrote the following in a paper published in 1850: "Now conceive any one line and any one column be struck out, we get... a square, one term less in breadth and depth than the original square; and by varying in every possible selection of the line and column excluded, we obtain, supposing the original square to consist of n lines and n columns, n^2 such minor squares, each of which will represent what I term a 'First Minor Determinant' relative to the principal or complete determinant."
Cofactor expansion is not the only method for expressing the determinant of a matrix in terms of determinants of lower order. For example, although it is not well known, the English mathematician Charles Dodgson, who was the author of Alice's *Adventures in Wonderland* and *Through the Looking Glass* under the pen name of Lewis Carroll, invented such a method, called "condensation." That method has recently been resurrected from obscurity because of its suitability for parallel processing on computers.

Instead of expanding along the first row, we can also expand along any other row, or even along any column.

Proposition 3.41: (General cofactor expansions)
Let $A = (a_{ij})$ be a $n \times n$–matrix, $n \geqslant 2$. Then

$$\begin{aligned} \det A &= \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(\widehat{A}_{ij}) \quad \textit{for any } i \qquad \textit{(expand in row } i\textit{)} \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(\widehat{A}_{ij}) \quad \textit{for any } j \qquad \textit{(expand in column } j\textit{)} \end{aligned}$$

PROOF. Omited here. The fact is that these are all equal to a formula which gives a big alternating sum of products of entries of A, which involves some concepts we cannot cover in this course. □

Slogan: The determinant of A is formed by taking all possible products of entries with exactly one from each row and each column, and adding them all together with plus or minus.

Example 3.42: Compute the cofactor expansion of $A = \begin{pmatrix} 1 & 5 & 0 \\ 2 & 4 & -1 \\ 0 & -2 & 0 \end{pmatrix}$ across the 3^{rd} column:

$$\begin{aligned} \det A &= (-1)^{1+3} \cdot 0 \cdot \begin{vmatrix} 2 & 4 \\ 0 & -2 \end{vmatrix} + (-1)^{2+3} \cdot (-1) \cdot \begin{vmatrix} 1 & 5 \\ 0 & -2 \end{vmatrix} + (-1)^{3+3} \cdot 0 \cdot \begin{vmatrix} 1 & 5 \\ 2 & 4 \end{vmatrix} \\ &= -2. \end{aligned}$$

So we see this *can* be very useful: we can make our work easier by picking rows or columns which have a lot of zeros in them.

Proposition 3.43: *If A contains a zero row or a zero column then* $\det A = 0$.

PROOF. Expand across the zero row or the zero column. □

As another useful consequence of the expansion across rows we get an easy formula to compute determinants of 3×3–matrices:

Formula 3.44: (Garden fence rule for 3×3 matrices)

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

Exercise 3.45: Verify this formula by expanding along the first row.

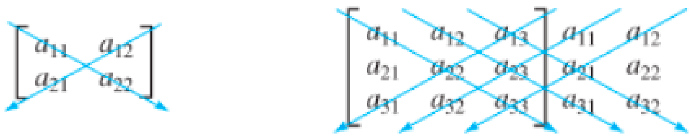


Figure 1. “garden fence” rule for 2×2 and 3×3 determinants,

Be aware that these simple methods only work for 2×2 and 3×3 matrices. For matrices of larger size there is no simple rule.

You can see here also the idea of “take one entry from each row and each column, take their product, and then add it all up, some with plus and some with minus”. It just happens that for bigger matrices, that idea doesn’t come out in as nice a pattern as this garden fence rule.

F. Properties of determinant

Because we can expand along rows or columns, transposing a matrix does not change the determinant:

Proposition 3.46: (Determinant of transpose)
Let $A \in \mathscr{M}_{n,n}$. Then

$$\det A^T = \det A$$

Proof. Since we can calculate the determinant by expanding in a row or a column, we get the same answer if we expand $\det A^T$ in row 1 and $\det A$ in column 1, say. $\square$

We can now show that the properties we calculated for 2×2 determinants also hold for a determinant of a $n \times n$ matrix:

Proposition 3.47: (Properties of determinant)
Let $a_1, \ldots, a_n$ and b_k be columns of $n \times n$ matrices (i.e. vectors in $\mathbb{R}^n$), and $a'_1, \ldots, a'_n$ and b'_k rows of $n \times n$ matrices. Then

(i) “scalar multiple comes out of a column or a row”:

$$\det \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & \lambda a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} = \lambda \det \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} \text{ and}$$

$$\det \begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \lambda a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix} = \lambda \det \begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix}$$

(ii) determinant is linear in columns and rows:

$$\det \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k + b_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} = \det \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} + \det \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & b_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} \text{ and}$$

$$\det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k+b'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix} = \det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix} + \det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & b'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix}$$

(iii) swapping columns or rows introduces a minus sign:

$$\det\begin{pmatrix} \uparrow & & \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_l & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \end{pmatrix} = -\det\begin{pmatrix} \uparrow & & \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_l & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \end{pmatrix} \textit{ and}$$

$$\det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_l & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix} = -\det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_l & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix}$$

(iv) a matrix with two columns or rows the same has determinant 0*:*

$$\det\begin{pmatrix} \uparrow & & \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \end{pmatrix} = 0 \qquad \textit{and} \qquad \det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix} = 0$$

(v) adding multiple of another column or row to a given column or row doesn't change the determinant:

$$\det\begin{pmatrix} \uparrow & & \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k+\lambda a_l & \cdots & a_l & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \end{pmatrix} = \det\begin{pmatrix} \uparrow & & \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_l & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \end{pmatrix} \textit{and}$$

$$\det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k+\lambda a'_l & \rightarrow \\ & \vdots & \\ \leftarrow & a'_l & \rightarrow \\ & \vdots & \\ \leftarrow & a'_n & \rightarrow \end{pmatrix} = \det\begin{pmatrix} \leftarrow & a'_1 & \rightarrow \\ & \vdots & \\ \leftarrow & a'_k & \rightarrow \\ & \vdots & \\ \leftarrow & a'_l & \rightarrow \\ & \vdots & \\ \leftrightarrow & a'_n & \rightarrow \end{pmatrix}$$

(vi) $\det I_n = 1$.

Proof. (i) You can do this by induction, or by choosing the right column/row to expand in.

Let $A = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$ and $B = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & \lambda a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$, i.e. $b_{ik} = \lambda a_{ik}$ for $i = 1, \ldots, n$. Then we expand in column k:

$$\begin{aligned} \det B &= \sum_{i=1}^{n} (-1)^{k+i} b_{ik} \det(\widehat{B}_{ik}) \\ &= \sum_{j=1}^{n} (-1)^{k+i} \lambda a_{ik} \det(\widehat{A}_{ik}) \\ &= \lambda \det A. \end{aligned}$$

For rows, this follows by $\det A^T = \det A$, Proposition 3.46. (Or we expand in the row instead.)

(ii) Exercise. Expand in row/column k.

(iii) (Stretch yourself Exercise). Can prove by induction and expanding in a row/column that is not one of the two we've swapped.

(iv) Follows from (iii): if two columns are the same, swapping them introduces a minus sign but also keeps the matrix the same.

(v) Combine (i), (ii) and (iv).
(vi) We compute the determinant of the identity matrix $I_n = (a_{ij})$ as

$$\begin{aligned}\det I_n &= (-1)^{1+1} \underbrace{a_{11}}_{=1} \det I_{n-1} + \sum_{j=2}^{n} (-1)^{1+j} \underbrace{a_{1j}}_{=0} (\widehat{I_n})_{1j} \\ &= 1 \cdot \det I_{n-1} = \ldots = \det I_2 = 1.\end{aligned}$$

□

The previous result shows how in some circumstances we can decompose a determinant into a sum of two determinants.
However, in general, the determinant of a sum of square matrices is not the sum of the determinants:

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

have determinants $\det A = 1, \det B = 5$ so that $\det A + \det B = 6$ but

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

has determinant $\det(A + B) = 0 \neq 6$.

Remark 3.48: In fact, the determinant is uniquely determined by the following properties:
A map $\det\colon \mathscr{M}_{n,n} \longrightarrow \mathbb{R}$ sending $A \longmapsto \det A$ is called **determinant** if the following holds:

(1) det is linear in each column. That is, if a_i are columns:
   (i) $\det(a_1, \ldots, a_{k-1}, a_k + b_k, a_{k+1}, \ldots, a_n) =$
   $\det(a_1, \ldots, a_{k-1}, a_k, a_{k+1}, \ldots, a_n) + \det(a_1, \ldots, a_{k-1}, b_k, a_{k+1}, \ldots, a_n)$
   (ii) $\det(a_1, \ldots, a_{k-1}, ra_k, a_{k+1}, \ldots, a_n) = r \det(a_1, \ldots, a_{k-1}, a_k, \ldots, a_n)$

(2) det is alternating: swapping two columns introduces a minus sign:

$$\begin{aligned}&\det(a_1, \ldots, a_{k-1}, a_k, a_{k+1}, \ldots, a_{l-1}, a_l, a_{l+1}, \ldots, a_n) = \\ -&\det(a_1, \ldots, a_{k-1}, a_l, a_{k+1}, \ldots, a_{l-1}, a_k, a_{l+1}, \ldots, a_n)\end{aligned}$$

(3) $\det I_n = 1$.

Proposition 3.49: (Determinant of upper triangular)
If A is upper triangular, i.e, if

$$A = \begin{pmatrix} a_{11} & * & \ldots & * \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & * \\ 0 & \ldots & 0 & a_{nn} \end{pmatrix}$$

then

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

A similar statement holds for lower triangular matrices.

Proof. Expand across the first column:

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

□

In general, expansion across rows or columns is not a useful tool to calculate a determinant. Apply this method only in the case that the matrix contains rows/columns with lots of zeros.

G. Evaluating determinants by row reduction

In this section we will show how to evaluate a determinant by reducing the associated matrix to row echelon form. In general, this method requires less computation than cofactor expansion and hence is the method of choice for large matrices, .

Proposition 3.50: (Determinants of elementary matrices)
Let $E \in \mathscr{M}_{n,n}$ be an elementary matrix. Then
- *(i) If E results from multiplying a row of I_n by a non-zero scalar λ, then* $\det(E) = \lambda$.
- *(ii) If E results from interchanging two rows of I_n, then* $\det E = -1$.
- *(iii) If E results from adding a multiple of one row of I_n to another, then* $\det E = 1$.

PROOF. This follows immediately from Proposition 3.47. □

So, more visually:

$$\det\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} = \lambda \qquad \det\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} = -1 \qquad \det\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 1 & \cdots & \lambda & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix} = 1$$

In particular, all elementary matrices have non-zero determinant.
Using this, we can rewrite some parts of Proposition 3.47 in terms of elementary matrices:

Proposition 3.51: (Determinant preserves product with elementary matrices.)
Let E be an elementary matrix, and A an $n \times n$ matrix.
- *(i) If E multiplies a row with $\lambda \in \mathbb{R}$ then*
$$\lambda \det A = \det(EA) = \det E \det A.$$
- *(ii) If E interchanges rows then*
$$-\det A = \det(EA) = \det E \det A.$$
- *(iii) If E adds a multiple of one row to another row then*
$$\det A = \det(EA) = \det E \det A.$$

PROOF. Recall from Prop. 3.19 that calculating EA has the effect of applying the elementary row operation corresponding to E to the matrix A. So this follows from the properties of determinant given in Proposition 3.47. □

Thus: if we only apply *elementary* row operations then we have control over the determinant! Therefore, we can use row reduction to compute the determinant of a matrix.

Example 3.52:

$$\underbrace{\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{vmatrix}}_{=6} \overset{\mathrm{II}+2\mathrm{I}}{=} \begin{vmatrix} 1 & 2 & 3 \\ 2 & 6 & 10 \\ 0 & 0 & 3 \end{vmatrix} = 3\begin{vmatrix} 1 & 2 \\ 2 & 6 \end{vmatrix} = 6$$

$$\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{vmatrix} \overset{\mathrm{II}/2}{=} 2 \begin{vmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 3 \end{vmatrix} = 2 \cdot 3$$

Take the factor 2 out of second row

$$\begin{vmatrix} 1 & 2 & 3 \\ 0 & 2 & 4 \\ 0 & 0 & 3 \end{vmatrix} \overset{\mathrm{I}\leftrightarrow\mathrm{III}}{=} (-1) \begin{vmatrix} 0 & 0 & 3 \\ 0 & 2 & 4 \\ 1 & 2 & 3 \end{vmatrix} = (-1) \cdot 3 \begin{vmatrix} 0 & 2 \\ 1 & 2 \end{vmatrix} = 6$$

Even with today's fastest computers it would take millions of years to calculate a determinant of matrix of size 25×25, unless it is very sparsely filled. Methods based on row reduction are often used for large determinants.

Example 3.53: To evaluate $\det A$ where

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

we use row reduction:

$$\det A \overset{\mathrm{I}\leftrightarrow\mathrm{II}}{=} - \begin{vmatrix} 3 & -6 & 9 \\ 0 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix} \overset{\mathrm{I}/3}{=} -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 2 & 6 & 1 \end{vmatrix}$$

$$\overset{\mathrm{III}-2\mathrm{I}}{=} -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 10 & -5 \end{vmatrix} \overset{\mathrm{III}-10\mathrm{II}}{=} -3 \begin{vmatrix} 1 & -2 & 3 \\ 0 & 1 & 5 \\ 0 & 0 & -55 \end{vmatrix} = (-3)(-55) = 165$$

where we used that the determinant of an upper triangular matrix is given as the product of the diagonal entries.

H. Determinant of a matrix product

Considering the complexity of the formulas for determinants and matrix multiplication, it would seem unlikely that a simple relationship should exist between them. This is what makes the simplicity of our next result so surprising. We will show that if A and B are square matrices of the same size, then $\det(AB) = \det A \det B$.
We already know from Proposition 3.51 that the product is preserved if one of the matrices is an elementary matrix:

If $E, B \in \mathscr{M}_{n,n}$ and E is an elementary matrix, then

$$\det(EB) = \det E \det B.$$

So applying that several times we get:

Corollary 3.54: *If $B \in \mathscr{M}_{n,n}$ and $E_1, \ldots, E_r \in \mathscr{M}_{n,n}$ are elementary matrices, then*

$$\det(E_1 E_2 \ldots E_r B) = \det(E_1) \ldots \det(E_r) \det(B).$$

In particular,

Theorem 3.55: *A square matrix A is invertible if and only if* $\det A \neq 0$.

Proof. Let R be the reduced row echelon form of A and $E_1, \ldots, E_r$ the elementary matrices which give $R = E_1 \ldots E_r A$. By the previous theorem we have

$$\det(R) = \det(E_1) \ldots \det(E_r) \det(A).$$

Since elementary matrices have non-zero determinant we see that $\det R = 0$ if and only if $\det A = 0$.
Now, if A is invertible then $R = I_n$ and $\det A \neq 0$ since $\det I_n = 1 \neq 0$.
Conversely, if $\det A \neq 0$ then $\det R \neq 0$ which tells us that R cannot have a row of zeros. Thus, $R = I_n$ and hence A is invertible. □

So we can add this to our list of equivalent invertibility conditions:

Theorem 3.56: (Invertibility conditions)
For a square matrix A, the following are equivalent:

(i) A is invertible.
(ii) $Ax = 0$ has only the trivial solution $x = 0$.
(iii) The reduced row echelon form of A is the identity I.
(iv) A is the product of elementary matrices.
(v) $\det A \neq 0$.

We are now ready for the main result concerning products of matrices.

Theorem 3.57: *If $A, B \in \mathscr{M}_{n,n}$ then*
$$\det(AB) = \det(A)\det(B)\,.$$

Proof. We divide the proof into two cases that depend whether or not A is invertible.
If A is singular then by Corollary 3.29 we see that AB is singular: if AB were invertible, then A would also be invertible. Thus by the previous theorem $\det(A) = \det(AB) = 0$ which shows the statement in this case.
If A is invertible, then A is the product of elementary matrices $E_1, \ldots, E_r$. Thus
$$\det(AB) = \det(E_1 \ldots E_r B) = \det(E_1) \ldots \det(E_r)\det(B) = \det(A)\det(B).$$
□

In 1815 the great French mathematician Augustin Cauchy published a landmark paper in which he gave the first systematic and modern treatment of determinants. It was in that paper that Theorem 3.57 was stated and proved in full generality for the first time. Special cases of the theorem had been stated and proved earlier, but it was Cauchy who made the final jump.

Example 3.58: Given
$$A = \begin{pmatrix} 3 & 1 \\ 2 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} -1 & 3 \\ 5 & 8 \end{pmatrix} \qquad \text{and} \qquad AB = \begin{pmatrix} 2 & 3 \\ 17 & 14 \end{pmatrix},$$
we have
$$\begin{aligned} \det A &= 3 - 2 = 1, \\ \det B &= -8 - 15 = -23, \\ \det(AB) &= 28 - 51 = -23 = \det A \cdot \det B. \end{aligned}$$

Proposition 3.59: *If A is invertible, then*
$$\det(A^{-1}) = \frac{1}{\det A}$$

Proof. Since $A^{-1}A = I_n$ we have
$$\det(A^{-1})\det(A) = \det(I_n) = 1$$
which shows the result. □

I. Determinants: Study guide

Concept review.

- ◇ Determinant
- ◇ Minor
- ◇ Cofactor
- ◇ Cofactor expansion
- ◇ Properties of determinant
- ◇ The effect of elementary row operations on the value of a determinant.
- ◇ The determinants of the three types of elementary matrices.
- ◇ Determinant test for invertibility.
- ◇ Determinant of product.
- ◇ Determinant of inverse.

Skills.

- ◇ Use cofactor expansion to evaluate the determinant of a square matrix.
- ◇ Use the garden fence rule to evaluate the determinant of a 2×2 or 3×3 matrix.

- ◇ Find the determinant of a matrix which contains a zero row or column.
- ◇ Find the determinant of an upper triangular, lower triangular, diagonal matrix by inspection.
- ◇ Find the determinant of a transpose matrix.
- ◇ Use row reduction to evaluate the determinant of a matrix.
- ◇ Combine the use of row reduction and cofactor expansion to evaluate the determinant of a matrix.
- ◇ Use the determinant to test a matrix for invertibility.
- ◇ Compute the determinant of products.
- ◇ Compute the determinant of an inverse.

CHAPTER 4

Vector Spaces

A. Vector space axioms and examples

Recall from Chapter 1 that the set $\mathbb{R}^n$ of all vectors with n real entries satisfies certain conditions that make it a *vector space*. We will now look at these conditions again and see more examples of such real vector spaces.

Definition 4.1: A **real vector space** is a set V equipped with addition and scalar multiplication by real numbers, satisfying the following axioms: for any $u,\ v,\ w \in V$ and $\lambda,\ \mu \in \mathbb{R}$,

VA0	$u + v \in V$	(closure under addition)
VA1	There exists a zero vector $0 \in V$ which satisfies $v + 0 = v = 0 + v$.	(zero vector)
VA2	There are negative vectors satisfying $v + (-v) = 0 = (-v) + v$.	(negative vectors)
VA3	$(u + v) + w = u + (v + w)$	(associativity of addition)
VA4	$u + v = v + u$	(commutativity of addition)
SM0	$\lambda v \in V$	(closure under scalar multiplication)
SM1	$1 \cdot v = v$	(unit scalar)
SM2	$\lambda \cdot (\mu v) = (\lambda \cdot \mu)v$	(associativity of scalar mult)
SM3	$(\lambda + \mu)\ v = \lambda v + \mu v$	(distributivity of scalar mult over real addition)
SM4	$\lambda(u + v) = \lambda u + \lambda v$	(distribuitivity of scalar mult over vector addition)

A **vector** is defined to be an element of a vector space.

Examples 4.2:

- ◇ We have seen that for any natural number n, $\mathbb{R}^n$ is a vector space, where vector addition and scalar multiplication is defined entry-wise.
- ◇ We have seen that the set of $m \times n$ matrices $\mathscr{M}_{m,n}$ is a vector space, where addition and scalar multiplication is also defined entry-wise.
- ◇ Consider the set P_n of real polynomials of degree less than or equal to n. So a general such polynomial is $a_0 + a_1x + a_2x^2 + a_3x^3 + \cdots + a_nx^n$, with real coefficients $a_i \in \mathbb{R}$. (Note that the coefficients are allowed to be 0, so this includes polynomials of smaller degree.)

  Given two such polynomials $p = a_0 + a_1x + \cdots + a_nx^n$ and $q = b_0 + b_1x + \cdots + b_nx^n$, we can add them:

$$p + q = (a_0 + b_0) + (a_1 + b_1)x + \cdots + (a_n + b_n)x^n$$

  which is another polynomial of degree at most n. So VA0 is satisfied: $p + q \in P_n$.

  The zero polynomial $0 = 0 + 0x + 0x^2 + \cdots + 0x^n$ satisfies $0 + p = 0 = p + 0$ for any other polynomial $p \in P_n$. So VA1 is satisfied.

  Given a polynomial p, $-p$ is also a polynomial of the same degree as p, so VA2 is satisfied.

  It is easy to check that VA3 and VA4 also hold.

  Given a polynomial p as above and a real number $\lambda \in \mathbb{R}$, then

$$\lambda p = (\lambda a_0) + (\lambda a_1)x + (\lambda a_2)x^2 + \cdots + (\lambda a_n)x^n$$

  is another polynomial of degree at most n, so SM0 holds. It is then also easy to check that SM1-SM4 hold as well. E.g., for SM2:

$$\begin{aligned}\lambda \cdot (\mu p) &= \lambda \cdot ((\mu a_0) + (\mu a_1)x + (\mu a_2)x^2 + \cdots + (\mu a_n)x^n) \\ &= (\lambda\mu a_0) + (\lambda\mu a_1)x + (\lambda\mu a_2)x^2 + \cdots + (\lambda\mu a_n)x^n \\ &= (\lambda\mu)p\end{aligned}$$

So the set P_n of polynomials of degree at most n forms a vector space.

- ◇ If we take just the set of polynomials of degree exactly n, they do not form a vector space: For example,
$$(3 + 2x + 5x^2) + (2 + x - 5x^2) = 5 + 3x$$
the sum of two degree 3 polynomials might have smaller degree. So VA0 is not satisfied.
- ◇ Let $C[0,1]$ be the set of continuous functions from $[0,1] \longrightarrow \mathbb{R}$, i.e. real valued continuous functions on the interval $[0,1]$. We can add such functions pointwise, meaning $(f+g)(x) = f(x) + g(x)$, and we can multiply such a function by a scalar: $(\lambda f)(x) = \lambda \cdot f(x)$. Then:
  - The sum of two continuous functions is again continuous, so VA0 holds.
  - The zero function $f(x) = 0$ is continuous, and satisfies VA1.
  - If f is continuous, then so is λf for any real number λ, so SM0 is satisfied. Taking $\lambda = -1$ gives a function which satisfies VA2.
  - VA3 and VA4, and SM1-4, can be easily checked. For example, SM3: To check two functions $[0,1] \longrightarrow \mathbb{R}$ are equal, we have to check that they do the same thing on each $x \in [0,1]$:
  $$((\lambda + \mu)f)(x) = (\lambda + \mu) \cdot f(x) = \lambda \cdot f(x) + \mu \cdot f(x)$$
  and
  $$(\lambda f + \mu f)(x) = (\lambda f)(x) + (\mu f)(x) = \lambda \cdot f(x) + \mu \cdot f(x)$$
  So $(\lambda + \mu)f = \lambda f + \mu f$.
- ◇ Let V be the set of all infinite real sequences, with entry-wise addition and scalar multiplication. I.e. $(a_i)_i + (b_i)_i = (a_i + b_i)_i$, and $\lambda(a_i)_i = (\lambda a_i)_i$. This also forms a vector space.
- ◇ If we take the set P of all real polynomials (regardless of degree), then we also get a vector space.

As you can see in these examples, the way you prove that some given set with two operations is a vector space is the following:

- ◇ Identify the set V whose elements are the vectors.
- ◇ Identify the operations of addition and scalar multiplication. (I.e. how are these defined in your example of V?)
- ◇ Verify that these operations satisfy VA0 and SM0. We then say that V is **closed under addition and scalar multiplication**.
- ◇ Verify that there is a zero vector (i.e. VA1) and that there are negative vectors (i.e. VA2).
- ◇ Verify all the other axioms.

Exercise 4.3: Show that the last two examples, infinite sequences and all polynomials, form a vector space.

We can have some slightly stranger vector spaces:

Example 4.4: (Unusual vector space) Let $V = \mathbb{R}^+$, the set of positive real numbers, and let $u + v = u \cdot v$, i.e. vector addition is multiplication of the real numbers, and let $\lambda u = u^\lambda$, i.e. scalar multiplication is exponentiation.
So, for example $1 + 1 = 1$ and $2 \cdot 1 = 1^2 = 1$.
Is this a vector space?

- ◇ We have identified the set V and the two operations.

- ◇ If u, v are positive real numbers, then their vector sum $u + v = u \cdot v$ is again a positive real number. So VA0 holds.
- ◇ If u is a positive real number, then $\lambda \cdot u = u^\lambda$, any power of u, is also a positive real number. So SM0 holds.
- ◇ The "zero vector" is the number 1: $1 + u = 1 \cdot u = u$, and $u + 1 = u \cdot 1 = u$.
- ◇ "Negative vectors" are the reciprocals: $u + \frac{1}{u} = u \cdot \frac{1}{u} = 1$, the zero vector.
- ◇ VA3 and VA4 hold, because for multiplication of real numbers we have $(u \cdot v) \cdot w = u \cdot (v \cdot w)$, and $u \cdot v = v \cdot u$.
- ◇ SM1: $1 \cdot v = v^1 = v$ so this is true.
- ◇ SM2: $\lambda \cdot (\mu v) = (\mu v)^\lambda = (v^\mu)^\lambda = v^{\lambda \cdot \mu} = (\lambda\mu)v$ by rules of indices.
- ◇ SM3: $(\lambda + \mu)v = v^{\lambda+\mu} = v^\lambda \cdot v^\mu = (\lambda v) + (\mu v)$ by rules of indices and definition of vector addition.
- ◇ SM4: $\lambda(u + v) = (u \cdot v)^\lambda = u^\lambda \cdot v^\lambda = (\lambda u) + (\lambda v)$.

So this is a vector space!

Example 4.5: Let $V = \mathbb{R}^2$ with the usual vector addition, but define scalar multiplication like this: for $v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathbb{R}^2$ and $\lambda \in R$, let

$$\lambda v = \begin{pmatrix} \lambda v_1 \\ 0 \end{pmatrix}.$$

Is this also a vector space?

- ◇ We have identified the set V, and the operations: vector addition as usual, and scalar multiplication as above.
- ◇ We verify VA0: $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \end{pmatrix} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \end{pmatrix} \in \mathbb{R}^2$. And SM0: $\lambda v = \begin{pmatrix} \lambda v_1 \\ 0 \end{pmatrix} \in \mathbb{R}^2$.
- ◇ As the vector addition is the usual one, the zero vector $0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ satisfies VA1. Similarly, $-v = \begin{pmatrix} -v_1 \\ -v_2 \end{pmatrix}$ satisfies VA2.
- ◇ Let's try to verify the other axioms:

  VA3 and VA4 only involve vector addition, and that is the same as for the normal $\mathbb{R}^2$, so they hold.

  SM1: $1 \cdot v = \begin{pmatrix} v_1 \\ 0 \end{pmatrix} \neq v$ so SM1 does not hold.

So this is not a vector space!

The point of defining such an abstract structure is that we can use just those axioms to prove things that will hold for *any* example of a vector space. For example, zero vectors and negative vectors behave as we expect them to:

Proposition 4.6: (Zeros and negatives in vector spaces)
Let V be a vector space. Then

(i) The zero vector is unique.
(ii) For any $v \in V$, $0v = 0$: zero times any vector gives the zero vector.
(iii) For any $\lambda \in \mathbb{R}$, $\lambda 0 = 0$: any multiple of the zero vector is still the zero vector.
(iv) If $\lambda v = 0$, then either $\lambda = 0 \in \mathbb{R}$, or $v = 0 \in V$.
(v) For any $v \in V$, the negative $-v$ is unique.
(vi) For any $v \in V$, $(-1)v = -v$, the negative of v.

You can view this as lots of examples on how we use axioms to prove simple statements.

PROOF. We don't know anything about V except that it satisfies the axioms VA0-4 and SM0-4. So in our proofs, that is the only thing we can use. We will record carefully which axioms are used where.

(i) Suppose we have two zero vectors 0 and $0'$ which both satisfy Axiom VA1. Then

$$\begin{aligned} 0' &= 0' + 0 && \text{using VA1 for } 0 \\ &= 0 && \text{using VA1 for } 0' \end{aligned}$$

(ii) We have

$$\begin{aligned} && v &= 1 \cdot v && \text{using SM1, unit scalar} \\ && &= (1+0) \cdot v && \text{calculating in } \mathbb{R} \\ && &= 1 \cdot v + 0 \cdot v && \text{using SM3, distributivity} \\ && &= v + 0 \cdot v && \text{using SM1, unit scalar} \\ \Leftrightarrow && (-v) + v &= (-v) + (v + 0 \cdot v) && \text{adding } -v \text{ to both sides} \\ \Leftrightarrow && (-v) + v &= ((-v) + v) + 0 \cdot v && \text{using VA3, associativity} \\ \Leftrightarrow && 0 &= 0 + 0 \cdot v && \text{using VA2, negative vectors} \\ \Leftrightarrow && 0 &= 0 \cdot v && \text{using VA1, zero vector} \end{aligned}$$

(iii) Exercise. There are different ways you can do this: using $0 + 0 = 0$ and the existence of negative vectors (i.e. using VA1 and VA2, with ideas similar to ii), or using vi.

(iv) If $\lambda = 0$ then we are done. Suppose $\lambda \neq 0$. So we want to show that $v = 0$. As $\lambda \neq 0$, we have $\frac{1}{\lambda} \in \mathbb{R}$. So

$$\begin{aligned} v &= 1 \cdot v && \text{using SM1, unit scalar} \\ &= (\frac{1}{\lambda}\lambda) \cdot v && \text{calculating in } \mathbb{R} \\ &= \frac{1}{\lambda}(\lambda v) && \text{using SM2} \\ &= \frac{1}{\lambda} \cdot 0 && \text{by assumption: } \lambda v = 0 \\ &= 0 && \text{using iii} \end{aligned}$$

(v) Suppose a given $v \in V$ has two negatives: u and w. So $v+w = 0 = w+v$ and $v+u = 0 = u+v$, i.e. both u and w satisfy VA2 for v. Then

$$\begin{aligned} u &= u + 0 && \text{using VA1, zero vector} \\ &= u + (v + w) && \text{using VA2 with } w \\ &= (u + v) + w && \text{using VA3, associativity} \\ &= 0 + w && \text{using VA2 with } u \\ &= w && \text{using VA1, zero vector.} \end{aligned}$$

So $u = w = -v$, the negative is unique.

This is exactly the same proof as "inverse matrices are unique". This is because negatives are the "inverse with respect to addition".

(vi) We have

$$\begin{aligned} v + (-1) \cdot v &= 1 \cdot v + (-1) \cdot v && \text{using SM1, unit scalar} \\ &= (1 + (-1)) \cdot v && \text{using SM3, distributivity} \\ &= 0 \cdot v && \text{calculating in } \mathbb{R} \\ &= 0 && \text{using ii} \end{aligned}$$

Similarly $(-1) \cdot v + v = 0$. So $(-1) \cdot v$ satisfies the condition to be the negative of v, so by uniqueness of negatives, $(-1)v = -v$. $\square$

This result tells us that the following example is possible:

Example 4.7: The smallest possible vector space is the **zero vector space**: $0 = \{0\}$. It has only one vector in it, the zero vector. It is closed under vector addition because $0 + 0 = 0$ by VA1. It is closed under scalar multiplication, because $\lambda \cdot 0 = 0$, by Proposition 4.6iii. All the other axioms are easy to check, because we have just one vector.

B. Subspaces

We already saw in Chapter 1 that the vector space $\mathbb{R}^n$ can have subspaces: smaller vector spaces that are inside $\mathbb{R}^n$. We will recap the definition, now applying it to any vector space V, not just $\mathbb{R}^n$.

Definition 4.8: Let V be a vector space. A **subspace** of V is a subset $W \subseteq V$ that is also a vector space, with the same vector addition and scalar multiplication as V. We write $W \leqslant V$ when W is a subspace of V.

Examples 4.9:

- Every vector space V has the zero space as a subspace, and itself as a subspace.

$$0 \leqslant V \qquad \text{and} \qquad V \leqslant V$$

- We have seen subspaces such as

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ 0 \end{pmatrix} \middle| x_1, x_2, x_3 \in \mathbb{R} \right\} \leqslant \mathbb{R}^4$$

and

$$\left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_1 - x_2 \\ 3x_3 \end{pmatrix} \middle| x_1, x_2, x_3 \in \mathbb{R} \right\} \leqslant \mathbb{R}^5$$

- As mentioned above, the set containing only 0 is a subspace of $\mathbb{R}$ with usual addition, and usual multiplication as scalar multiplication. (I.e. viewing $\mathbb{R}$ as $\mathbb{R}^1$, vectors with one entry.)
- The "unusual vector space" $\mathbb{R}^+$ with $u + v = u \cdot v$ and $\lambda v = v^\lambda$ is *not* a subspace of $\mathbb{R}$: while positive real numbers are a subset of all real numbers, the vector addition and scalar multiplication are not the same in the two vector spaces.
- In fact the only subspaces of $\mathbb{R}$ are 0 and $\mathbb{R}$: if a subspace $W \leqslant \mathbb{R}$ has some element $w \in W$, $w \neq 0$, then as W is closed under scalar multiplication, we get $\frac{1}{w} \cdot w = 1 \in W$. Then for any $x \in \mathbb{R}$, we get $x \cdot 1 \in W$, so W is all of $\mathbb{R}$.

Proposition 4.10: (Subspace conditions)
A subset $W \subseteq V$ of a vector space V is a subspace if and only if it satisfies:

- $0 \in W$ *(zero vector is in the set)*
- *for any $u, v \in W$, $u + v \in W$* *(closed under vector addition)*
- *for any $v \in W$ and any $\lambda \in \mathbb{R}$, $\lambda v \in W$* *(closed under scalar mult)*

Proof. If we check these three, then we automatically get negative vectors: $-v = (-1) \cdot v$ by Proposition 4.6.
The other axioms VA3-4 and SM1-4 are inherited from V.
Conversely, if W is a subspace, then it is a vector space with same addition and scalar multiplication, so by definition it is closed under addition and scalar multiplication. We also have the

existance of a zero vector 0_W which satisfies $0_W + w = w$ for all $w \in W$. How do we know that this zero vector for W is the same as the zero vector for V?
We know from Proposition 4.6 that in any vector space, zero times any vector gives the zero vector. So if $w \in W$, we have $0 \cdot w = 0_W$. But w is also in V, so $0 \cdot w = 0_V$ also. So we must have $0_W = 0_V$: the zero vector is the same for the subspace as for the bigger space. $\square$

Let's see some subspaces of the vector space of all $n \times n$ matrices $\mathscr{M}_{n,n}$.

Examples 4.11:
- ◇ The set of diagonal $n \times n$ matrices forms a subspace of $\mathscr{M}_{n,n}$.
  - The zero matrix is diagonal.
  - Adding two diagonal matrices gives another diagonal matrix.
  - The scalar multiple of a diagonal matrix is another diagonal matrix.
- ◇ The set of upper triangular $n \times n$ matrices forms a subspace of $\mathscr{M}_{n,n}$.
  - The zero matrix is upper triangular.
  - Adding two upper triangular matrices gives another upper triangular matrix. Essentially this is because in the entries below the diagonal we have $0 + 0 = 0$.
  - The scalar multiple of an upper triangular matrix is upper triangular. This is because in the entries below the diagonal, we have $\lambda \cdot 0 = 0$.
- ◇ In the same way, the set of lower triangular $n \times n$ matrices forms a subspace of $\mathscr{M}_{n,n}$.

We can have some more:

Definition 4.12: A square matrix A is called **symmetric** if $A = A^T$. It is called **skew-symmetric** or **anti-symmetric** if $A = -A^T$.

Examples 4.13: These matrices are symmetric:

◇ $\begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$ ◇ $\begin{pmatrix} -23 & 19 \\ 19 & 37 \end{pmatrix}$ ◇ $\begin{pmatrix} 3 & 4 & 5 \\ 4 & 8 & 2 \\ 5 & 2 & 9 \end{pmatrix}$

These matrices are anti-symmetric:

◇ $\begin{pmatrix} 0 & 3 \\ -3 & 0 \end{pmatrix}$ ◇ $\begin{pmatrix} 0 & -19 \\ 19 & 0 \end{pmatrix}$ ◇ $\begin{pmatrix} 0 & 4 & 5 \\ -4 & 0 & -2 \\ -5 & 2 & 0 \end{pmatrix}$

Another way to view symmetric matrices is to say that A is symmetric if $a_{ij} = a_{ji}$ for all i, j. And an anti-symmetric matrix has to satisfy $a_{ij} = -a_{ij}$.
Notice that the diagonal elements of a symmetric matrix are completely unrestricted, but the diagonal elements of anti-symmetric matrices all have to be 0. That is because $a_{ii} = -a_{ii}$ in an anti-symmetric matrix.

Example 4.14: The set Sym_n of symmetric $n \times n$ matrices forms a subspace of $\mathscr{M}_{n,n}$.

- ◇ The zero matrix 0 satisfies $0^T = 0$, so it is symmetric.
- ◇ If A, B are symmetric, then $(A + B)^T = A^T + B^T = A + B$, so the sum is symmetric.
- ◇ If A is symmetric and $\lambda \in \mathbb{R}$, then $(\lambda A)^T = \lambda A^T = \lambda A$, so it is also symmetric.

Exercise 4.15: Show that the set ASym_n of anti-symmetric matrices form a subspace of $\mathscr{M}_{n,n}$.

If we have two subspaces $U \leqslant V$ and $W \leqslant V$, we can form their intersection, and we can form their union.

Proposition 4.16: (Intersection and union of subspaces)
Let V be a vector space, and U, W be subspaces of V. Then
(i) The intersection $U \cap W$ is also a subspace of V.
(ii) The union $U \cup W$ is a subspace of V if and only if either $U \subseteq W$ or $W \subseteq U$.

Proof. Exercise. □

So the second statement here shows that unions are hardly ever subspaces. The right way to "combine" subspaces is not by using union, but by using sum.

Definition 4.17: Let U, W be two subspaces of the vector space V. The **vector space sum** $U + W$ is the set of vectors

$$U + W = \{u + w \mid u \in U,\ w \in W\}.$$

Proposition 4.18: (Sum of subspaces)
Let U, W be subspace of a vector space V. Then the sum $U + W$ is another subspace of V which contains both U and W: $U \leqslant U + W$ and $W \leqslant U + W$.

Proof.
- ◇ $0 = 0 + 0$, with $0 \in U$ and $0 \in W$, so $0 \in U + W$.
- ◇ If $u_1 + w_1$ and $u_2 + w_2$ are two elements of $U + W$, then their sum is $(u_1 + u_2) + (w_1 + w_2)$, with $u_1 + u_2 \in U$ and $w_1 + w_2 \in W$, because both subspaces are closed under addition. So the sum is again in $U + W$.
- ◇ If $u + w \in U + W$ and $\lambda \in \mathbb{R}$, then $\lambda(u + w) = \lambda u + \lambda w$, with $\lambda u \in U$ and $\lambda w \in W$, because both subspaces are closed under scalar multiplication. So $\lambda(u + w) \in U + W$.

So $U + W \leqslant V$.
For any $u \in U$, we have $u = u + 0 \in U + W$, because $0 \in W$. Similarly, for any $w \in W$, we have $w = 0 + w \in U + W$. So $U + W$ contains both U and W. □

Examples 4.19:
- ◇ Let $U = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} \,\middle|\, x \in \mathbb{R} \right\}$ and $W = \left\{ \begin{pmatrix} 0 \\ y \end{pmatrix} \,\middle|\, y \in \mathbb{R} \right\}$ be subspaces of $\mathbb{R}^2$. Then $U + W = \left\{ \begin{pmatrix} x \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ y \end{pmatrix} \,\middle|\, x,\ y \in \mathbb{R} \right\} = \mathbb{R}^2$.

  This shows you the difference of union and sum: the union of the x-axis and the y-axis is just the set of the two lines, but the sum of the x-axis and the y-axis gives the whole plane. This is what we need when we are combining vector spaces.

  The sum of two subspaces is the smallest subspace that contains them both.
- ◇ Let $U = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \,\middle|\, x,\ y \in \mathbb{R} \right\}$ and $W = \left\{ \begin{pmatrix} 0 \\ y \\ z \end{pmatrix} \,\middle|\, y, z \in \mathbb{R} \right\}$. Then $U + W = \mathbb{R}^3$, but there is some intersection: $U \cap W = \left\{ \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix} \,\middle|\, y \in \mathbb{R} \right\}$.

We see that if $U \cap W$ is not just the zero space, then there is some "superfluity" in the sum. Sometimes it might be nice to distinguish between these situations.

Definition 4.20: Given subspaces U, $W \leqslant V$, we say the sum $U + W$ is a **direct sum**, and write $U \oplus W$, if $U \cap W = 0$, the intersection is just the zero space.

Examples 4.21: In the previous examples, the first one is a direct sum, but the second one isn't.

Proposition 4.22: (Symmetric and anti-symmetric matrices)
The vector space $\mathscr{M}_{n,n}$ of $n \times n$ matrices is the direct sum of the subspaces of symmetric and anti-symmetric matrices.

$$\mathscr{M}_{n,n} = \mathrm{Sym}_n \oplus \mathrm{ASym}_n$$

PROOF. We know that $\mathrm{Sym}_n + \mathrm{ASym}_n \subseteq \mathscr{M}_{n,n}$. So we have to show that any $n \times n$ matrix A can be written as the sum of a symmetric matrix and an anti-symmetric matrix.
Given A, note that $A + A^T$ is symmetric and $A - A^T$ is anti-symmetric. And we have

$$A = \frac{1}{2}(A + A^T) + \frac{1}{2}(A - A^T),$$

so $\mathscr{M}_{n,n} = \mathrm{Sym}_n + \mathrm{ASym}_n$.
To show that this is a direct sum, we have to show that $\mathrm{Sym}_n \cap \mathrm{ASym}_n = 0$. Suppose A is symmetric and anti-symetric. Then $A = A^T$ and $A = -A^T$, so $A = -A$, so $A = 0$.
So $\mathscr{M}_{n,n} = \mathrm{Sym}_n \oplus \mathrm{ASym}_n$. □

C. Column space and Null space

We can view the subspace conditions (Prop. 4.10) as saying that a subspace is a non-empty subset which is closed under linear combinations. So we should expect that "taking linear combinations of some vectors" should give us a subspace.

Definition 4.23: Given a non-empty set of vectors $S = \{w_1, w_2, \ldots, w_r\}$ in a vector space V, then the **span of** S, written

$$\mathrm{Span}(S) = \mathrm{Span}(w_1, w_2, \ldots, w_r) = \langle w_1, w_2, \ldots, w_r \rangle$$

is the set of all linear combinations of the vectors in S.

$$\langle w_1, w_2, \ldots, w_r \rangle = \{\lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_r w_r \mid \lambda_1, \ldots, \lambda_r \in \mathbb{R}\}$$

The span of the empty set is the zero space:

$$\mathrm{Span}(\varnothing) = 0.$$

Note that in our notation, the set S is finite. We can form spans of infinite sets: in that case, a linear combination of an infinite number of vectors can only have finitely many of the λ being non-zero. But mostly we'll consider spans of finite sets.

Proposition 4.24: (Span gives subspace)
The span of any non-empty subset S of V is a subspace of V.
Moreover, this is the smallest subspace of V which contains all elements of S.

The second sentence means: if W is any other subspace of V with $S \subset W$, then $\mathrm{Span}(S) \subseteq W$.

PROOF. Let $S = \{w_1, w_2, \ldots, w_r\}$.

- ◇ $0 \in \mathrm{Span}(S)$ because $0 = 0 \cdot w_1 + 0 \cdot w_2 + \cdots + 0 \cdot w_r$ is a linear combination of the vectors in S.
- ◇ Given two linear combinations $u = \lambda_1 w_1 + \cdots + \lambda_r w_r$ and $v = \mu_1 w_1 + \cdots + \mu_r w_r$, their sum is again a linear combination of the vectors in S:
$$u + v = (\lambda_1 + \mu_1)w_1 + (\lambda_2 + \mu_2)w_2 + \cdots + (\lambda_r + \mu_r)w_r.$$
So $u, v \in \mathrm{Span}(S) \Rightarrow u + v \in \mathrm{Span}(S)$.
- ◇ Given $u = \lambda_1 w_1 + \cdots + \lambda_r w_r \in \mathrm{Span}(S)$ and a scalar $\mu \in R$, then
$$\mu u = (\mu\lambda_1)w_1 + (\mu\lambda_2)w_2 + \cdots + (\mu\lambda_r)w_r \in \mathrm{Span}(S).$$

So $\mathrm{Span}(S) \leqslant V$.
Now any subspace of V is closed under taking linear combinations. So if we want all elements of S to be in a subspace, then all linear combinations of elements of S must also be in that subspace. So taking just the linear combinations of elements of S is the smallest subspace of V which contains all elements of S. □

Sometimes the span of a set might give the whole vector space V.

Definition 4.25: We say that a set $S = \{w_1, w_2, \ldots, w_r\}$ **spans** V if $\mathrm{Span}(S) = V$. If S spans V, we also call S a **spanning set** for V.

So if S spans V, then any vector $v \in V$ can be written (in at least one way) as a linear combination of vectors from S.

Examples 4.26:

◇ The set $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ spans $\mathbb{R}^2$:

$$\begin{pmatrix} x \\ y \end{pmatrix} = x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

so any vector in $\mathbb{R}^2$ is a linear combination of those two vectors.

◇ The set $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\}$ also spans $\mathbb{R}^2$: the linear combinations need not be unique.

$$\begin{aligned}
\begin{pmatrix} x \\ y \end{pmatrix} &= x \begin{pmatrix} 1 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0 \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= 0 \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (y - x) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + x \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= (x - y) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + 0 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + y \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= (-y) \begin{pmatrix} 1 \\ 0 \end{pmatrix} + (-x) \begin{pmatrix} 0 \\ 1 \end{pmatrix} + (x + y) \begin{pmatrix} 1 \\ 1 \end{pmatrix} \\
&= \ldots
\end{aligned}$$

◇ The set $\left\{ \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ spans $\mathbb{R}^3$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \left(\frac{x+y}{2} - z \right) \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \left(\frac{y-x}{2} \right) \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} + z \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

We will learn later how to find these coefficients.

◇ In $\mathbb{R}^n$, let $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$, i.e. e_k is the vector with zeros everywhere except a 1 in the kth entry. Then $e_1, \ldots, e_n$ span $\mathbb{R}^n$.

$$\begin{aligned}
\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} &= x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \cdots + x_{n-1} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix} + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \\
&= x_1 e_1 + x_2 e_2 + \cdots + x_{n-1} e_{n-1} + x_n e_n
\end{aligned}$$

These vectors are very useful vectors and come back again and again. Let's call them the **standard unit vectors**.

◇ The vectors $\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}$ span the x, y-plane in $\mathbb{R}^3$. But so do the vectors $\begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix}$.

So we have

$$\left\langle \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix} \right\rangle = \left\langle \begin{pmatrix}1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\end{pmatrix} \right\rangle .$$

◇ The monomials $\{1,\ x,\ x^2, \ldots,\ x^n\}$ span the polynomial space P_n: any poly is a linear combination of these monomials.

$$p = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_1 x + a_0 \cdot 1$$

We've seen that different sets can span the same subspace.

Proposition 4.27: (Equal spans)
Given two non-empty subsets S and S' of a vector space V, then $\mathrm{Span}(S) = \mathrm{Span}(S')$ *if and only if every vector in S is a linear combination of vectors in S', and every vector in S' is a linear combination of vectors in S.*

PROOF. Exercise. (See steps in workbook.) □

Given any $m \times n$ matrix A, the n columns of A are vectors in $\mathbb{R}^m$:

$$A = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} \qquad \text{with } a_1,\ a_2, \ldots,\ a_n \in \mathbb{R}^m$$

So we can consider the span of the columns of A, and this is a subspace of $\mathbb{R}^m$.

Definition 4.28: Given an $m \times n$ matrix A, the **column space** of A is the space spanned by the columns of A.

$$\text{For } A = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}, \qquad \mathrm{Col}(A) = \langle a_1, \ldots, a_n \rangle .$$

This has a link to linear systems: recall that

$$Ax = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = x_1 \begin{pmatrix} \uparrow \\ a_1 \\ \downarrow \end{pmatrix} + x_2 \begin{pmatrix} \uparrow \\ a_2 \\ \downarrow \end{pmatrix} + \cdots + x_{n-1} \begin{pmatrix} \uparrow \\ a_{n-1} \\ \downarrow \end{pmatrix} + x_n \begin{pmatrix} \uparrow \\ a_n \\ \downarrow \end{pmatrix}$$

So for any given vector $x \in \mathbb{R}^n$, Ax is a linear combination of the columns of A, so $Ax \in \mathrm{Col}(A)$, it is in the column space of A. This gives us

Proposition 4.29: *Let A be an $m \times n$ matrix and $b \in \mathbb{R}^m$. Then*

$Ax = b$ is consistent $\Leftrightarrow$ *b is a linear combination of the columns of A* $\Leftrightarrow$ $b \in \mathrm{Col}(A)$.

PROOF. The second equivalence is by definition of $\mathrm{Col}(A)$. The first equivalence holds because Ax is a linear combination of the columns of A, as seen above. □

We can also saying

Proposition 4.30: *Let A be an $m \times n$ matrix. Then the following are equivalent:*

(i) For every $b \in \mathbb{R}^m$, the linear system $Ax = b$ has a solution.
(ii) Every $b \in \mathbb{R}^m$ is in $\mathrm{Col}(A)$ *(i.e. is a linear combination of the columns of A).*
(iii) $\mathrm{Col}(A) = \mathbb{R}^m$*: the columns of A span all of $\mathbb{R}^m$.*

PROOF. i $\Leftrightarrow$ii follows from Proposition 4.29.
ii $\Leftrightarrow$iii: We know $\text{Col}(A) \leqslant \mathbb{R}^m$, so the two spaces are equal if and only if every vector $b \in \mathbb{R}^m$ is in $\text{Col}(A)$. But that is exactly what ii says. □

These two results give us a method to find out if a given vector is in a given span.

Corollary 4.31: (Finding linear combinations)
Let $S = \{a_1, \ldots, a_n\}$ be a finite set of vectors in a vector space V. Then $b \in V$ is a linear combination of the vectors $a_1, \ldots, a_n$ (i.e. $b \in \text{Span}(S)$) if and only if $Ax = b$ has a solution.

Moreover, any such solution $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ provides the coefficients for the linear combination:

$$b = x_1 a_1 + x_2 a_2 + \cdots + x_n a_n$$

So

Finding a linear combination To work out whether $b \in \mathbb{R}^m$ is a linear combination of $a_1, \ldots, a_n \in \mathbb{R}^m$:

- ◇ Try to solve the linear system $Ax = b$.
- ◇ If it has a solution, then that solution gives the coefficients of the linear combination.

Example 4.32: Is $b = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}$ a linear combination of $a_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$ and $a_2 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$?

Try to solve $\begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix}$.

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

So we have the solution $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. So $b = 2 \cdot a_1 + 1 \cdot a_2$. (It is a good idea to check it:)

$$\begin{pmatrix} 3 \\ 3 \\ 1 \end{pmatrix} = 2 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

We can also use this technique to work out if a given set spans all of $\mathbb{R}^n$.

Examples 4.33: ◇ Do the vectors

$$a_1 = \begin{pmatrix} 1 \\ -4 \\ -3 \end{pmatrix}, \; a_2 = \begin{pmatrix} 3 \\ 2 \\ -2 \end{pmatrix}, \; a_3 = \begin{pmatrix} 4 \\ -6 \\ -7 \end{pmatrix}$$

span all of $\mathbb{R}^3$? The vectors will span all of $\mathbb{R}^3$ if we can write an arbitrary vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3$ as a linear combination of $a_1, \ a_2, \ a_3$.

So we want to solve the linear system

$$Ax = \begin{pmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

$$\left(\begin{array}{ccc|c} 1 & 3 & 4 & x \\ -4 & 2 & -6 & y \\ -3 & -2 & -7 & z \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 3 & 4 & x \\ 0 & 14 & 10 & y+4x \\ 0 & 7 & 5 & z+3x \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 3 & 4 & x \\ 0 & 7 & 5 & \frac{y+4x}{2} \\ 0 & 0 & 0 & z+3x-\frac{y+4x}{2} \end{array}\right)$$

We can see that this system is in general inconsistent. (It is only consistent for some special constellations of $x,\ y,\ z$, not for all $x,\ y,\ z$). So these three vectors do not span all of $\mathbb{R}^3$. For example, $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ is not in the span of the three vectors.

◇ Do the vectors $a_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$, $a_2 = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and $a_3 = \begin{pmatrix} -1 \\ 4 \end{pmatrix}$ span all of $\mathbb{R}^2$?

Solve

$$Ax = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

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

This system has a solution (in fact infinitely many), so the given three vectors do span $\mathbb{R}^2$.

For example, choosing $x_3 = 0$, we can write

$$\begin{pmatrix} x \\ y \end{pmatrix} = \left(-\frac{1}{3}x + \frac{2}{3}y\right) \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \left(\frac{2}{3}x - \frac{1}{3}y\right) \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$$

◇ Do the vectors

$$a_1 = \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix},\ a_2 = \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix},\ a_3 = \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}$$

span all of $\mathbb{R}^3$?

Solve

$$Ax = \begin{pmatrix} 1 & 1 & 2 \\ 2 & 1 & 3 \\ -3 & 3 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

$$\left(\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 2 & 1 & 3 & y \\ -3 & 3 & 3 & z \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 0 & -1 & -1 & y-2x \\ 0 & 6 & 9 & z+3x \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 0 & 1 & 1 & 2x-y \\ 0 & 0 & 3 & z+3x-(12x-6y) \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 0 & 1 & 1 & 2x-y \\ 0 & 0 & 3 & z+3x-(12x-6y) \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 2 & x \\ 0 & 1 & 1 & 2x-y \\ 0 & 0 & 1 & \frac{z+6y-9x}{3} \end{array}\right)$$

$$\longrightarrow \left(\begin{array}{ccc|c} 1 & 1 & 0 & x-(\frac{2}{3}z+4y-6x) \\ 0 & 1 & 0 & 2x-y-(\frac{1}{3}z+2y-3x) \\ 0 & 0 & 1 & \frac{1}{3}z+2y-3x \end{array}\right) \longrightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 7x-4y-\frac{2}{3}z-(5x-3y-\frac{1}{3}z) \\ 0 & 1 & 0 & 5x-3y-\frac{1}{3}z \\ 0 & 0 & 1 & \frac{1}{3}z+2y-3x \end{array}\right)$$

So, as this system has a solution, the three vectors do span $\mathbb{R}^3$. And we have

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \left(2x - y - \frac{1}{3}z\right) \begin{pmatrix} 1 \\ 2 \\ -3 \end{pmatrix} + \left(5x - 3y - \frac{1}{3}z\right) \begin{pmatrix} 1 \\ 1 \\ 3 \end{pmatrix} + \left(\frac{1}{3}z + 2y - 3x\right) \begin{pmatrix} 2 \\ 3 \\ 3 \end{pmatrix}.$$

Important: we will be solving linear systems a lot in the rest of the course. It is crucial that you keep in mind the *meaning* of what you are doing: we usually solve a linear system for a reason, to find out something else we want to know. When you've solved a linear system, you then have to translate it back into what it means for the question you were trying to answer.

There is an alternative method in the case when we have a square matrix: Recall from invertible linear systems (Proposition 3.30) that an $n \times n$ matrix A is invertible if and only if $Ax = b$ is consistent for all $b \in \mathbb{R}^n$. So given n vectors $a_1, \ldots, a_n \in \mathbb{R}^n$, we know that

$$\langle a_1, \ldots, a_n \rangle = \mathbb{R}^n \quad \Leftrightarrow \quad A = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ a_1 & \cdots & a_k & \cdots & a_n \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix} \text{ is invertible.}$$

So the above links spans and the column space of a matrix to the existence of solutions of certain linear systems. We can also look at the set of solutions of certain linear systems.
We already mentioned in Chapter 2 that the solutions of a *homogeneous* linear system form a subspace.

Recall: The set of solutions to a homogeneous linear system $Ax = 0$ with n variable forms a subspace of $\mathbb{R}^n$. See Proposition 2.29.

This has a name:

Definition 4.34: For any $m \times n$ matrix A, the set of vectors $\{x \in \mathbb{R}^n \mid Ax = 0\}$ is called the **null space of** A, written $\operatorname{Null} A$. Another name for the null space is the **kernel of** A, written $\operatorname{Ker} A$.

We will learn more about kernels in the second semester.

Examples 4.35:

◇ Any invertible matrix has $\operatorname{Null} A = 0$: $Ax = 0$ has only the trivial solution $x = 0$, so the null space of A is the zero space.

A square matrix is invertible if and only if $\operatorname{Null} A = 0$.

◇ The null space of $A = \begin{pmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{pmatrix}$ is the set of solutions to $Ax = 0$:

$$\begin{pmatrix} 1 & 3 & 4 \\ -4 & 2 & -6 \\ -3 & -2 & -7 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 4 \\ 0 & 14 & 10 \\ 0 & 7 & 5 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 3 & 4 \\ 0 & 1 & \frac{5}{7} \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & \frac{13}{7} \\ 0 & 1 & \frac{5}{7} \\ 0 & 0 & 0 \end{pmatrix}$$

So

$$\operatorname{Null} A = \left\{ t \begin{pmatrix} -13 \\ -5 \\ 7 \end{pmatrix} \middle| \; t \in \mathbb{R} \right\}$$

◇ A non-square matrix can have the null space being 0 without being invertible (recall a non-square matrix can never be invertible):

Let $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$. Find null space:

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \\ 3 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 \\ 0 & -3 \\ 0 & -5 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \qquad \text{so} \qquad \operatorname{Null} A = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\} = 0.$$

D. Vector Spaces: Study guide

Concept review.

◇ Vector space, vector space axioms
◇ Space of matrices
◇ Space of polynomials
◇ Properties of zero and negatives in a vector space
◇ Subspace, subspace conditions
◇ Symmetric and skew-symmetric/anti-symmetric matrices
◇ Several matrix subspaces
◇ Intersection and union of subspaces
◇ Sum of subspaces
◇ Direct sum
◇ Span of vectors, spanning set for vector space
◇ Span is a subspace
◇ Column space of a matrix
◇ Null space of a matrix

Skills.

◇ Show a given set with operations is a vector space.
◇ Determine whether a given set with operations is a vector space or not.
◇ Prove simple results using vector space axioms.
◇ Show a given subset is a subspace.
◇ Determine whether a given subset is a subspace or not.
◇ Form intersection and sum of given subspaces.
◇ Determine whether a sum of subspaces is direct or not.
◇ Determine whether a given vector is in the span of a given set of vectors.
◇ Determine whether a set of vectors spans the whole vector space.
◇ Find the coefficients to write a given vector (or a general vector) as the linear combination of a set of vectors.
◇ Determine whether a given vector is in the column space of a matrix.
◇ Determine the null space of a matrix.

CHAPTER 5

Linear Independence and Bases

A. Linearly independent sets

We have seen that when we write a given vector as the linear combination of a set of vectors, sometimes we only have one way of doing this and sometimes we have several different ways of doing this. For example,

$$\begin{aligned}\begin{pmatrix} x \\ y \end{pmatrix} &= x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix} + 0\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= 0\begin{pmatrix} 1 \\ 0 \end{pmatrix} + (y-x)\begin{pmatrix} 0 \\ 1 \end{pmatrix} + x\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \dots\end{aligned}$$

has many options, but when using just the first two vectors

$$\begin{pmatrix} x \\ y \end{pmatrix} = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

there is only one option. We will now investigate the relationship between the vectors which determines which of these cases happens.
To do this, we focus on the number of ways we can write the zero vector as a linear combination: Given vectors $v_1, \dots, v_k \in V$, the equation

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k = 0$$

always has at least one solution: $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$. We call this the **trivial solution**, as we did for linear systems. The question now is whether there are other solutions for the λ_i which also give the zero vector.

Definition 5.1: A subset $S = \{v_1, \dots, v_k\} \subset V$ of a vector space V is called **linearly independent** if

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k = 0$$

with $\lambda_i \in \mathbb{R}$ implies that $\lambda_1 = \lambda_2 = \dots = \lambda_k = 0$.
If there are λ_i not all zero such that

$$\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k = 0$$

then the vectors are called **linearly dependent**.
We call such a linear combination with non-zero coefficents a **dependence relation**.

In words: a set of vectors is linearly independent exactly when the only way to get 0 as a linear combination of them is to set all coefficients to zero.

Examples 5.2: ◇ The vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \end{pmatrix}$ are linearly dependent. If you draw them into a coordinate system, they both lie on the x-axis, so we see they are parallel.

Two vectors v_1, v_2 are linearly dependent exactly when they are parallel:

$$\lambda_1 v_1 + \lambda_2 v_2 = 0 \text{ with } \lambda_1 \neq 0 \quad \Rightarrow \quad v_1 = -\frac{\lambda_2}{\lambda_1} v_2$$
$$v_1 = \mu v_2 \quad \Rightarrow \quad 1 \cdot v_1 - \mu v_2 = 0$$

For example, for these two vectors, $2\begin{pmatrix}1\\0\end{pmatrix} - 1 \cdot \begin{pmatrix}2\\0\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}$ is a dependence relation.

- $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$ are linearly independent. Draw them on a coordinate system.
- The vectors $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}-1\\0\end{pmatrix}$ are linearly dependent.
- It helps your intuition to picture two vectors in the x, y-plane and decide if they are linearly independent or not. If they lie on the same line, they are dependent. If they do not lie on the same line, i.e. are not parallel, then they are linearly independent.
- Are the vectors $\begin{pmatrix}1\\\frac{1}{2}\end{pmatrix}$, $\begin{pmatrix}1\\2\end{pmatrix}$ linearly independent?

  We want to know if there are any non-zero coefficients such that

$$\lambda_1 \begin{pmatrix}1\\\frac{1}{2}\end{pmatrix} + \lambda_2 \begin{pmatrix}1\\2\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}.$$

  To find this out, we could solve the linear system:

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

  So the ony solution is $\lambda_1 = 0 = \lambda_2$. So these two vectors are linearly independent.
- The standard unit vectors $e_1 = \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}$, $e_2 = \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}$, $\ldots$, $e_n = \begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix}$ are linearly independent in $\mathbb{R}^n$:

$$\lambda_1 \begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix} + \lambda_2 \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix} + \cdots + \lambda_n \begin{pmatrix}0\\\vdots\\0\\1\end{pmatrix} = \begin{pmatrix}0\\\vdots\\0\\0\end{pmatrix}$$

  has unique solution $\lambda_1 = \lambda_2 = \cdots = \lambda_n = 0$.
- The monomials 1, x, $\ldots$, x^n are linearly independent in P_n:

$$\lambda_0 \cdot 1 + \lambda_1 \cdot x + \cdots + \lambda_n \cdot x^n = 0$$

  has unique solution $\lambda_0 = \lambda_1 = \cdots = \lambda_n = 0$.
- The polynomials $p_1 = 1 + x$, $p_2 = 1 + x^2$ and $p_3 = 2 + x + x^2$ are not linearly independent (so they are linearly dependent):

$$1 \cdot p_1 + 1 \cdot p_2 - 1 \cdot p_3 = (1 + x) + (1 + x^2) - (2 + x + x^2) = 0$$

  So there is a non-trivial linear combination of these three polynomials which gives the zero polyomial. (This is a dependence relation.)
- The vectors $\begin{pmatrix}1\\1\end{pmatrix}$ and $\begin{pmatrix}1\\-1\end{pmatrix}$ are linearly independent in $\mathbb{R}^2$: To check whether

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

has non-trivial solutions, we solve the linear system

$$\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

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

so the only solution is $\lambda_1 = 0,\ \lambda_2 = 0$. So these two vectors are linearly independent.

- The vectors $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, $v_3 = \begin{pmatrix} 4 \\ 5 \\ 4 \end{pmatrix}$ are not linearly independent: for example,

$$3 \cdot \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} - 1 \cdot \begin{pmatrix} 4 \\ 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

So there is a non-trivial linear combination of the vectors which gives the zero-vector.

How could we find such coefficients? Solving

$$\lambda_1 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix} + \lambda_3 \begin{pmatrix} 4 \\ 5 \\ 4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

gives us:

$$\begin{pmatrix} 1 & 1 & 4 \\ 1 & 2 & 5 \\ 1 & 1 & 4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 4 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 3 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

So there are infinitely many solutions, not just $\lambda_1 = \lambda_2 = \lambda_3 = 0$. So the set is linearly dependent. Choosing $\lambda_3 = -1$ gives the linear combination we wrote down above.

We can see from these examples that when we have a set of vectors in $\mathbb{R}^n$, we have a method of finding out if they are linearly independent.

To find out whether a set $\{a_1,\ a_2, \ldots, a_n\}$ of vectors in $\mathbb{R}^m$ is linearly independent:

- Form the $m \times n$ matrix A with the vectors as columns.
- Solve the homogeneous linear system $Ax = 0$.
- If $x = 0$ is the only solution, the vectors are linearly independent. If there are other solutions, then the vectors are linearly dependent.

How do we check in other vector spaces whether a given set of vectors is linearly independent? For example in a polynomial space or in a matrix space?

This will depend on the example, and we will learn some methods later. With a bit of thinking, you can make a linear system out of these situations. For example:

- In a matrix space, you get one equation per entry of the matrix, which you can put into a linear system.
- In a polynomial space, you can compare coefficients, i.e. you get one equation by looking at coefficients of x^n, another from coefficients of x^{n-1} and so on. so you can put these equations into a linear system.

Example 5.3:

- Are the matrices $A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$, $A_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $A_3 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ linearly independent in $\mathscr{M}_{2,2}$? (They are also all symmetric matrices, so you can think of them in the vector space Sym_2 as well.) We want to solve

$$\lambda_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + \lambda_3 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Looking at each of the four entries, this gives us four equations:

$$\begin{aligned}1\cdot\lambda_1+0\cdot\lambda_2+0\cdot\lambda_3&=0\\0\cdot\lambda_1+1\cdot\lambda_2+0\cdot\lambda_3&=0\\0\cdot\lambda_1+1\cdot\lambda_2+0\cdot\lambda_3&=0\\0\cdot\lambda_1+0\cdot\lambda_2+1\cdot\lambda_3&=0\end{aligned}$$

or

$$\begin{pmatrix}1&0&0\\0&1&0\\0&1&0\\0&0&1\end{pmatrix}\begin{pmatrix}\lambda_1\\\lambda_2\\\lambda_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\\0\end{pmatrix}.$$

We can see that the only solution is $\lambda_1=\lambda_2=\lambda_3=0$, so these three matrices are linearly independent in $\mathscr{M}_{2,2}$ (and in Sym_2).

- Are the polynomials $p_1=x^2-3x+1$, $p_2=2x^2+x-2$ and $p_3=x^2+4x-3$ linearly independent in P_2?

  We want to solve

$$\lambda_1(x^2-3x+1)+\lambda_2(2x^2+x-2)+\lambda_3(x^2+4x-3)=0.$$

  Looking at the coefficients of x^2, we need

$$\lambda_1+2\lambda_2+\lambda_3=0.$$

  The coefficients of x tell us that we need

$$-3\lambda_1+\lambda_2+4\lambda_3=0.$$

  The constant coefficients give

$$\lambda_1-2\lambda_2-3\lambda_3=0.$$

  So we have to solve

$$\begin{pmatrix}1&2&1\\-3&1&4\\1&-2&-3\end{pmatrix}\begin{pmatrix}\lambda_1\\\lambda_2\\\lambda_3\end{pmatrix}=\begin{pmatrix}0\\0\\0\end{pmatrix}.$$

  The reduced row echelon form of this matrix is

$$\begin{pmatrix}1&0&-1\\0&1&1\\0&0&0\end{pmatrix}$$

  so there are non-trivial solutions. So p_1, p_2, p_3 are not linearly independent.

  For example, this is a non-trivial linear combiation of them which gives 0:

$$p_1-p_2+p_3=(x^2-3x+1)-(2x^2+x-2)+(x^2+4x-3)=0.$$

As well as having the algebraic way of working out whether a set of vectors is linearly independent, it's also good to have an intuition.

Intuition about linear independence

- For 2 vectors: dependent $\Leftrightarrow$ parallel
- For 3 vectors: dependent means they span a plane. Independent means they span a 3-dimensional space.

We phrased the definition of linear independence and dependence by looking at a linear combination of all the vectors, all on one side of the equation. In most situations, this is the most useful way to look at this, because that means we have a linear system we can solve, as we've seen above.

However, now and then it can be useful to think of it this way:

Proposition 5.4: *Let $S = \{v_1, v_2, \dots, v_k\}$ be a set of vectors in a vector space V. Then*
(i) S is linearly dependent if and only if one of the vectors in S can be written as a linear combination of the remaining vectors in S.
(ii) S is linearly independent if none of the vectors in S can be written as a linear combination of the other vectors in S.

PROOF. Suppose S is linearly dependent, so we have a dependence relation

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = 0$$

with not all the λ_i being zero. So at least one of the λ_i is non-zero, say $\lambda_l \neq 0$. Then we can write

$$v_l = -\frac{\lambda_1}{\lambda_l} v_1 - \frac{\lambda_2}{\lambda_l} v_2 - \cdots - \frac{\lambda_{l-1}}{\lambda_l} v_{l-1} - \frac{\lambda_{l+1}}{\lambda_l} v_{l+1} + \cdots - \frac{\lambda_k}{\lambda_l} v_k.$$

So we can write v_l as a linear combination of the other vectors in S.
Conversely, suppose we can write v_l as a linear combination of the other vectors,

$$v_l = \mu_1 v_1 + \cdots + \mu_{l-1} v_{l-1} + \mu_{l+1} v_{l+1} + \cdots + \mu_k v_k$$

then we can put everything on one side to get the dependence relation

$$\mu_1 v_1 + \cdots + \mu_{l-1} v_{l-1} - v_l + \mu_{l+1} v_{l+1} + \cdots + \mu_k v_k = 0.$$

So S is linearly dependent.
The second statement follows immediately from the first: If all the λ_i have to be zero to get

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = 0,$$

then we can't put any of the vectors on the other side. $\square$

This immediately shows us:

Corollary 5.5: (Easy to spot dependent sets)
(i) Any set containing the zero vector is linearly dependent.
(ii) A set with exactly one vector is linearly independent if and only if that vector is not the zero vector.
(iii) A set with exactly two vectors is linearly independent if and only if neither vector is a scalar multiple of the other.

PROOF. (i) If $S = \{v_1, v_2, \dots, v_k\}$ with say $v_k = 0$, then we have a dependence relation

$$0v_1 + 0v_2 + \cdots + 0v_{k-1} + 1 \cdot v_k = 0,$$

so the set is linearly dependent.
(ii) If we have only one vector v_1, then either $v_1 = 0$, in which case $1 \cdot v_1 = 0$ is a dependence relation, or $v_1 \neq 0$, in which case the only way to get zero is $0 \cdot v_1$.
(iii) We've seen this already:

$$\begin{aligned} \lambda_1 v_1 + \lambda_2 v_2 = 0 \text{ with } \lambda_1 \neq 0 \quad &\Rightarrow \quad v_1 = -\frac{\lambda_2}{\lambda_1} v_2 \\ v_1 = \mu v_2 \quad &\Rightarrow \quad 1 \cdot v_1 - \mu v_2 = 0 \end{aligned}$$

$\square$

Exercise 5.6: Determine if the following sets are linearly independent or not.

◇ $\{0\}$ in $\mathbb{R}^4$. ◇ $\left\{\begin{pmatrix}1\\0\\0\end{pmatrix}\right\}$ in $\mathbb{R}^3$. ◇ $\left\{\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\-1\end{pmatrix}\right\}$ in $\mathbb{R}^2$.

◇ $\left\{\begin{pmatrix}1\\-1\\0\end{pmatrix}, \begin{pmatrix}-4\\4\\0\end{pmatrix}\right\}$ in $\mathbb{R}^3$. ◇ $\{p_1 = x,\ p_2 = x^2,\ p_3 = 0\}$ in P_2.

Our intuition about a set of two or three vectors being linearly independent or not should tell us: If I have 3 vectors in $\mathbb{R}^2$, then they must be linearly dependent.
In fact, we can prove this in more generality:

Proposition 5.7: (Too many vectors)
A set of more than n vectors in $\mathbb{R}^n$ is always linearly dependent.

Proof. Let $S = \{v_1, v_2, \ldots, v_k\}$ be a set of vectors in $\mathbb{R}^n$, with $k > n$. Let A be the matrix with the vectors $v_1, \ldots, v_k$ as columns. So A is an $n \times k$ matrix.

$$A = \begin{pmatrix} \uparrow & & \uparrow & & \uparrow \\ v_1 & \cdots & v_l & \cdots & v_k \\ \downarrow & & \downarrow & & \downarrow \end{pmatrix}$$

Then the system $Ax = 0$, with $x = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{pmatrix}$, has more variables than equations. So the reduced row echelon form of A definitely has some stuff columns, so the system has a non-trivial solution (Proposition 2.34). So the set is linearly dependent. □

B. Bases

We have now learnt the concepts of spanning set and of a linearly independent set. Putting them together gives one of the most important concepts of Linear Algebra.

Definition 5.8: A **basis** in a vector space V is a set of vectors that is

- ⋄ linearly independent and
- ⋄ spans V.

Note that the singular is *basis*, not *base*. A "base" is not a term that has a special definition in Linear Algebra. The plural of basis is bases, said with a long ee. So we have one basis and many "basees".

A good way to think about a basis is as a coordinate system for a vector space. The first condition tells us that there is no inter-reelation between the basis vectors, and the second tell us that we have enough basis vectors to give coordinates for every vector in V.

Examples 5.9: (Bases)
Since we have already seen examples of spanning sets and examples of linearly independent sets, we can put things together.

a) In $\mathbb{R}^n$, the set of standard unit vectors $\{e_1, e_2, \ldots, e_n\}$ is a basis: we call it the **standard basis of $\mathbb{R}^n$**.

b) In $\mathbb{R}^2$, the set $\left\{ v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \ v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$ is a basis:
   - ⋄ We have seen that this set is linearly independent.
   - ⋄ It also spans $\mathbb{R}^2$:

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{x+y}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{x-y}{2} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

c) The monomials $\{1, x, x^2, \ldots, x^n\}$ are a basis of P_n. We call this the **standard basis of P_n**.

Exercise 5.10: Prove that the following are bases of the given vector space.

a) $\{v_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), \ v_2 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\}$ in $\mathbb{R}^2$.
b) $\{w_1 = \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right), \ w_2 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\}$ in $\mathbb{R}^2$.

c) $\left\{ v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \; v_2 = \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \; v_3 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\}$ in $\mathbb{R}^3$. (See earlier examples.)

d) $\left\{ v_1 = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}, \; v_2 = \begin{pmatrix} 2 \\ 9 \\ 0 \end{pmatrix}, \; v_3 = \begin{pmatrix} 3 \\ 3 \\ 4 \end{pmatrix} \right\}$ in $\mathbb{R}^3$.

e) The set of matrices $\left\{ E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \; E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \; E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \; E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$ in $\mathscr{M}_{2,2}$.

f) More generally, let E_{ij} be the $m \times n$ matrix with zeros everywhere, except a 1 in the i,jth entry. Show that $E_{11}, E_{12}, \ldots, E_{1n}, E_{21}, E_{22}, \ldots, E_{2n}, E_{31}, \ldots, E_{m1}, \ldots, E_{mn}$ form a basis of $\mathscr{M}_{m,n}$.

You see that

a vector space can have several different bases.

We will learn more about bases, and later find a way of determining especially good bases for a given situation.
In all our examples so far, we've had a finite set being a basis of a vector space. However, this is not always possible.

Proposition 5.11: *The vector space P of all polynomials has no finite spanning set.*

Proof. Suppose we have a finite spanning set. Then let k be the maximum degree of all the polynomials in this spanning set. Then the polynomial $x^{k+1} \in P$ cannot possibly be written as a linear combination of the spanning set, because we can't create something of larger degree. So the set does not span P after all. $\square$

Some of our results only work for vector spaces which have a finite basis.

Definition 5.12: A vector space V is said to be **finite-dimensional** if it has a finite spanning set. If V has no finite spanning set, we say it is **infinite-dimensional**.

You see that we have defined something to do with dimensions without yet knowing what dimension means formally. We need a little bit more before we can properly define dimension, other than saying that something is finite-dimensional or not.

C. Coordinates Relative to a Basis

Let's have a little summary:

Let V be a (real) vector space. A set $\{v_1, v_2, \ldots, v_n\} \subset V$ of vectors in V

- ⋄ is **linearly independent** if $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = 0$ for $\lambda_i \in \mathbb{R}$ implies $\lambda_i = 0$ for all i;
- ⋄ **spans** V if very vector in V can be written as some linear combination of the v_i;
- ⋄ is a **basis** of V if it is linearly independent and spans V.

We said that we can think of a basis as something like a coordinate system. Let's make this more precise.

Theorem 5.13: (Uniqueness of basis representation)
If $S = \{v_1, \ldots, v_n\}$ is a basis for a vector space V, then every vector $v \in V$ can be expressed in exactly one way as a linear combination of S.

PROOF. As S is a basis, it spans V. So given any $v \in V$ there exist $x_1, \ldots, x_n \in \mathbb{R}$ with

$$v = x_1 v_1 + \ldots + x_n v_n .$$

Assume that for $y_1, \ldots, y_n \in \mathbb{R}$ we have

$$v = y_1 v_1 + \ldots + y_n v_n .$$

i.e. there is some other linear combination which gives v.
We need to show that $x_1 = y_1, \ldots, x_n = y_n$, i.e. that they are actually the same linear combination. Subtracting one from the other we get

$$0 = (x_1 - y_1)v_1 + \ldots + (x_n - y_n)v_n .$$

Since S is linearly independent this gives $x_1 - y_1 = \ldots = x_n - y_n = 0$ which shows the claim. □

You can think of it this way:

- ◇ A basis spans $V \Rightarrow$ every $v \in V$ can be written as a linear combination of basis in *at least one* way.
- ◇ A basis is linearly independent $\Rightarrow$ every $v \in V$ can be written as a linear combination it *at most one* way.

Thus, given a basis $S = \{v_1, \ldots, v_n\}$ of a vector space V, we have a unique representation of every vector in V by a vector in $\mathbb{R}^n$.

Definition 5.14: If $S = \{v_1, \ldots, v_n\}$ is a basis for a vector space V and $v = x_1 v_1 + \cdots + x_n v_n$, then the scalars $x_1, \ldots, x_n$ are called the **coordinates** of v relative to the basis S. The vector

$$[v]_S = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

is called the **coordinate vector** of v relative to S.

Put differently, there is a one-to-one map from V to $\mathbb{R}^n$, assigning to each vector in V a unique vector in $\mathbb{R}^n$.

Example 5.15: Let $V = \mathbb{R}^2$ and let $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ be the standard basis E. Then the coordinate vector of $v = x_1 \cdot e_1 + x_2 \cdot e_2$ is

$$[v]_E = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

Now let B be the basis consisting of $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. Then

$$\begin{pmatrix} x \\ y \end{pmatrix} = y \begin{pmatrix} 1 \\ 1 \end{pmatrix} + (x - y) \begin{pmatrix} 1 \\ 0 \end{pmatrix},$$

so the coordinate vector of $v = \begin{pmatrix} x \\ y \end{pmatrix}$ with respect to B is

$$[v]_B = \begin{pmatrix} y \\ x - y \end{pmatrix}.$$

Exercise: Find the coordinate vector of v with respect to the basis $\tilde{S} = \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \end{pmatrix} \right\}$.

So how do we find these coefficients to make a coordinate vector? We learnt how to do this in "Finding linear combinations", Corollary 4.31.

Finding coordinate vector
Given a basis $B = \{v_1, v_2, \ldots, v_n\}$ of a vector space V, we find the coordinate vector $[v]_B$ of a vector $v \in V$ by solving the linear system

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n = v.$$

If $V = \mathbb{R}^n$, then:

- ◇ Put the basis vectors $v_1, \ldots, v_n$ next to each other into a matrix A.
- ◇ Solve the system $Ax = v$ for the given vector v, which may be a particular vector, or a general $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

If V is not $\mathbb{R}^n$, you have to use the methods mentioned earlier to set up your linear system to solve. In all cases:

- ◇ The solution of this sytem gives the coefficients $\lambda_1, \lambda_2, \ldots, \lambda_n$.
- ◇ Then $[v]_B = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}$.

Exercise 5.16: Find the coordinate vector for the polynomial

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n$$

with respect to the standard basis of P_n.

Exercise 5.17: Let $v, w \in V$, $\lambda, \mu \in \mathbb{R}$ and $v_1, v_2, \ldots, v_n$ be a basis B of V. Show that

$$[\lambda v + \mu w]_B = \lambda [v]_B + \mu [w]_B.$$

In words: the coordinate vector of a linear combination is the linear combination of the coordinate vectors.

D. Linear Independence and Bases: Study guide

Concept review.

- ◇ Linearly indepdendent set, linearly dependent set.
- ◇ Intuition about sets of 1, 2, 3 vectors being linearly independent.
- ◇ Sets of more than n vectors in $\mathbb{R}^n$.
- ◇ Basis of a vector space.
- ◇ Finite-dimensional vector space.
- ◇ Uniqueness of basis representation.
- ◇ Coordinate vectors.

Skills.

- ◇ Determine whether a set is linearly independent.
- ◇ Determine whether a set is a basis.
- ◇ Find coordinate vector of some vector with respect to some basis.

CHAPTER 6

Bases and Dimension

The results in this chapter are very crucial to the understanding of Linear Algebra, and much of the theory relies on them.
We are interested in the relationship of linearly indepenent sets, spanning sets and bases, their sizes, and if we can turn one into another in certain circumstances. We need to understand these relationships in order to be able to properly define *dimension*, which we know intuitively as "the degrees of freedom".

A. Dimension

First notice the following:
We showed in the previous chapter that the standard basis of $\mathbb{R}^n$ has n vectors. So the standard basis for $\mathbb{R}^3$ has three vectors, the standard basis for $\mathbb{R}^2$ has two vectors, and the standard basis for $\mathbb{R}^1 = \mathbb{R}$ has one vector. Since we think of space as three dimensional, a plane as two dimensional, and a line as one dimensional, there seems to be a link between the number of vectors in a basis and the dimension of a vector space. We will develop this idea in this section.

Proposition 6.1: (Bases of one space have the same size.)
Let V be a finite-dimensional vector space, and let $S = \{v_1, \dots, v_n\}$ be any basis.

(i) If a set has more than n vectors, then it is linearly dependent.
(ii) If a set has fewer than n vectors, then it does not span V.

PROOF. (i) Let $\tilde{S} = \{w_1, \dots, w_k\}$, $k > n$. Since S is a basis, we can represent each vector w_i by its coordinate vector $[w_i]_S$. To show that $\tilde{S}$ is linearly dependent, consider

$$\lambda_1 w_1 + \cdots + \lambda_k w_k = 0.$$

Using the coordinate vectors with respect to S, this is equivalent to

$$\lambda_1 [w_1]_S + \cdots + \lambda_k [w_k]_S = 0\,.$$

But $[w_1]_S, \dots, [w_k]_S$ are k vectors in $\mathbb{R}^n$ with $k > n$, therefore by Proposition 5.7 we have too many vectors, and we know that $\{[w_1]_S, \dots, [w_k]_S\}$ is linearly dependent. So there are $\lambda_1, \dots, \lambda_k$ not all zero, such that

$$\lambda_1 w_1 + \cdots + \lambda_k w_k = 0\,.$$

Therefore, $\tilde{S}$ is linearly dependent.

Summary: We turn the vectors w_i into coordinate vectors so that we can use knowledge from $\mathbb{R}^n$.

(ii) Let $\tilde{S} = \{w_1, \dots, w_m\}$, $m < n$. If we assume $\mathrm{Span}\tilde{S} = V$, we can represent each vector $v_i \in S$ as a linear combination of $\tilde{S}$, i.e.,

$$\begin{aligned} v_1 &= a_{11} w_1 + \cdots + a_{m1} w_m \\ \vdots &\quad \vdots \\ v_n &= a_{1n} w_1 + \cdots + a_{mn} w_m \end{aligned}$$

Consider the equation $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$. Substituting the expresssions fo v_i into this equation, and reordering we have

$$(\lambda_1 a_{11} + \cdots + \lambda_n a_{1n})w_1 + \cdots + (\lambda_1 a_{m1} + \cdots + \lambda_n a_{mn})w_m = 0\,.$$

Let's call it

$$\mu_1 w_1 + \cdots + \mu_m w_m = 0,$$

i.e. $\lambda_1 a_{11} + \cdots + \lambda_n a_{1n} = \mu_1$ and so on. Now consider the homogeneous linear system

$$\begin{aligned} \lambda_1 a_{11} + \cdots + \lambda_n a_{1n} &= 0 \\ \vdots \quad & \quad \vdots \\ \lambda_1 a_{m1} + \cdots + \lambda_n a_{mn} &= 0 \end{aligned}$$

This is a homogeneous linear system with n unknowns and m equations where $n > m$. By Proposition 2.34, as this system has more unknowns than equations, there are non-trivial solutions $\lambda_1, \ldots, \lambda_n$ to this equation. But then the coefficients μ_i in the equation

$$\mu_1 w_1 + \cdots + \mu_m w_m = 0$$

all vanish (with some non-zero λ_i), so if we rearrange it back to

$$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$$

we see that this has a non-trivial solution. This contradicts our assumption that S was a basis: in particular, S is linearly independent. □

Summary: Assume that $\tilde{S}$ spans V, then show that this implies that S is linearly dependent. Contradiction. So $\tilde{S}$ does not span V.

Therefore, any basis of a vector space must have the same number of vectors.

Definition 6.2: The **dimension** of a finite-dimensional vector space V is denoted by $\dim(V)$ or $\dim V$ and is defined to be the number of vectors in a basis for V. In addition, the zero vector space is defined to have dimension zero.

Example 6.3: $\dim \mathbb{R}^n = n$, $\dim P_n = n + 1$, $\dim \mathscr{M}_{mn} = mn$.

Example 6.4: If $S = \{v_1, \ldots, v_k\}$ is linearly independent, then S is a basis for $\operatorname{Span}(S)$. This implies that $\dim \operatorname{Span}(S) = k$.
In words: the dimension of the space spanned by a linearly independent set of vectors is equal to the number of vectors in that set.
For example:

- ◇ $v_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$ in $\mathbb{R}^3$ is the basis of the one-dimensional subspace $\operatorname{Span}(v_1) = \langle v_1 \rangle$.
- ◇ $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$ in $\mathbb{R}^4$ are the basis of the two-dimensional subspace $W = \langle v_1, v_2 \rangle$ they span.
- ◇ $v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}, v_3 = \begin{pmatrix} 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}$ are *not* linearly independent (as $v_3 = v_1 + v_2$). We have $\langle v_1, v_2, v_3 \rangle = \langle v_1, v_2 \rangle = W$, but v_1, v_2, v_3 do *not* form a basis of W. Any two of them do, so v_1, v_2 form a basis of W, and v_1, v_3 also form a basis of W, and v_2, v_3 also.

Example 6.5: It is sometimes useful to think of dimension as "degrees of freedom" or "how many numbers can I choose till a vector of that vector space is determined". For example:
The vector space of symmetric $n \times n$ matrices has dimension $n + \frac{1}{2}(n^2 - n) = \frac{1}{2}(n^2 + n)$. We can see this in different ways:

- ◇ We can choose all the entries on the diagonal, that is n entries; then of the remaining $n^2 - n$ entries, we can choose half, because that determines the other half (since the matrix is symmetric).

- ◇ OR: Let's say we choose the lower half of the matrix (including the diagonal). So in the first row, we can choose 1 entry, in the second row we choose 2 entries, in the third row 3 entries, etc., up to n entries in row n. So we have chosen $1+2+3+\cdots+n = \frac{1}{2}n(n+1)$ entries.

That is a good way to think of it. How do we prove that this is really correct, in terms of the definition of dimension as the size of a basis? We have to give some basis of the symmetric matrices. Essentially, for each entry we are allowed to choose, we can make a basis vector (or matrix in this case).
For symmetric 2×2 matrices, a nice basis is

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, S_{12} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

For 4×4 matrices, we can give

$$E_1 = \begin{pmatrix} 1&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0&0&0&0\\0&1&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0&0&0&0\\0&0&0&0\\0&0&1&0\\0&0&0&0 \end{pmatrix}, \quad E_4 = \begin{pmatrix} 0&0&0&0\\0&0&0&0\\0&0&0&0\\0&0&0&1 \end{pmatrix},$$

$$S_{12} = \begin{pmatrix} 0&1&0&0\\1&0&0&0\\0&0&0&0\\0&0&0&0 \end{pmatrix}, \quad S_{13} = \begin{pmatrix} 0&0&1&0\\0&0&0&0\\1&0&0&0\\0&0&0&0 \end{pmatrix}, \quad S_{14} = \begin{pmatrix} 0&0&0&1\\0&0&0&0\\0&0&0&0\\1&0&0&0 \end{pmatrix},$$

$$S_{23} = \begin{pmatrix} 0&0&0&0\\0&0&1&0\\0&1&0&0\\0&0&0&0 \end{pmatrix}, \quad S_{24} = \begin{pmatrix} 0&0&0&0\\0&0&0&1\\0&0&0&0\\0&1&0&0 \end{pmatrix}, \quad S_{34} = \begin{pmatrix} 0&0&0&0\\0&0&0&0\\0&0&0&1\\0&0&1&0 \end{pmatrix}$$

Exercise: Write down a basis for the 3×3 symmetric matrices.

You can use similar ideas to find the basis of the null space of a matrix (or the solution space of a homogeneous linear system):

Examples 6.6: ◇ Let $A = \begin{pmatrix} 1&5&3\\2&7&9\\1&2&6 \end{pmatrix}$. We want to find a basis for $\operatorname{Null} A$, which is the set of solutions to $Ax=0$.

We know how to solve this system and write down the set of solutions:

$$\begin{pmatrix} 1&5&3\\2&7&9\\1&2&6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1&5&3\\0&-3&3\\0&-3&3 \end{pmatrix} \longrightarrow \begin{pmatrix} 1&5&3\\0&1&-1\\0&0&0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1&0&8\\0&1&-1\\0&0&0 \end{pmatrix}$$

So $\operatorname{Null} A = \left\{ t \begin{pmatrix} -8\\1\\1 \end{pmatrix} \right\}$.

We have one free variable in the solution, so we have "one degree of freedom". We can see that $\operatorname{Null} A = \left\langle \begin{pmatrix} -8\\1\\1 \end{pmatrix} \right\rangle$, and as this is a set of one non-zero vector, it is linearly independent. So $\begin{pmatrix} -8\\1\\1 \end{pmatrix}$ is a basis for $\operatorname{Null} A$. So $\dim(\operatorname{Null} A) = 1$.

◇ Let $A = \begin{pmatrix} 1 & 4 & 3 \\ -1 & -4 & -3 \\ 2 & 8 & 6 \end{pmatrix}$. This matrix has reduced row echelon form $\begin{pmatrix} 1 & 4 & 3 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, so

$$\text{Null}\, A = \left\{ s \begin{pmatrix} -4 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right\} = \left\langle \begin{pmatrix} -4 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} \right\rangle.$$

Our intuition is that we have "2 degrees of freedom", so these two vectors should form a basis of Null A. How can we be sure?

Clearly they span Null A, so we just have to show they are linearly independent. The system

$$\lambda_1 \begin{pmatrix} -4 \\ 1 \\ 0 \end{pmatrix} + \lambda_2 \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

has only solution $\lambda_1 = \lambda_2 = 0$, which we can see from the second and third rows.

So these two vectors form a basis for Null A, and $\dim(\text{Null}\, A) = 2$.

It will always happen like this when we have two free variables: if say the second and third variable have "stuff columns", then we can choose them independently of each other. That's why it is so useful to separate the solution out as $s \begin{pmatrix} -4 \\ 1 \\ 0 \end{pmatrix} + t \begin{pmatrix} -3 \\ 0 \\ 1 \end{pmatrix}$ rather than write it all together as $\begin{pmatrix} -4s - 3t \\ s \\ t \end{pmatrix}$.

So we see that

$$\dim(\text{Null}(A)) = \text{ number of stuff columns in reduced REF of } A.$$

We will talk more about bases and dimension of subspaces later in the chapter.
We can also build new vector spaces by pairing given vector spaces together:

Definition 6.7: Given vector spaces V and W, the **cartesian product** (also called **vector space product**) of V and W is the set of pairs of vectors from V and W, with vector addition and scalar multiplication defined entry-wise:

$$V \times W = \{(v, w) \mid v \in V, w \in W\},$$

with $(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2)$ and $\lambda(v, w) = (\lambda v, \lambda w)$.

Examples 6.8:

◇ $\mathbb{R} \times \mathbb{R}$ is the vector space $\mathbb{R}^2$ that we know: whether we write the pairs horizontally as (x, y) or as vertical vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ does not matter that much.

◇ $\mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^{n+m}$. (We will see later what this "=" really means precisely.)

Proposition 6.9: (Dimension of vector space product)
Given vector spaces V and W, their cartesian product $V \times W$ is again a vector space. If V and W are finite-dimensional, then

$$\dim(V \times W) = \dim V + \dim W.$$

Proof. **Exercise:** Prove that $V \times W$ with the given addition and scalar multiplication satisfies the vector space axioms. (You cannot use the subspace ones: it is not contained in a larger vector space, but an entirely new one.)
Let $v_1, \ldots, v_n$ be a basis of V and $w_1, \ldots, w_m$ a basis of W. We show that

$$(v_1, 0), \ldots, (v_n, 0), (0, w_1), \ldots, (0, w_m)$$

is a basis for $V \times W$. (These two sentences are a summary of the proof.)
First we show these vectors span the product: given $(v, w) \in V \times W$, we have $v \in V$, and so $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ for some $\lambda_i \in \mathbb{R}$, because the v_i are a basis of V. Similarly $w \in W$ and $w = \mu_1 w_1 + \cdots + \mu_m w_m$ for some $\mu_i \in \mathbb{R}$. Then

$$\begin{aligned}(v, w) &= (\lambda_1 v_1 + \cdots + \lambda_n v_n, \mu_1 w_1 + \cdots + \mu_m w_m) \\ &= (\lambda_1 v_1, 0) + \cdots + (\lambda_n v_n, 0) + (0, \mu_1 w_1) + \cdots + (0, \mu_m w_m) \\ &= \lambda_1(v_1, 0) + \cdots + \lambda_n(v_n, 0) + \mu_1(0, w_1) + \cdots + \mu_m(0, w_m),\end{aligned}$$

so (v, w) is a linear combination of the given vectors.
Now we show the vectors are linearly independent: consider

$$\lambda_1(v_1, 0) + \cdots + \lambda_n(v_n, 0) + \mu_1(0, w_1) + \cdots + \mu_m(0, w_m) = 0.$$

Then, using the same steps as above but backwards, we get

$$(\lambda_1 v_1 + \cdots + \lambda_n v_n, \mu_1 w_1 + \cdots + \mu_m w_m) = (0, 0),$$

which gives

$$\begin{aligned}\lambda_1 v_1 + \cdots + \lambda_n v_n &= 0 \\ \mu_1 w_1 + \cdots + \mu_m w_m &= 0\end{aligned}$$

separately. As $v_1, \ldots, v_n$ are a basis of V, we have $\lambda_i = 0$ for all i, and similarly $\mu_i = 0$ for all i. So the set

$$(v_1, 0), \ldots, (v_n, 0), (0, w_1), \ldots, (0, w_m)$$

is linearly independent.
So as the set spans $V \times W$ and is linearly independent, it is a basis for $V \times W$. Therefore $\dim(V \times W) = n + m$, the size of the basis, and so

$$\dim(V \times W) = \dim V + \dim W. \qquad \square$$

Remark 6.10: This shows how you can build up the standard basis of $\mathbb{R}^n$ step by step from the basis 1 of $\mathbb{R}$, by repeatedly taking such cartesian products.

B. Plus/Minus Theorem

Next, we consider a result which allows us to enlarge linearly independent sets and reduce spanning sets. The consequences of this result are very important for most of what we continue to do.

Theorem 6.11: (Plus/Minus Theorem)
Let S be a non-empty finite set of vectors in a vector space V.
- *(i) If S is linearly independent and $v \in V$ is not in the span of S, then $S \cup \{v\}$ is still linearly independent.*
- *(ii) If S spans V, and $v \in S$ can be expressed as a linear combination of other vectors in S, then $S \backslash \{v\}$ still spans V.*

Proof. (i) Let $S = \{v_1, v_2, \ldots, v_k\}$. We want to show that $S \cup \{v\}$ is linearly independent. Consider

$$\lambda_1 v_1 + \cdots \lambda_k v_k + \lambda v = 0.$$

If we have $\lambda \neq 0$, then $v = -\frac{1}{\lambda}(\lambda_1 v_1 + \cdots \lambda k v_k)$. But we know v is not in the span of S, so we must have $\lambda = 0$. So with $\lambda = 0$, the linear combination reduces to

$$\lambda_1 v_1 + \cdots \lambda_k v_k = 0.$$

But S is linearly independent, so all $\lambda_i = 0$. So $S \cup \{v\}$ is linearly independent.

Summary: look at v and S separately when showing linear independence.

(ii) We know $S = \{v_1, v_2, \ldots, v_k\}$ spans V, and one of the vectors in S can be expressed as a linear combination of the others. Without loss of generality (meaning we can reorder them if necessary), we have $v_k = \lambda_1 v_1 + \cdots + \lambda_{k-1} v_{k-1}$. We want to show that $\{v_1, \ldots, v_{k-1}\}$ still spans. Given any $w \in S$, as S spans, there are some $\mu_i \in F$ such that

$$w = \mu_1 v_1 + \cdots + \mu_{k-1} v_{k-1} + \mu_k v_k.$$

Then we can substitute for v_k and get

$$\begin{aligned} w &= \mu_1 v_1 + \cdots + \mu_{k-1} v_{k-1} + \mu_k(\lambda_1 v_1 + \cdots + \lambda_{k-1} v_{k-1}) \\ &= (\mu_1 + \mu_k \lambda_1) v_1 + \cdots + (\mu_{k-1} + \mu_k \lambda_{k-1}) v_{k-1} \end{aligned}$$

so w can be written as a linear combitation of $v_1, \ldots, v_{k-1}$. As this is possible for any w, $\{v_1, \ldots, v_{k-1}\} = S \backslash \{v_k\}$ still spans S. □

Summary: replace v_k by its linear combination.

We will need this theorem for some other crucial results, but here is also a smaller example of how it can be used.

Examples 6.12: a) Given the vectors $v_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 2 \\ 3 \\ 1 \end{pmatrix}$, we can see that

- ◇ v_1 and v_2 are linearly independent: they are just two vectors and they are not parallel.
- ◇ v_1 and v_2 have 0 in the third entry, so they are in the x, y-plane.
- ◇ v_3 is not in the span of $\{v_1, v_2\}$: it has a non-zero entry in the third place.

So using the Plus/Minus Theorem (Theorem 6.11), we can see that $\{v_1, v_2, v_3\}$ are linearly independent.

b) We can use the theorem similarly for the polynomials $p_1 = 1 - x^2$, $p_2 = 2 - x^2$, $p_3 = x^3$. Can you explain why p_3 is not in the span of $\{p_1, p_2\}$?

c) If v_1, v_2, v_3, v_4 span a vector space V and $v_3 = v_1 - 4v_2 + 3v_4$, or some other linear combination like this, then we can delete v_3, and v_1, v_2, v_4 still span all of V.

d) If v_1, v_2, v_3, v_4 span a vector space V and $v_2 = 2v_4$, then we can delete either v_2 or v_4 (but not both!), so v_1, v_3, v_4 span V, and also v_1, v_2, v_3 span V.

One of the nice consequences of the Plus/Minus Theorem is that, when a potential basis has the correct size, i.e. the same size as the dimension of the vector space, we can get away with checking only one of the two defining properties of a basis.

Proposition 6.13: (Check one get one free for bases)
Let V be a vector space of dimension n. Then

(i) any linearly independent set of size n automatically spans V, and
(ii) any spanning set of size n is automatically linearly independent.

PROOF. Let $\{v_1, \ldots, v_n\}$ be linearly independent. Suppose it does not span V: then there is some $v \in V$ which is not in the span of $\{v_1, \ldots, v_n\}$. So by the Plus/Minus Theorem (Theorem 6.11), $\{v_1, \ldots, v_n, v\}$ is still linearly independent. But this is a set of more than n vectors, so can't be linearly independent (see Theorem 6.1). So in fact $\{v_1, \ldots, v_n\}$ spans V.

Suppose $\{v_1, \ldots, v_n\}$ spans V but is linearly dependent. Then

$$\lambda_1 v_1 + \ldots + \lambda_n v_n = 0$$

with not all $\lambda_i = 0$. Without loss of generality (meaning we can reorder if necessary), we have $\lambda_n \neq 0$. But then v_n can be written as a linear combitation of $v_1, \ldots, v_{n-1}$, so by the Plus/Minus Theorem (Theorem 6.11), the set $\{v_1, \ldots, v_{n-1}\}$ still spans V. But this set has fewer than n vectors, so it cannot span V (see Theorem 6.1). So in fact $\{v_1, \ldots, v_n\}$ is linearly independent. □

So when we know (through some other way) that V has dimension n, and we have a set of n vectors that might be a basis, it is enough to check *one of* linearly independent and spanning, we don't have to do both.

Examples 6.14: (a) We can see without any calculation that

$$v_1 = \begin{pmatrix} -3 \\ 7 \end{pmatrix}, \ v_2 = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$$

form a basis of $\mathbb{R}^2$. These two vectors are clearly not parallel, so they are linearly independent, and $\mathbb{R}^2$ has dimension 2, so any linearly independent set of 2 vectors forms a basis.

(b) We can see without any calculation that

$$v_1 = \begin{pmatrix} 2 \\ 0 \\ -1 \end{pmatrix}, \ v_2 = \begin{pmatrix} 4 \\ 0 \\ 7 \end{pmatrix}, \ v_3 = \begin{pmatrix} -1 \\ 1 \\ 4 \end{pmatrix}$$

form a basis of $\mathbb{R}^3$. The first two vectors lie in the x, z-plane (can you explain why?) and are clearly not parallel. The last vector v_3 is not in the x, z-plane, so it is not in the span of $\{v_1, v_2\}$. So by the Plus/Minus Theorem (Theorem 6.11), the set $\{v_1, v_2, v_3\}$ is linearly independent. So as $\mathbb{R}^3$ has dimension 3, it is automatically a basis.

(c) Let P_n be the space of polynomials of degree at most n. We know that $1, x, x^2, \ldots, x^n$ is a basis of P_n, see Example 5.9 c. So $\dim P_n = n + 1$. So to check that

$$p_0 = 1, \ p_1 = x + 1, \ p_2 = x^2 + x + 1, \ \ldots, \ p_n = x^n + x^{n-1} + \cdots + x + 1$$

is a basis of P_n, it is enough to check *one of* linearly independent and spanning.

Exercise Check that $p_0, \ldots, p_n$ is a basis of P_n. (This is 6.13c) in the workbook.)

(d) (This is 6.12c) from the workbook.) Let $e_1, e_2, \ldots, e_n$ be the standard basis of $\mathbb{R}^n$, and consider

$$v_1 = e_1 + e_2, \ v_2 = e_2 + e_3, \ \ldots, \ v_k = e_k + e_{k+1}, \ \ldots, \ v_n = e_n + e_1.$$

We will check in lectures whether these form a basis of $\mathbb{R}^n$ or not. The answer will depend on whether n is odd or even. When n is even we will show that there is some dependence relation. When n is odd, we will show they form a basis by showing that they span. The easiest way to show that these particular vectors span is to give each standard basis vector as a linear combination of one of these. As soon as we can get all the standard basis vectors, we can also get any other vector.

The next result is going to be used a lot throughout the course.

Theorem 6.15: (Extend to a basis)
Let V be a finite-dimensional vector space. Then

(i) any linearly independent set in V can be extended to a basis of V, and
(ii) any spanning set of V contains a basis of V.

Proof. (i) Suppose $\{v_1, \ldots, v_k\}$ is a linearly independent set in V with $k < n = \dim V$. We want to show that we can add vectors to this set to get a basis of V.

As a set of fewer than n vectors cannot span, there is some $v_{k+1} \in V$ which is not in the span of $\{v_1, \ldots, v_k\}$. So by the Plus/Minus Theorem (Theorem 6.11), $\{v_1, \ldots, v_k, v_{k+1}\}$ is still linearly independent. If $k + 1 = n$, we have a linearly independent set of n vectors, so it is a basis (by Proposition 6.13). If $k + 1 < n$, then $\{v_1, \ldots, v_{k+1}\}$ still cannot span V, so there is some v_{k+2} not in the span, so $\{v_1, \ldots, v_{k+1}, v_{k+2}\}$ is still linearly independent. We continue this way until we have a set of size n which is still linearly independent, and so a basis of V.

(ii) Let $\{v_1, \ldots, v_m\}$ be a spanning set, with $m > n$. Then the set cannot be linearly independent, so we have

$$\lambda_1 v_1 + \ldots + \lambda_m v_m = 0$$

with not all $\lambda_i = 0$; wlog (without loss of generality) $\lambda_m \neq 0$. So v_m can be written as a linear combination of the other vectors, so by the Plus/Minus Theorem (Theorem 6.11), $\{v_1, \ldots, v_{m-1}\}$ still spans V. If $m-1 = n$, this set is automatically a basis (by Proposition 6.13: it has the right size). If $m-1 > n$, we can repeat the previous step and remove another vector from the set to still have a spanning set $\{v_1, \ldots, v_{m-2}\}$. We continue like this until we are left with a spanning set of n vectors, which is automatically a basis. □

Examples 6.16: (a) Given the linearly independent vectors

$$v_1 = \begin{pmatrix} 0 \\ 3 \\ 1 \end{pmatrix}, \; v_2 = \begin{pmatrix} 0 \\ 2 \\ 6 \end{pmatrix},$$

how can we extend them to a basis of $\mathbb{R}^3$? We can see that v_1, v_2 cannot span any vector which has a non-zero entry in the first coordinate. So adding

$$v_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

extends v_1, v_2 to a basis v_1, v_2, v_3 of $\mathbb{R}^3$. This added vector is not unique! We can choose any vector that is not in the span of the first two, for example $v_3 = \left(\begin{smallmatrix} 4 \\ 3 \\ -5 \end{smallmatrix}\right)$, or anything else that does not have a 0 in the first entry. It is however useful to keep it simple and look, if possible, for standard basis vectors.

(b) Let's try to find which standard basis vectors we can add to the two vectors below to extend to a basis of $\mathbb{R}^4$.

$$v_1 = \begin{pmatrix} 1 \\ 3 \\ 6 \\ -3 \end{pmatrix}, \; v_2 = \begin{pmatrix} -1 \\ -2 \\ -4 \\ 2 \end{pmatrix}.$$

The vectors are not as nice as in the previous example, but we can use column operations to turn them into vectors which span the same subspace but are nicer.

Performing elementary column operations does not change the column space of a matrix.

$$\begin{pmatrix} 1 & -1 \\ 3 & -2 \\ 6 & -4 \\ -3 & 2 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 3 & 1 \\ 6 & 2 \\ -3 & -1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 2 \\ 0 & -1 \end{pmatrix}$$

So we see that e_1 is in the span of those two vectors, so we definitely don't want to add that one. We know we need to add two more vectors to get a basis for $\mathbb{R}^4$. We can see that neither e_2, e_3 or e_4 are in the span of our two vectors (because the column operations did not change the span!!!!). So by the Plus/Minus Theorem we can add any one of them; let's say we add e_2. We can now repeat the "make it easier through column operations" step:

$$\begin{pmatrix} 1 & -1 & 0 \\ 3 & -2 & 1 \\ 6 & -4 & 0 \\ -3 & 2 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 1 \\ 6 & 2 & 0 \\ -3 & -1 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & -1 & 0 \end{pmatrix}$$

From this we see that neither e_3 nor e_4 are in the span of those three vectors, so we can add either of those to get a bigger linearly independent set.

In summary: we have found out by using column operations and the Plus/Minus Theorem that

$$v_1, v_2, e_2, e_3$$

form a basis of $\mathbb{R}^4$.

(c) Not lectured: an extra example for you to work through to help you understand. In the space P_3 of polynomials of degree at most 3, consider

$$p_1 = x - 1,\ p_2 = x + 3.$$

We can see that this is a linearly independent set. How can we extend it to a basis of the 4-dimensional space P_3? We can see that p_1, p_2 can't combine to give a polynomial of degree 2, so we can add $p_3 = x^2$. Then by the Plus/Minus Theorem, p_1, p_2, p_3 are still linearly independent. We now still can't make polynomials of degree 3, so we add $p_4 = x^3$, and then p_1, p_2, p_3, p_4 is a basis of P_3.

Again these choices are not unique. We can use any polynomial with some x^2 term as the second one, and any polynomial with some x^3 term for the third one. Or we could even use $q_3 = x^2 + x^3$ for the first one and $q_4 = x^2 - x^3$ for the third one: $x^2 + x^3$ is not in the span of the first two, and then we have to find another vector that is not in the span of the first three. Using the first three, we can only get a polynomial which has the same coefficient in front of x^2 and x^3, so the q_4 I suggested is not in the span of p_1, p_2, q_3.

Exercise 6.17: Using the result "equal spans" (Prop. 4.27), explain why performing elementary column operations on a matrix does not change the column space (= span of the column vectors).

Let's write down the method we have seen in these examples:

Extending to a basis

Given some linearly independent set of vectors $\{v_1, \dots, v_k\}$ in a vector space V of dimension n, we want to extend this set to a basis of V.

- If $V = \mathbb{R}^n$, we already have column vectors $v_1, \dots, v_k$, and we will want to add some of the standard basis vectors $e_1, \dots, e_n$.
- If V is not $\mathbb{R}^n$, then pick a basis B of V (usually some nice one, like standard basis of polynomials, etc) and write each of the v_i as a coordinate vector: so now we can treat them as vectors in $\mathbb{R}^n$ after all.
- Write all the vectors $v_1, \dots, v_k$ next to each other as the columns of a matrix A.
- Perform **elementary column operations** on the matrix A, to get the columns into as easy a form as possible. (The column version of echelon form, or reduced echelon form.)
- You should now be able to see easily which of the standard basis vectors $e_1, \dots, e_n$ are already in the span of $v_1, \dots, v_k$. So pick some e_i which is *not* in the span, and add it. (By Plus/Minus Theorem, these $k + 1$ vectors are still linearly independent.)
- Add that vector to the end of the reduced version of matrix A, and reduce this bigger matrix again (if possible). If possible, add another standard basis vector which is not in the span of the given vectors.
- Continue like this until you have n vectors. (By Plus/Minus Theorem, these are still linearly independent, and so by chogof for bases, they form a basis of V.)

Reducing to a basis or **Finding basis for column space**

Given some spanning set $\{v_1, \dots, v_l\}$ in a vector space V of dimension n, we want to reduce it to a basis of V.

(Or given a matrix A, we want to find a basis for the column space of A. For this start at the fourth point.)

- If $V = \mathbb{R}^n$, we already have column vectors $v_1, \dots, v_l$.
- If V is not $\mathbb{R}^n$, then pick a basis B of V (usually some nice one, like standard basis of polynomials, etc) and write each of the v_i as a coordinate vector: so now we can treat them as vectors in $\mathbb{R}^n$ after all.
- Put all the vectors next to each other to form the columns of a matrix A.
- Perform **elementary row operations** on A (to solve the system $Ax = 0$).
- Delete all columns of A that turn into stuff columns at the end.

◇ The remaining columns are your basis. (Either of V, or of the column space of A, depending on what you started with).

Exercise 6.18: In the calculations of getting A into reduced row echelon form, in every step delete all the columns which at the end turn into stuff columns. Explain why the remaining original columns of A are linearly independent. (C.f. Workbook Question 6.9)

This "reducing to a basis" also shows us the following:

Corollary 6.19: *Every finite-dimensional vector space has a (finite) basis.*

Proof. Let V be finite-dimensional, so it has a finite spanning set. Then this spanning set contains a basis (by Theorem 6.15ii). Clearly this basis is a finite set. $\square$

This result also allows us to look at subspaces more carefully.

C. Bases and dimension of subspaces

Proposition 6.20: (Dimensions of subspaces)
If W is a subspace of a finite-dimensional vector space V, then:

(i) W *is finite-dimensional.*
(ii) $\dim W \leqslant \dim V$.
(iii) $W = V$ *if and only if* $\dim W = \dim V$.

Proof. (i) As V is finite-dimensional, it has some finite basis. Let's say the basis has n elements, so $\dim V = n$.

Now consider W. We want to show that it has a finite spanning set. If $W = 0$, then it is finite-dimensional as $\{0\}$ is a spanning set. So assume $W \neq 0$. Take some $w_1 \neq 0 \in W$. If $\langle w_1 \rangle = W$ (i.e. w_1 spans all of W), then we have found a finite spanning set, so we are done.

If $\langle w_1 \rangle \neq W$, then there is some $w_2 \in W$ which is not in this span: $w_2 \notin \langle w_1 \rangle$. Then by the Plus/Minus Theorem (Theorem 6.11), $\{w_1, w_2\}$ is linearly independent. If $\langle w_1, w_2 \rangle = W$, we are done. Otherwise we can find $w_3 \in W$, $w_3 \notin \langle w_1, w_2 \rangle$, and so on.

As the set we build always stays linearly independent (by the Plus/Minus Theorem), it cannot get bigger than n elements, because any set of more than n vectors in V must be linearly dependent (Prop. 6.1). So this process must stop, giving a finite spanning set for W.

(ii) **Exercise.** Note that a basis of W is a linearly independent set in V.

(iii) If $W = V$, then clearly $\dim W = \dim V$.

Conversely, suppose $\dim W = \dim V = n$ and $W \leqslant V$. Suppose there is some $v \in V$, $v \notin W$. Let $w_1, w_2, \ldots, w_n$ be a basis for W (which exists, because $W \leqslant V$ and V finite-dimensional, so W is finite-dimensional, so it has a finite basis). If $v \notin W$, then by Plus/Minus Theorem (Theorem 6.11), $w_1, w_2, \ldots, w_n, v$ is linearly independent in V. But this is a set of $n+1$ linearly independent vectors in a space of dimension n, which is not possible (Prop. 6.1). So in fact, there is no $v \in V$ which is not in W, so $V = W$.

$\square$

Careful! If W is not a subspace of V, then it is not true that $\dim W = \dim V$ implies $W = V$. For example, the space P_2 of polynomials of degree at most 2 has dimension 3, but it is not the same vector space as $\mathbb{R}^3$.

Now while we have $\dim W \leqslant \dim V$, so we have a simple relation for the dimensions, we do have to be a lot more careful about *bases* of subspaces.

Fact 6.21: (Bases of subspaces)
If $v_1, v_2, \ldots, v_n$ is a basis B of V, and $W \leqslant V$ is a subspace of V, there is no reason to assume that a subset of the basis B can be found to give a basis of W. For example:

- ◇ Take the standard basis $E = \left\{ e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ of $\mathbb{R}^2$, and let $W = \left\langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right\rangle$ be the subspace which is the line spanned by $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Then no subset of E which gives a basis of W: any basis of W is a single vector of the form $\begin{pmatrix} x \\ x \end{pmatrix}$. (For example $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ or $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$ or $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$.)
- ◇ Let W be the space of symmetric 2×2 matrices, which is a subspace of $V = \mathscr{M}_{2\times 2}$. The standard basis of $\mathscr{M}_{2\times 2}$ is

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

  We saw above that W has a basis of size 3, so has dimension 3. But there are only two symmmetric matrices in this standard basis of $\mathscr{M}_{2\times 2}$, so there is no subset of 3 vectors which could give us a basis for W.

How do we get round this?

When working with subspaces $W \leqslant V$ and bases, we *always* start with a basis B_W of the subspace. This is a linearly independent set in V, so we can extend it to a basis B_V of V. This way we get a basis of V which contains a basis of the subspace.
But we always have to start with a basis for the smaller space for this to work!

We know that given two subspaces $U, W \leqslant V$, their intersection $U \cap W = \{v \in V \mid v \in U \text{ and } v \in W\}$ and their sum $U + W = \{u + w \mid u \in U, w \in W\}$ are again vector spaces (Propositions 4.16 and 4.18). Using this trick of extending a basis for the smaller space, we can now prove a relationship of their dimensions.

Proposition 6.22: (Dimension of sum of subspaces)
Let V be a finite-dimensional vector space, and $U, W \leqslant V$ two subspaces. Then

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

Proof. Suppose $\dim U = k$ and $\dim W = l$, and $\dim U \cap W = m$. So $m \leqslant k$ and $m \leqslant l$, as $U \cap W$ is a subspace of both U and W (see Theorem 6.20). Let $u_1, \ldots, u_m$ be a basis $B_{U\cap W}$ of $U \cap W$ (we start with a basis of the very smallest space), and extend it to a basis $B_U = \{u_1, \ldots, u_m, u_{m+1}, \ldots, u_k\}$ of U, and to a basis $B_W = \{u_1, \ldots, u_m, w_{m+1}, \ldots, w_l\}$ of W. If $U \cap W$ happens to be the zero vector space, then we don't have any basis for $U \cap W$ and just take bases for U and W separately. The rest of the proof still makes sense for that case, i.e. when $m = 0$.
We now show that $B_U \cup B_W = \{u_1, \ldots, u_m, u_{m+1}, \ldots, u_k, w_{m+1}, \ldots, w_k\}$ forms a basis for $U + W$.
These vectors span $U + W$:
given $u + w \in U + W$ with $u \in U, w \in W$, then $u = \lambda_1 u_1 + \cdots + \lambda_m u_m + \lambda_{m+1} u_{m+1} + \cdots \lambda_k u_k$ for some $\lambda_i \in \mathbb{R}$, because B_U is a basis of U, and similarly $w = \mu_1 u_1 + \cdots + \mu_m u_m + \mu_{m+1} w_{m+1} + \cdots + \mu_l w_l$ for some $\mu_i \in \mathbb{R}$. So

$$u + w = (\lambda_1 + \mu_1) u_1 + \cdots + (\lambda_m + \mu_m) u_m + \lambda_{m+1} u_{m+1} + \cdots + \lambda_k u_k + \mu_{m+1} w_{m+1} + \cdots + \mu_l w_l.$$

To show the vectors are linearly indepdent, consider

$$\lambda_1 u_1 + \cdots + \lambda_m u_m + \lambda_{m+1} u_{m+1} + \cdots + \lambda_k u_k + \mu_{m+1} w_{m+1} + \cdots + \mu_l w_l = 0.$$

We can rearrange this to get

$$\lambda_1 u_1 + \cdots + \lambda_m u_m + \lambda_{m+1} u_{m+1} + \cdots + \lambda_k u_k = -\mu_{m+1} w_{m+1} - \cdots - \mu_l w_l.$$

Let's call this vector v. The LHS expression shows that $v \in U$, and the RHS shows that $v \in W$. So $v \in U \cap W$. So, as $B_{U\cap W}$ is a basis of $U \cap W$, there are some $\nu_i \in \mathbb{R}$ such that $v = \nu_1 u_1 + \cdots + \nu_m u_m$.

Then, using the RHS expression for v, we get

$$\begin{aligned} & nu_1u_1 + \cdots + \nu_m u_m = -\mu_{m+1}w_{m+1} - \cdots - \mu_l w_l \\ \Leftrightarrow \quad & nu_1u_1 + \cdots + \nu_m u_m + \mu_{m+1}w_{m+1} + \cdots + \mu_l w_l = 0. \end{aligned}$$

But those vectors are the basis B_W, so they are linearly independent, so $\nu_i = 0$ for all i and $\mu_j = 0$ for all j.
This implies $v = 0$, so then using the LHS expresion for v, we get

$$\lambda_1 u_1 + \cdots + \lambda_m u_m + \lambda_{m+1}u_{m+1} + \cdots + \lambda_k u_k = 0.$$

Those vectors are the basis B_U, so they are linearly independent, so $\lambda_i = 0$ for all i.
So we have shown that $\{u_1, \ldots, u_m, u_{m+1}, \ldots, u_k, w_{m+1}, \ldots, w_k\}$ is a linearly independent set.
So as it is linearly independent and spans $U + W$, it is a basis of $U + W$.
Therefore $\dim(U + W) = k + (l - m)$, the size of the basis. This gives

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W)$$

as required. □

So this means if we know the dimensions of U, W and $U \cap W$, we know the dimension of $U + W$.
The proof also gives us a way to find nice bases for these spaces that work well together.
Sometimes it is easy to see what $U \cap W$ is.

Examples 6.23: ⋄ $V = \mathbb{R}^3$, $U = \left\langle \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix} \right\rangle$, the x, y-plane, $W = \left\langle \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix} \right\rangle$.

So $U \cap W = \left\langle \begin{pmatrix}1\\0\\0\end{pmatrix} \right\rangle$. So $\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}1\\0\\1\end{pmatrix}$ is a basis for $U + W = \mathbb{R}^3$.

⋄ $V = \mathbb{R}^4$, $U = \left\langle \begin{pmatrix}1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\\0\end{pmatrix} \right\rangle$, $W = \left\langle \begin{pmatrix}1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\\0\end{pmatrix} \right\rangle$.

$U \cap W$ has basis $\begin{pmatrix}1\\1\\1\\0\end{pmatrix}$. So $U + W$ has basis $\begin{pmatrix}1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}, \begin{pmatrix}1\\1\\0\\0\end{pmatrix}$.
$\dim(U + W) = \dim(U) + \dim(W) - \dim(U \cap W) = 2 + 2 - 1 = 3$.
So $U + W$ is not all of $\mathbb{R}^4$. (We can see that the last entry is always 0.)

There is another situation when it is reasonably easy to work out what the intersection is:

Examples 6.24: Let $V = \mathbb{R}^3$, and

$$U = \left\{ \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \middle| \; x_1 + x_2 + x_3 = 0 \right\}, \quad W = \left\{ \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \middle| \; x_1 + 2x_2 + x_3 = 0 \right\}.$$

Here both U and W are given as a solution set to a linear system. So the vectors in U are all vectors which satisfy the first equation, and the vectors in W are all vectors which satisfy the second equation. So it is easy to see that the vectors in $U \cap W$ are all vectors which satisfy both equations. So we simply solve the linear system

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ x_1 + 2x_2 + x_3 &= 0 \end{aligned}$$

to find $U \cap W$. We get

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

so the set of solutions is $U \cap W = \left\{ t \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} \right\}$. And so $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$ is a basis of $U \cap W$.

We can use the same technique when the subspaces are given as nullspaces of some matrix A, as this can also be viewed as solutions to a linear system $Ax = 0$.

Exercise 6.25: Write down bases of U and of W that contain this basis of $U \cap W$, and write down a basis of $U + W$.

But sometimes it is not so easy to see what the intersection is.

Example 6.26: Find a basis for the intersection of

$$U = \left\langle \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix} \right\rangle, \quad W = \left\langle \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix} \right\rangle.$$

Let $v \in U \cap W$. This gives us

$$v = \begin{pmatrix} \lambda_1 \\ \lambda_1 \\ 0 \end{pmatrix} + \begin{pmatrix} \lambda_2 \\ 0 \\ 2\lambda_2 \end{pmatrix} = \begin{pmatrix} 2\mu_1 \\ \mu_1 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 2\mu_2 \end{pmatrix}.$$

The second row gives $\lambda_1 = \mu_1$, and the third row gives $\lambda_2 = \mu_2$. So now we have

$$v = \begin{pmatrix} \lambda_1 + \lambda_2 \\ \lambda_1 \\ 2\lambda_2 \end{pmatrix} = \begin{pmatrix} 2\lambda_1 \\ \lambda_1 \\ 2\lambda_2 \end{pmatrix}.$$

Then the first row gives $\lambda_1 = \lambda_2$. So

$$v = \begin{pmatrix} 2\lambda_1 \\ \lambda_1 \\ 2\lambda_1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$$

So choosing $\lambda_1 = 1$ gives us the vector $\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$ which forms a basis for $U \cap W$.

So we can note down:

Finding a basis of an intersection

If U and W are given as solution set of linear systems (including a nullspace of a matrix):

- ◇ Put all equations of both linear systems together and solve the bigger linear system.
- ◇ The solutions of this bigger linear system is the intersection: find a basis like you do when giving a basis of a nullspace.

If U and W are given as spans, or with bases:

Let U have basis $u_1, u_2, \ldots, u_k$ and W have basis $w_1, w_2, \ldots, w_l$. Then

- ◇ Take $v \in U \cap W$, so $v = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = \mu_1 w_1 + \cdots + \mu_l w_l$.
- ◇ Solve the system $\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k - \mu_1 w_1 - \cdots - \mu_l w_l = 0$.
- ◇ Use the solution to this system to write down v, and find a basis.

Example 6.27: Let $V = \mathbb{R}^4$, $U = \left\langle u_1 = \begin{pmatrix}1\\1\\1\\0\end{pmatrix}, u_2 = \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}\right\rangle$, $W = \left\langle w_1 = \begin{pmatrix}3\\1\\1\\0\end{pmatrix}, w_2 = \begin{pmatrix}1\\1\\0\\0\end{pmatrix}\right\rangle$.

Let $v \in U \cap W$, so

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

as $v \in U$, and also

$$v = \mu_1 \begin{pmatrix}3\\1\\1\\0\end{pmatrix} + \mu_2 \begin{pmatrix}1\\1\\0\\0\end{pmatrix}$$

because $v \in W$. So

$$v = \begin{pmatrix}\lambda_1+\lambda_2\\ \lambda_1-\lambda_2\\ \lambda_1\\ 0\end{pmatrix} = \begin{pmatrix}3\mu_1+\mu_2\\ \mu_1+\mu_2\\ \mu_1\\ 0\end{pmatrix}.$$

or

$$\begin{pmatrix}\lambda_1+\lambda_2\\ \lambda_1-\lambda_2\\ \lambda_1\\ 0\end{pmatrix} - \begin{pmatrix}3\mu_1+\mu_2\\ \mu_1+\mu_2\\ \mu_1\\ 0\end{pmatrix} = 0.$$

This gives us a system with 4 equations and 4 variables, which we can solve.

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

$$\longrightarrow \begin{pmatrix}1&1&0&2\\0&1&0&1\\0&0&1&1\\0&0&0&0\end{pmatrix} \longrightarrow \begin{pmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\\0&0&0&0\end{pmatrix}$$

So if we say $\mu_2 = t$, then $\mu_1 = -t$, $\lambda_1 = \lambda_2 = -t$. So our v is

$$v = \begin{pmatrix}-t-t\\ -t-(-t)\\ -t\\ 0\end{pmatrix} = -t\begin{pmatrix}2\\0\\1\\0\end{pmatrix}.$$

(If it happens to be a reasonably simple system where you can just find a solution directly from looking at the two expressions for v, that's ok.)

So $v_0 = \begin{pmatrix}2\\0\\1\\0\end{pmatrix}$ is a basis for $U \cap W$.

Now we have to extend this to a basis of U and a basis of W. We can use the Plus/Minus Theorem: we know that $u_1 \in U$ and u_1 is not in the span of v_0, so $\{v_0, u_1\}$ is linearly independent, so it is a basis of U (since $\dim U = 2$). Similarly $\{v_0, w_1\}$ is a basis of W.

(Note we could also choose $\{v_0, u_2\}$ for U and/or $\{v_0, w_2\}$ for W. We've already seen that "extending a basis" is not unique.)
So $\{v_0, u_1, w_1\}$ is a basis for $U + W$, which has dimension 3.

We know how to go from a system of equations to a basis of the solution set (or from a matrix to a basis of its nullspace), but is there an easy "algorithmic" way to go back the other way?

Proposition 6.28: (Subspace as solution set) *Every subspace of $\mathbb{R}^n$ is the solution space of some homogeneous linear system (i.e. the nullspace of some matrix).*

Proof. Let $U \leqslant \mathbb{R}^n$ have a basis $u_1, \ldots, u_k$, i.e. $\dim(U) = k$. Form a matrix A with the vectors u_i as rows:

$$A = \begin{pmatrix} \leftarrow & u_1^T & \rightarrow \\ \leftarrow & u_2^T & \rightarrow \\ & \vdots & \\ \leftarrow & u_k^T & \rightarrow \end{pmatrix}$$

As the u_i are a basis of U, they are linearly independent. This means that the reduced row echelon form of A has no zero rows, since a zero row occurs when one row is some linear combination of other rows. So the reduced row echelon form of A has k leading 1s, so $n-k$ stuff columns.
Therefore its nullspace $W = \{x \in \mathbb{R}^n \mid Ax = 0\}$ has dimension $n-k$. Let $w_1, \ldots, w_{n-k}$ be a basis of W, and make a matrix B with the w_j as rows:

$$B = \begin{pmatrix} \leftarrow & w_1^T & \rightarrow \\ \leftarrow & w_2^T & \rightarrow \\ & \vdots & \\ \leftarrow & w_{n-k}^T & \rightarrow \end{pmatrix}$$

By the same argument, the reduced row echelon form of B has $n-k$ leading 1s, and k stuff columns. So the nullspace $\{y \in \mathbb{R}^n \mid By = 0\}$ has dimension k. We will show that this nullspace of B is in fact U, the subspace we started with.
We first show that $U \subseteq \{y \in \mathbb{R}^n \mid By = 0\}$. Let $u \in U$. Then as the u_i form a basis of U, there are λ_i such that $u = \lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_k u_k$. So if $Bu_i = 0$ for all i, then $Bu = B(\lambda_1 u_1 + \cdots + \lambda_k u_k) = \lambda_1 Bu_1 + \cdots + \lambda_k Bu_k = 0$ as well. So we only have to show that $Bu_i = 0$ for any i.
Now Bu_i is a vector with the jth entry $w_j^T u_i$, i.e. row j of B times u_i. But we know that $u_i^T w_j = 0$, because w_j is in W, the nullspace of A, which has rows u_i^T. But then also $(u_i^T w_j)^T = 0$, i.e. $w_j^T u_i = 0$, so $Bu_i = 0$. Written more compactly:

$$Bu_i = \begin{pmatrix} w_1^T u_i \\ w_2^T u_i \\ \vdots \\ w_{n-k}^T u_i \end{pmatrix} = \begin{pmatrix} (u_i^T w_1)^T \\ (u_i^T w_2)^T \\ \vdots \\ (u_i^T w_{n-k})^T \end{pmatrix} = \begin{pmatrix} 0^T \\ 0^T \\ \vdots \\ 0^T \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

So $u_i \in \{y \in \mathbb{R}^n \mid By = 0\}$.
So we know that $U \subseteq \{y \in \mathbb{R}^n \mid By = 0\}$. But we also know that $\dim(U) = k = \dim(\{y \in \mathbb{R}^n \mid By = 0\})$.
So by Theorem 6.20(iii), we get $U = \{y \in \mathbb{R}^n \mid By = 0\}$, as required.
So U is the set of solutions of the homogeneous linear system $By = 0$ (i.e. the nullspace of B). $\square$

So if you want to, you can use this technique to turn a "subspace given as a span/with a basis" into a "subspace given as a solution to a linear system", and then use the first method for finding a basis of an intersection.

Exercise 6.29: Let $V = \mathbb{R}^4$, $U = \left\langle u_1 = \begin{pmatrix}1\\1\\1\\0\end{pmatrix}, u_2 = \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}\right\rangle$, $W = \left\langle w_1 = \begin{pmatrix}3\\1\\1\\0\end{pmatrix}, w_2 = \begin{pmatrix}1\\1\\0\\0\end{pmatrix}\right\rangle$, as in the previous example.
Use the technique from Prop. 6.28 to write U and W as solution spaces of linear systems, and then use the "linear systems" method to find a basis for $U \cap W$. Check that you get the same answer as the example above.

Recall that If $U, W \leqslant V$ are two subspaces with intersection $U \cap W = 0$ being just the zero vector, then we call $U + W$ a **direct sum** and write $U \oplus W$.

Corollary 6.30: (Dimension of direct sum)

$$\dim(U \oplus W) = \dim U + \dim W.$$

PROOF. Recall that $\dim 0 = 0$. ☐

Examples 6.31:

- $V = \mathbb{R}^3$, $U = x$-y-plane $= \left\langle \begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}\right\rangle$, $W = z$-axis $= \left\langle \begin{pmatrix}0\\0\\1\end{pmatrix}\right\rangle$.
  Then $U \cap W = 0$, so $U \oplus W$ is a direct sum. Here we have $U \oplus W = \mathbb{R}^3$.
- $V = \mathbb{R}^4$, $U = \left\langle \begin{pmatrix}1\\1\\1\\0\end{pmatrix}, \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}\right\rangle$, $W = \left\langle \begin{pmatrix}1\\1\\0\\0\end{pmatrix}\right\rangle$. We can check that $U \cap W = 0$, so $U \oplus W$ is a direct sum. But $U \oplus W \neq \mathbb{R}^4$: $\dim(U \oplus W) = \dim(U) + \dim(W) - \dim(U \cap W) = 2 + 1 - 0 = 3$.
  How do we show that the intersection is just 0: take a vector $v \in U \cap W$. Then $v \in U$, so
  $$v = \lambda_1 \begin{pmatrix}1\\1\\1\\0\end{pmatrix} + \lambda_2 \begin{pmatrix}1\\-1\\0\\0\end{pmatrix}.$$
  But $v \in W$, so also
  $$v = \mu \begin{pmatrix}1\\1\\0\\0\end{pmatrix}.$$
  So we have
  $$\begin{pmatrix}\lambda_1 + \lambda_2\\ \lambda_1 - \lambda_2\\ \lambda_1\\ 0\end{pmatrix} = \begin{pmatrix}\mu\\ \mu\\ 0\\ 0\end{pmatrix}$$
  Looking at the third row, we get $\lambda_1 = 0$. Then the first two rows give $\mu = \lambda_2 = -\lambda_2$, so $\mu = \lambda_2 = 0$. So $v = 0$.

D. Bases and Dimension: Study guide

Concept review.

- Dimension of a vector space.

- ◇ Vector space product (or cartesian product), dimension of vector space product.
- ◇ Possible sizes of linear independent sets, spanning sets, bases, in a given vector space.
- ◇ Relationships among the concepts of linear independence, spanning set, basis, and dimension.
- ◇ Plus/Minus Theorem.
- ◇ Check one get one free for bases.
- ◇ Extending to a basis, reducing to a basis.
- ◇ Effect of elementary column operations on span of columns.
- ◇ Dimensions of subspaces.
- ◇ Dimensions of sums of subspaces, dimensions of direct sum of subspaces.

Skills.

- ◇ Find dimension of a vector space (or subspace).
- ◇ Use dimension to determine whether a set of vectors is a basis for a finite-dimensional vector space.
- ◇ Extend a linearly independent set to a basis.
- ◇ Reduce a spanning set to a basis.
- ◇ Find a basis for the column space of a matrix.
- ◇ Find the basis for the intersection of two subspaces.
- ◇ Find the dimension of a sum of subspaces.

CHAPTER 7

Detour: Complex Numbers

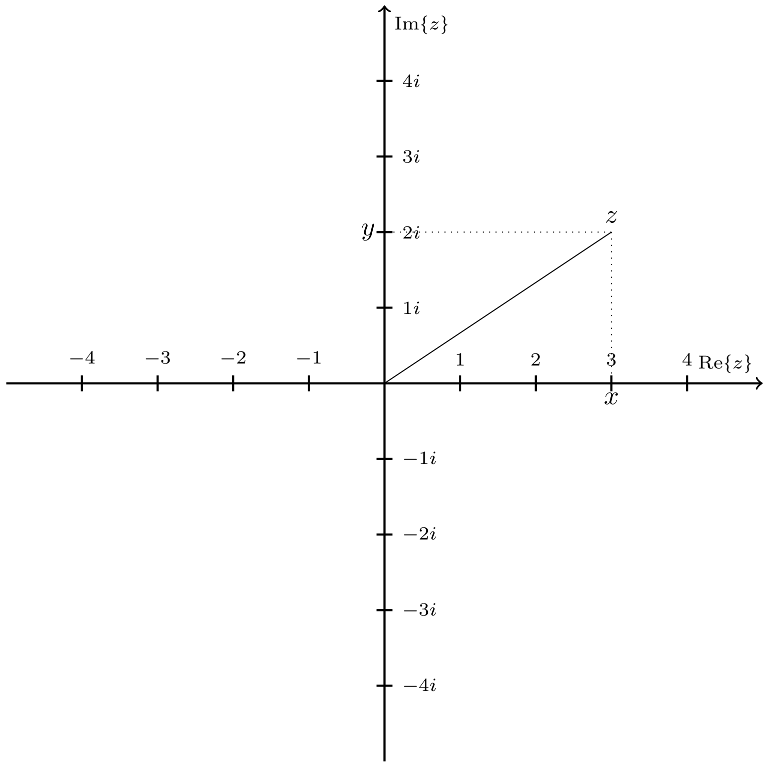
A. Complex Numbers

Some of you will know complex numbers from school. We will go through the main points.

The **complex numbers** $\mathbb{C}$ are numbers $z = x + iy$ with $x, y \in \mathbb{R}$. Here $i^2 = -1$, giving, for $z_k = x_k + iy_k$,

$$\text{addition:} \qquad z_1 + z_2 = x_1 + x_2 + i(y_1 + y_2)$$
$$\text{and multiplication:} \qquad z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = x_1x_2 - y_1y_2 + i(y_1x_2 + x_1y_2).$$

Given $z = x + iy$, we call $\mathrm{Re}(z) = x$ the **real part** and $\mathrm{Im}(z) = y$ the **imaginary part** of z. $|z| = \sqrt{x^2 + y^2}$ is the **modulus** of z, and $\overline{z} = x - iy$ is the **complex conjugate** of z.
We have $z\overline{z} = x^2 + y^2 = |z|^2$.
We can depict complex numbers in an Argand diagram:



We will draw more into the argand diagram in lectures.
The angle θ which the line to z makes with the positive x-axis is called the **argument** of z. We have

$$\mathrm{Re}(z) = |z| \cos\theta$$
$$\mathrm{Im}(z) = |z| \sin\theta.$$

We can write z in **polar coordinates** as $z = |z|(\cos\theta + i\sin\theta) = |z|e^{i\theta}$.

Complex numbers have the very useful property that we can factorise any polynomial into linear factors. For example

- $x^2 - 1 = (x+1)(x-1)$ can be factorised in real numbers, but
- x^2+1 has no real roots. But $x^2+1 = (x+i)(x-i)$ can be factorised in complex numbers, so it has complex roots.

This will become important, particularly when we study eigenvalues.

Exercise 7.1: Show that a complex number z is real if and only if $z = \overline{z}$.

If you are not that familiar with complex numbers and want to practice a bit, I recommend `https://nrich.maths.org/1403`, the section "The Basics of Complex Numbers" has a few exercises you can do to get some practice.

B. Complex vector spaces

We saw the definition of a real vector space last semester; the same definition gives complex vector spaces if we take the scalars to be complex numbers instead. When we want to talk about real and complex vector spaces at the same time, we often write F, where F can be $\mathbb{R}$ or $\mathbb{C}$. Later on you might meet vector spaces over other fields such as $\mathbb{Z}_p$, the integers modulo a prime, but we won't do that in this course.

Examples 7.2:
- $\mathbb{C}^n$ is a complex vector space for any natural number n.
  e.g. $\mathbb{C}^3$ has vectors $\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix}$ with entries $z_j \in \mathbb{C}$, addition is entry-wise
  $$\begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} + \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} z_1+w_1 \\ z_2+w_2 \\ z_3+w_3 \end{pmatrix}$$
  and for a complex scalar $\lambda \in \mathbb{C}$, we have scalar multiplication
  $$\lambda \begin{pmatrix} z_1 \\ z_2 \\ z_3 \end{pmatrix} = \begin{pmatrix} \lambda z_1 \\ \lambda z_2 \\ \lambda z_3 \end{pmatrix},$$
  just like in $\mathbb{R}^3$.
- $\mathscr{M}_{m,n}$, the set of $m \times n$ matrices, is a complex vector space if we take complex matrices, i.e. matrices with complex numbers as entries.
- The set of polynomials with complex coefficients is a complex vector space.
- The set $\mathbb{C}$ of complex numbers can be viewed as
  - a complex vector space of dimension 1: "vectors" are complex numbers z, and scalars are also complex numbers (This is analogous to how $\mathbb{R}$ is a one-dimensional real vector space.) Think of this as one complex degree of freedom: for each "vector", I can choose one complex number.
  - a *real* vector space of dimension 2: any complex number z can be written as $z = x+iy$ with $x, y \in \mathbb{R}$, the real and imaginary part. Think of this as two real degrees of freedom: for each "vector", I can choose two real numbers.

Exercise 7.3: Make sure you know how to verify the vector space axioms for all these examples. In particular, you should show that $\mathbb{C}$ is a real vector space as given in the very last example.

Remark 7.4: The vector space axioms are exactly the same for real and complex vector spaces. The difference is which scalars we use. For example,

- for a linear combination in a *real* vector space, we use real numbers as scalars,
- for a linear combination in a *complex* vector space, we use complex numbers as scalars, so more scalars are possible.

For example, as we said in the last example above, $\mathbb{C}$ can be viewed as a real or complex vector space. Here is one particular difference:

- As a real vector space, we have

$$\begin{aligned} & \lambda \cdot 1 + \mu \cdot i = 0 \quad \text{with } \lambda, \mu \in \mathbb{R} \\ \Leftrightarrow \quad & \lambda = \mu = 0. \end{aligned}$$

  We can't cancel out the i with any real numbers. So 1 and i are linearly independent in the real vector space $\mathbb{C}$.
- But as a complex vector space, we have a linear combination

$$\lambda \cdot 1 + \mu \cdot i = 0 \quad \text{with } \lambda = i, \mu = -1,$$

  because now *complex* scalars are allowed. So in the complex vector space $\mathbb{C}$, 1 and i are linearly dependent.

Remark 7.5: The definition of a subspace and the results on how to check something is a subspace carry over just the same to complex vector spaces.

C. Real and imaginary parts of matrices

Given a vector or matrix with complex entries, we can also form real and imaginary parts: Let $z_j = x_j + iy_j$ with $x_j, y_j \in \mathbb{R}$, and $v = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$. Then $\mathrm{Re}(v) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ and $\mathrm{Im}(v) = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$, and we have $v = \mathrm{Re}(v) + i\,\mathrm{Im}(v)$. Similarly for a matrix A with entries $z_{jk} = x_{jk} + iy_{jk}$, we have $\mathrm{Re}(A) = (x_{jk})$, $\mathrm{Im}(A) = (y_{jk})$ and $A = \mathrm{Re}(A) + i\,\mathrm{Im}(A)$.

$$2 \times 2 \text{ example:} \qquad \begin{pmatrix} 2+3i & \sqrt{3}+i \\ 8i & 6 \end{pmatrix} = \begin{pmatrix} 2 & \sqrt{3} \\ 0 & 6 \end{pmatrix} + i \begin{pmatrix} 3 & 1 \\ 8 & 0 \end{pmatrix}$$

We can also form complex conjugates componentwise:
if $v = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$ then $\overline{v} = \begin{pmatrix} \overline{z_1} \\ \overline{z_2} \\ \vdots \\ \overline{z_n} \end{pmatrix} = \mathrm{Re}(v) - i\,\mathrm{Im}(v)$, and similarly for matrices.

$$\begin{aligned} 2 \times 2 \text{ example:} \qquad A &= \begin{pmatrix} 2+3i & \sqrt{3}+i \\ 8i & 6 \end{pmatrix} \\ \overline{A} &= \begin{pmatrix} 2-3i & \sqrt{3}-i \\ -8i & 6 \end{pmatrix} = \begin{pmatrix} 2 & \sqrt{3} \\ 0 & 6 \end{pmatrix} - i \begin{pmatrix} 3 & 1 \\ 8 & 0 \end{pmatrix} \end{aligned}$$

Proposition 7.6: (Properties of complex conjugation)
Let $u, v \in \mathbb{C}^n$ and let A be a $m \times k$ complex matrix, B a $k \times n$ complex matrix, and $\lambda \in \mathbb{C}$. Then
(a) $\overline{\overline{u}} = u$ *and* $\overline{\overline{A}} = A$.
(b) $\overline{\lambda u} = \overline{\lambda}\,\overline{u}$.
(c) $\overline{u + v} = \overline{u} + \overline{v}$.
(d) $\overline{(A^T)} = (\overline{A})^T$.
(e) $\overline{(AB)} = (\overline{A})(\overline{B})$.

PROOF. Exercise. (For example, consider individual entries.) □

For the rest of this course, when we say "vector space" we mean "real or complex vector space", unless otherwise specified.

D. Complex roots of real polynomials

As we said earlier, one of the main (algebraic) properties of complex numbers is that any polynomial has a root in $\mathbb{C}$.

Example 7.7: The polynomial $x^2 + 8x + 20$ has roots

$$\begin{aligned} x_{1,2} &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \\ &= \frac{-8}{2} \pm \frac{\sqrt{64 - 80}}{2} \\ &= -4 \pm \frac{4i}{2} \\ &= -4 \pm 2i. \end{aligned}$$

Let's check:

$$\begin{aligned} (x + 4 - 2i)(x + 4 + 2i) &= (x + 4)^2 - (2i)^2 \\ &= x^2 + 8x + 16 + 4 \\ &= x^2 + 8x + 20 \end{aligned}$$

so this indeed factorises our quadratic polynomial.

Notice that the roots come out as a complex conjugate pair. Looking at the formula for the roots of a quadratic, we see that if we do get a complex number as the root of a real polynomial, the i can only come from the discriminant. And in that case we get the two roots being complex conjugates.

On the other hand, we can also check

$$(x - z_1)(x - \overline{z_1}) = x^2 - (z_1 + \overline{z_1})x + z_1\overline{z_1}.$$

Here

$$\begin{aligned} z_1 + \overline{z_1} &= (x_1 + iy_1) + (x_1 - iy_1) = 2x = 2\mathrm{Re}(z_1) \\ z_1 \cdot \overline{z_1} &= (x_1 + iy_1)(x_1 - iy_1) = x_1^2 - i^2y_1^2 = x_1^2 + y_1^2 = |z_1|^2 \end{aligned}$$

i.e.

$$(x - z_1)(x - \overline{z_1}) = x^2 - 2\mathrm{Re}(z_1)x + |z_1|^2,$$

where $2\mathrm{Re}(z_1)$ and $|z_1|^2$ are real numbers.

Slogan: If a real poly has complex roots, then the roots come in complex conjugate pairs.

E. Complex Numbers: Study guide

Concept review.

- ◇ Complex numbers, real and imaginary part, complex conjugation.
- ◇ Complex vector spaces.
- ◇ Vectors and matrices with complex entries, real and imaginary part of those.

Skills.

- ◇ Determine the real and imaginary part of a complex number or a complex vector or matrix.
- ◇ Determine the complex conjugate of a complex number, or complex vector or matrix.
- ◇ Verify something is a complex vector space.
- ◇ Find (possibly complex) roots of a real polynomial.

CHAPTER 8

Linear Maps

A. Definition and basic properties

Recall the following definition from Chapter 1:

Definition 8.1: Given $m \times n$ matrix A, the function $T_A \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ which sends $v \longmapsto Av$ is called a **matrix transformation**. We may sometimes just write $A \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$.

We have seen that the properties of matrix multiplication give us:

$$\begin{aligned} A(u+v) &= Au + Av && \text{for any } u, v \in \mathbb{R}^n; \\ A(\lambda v) &= \lambda \cdot Av && \text{for any } v \in \mathbb{R}^n, \lambda \in \mathbb{R}. \end{aligned}$$

Example 8.2: (Not lectured, just a reminder.)
Consider the map

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}.$$

Then we have

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 + y_1 \\ x_2 + y_2 \end{pmatrix} = \begin{pmatrix} a(x_1+y_1) + b(x_2+y_2) \\ c(x_1+y_1) + d(x_2+y_2) \end{pmatrix} \\ &= \begin{pmatrix} ax_1 + ay_1 + bx_2 + by_2 \\ cx_1 + cy_1 + dx_2 + dy_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} + \begin{pmatrix} ay_1 + by_2 \\ cy_1 + dy_2 \end{pmatrix} \end{aligned}$$

and

$$\begin{aligned} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \left(\lambda \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) &= \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix} = \begin{pmatrix} a\lambda x_1 + b\lambda x_2 \\ c\lambda x_1 + d\lambda x_2 \end{pmatrix} \\ &= \lambda \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} \end{aligned}$$

This property of "preserving addition and scalar multiplication" is a very important one, and we call such a map linear:

Definition 8.3: Given two vector spaces V, W, a function $T \colon V \longrightarrow W$ is called a **linear map** (or **linear transformation**) if

$$T(\lambda v + \mu u) = \lambda T(v) + \mu T(u) \qquad \text{for all } u, v \in V, \lambda, \mu \in F.$$

(Here $F = \mathbb{R}$ or $\mathbb{C}$.)
The space V is called the **domain** (or **source**) of T, and W is called **codomain** (or **target**) of T.

In words we would say *linear maps preserve linear combinations.*

Consequences 8.4:
- ◇ Setting $\mu = 0$ we get $T(\lambda v) = \lambda T(v)$: you can take out scalars.
- ◇ Setting $\lambda = \mu = 1$ we get $T(v+u) = T(v) + T(u)$: linear maps preserve addition.
- ◇ We could define a linear map by asking for these two properties, as we can put them together to give the one we use in the definition. So it does not matter whether you check preservation of linear combinations in one go as given in the definition, or separately check addition and scalar multiplication (as long as you do both of those).

Examples 8.5: a) **[Matrix transformations]** As we saw in Chapter 1, any real $m \times n$ matrix A gives a linear map $T_A\colon \mathbb{R}^n \longrightarrow \mathbb{R}^m$ via $v \longmapsto Av$. So
- ◇ the domain of T_A is $\mathbb{R}^n$: to make the matrix multiplication work, we need the vector v to have as many entries as A has columns;
- ◇ the codomain of T_A is $\mathbb{R}^m$: the vector Av has as many entries as A has rows.

If $e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ldots, e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$ is the standard basis of $\mathbb{R}^n$, we notice that
- ◇ Ae_1 gives the first column of A,
- ◇ Ae_2 gives the second column of A,
- ◇ ...
- ◇ Ae_n gives the last column of A.

So the columns of A are the images of the standard basis vectors:

$T_A(e_k) = (k\text{th column of } A)$.

Then using the linearity, we can see things like
- ◇ $T_A(e_1 + e_2) = T_A(e_1) + T_A(e_2) =$ sum of first two columns of A.
- ◇ $T_A(e_1 - 2e_n) = T_A(e_1) - 2T_A(e_n) =$ first column minus two times last column.

Similarly any complex $m \times n$ matrix gives a linear map $A\colon \mathbb{C}^n \longrightarrow \mathbb{C}^m$.

b) **[Dilations]** For any $r \in \mathbb{R}$, the function $T\colon \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \longmapsto \begin{pmatrix} rx_1 \\ \vdots \\ rx_n \end{pmatrix}$ is linear: if we write $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and $u = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, then

$$\begin{aligned} T(\lambda v + \mu u) &= r(\lambda v + \mu u) = \lambda r v + \mu r u && \text{by vector space axioms} \\ &= \lambda T(v) + \mu T(v). \end{aligned}$$

There are two particularly important cases of this example:
- ◇ when $r = 0$, we get the zero map $0\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$, which sends every vector to 0;
- ◇ when $r = 1$, we get the identity map $\mathrm{id}\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$, which sends every vector to itself.

c) **[Zero map]** For any two vector spaces V, W, there is always a **zero map** $0\colon V \longrightarrow W$ which sends every vector $v \in V$ to $0 \in W$, and this is linear.

d) **[Identity]** For any vector space V, there is always the **identity map** $\mathrm{id}\colon V \longrightarrow V$ which sends every vector $v \in V$ to itself, and this is linear.

e) **[Projections]** The function $S\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ sending $\binom{x_1}{x_2} \longmapsto x_1$ is linear:

$$\begin{aligned} S\left(\lambda\tbinom{x_1}{x_2} + \mu\tbinom{y_1}{y_2}\right) &= S\left(\begin{pmatrix} \lambda x_1 + \mu y_1 \\ \lambda x_2 + \mu y_2 \end{pmatrix}\right) = \lambda x_1 + \mu y_1 \\ &= \lambda S\left(\tbinom{x_1}{x_2}\right) + \mu S\left(\tbinom{y_1}{y_2}\right) \end{aligned}$$

f) The function $P\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ sending $\binom{x_1}{x_2} \longmapsto \binom{x_1}{x_2} + \binom{5}{3}$ is *not* linear:

$$\begin{aligned} & P\left(\tbinom{x_1}{x_2} + \tbinom{y_1}{y_2}\right) = \tbinom{x_1+y_1}{x_2+y_2} + \tbinom{5}{3} \\ \text{but} \quad & P\left(\tbinom{x_1}{x_2}\right) + P\left(\tbinom{y_1}{y_2}\right) = \tbinom{x_1}{x_2} + \tbinom{y_1}{y_2} + \tbinom{10}{6}. \end{aligned}$$

"Translations are not linear."

g) The function $Q\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ sending $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) \longmapsto x_1x_2$ is *not* linear:

$$\begin{aligned} Q\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) + \left(\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right)\right) &= (x_1+y_1)(x_2+y_2) = x_1x_2 + y_1x_2 + x_1y_2 + y_1y_2 \\ \text{so e.g.} \qquad Q\left(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\right) &= 4 \\ \text{but} \qquad Q\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\right) + Q\left(\left(\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right)\right) &= x_1x_2 + y_1y_2 \\ \text{so} \qquad Q\left(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\right) + Q\left(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\right) &= 2. \end{aligned}$$

h) Let P be the space of all real polynomials. The function $T\colon P \longrightarrow P$ defined by $T(p) = xp$ is linear: for polynomials $p, q \in P$, we have

$$T(\lambda p + \mu q) = x(\lambda p + \mu q) = \lambda xp + \mu xq.$$

i) Let V be the space of infinitely differentiable functions on the interval $(0, 1)$. Then differntiation is a linear map $D\colon V \longrightarrow V$:

$$D(\lambda f + \mu g) = (\lambda f + \mu g)' = \lambda f' + \mu g'.$$

Exercise 8.6: Show that if T is linear, then

$$T(\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n) = \lambda_1 T(v_1) + \cdots + \lambda_n T(v_n).$$

We will now see some of the basic properties of linear maps.

Proposition 8.7: (Linear maps preserve 0 and differences.)
If $T\colon V \longrightarrow W$ is a linear map, then

(i) $T(0) = 0$, and
(ii) $T(u - v) = T(u) - T(v)$.

Proof. Exercise. □

We know that multiplying matrices of matching sizes gives another matrix. This corresponds to the composition of linear maps.

Proposition 8.8: (Composite of linear maps is linear.)
Let $T\colon U \longrightarrow V$ and $S\colon V \longrightarrow W$ be linear maps. Then the composite $S \circ T\colon U \longrightarrow W$ given by $S \circ T(u) = S(T(u))$ is linear.

Proof.

$$\begin{aligned} S \circ T(\lambda_1 u_1 + \lambda_2 u_2) &= S(T(\lambda_1 u_1 + \lambda_2 u_2)) \\ &= S(\lambda_1 T(u_1) + \lambda_2 T(u_2)) && \text{because } T \text{ is linear} \\ &= \lambda_1 S(T(u_1)) + \lambda_2 S(T(u_2)) && \text{because } S \text{ is linear} \\ &= \lambda_1 S \circ T(u_1) + \lambda_2 S \circ T(u_2) \end{aligned}$$ □

Example 8.9: If A is an $m \times k$ matrix and B is a $k \times n$ matrix, then $T_A \circ T_B = T_{AB}$: composing the linear maps given by matrix transformations with A and B gives the matrix transformation with the matrix product AB. Notice the order: we do T_B first and then T_A, as is usual in functions: $(T_A \circ T_B)(v) = T_A(T_B(v))$.

$$\begin{tikzcd} \mathbb{R}^n \arrow[r, "T_B"] \arrow[dr, "T_{AB}"'] & \mathbb{R}^k \arrow[d, "T_A"] \\ & \mathbb{R}^m \end{tikzcd}$$

As we are used to from the identity matrix, composing a map on either side with the identity map does not change the map.

Proposition 8.10: (Composition with identity)
For any linear map $T\colon U \longrightarrow V$, we have

(i) $T \circ \mathrm{id}_U = T\colon U \longrightarrow V$ and
(ii) $\mathrm{id}_V \circ T = T\colon U \longrightarrow V$.

$$\begin{tikzcd} U \arrow[r, "\mathrm{id}_U"] \arrow[dr, "T"'] & U \arrow[d, "T"] \arrow[dr, "T"] & \\ & V \arrow[r, "\mathrm{id}_V"'] & V \end{tikzcd}$$

Proof. Exercise. □

Exercise 8.11: In general, composition of maps is associative. This means, if $f\colon X \longrightarrow Y$, $g\colon Y \longrightarrow Z$ and $h\colon Z \longrightarrow W$, then $(h\circ g)\circ f = h\circ(g\circ f)\colon X \longrightarrow W$. Show this by looking at how these two composites act on a single element $a \in X$.

One very nice property of linear maps is that they are determined by their values on a basis.

Proposition 8.12: (Linear maps are determined by values on a basis.)
Let $T\colon V \longrightarrow W$ be a linear map and $v_1, \cdots, v_n$ a basis of V. Then for any $v \in V$ we have

$$T(v) = \lambda_1 T(v_1) + \cdots \lambda_n T(v_n),$$

where the λ_i are scalars such that $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$.
Also, any choice of images $w_1, \ldots w_n \in W$ such that $T(v_i) = w_i$ gives rise to a linear map in this way.

Proof. As a basis spans, any $v \in V$ can be written as $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$. So by linearity, $T(v) = T(\lambda_1 v_1 + \cdots + \lambda_n v_n) = \lambda_1 T(v_1) + \cdots + \lambda_n T(v_n)$.
For the second part, given $T(v_i) = w_i$, we can define $T(v) = \lambda_1 T(v_1) + \cdots + \lambda_n T(v_n)$. This is unambiguous, because $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ for *unique* coefficients λ_i, because a basis is linearly independent.
We need to check that if T is defined in this way, it really is linear. Let $u, v \in V$, then $u = \mu_1 v_1 + \ldots + \mu_n v_n$ and $v = \lambda_1 v_1 + \ldots + \lambda_n v_n$ for unique $\mu_i, \lambda_i \in F$. Then for any scalars $\nu_1, \nu_2 \in F$, we have

$$\begin{aligned}\nu_1 u + \nu_2 v &= \nu_1(\mu_1 v_1 + \ldots + \mu_n v_n) + \nu_2(\lambda_1 v_1 + \ldots + \lambda_n v_n)\\ &= (\nu_1\mu_1 + \nu_2\lambda_1)v_1 + (\nu_1\mu_2 + \nu_2\lambda_2)v_2 + \ldots + (\nu_1\mu_n + \nu_2\lambda_n)v_n.\end{aligned}$$

So

$$\begin{aligned} T(\nu_1 u + \nu_2 v) &= (\nu_1\mu_1 + \nu_2\lambda_1)T(v_1) + (\nu_1\mu_2 + \nu_2\lambda_2)T(v_2) + \ldots + (\nu_1\mu_n + \nu_2\lambda_n)T(v_n)\\ \text{and}\quad \nu_1 T(u) + \nu_2 T(v) &= \nu_1\left[\mu_1 T(v_1) + \ldots + \mu_n T(v_n)\right] + \nu_2\left[\lambda_1 T(v_1) + \ldots + \lambda_n T(v_n)\right]\\ &= (\nu_1\mu_1 + \nu_2\lambda_1)T(v_1) + (\nu_1\mu_2 + \nu_2\lambda_2)T(v_2) + \ldots + (\nu_1\mu_n + \nu_2\lambda_n)T(v_n),\end{aligned}$$

so $T(\nu_1 u + \nu_2 v) = \nu_1 T(u) + \nu_2 T(v)$, and T is linear. □

Examples 8.13: a) The matrix $A = \left(\begin{smallmatrix}3&1\\5&6\end{smallmatrix}\right)$ sends the standard basis vectors $e_1 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ and $e_2 = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ to $\left(\begin{smallmatrix}3\\5\end{smallmatrix}\right)$ and $\left(\begin{smallmatrix}1\\6\end{smallmatrix}\right)$ respectively. Then for any $v = \left(\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right)$ we have

$$v = x_1\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right) + x_2\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$$

so

$$Av = x_1\left(\begin{smallmatrix}3\\5\end{smallmatrix}\right) + x_2\left(\begin{smallmatrix}1\\6\end{smallmatrix}\right) = \left(\begin{smallmatrix}3x_1+x_2\\5x_1+6x_2\end{smallmatrix}\right).$$

b) Consider the basis

$$v_1 = \begin{pmatrix}1\\1\\1\end{pmatrix},\ v_2 = \begin{pmatrix}1\\1\\0\end{pmatrix},\ v_3 = \begin{pmatrix}1\\0\\0\end{pmatrix}$$

of $\mathbb{R}^3$. We can define a linear map $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ by setting

$$T(v_1) = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right),\ T(v_2) = \left(\begin{smallmatrix}2\\-1\end{smallmatrix}\right),\ T(v_3) = \left(\begin{smallmatrix}4\\3\end{smallmatrix}\right).$$

Given any $v = \begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} \in \mathbb{R}^3$, to determine $T(v)$, we first find the coordinates $\lambda_1, \lambda_2, \lambda_3$ such that $v = \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$:

$$\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = \lambda_1\begin{pmatrix}1\\1\\1\end{pmatrix} + \lambda_2\begin{pmatrix}1\\1\\0\end{pmatrix} + \lambda_3\begin{pmatrix}1\\0\\0\end{pmatrix} = \begin{pmatrix}\lambda_1+\lambda_2+\lambda_3\\ \lambda_1+\lambda_2\\ \lambda_1\end{pmatrix}$$

So $\lambda_1 = x_3$, $\lambda_2 = x_2 - x_3$ and $\lambda_3 = x_1 - x_2$. So

$$T(v) = x_3\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right) + (x_2 - x_3)\left(\begin{smallmatrix}2\\-1\end{smallmatrix}\right) + (x_1 - x_2)\left(\begin{smallmatrix}4\\3\end{smallmatrix}\right).$$

This completely determines the map T.

Find formula for linear map from values on a basis
Suppose we know $v_1, \ldots, v_n$ is a basis for V, and we have $T(v_i) = w_i \in W$, i.e. we know the values of a linear map T on this basis. How do we work out what T does on a general vector $v \in V$? In the case where $V = \mathbb{R}^n$:

- Work out the coefficients $\lambda_1, \ldots, \lambda_n$ such that
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n.$$
(Either see if you can just write them down, or solve the inhomogeneous linear system.)
- Then $T\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right) = \lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_n w_n$.

Study guide.
Concept review

- Linear map.
- Domain and codomain.
- Composition of maps.
- Zero map, identity map.
- Linear map is determined by values on a basis.
- Build up a repertoire of and intuition for linear maps.

Skills

- Determine whether a function is a linear map.
- Find a formula for a linear map given values on a basis.

B. Kernels and Images

In Chapter 4, we learnt about the column space and the null space of a matrix. These concepts also exist for general linear maps.

Definition 8.14: Given a linear map $T : U \longrightarrow V$, the **kernel** of T is
$$\operatorname{Ker} T = \{u \in U \mid T(u) = 0\} \qquad \text{everything which is mapped to } 0$$
and the **image** of T is
$$\operatorname{Im} T = \{v \in V \mid \exists u \in U \text{ with } T(u) = v\} \qquad \text{everything in } V \text{ that is reached by } T$$

Examples 8.15: a) If T is a matrix transformation $T(v) = Av$, then the kernel of T is exactly the null space of the matrix A, and the image of T is exactly the column space of A. This is because the columns are the images of the standard basis vectors
$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \ \ldots, \ e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix},$$
so as a linear map is determined by values on a basis (Proposition 8.12), the linear combinations of the columns are exactly images under T: if A has columns $a_1, \ldots, a_n$, then
$$x_1 a_1 + \cdots + x_n a_n = x_1 T(e_1) + \cdots + x_n T(e_n) = T\left(\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}\right).$$

b) Recall the projection $S\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ sending $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ to x_1.

$$\operatorname{Ker} S = \{\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right) \mid x_1 = 0\} = \left\{\left(\begin{smallmatrix} 0 \\ x_2 \end{smallmatrix}\right) \mid x_2 \in \mathbb{R}\right\}$$
$$\operatorname{Im} S = \mathbb{R}$$

because for any $x \in \mathbb{R}$, we have for example $\left(\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right) \in \mathbb{R}^2$ with $S(\left(\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right)) = x$.

c) The map $T\colon P \longrightarrow P$ sending a polynomial p to $T(p)(x) = xp(x)$ has

$$\operatorname{Ker} T = \{0\} \qquad \text{— only 0 is mapped to 0}$$
$$\operatorname{Im} T = \{\text{polynomials with constant term } 0\}.$$

d) The differentiation map $D\colon V \longrightarrow V$, where V is the space of infinitely differentiable functions on $(0,1)$, has

$$\operatorname{Ker} D = \{\text{constant functions}\}$$
$$\operatorname{Im} D = V$$

Exercise 8.16: Determine the kernel and image of the zero map and the identity, and their respective dimensions in terms of the dimension of V.

Proposition 8.17: (Kernels and images are subspaces.)
Let $T\colon U \longrightarrow V$ be a linear map. Then $\operatorname{Ker} T$ is a subspace of the domain U and $\operatorname{Im} T$ is a subspace of the codomain V.

Proof. We must check:

(i) $0 \in \operatorname{Ker} T$: Yes, as $T(0) = 0$ by Proposition 8.7.
(ii) If $u, v \in \operatorname{Ker} T$, then $T(u + v) = T(u) + T(v) = 0 + 0 = 0$. So also $u + v \in \operatorname{Ker} T$.
(iii) If $u \in \operatorname{Ker} T$ and $\lambda \in F$, then $T(\lambda u) = \lambda T(u) = \lambda \cdot 0 = 0$. So also $\lambda u \in \operatorname{Ker} T$.

So $\operatorname{Ker} T$ is a subspace.
Exercise Check the corresponding properties for $\operatorname{Im} T$. □

As they are subspaces, the kernel and image of a linear map have dimensions.

Definition 8.18: Given a linear map $T\colon U \longrightarrow V$, the dimension of the image of T is called **rank** and written $\operatorname{rank}(T)$ or $\operatorname{r}(T)$. The dimension of the kernel is called **nullity** and written $\operatorname{nullity}(T)$ or $\operatorname{n}(T)$.

These two quantities are related.

Theorem 8.19: (Rank-nullity)
If $T\colon V \longrightarrow W$ is a linear map with V finite dimensional, then

$$\operatorname{n}(T) + \operatorname{r}(T) = \dim V.$$

Proof. Let $\dim V = n$. Let $v_1, \ldots, v_k$ be a basis for $\operatorname{Ker} T$, and extend it to a basis $v_1, \ldots, v_n$ for V (possible by "extend to basis" Theorem 6.15). Then $T(v_1) = T(v_2) = \cdots = T(v_k) = 0$, as these vectors are in the kernel of T. We will show that $T(v_{k+1}), T(v_{k+2}), \ldots, T(v_n)$ form a basis for $\operatorname{Im} T$.
First of all, these vectors span $\operatorname{Im} T$: for $w \in \operatorname{Im} T$, there exists $v \in V$ such that $w = T(v)$. As $v_1, \ldots, v_n$ is a basis for V,

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n \qquad \text{for some } \lambda_i \in F.$$

So

$$\begin{aligned} w &= T(v) = T(\lambda_1 v_1 + \cdots + \lambda_n v_n) \\ &= \lambda_1 T(v_1) + \cdots + \lambda_n T(v_n) && \text{by linearity} \\ &= 0 + \cdots + 0 + \lambda_{k+1} T(v_{k+1}) + \cdots + \lambda_n T(v_n). \end{aligned}$$

So $T(v_{k+1},)\dots,T(v_n)$ span $\operatorname{Im} T$.
To show they are linearly independent, suppose

$$\mu_{k+1} T(v_{k+1}) + \cdots + \mu_n T(v_n) = 0 \quad \text{for some } \mu_{k+1}, \dots, \mu_n \in F.$$

Then by linearity $T(\mu_{k+1} v_{k+1} + \cdots + \mu_n v_n) = 0$, so $\mu_{k+1} v_{k+1} + \cdots + \mu_n v_n \in \operatorname{Ker} T$. But $v_1, \cdots, v_k$ is a basis for $\operatorname{Ker} T$, so

$$\mu_{k+1} v_{k+1} + \cdots + \mu_n v_n = \mu_1 v_1 + \cdots \mu_k v_k.$$

But then

$$\mu_1 v_1 + \cdots + \mu_k v_k - \mu_{k+1} v_{k+1} - \cdots - \mu_n v_n = 0.$$

But $v_1, \dots, v_n$ are a basis of V, so they are linearly independent, so all the $\mu_j = 0$. This shows that $T(v_{k+1}), \dots, T(v_n)$ are linearly independent, so $\dim(\operatorname{Im} T) = n - k$. So

$$n = \dim V = k + (n-k) = \mathrm{n}(T) + \mathrm{r}(T).$$

□

Examples 8.20: Go through any of the linear maps you have met so far, either as examples in the notes, or from the Workbook, or other practice e.g. from Wiley+, and check that their rank and nullity are related in this way. Start on the 4 examples after the definition of kernel and image.

Study guide.

Concept review

- ◇ Kernel and image of linear map.
- ◇ Properties of kernel and image.
- ◇ Rank and nullity of linear map.
- ◇ Relationship between rank and nullity of a linear map.

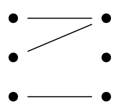
Skills

- ◇ Find the kernel and image of a linear map.
- ◇ Find bases for kernel and image of a linear map.
- ◇ Find rank and nullity of a linear map.

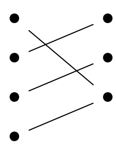
C. Surjective and injective functions

We have have seen that for a linear map T, we always have $T(0) = 0$. In examples, we've seen that sometimes other vectors are mapped to 0 as well, and sometimes 0 is the only vector mapped to 0. We have also seen that the image is always a subset of the codomain, and sometimes it is the whole of the codomain and sometimes it is not. These special properties get special names, which are actually applied to all functions, not just linear maps.

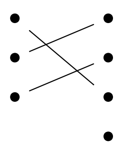
Definition 8.21: A function $f\colon X \longrightarrow Y$ is called **injective** if for any $a, b \in X$, $f(a) = f(b)$ implies $a = b$.
Slogan: "Different elements have different images."
f is called **surjective** if $\operatorname{Im}(f) = Y$.
Slogan: "f reaches everything in the codomain."



neither injective nor surjective



surjective, not injective



injective, not surjective

Examples 8.22: Look back at the examples after the definition of linear map.

b) [Dilations] If $r \neq 0$, then $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ defined by $T(v) = rv$ is injective:

$$\begin{aligned} \text{If} \qquad rv &= rw \\ \Rightarrow \qquad r(v-w) &= 0 \\ \Rightarrow \qquad v-w &= 0 \\ \Rightarrow \qquad v &= w. \end{aligned}$$

T is also surjective: For any $w \in \mathbb{R}^n$, also $\frac{1}{r}w \in \mathbb{R}^n$, and $r \cdot (\frac{1}{r}w) = w$.

c) The zero map $0\colon V \longrightarrow W$ is
not injective as long as $V \neq \{0\}$;
not surjective as long as $W \neq \{0\}$.

d) The identity $\mathrm{id}\colon V \longrightarrow V$ is both injective and surjective.

e) [Projection] $S\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ with $S\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\right) = x_1$ is surjective:
for any $x \in \mathbb{R}$, $S\left(\left(\begin{smallmatrix} x \\ 0 \end{smallmatrix}\right)\right) = x$.
It is not injective: $S\left(\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\right) = S\left(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\right) = 1$.

h) $T\colon P \longrightarrow P$ with $T(p) = xp$ is injective:
Suppose $p = a_0 + a_1x + \cdots + a_nx^n$ and $q = b_0 + b_1x + \cdots + b_mx^m$ and $xp = xq$.

$$\begin{aligned} \Rightarrow \qquad & a_0x + a_1x^2 + \cdots a_nx^{n+1} = b_0x + \cdots + b_mx^{m+1} \\ \Rightarrow \qquad & n = m, \quad \text{and} \quad a_k = b_k \text{ for all } k, \text{ by comparing coefficients.} \\ \Rightarrow \qquad & p = q. \end{aligned}$$

So T is injective.

But T is not surjective: the polynomial $p = 1$ has no preimage (there is no poly q with $xq = 1$).

Proposition 8.23: (Composing injective or surjective functions)
Consider functions $f\colon X \longrightarrow Y$ and $g\colon Y \longrightarrow Z$.
(i) If f and g are both injective, then the composite $g \circ f\colon X \longrightarrow Z$ is injective.
(ii) If f and g are both surjective, then the composite $g \circ f\colon X \longrightarrow Z$ is surjective.

Proof. (i) Suppose $(g \circ f)(a) = (g \circ f(b)$. Then

$$\begin{aligned} g(f(a)) &= g(f(b)) \\ \Rightarrow \qquad f(a) &= f(b) && \text{because } g \text{ injective} \\ \Rightarrow \qquad a &= b && \text{because } f \text{ injective.} \end{aligned}$$

(ii) Given $z \in Z$, there is some $y \in Y$ with $g(y) = z$ because g is surjective. Then there is some $x \in X$ with $f(x) = y$ because f is surjective. So $g(f(x)) = z$, so $g \circ f$ is surjective.

□

If we look at linear maps rather than general functions, we can check if a linear map is injective just by looking at the kernel.

Proposition 8.24: (Injectivity via kernels)
If $T\colon V \longrightarrow W$ is a linear map between vector spaces, then T is injective if and only if $\operatorname{Ker} T = 0$.

Proof. If T is injective, then

$$v \in \operatorname{Ker} T \quad \Rightarrow \quad T(v) = 0 \quad \Rightarrow \quad T(v) = T(0) \quad \Rightarrow \quad v = 0.$$

So $\operatorname{Ker} T = \{0\}$.
Conversely, if $\operatorname{Ker} T = 0$, suppose $T(v) = T(w)$. Then

$$T(v) - T(w) = 0 \quad \Rightarrow \quad T(v-w) = 0 \quad \Rightarrow \quad v - w \in \operatorname{Ker} T \quad \Rightarrow \quad v - w = 0 \quad \Rightarrow \quad v = w.$$

So T is injective. □

Exercise 8.25: Given linear maps $T\colon U \longrightarrow V$ and $S\colon V \longrightarrow U$, show that if $S \circ T = \mathrm{id}_U$, (that is, $S(T(u)) = u$ for all $u \in U$), then

$$\begin{tikzcd} U \arrow[r, "T"] \arrow[dr, "\mathrm{id}_U"'] & V \arrow[d, "S"] \\ & U \end{tikzcd}$$

(i) T is injective, and
(ii) S is surjective.

Study guide.
Concept review

- ⋄ Injective and surjective functions.
- ⋄ Behaviour of injectivity and surjectivity under composition.

Skills

- ⋄ Determine whether a map is injective or surjective.

D. Isomorphisms

Definition 8.26: A linear map $T\colon V \longrightarrow W$ which is both injective and surjective is called a **(linear) isomorphism**. If there exists some linear isomorphism between V and W, we say the two spaces are **isomorphic** and write $V \cong W$.

Examples 8.27: a) If T is a matrix transformation $T(v) = Av$, then T is an isomorphism exactly when A is an invertible matrix.
b) The identity map $\mathrm{id}\colon V \longrightarrow V$ is an isomorphism for any vector space V.
c) The dilation map $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by $T(v) = rv$ with $r \neq 0$ is an isomorphism.
d) The projection map $T\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ we saw before is not an isomorphism: it is not injective.
e) Let P_n be the space of complex polynomials of degree up to n, and let $S\colon P_n \longrightarrow \mathbb{C}^{n+1}$ be the map defined by $S(a_0 + a_1x + \cdots + a_nx^n) = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}$. This is an isomorphism. (**Exercise:** check it really is a linear map, and check that it is injective and surjective.)

We've seen that a map can be injective without being surjective, or surjective without being injective. But when looking at a linear map from one space to itself, the rank-nullity theorem restricts the possibilities.

Proposition 8.28: (Check one get one free for isos)
If $T\colon V \longrightarrow V$ is a linear map from a finite-dimensional vector space to itself, then the following are equivalent:
(i) T is injective.
(ii) T is surjective.
(iii) T is an isomorphism.

Proof.

$$\begin{array}{lcr} & T \text{ injective} & \\ \Leftrightarrow & \operatorname{Ker} T = 0 & \text{by Injectivity via Kernels, Prop. 8.24} \\ \Leftrightarrow & \mathrm{n}(T) = 0 & \\ \Leftrightarrow & \mathrm{r}(T) = \dim V & \text{by the Rank-Nullity Theorem, Theorem 8.19} \\ \Leftrightarrow & \operatorname{Im} T = V & \text{because } \operatorname{Im} T \text{ subspace of } V \text{ (see Theorem 6.20)} \\ \Leftrightarrow & T \text{ surjective.} & \end{array}$$

This proves "(i)⇔(ii)", i.e. as soon as T is one of injective or surjective, it is both, making it an isomorphism. □

Similarly to matrix inverses, isomorphisms have inverse linear maps.

Definition 8.29: Given a linear map $T: U \longrightarrow V$, an **inverse** of T is a linear map $S: V \longrightarrow U$ in the other direction such that

$$S \circ T = \mathrm{id}_U : U \longrightarrow U \qquad \text{and}$$
$$T \circ S = \mathrm{id}_V : V \longrightarrow V.$$

$$\begin{tikzcd}
U \arrow[r, "T"] \arrow[dr, "\mathrm{id}_U"'] & V \arrow[d, "S"] \arrow[dr, "\mathrm{id}_V"] & \\
 & U \arrow[r, "T"'] & V
\end{tikzcd}$$

Example 8.30: If A is an invertible $n \times n$ matrix, then $(T_A)^{-1} = T_{A^{-1}}$: the inverse transformation uses the inverse matrix. This is because

$$(T_{A^{-1}} \circ T_A)(v) = A^{-1}(Av) = I_n v = v$$

and

$$(T_A \circ T_{A^{-1}})(v) = A(A^{-1}v) = I_n v = v.$$

Proposition 8.31: (Inverses are unique.)
If a linear map $T: U \longrightarrow V$ has an inverse, then this inverse is unique.

Proof. Suppose S_1 and S_2 are both inverses to T. Then

$$\begin{aligned}
S_1 &= S_1 \circ \mathrm{id}_V && \text{because composition with the identity does not change the map,}\\
&= S_1 \circ (T \circ S_2) && \text{because } S_2 \text{ is an inverse to } T,\\
&= (S_1 \circ T) \circ S_2 && \text{because composition of functions is associative,}\\
&= \mathrm{id}_U \circ S_2 && \text{because } S_1 \text{ is an inverse to } T,\\
&= S_2 && \text{because composition with the identity does not change the map,}
\end{aligned}$$

so they are the same map. □

Notation 8.32: As inverses are unique, it is unambiguous to write T^{-1} for the inverse of T, if it exists.

Fact 8.33: A linear map $T: U \longrightarrow V$ has an inverse if and only if it is an isomorphism.

Proof. [Not lectured] Exercise 8.25 shows that if T has an inverse, then it is both injective and surjective, so an isomorphism.
Conversely, suppose T is both injective and surjective. We can define the inverse $T^{-1}: V \longrightarrow U$ by setting $T^{-1}(v) = u$ where $u \in U$ is such that $T(u) = v$. Such a u exists because T is surjective, and it is unique because T is injective. It is clear that then $T \circ T^{-1} = \mathrm{id}_V$ and $T^{-1} \circ T = \mathrm{id}_U$. The harder part is to show that this T^{-1} is also a linear map. We won't do it here. □

Corollary 8.34: *The inverse of an isomorphism is also an isomorphism.*

Examples 8.35: Looking back to the examples after the definition of isomorphism:

a) A matrix transformation T_A is an isomorphism if and only if A is invertible: we've already seen that then $(T_A)^{-1} = T_{A^{-1}}$.
b) The identity map $\mathrm{id}: V \longrightarrow V$ is its own inverse.
c) The dilation map $T: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ given by $T(v) = rv$ with $r \neq 0$ has inverse $T^{-1}(v) = \frac{1}{r}v$.
d) The map $S: P_n \longrightarrow \mathbb{C}^{n+1}$ defined by $S(a_0 + a_1x + \cdots + a_nx^n) = \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}$ has inverse $S^{-1}\left(\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}\right) = a_0 + a_1x + \cdots + a_nx^n$.

Since composing two injective linear maps gives another injective linear map, and composing two surjective linear maps gives another surjective linear map (see Proposition 8.23), it is immediately clear that composing two isomorphisms gives another isomorphism. But we also have an easy way to find the inverse of the composite.

Lemma 8.36: (Socks and Shoes)
Given two invertible linear maps $T\colon U \longrightarrow V$ and $S\colon V \longrightarrow W$, then $(S\circ T)^{-1} = T^{-1}\circ S^{-1}$.

If you first put on your socks and then your shoes, to undo the process you have to first take off your shoes before you can take off your socks.

$$\begin{tikzcd}
U \arrow[r, "T"] \arrow[ddrr, bend right, "\mathrm{id}_U"'] & V \arrow[d, "S"] \arrow[dr, "\mathrm{id}_V"] & \\
& W \arrow[r, "S^{-1}"'] & V \arrow[d, "T^{-1}"] \\
& & U
\end{tikzcd}$$

PROOF. As inverses are unique, it is enough to show that $(S\circ T)\circ(T^{-1}\circ S^{-1}) = \mathrm{id}_W$ and $(T^{-1}\circ S^{-1})\circ(S\circ T) = \mathrm{id}_U$. We have, for any $u \in U$,

$$(T^{-1}\circ S^{-1})\circ(S\circ T) = T^{-1}\circ(S^{-1}\circ S)\circ T = T^{-1}\circ\mathrm{id}_V\circ T = T^{-1}\circ T = \mathrm{id}_U$$

and similarly for the other direction. □

Isomophisms are special because they have inverses, and also because they send bases to bases.

Proposition 8.37: (Isos perserve bases.)
If $T\colon V \longrightarrow W$ is an isomorphism between finite-dimensional vector spaces and $v_1, \ldots, v_n$ is a basis of V, then $T(v_1), \ldots, T(v_n)$ is a basis of W.

PROOF. First let us work out the dimension of W. As T is an isomorphism, it is injective, so $\mathrm{n}(T) = 0$. And it is surjective, so $\operatorname{Im} T = W$, so $\dim W = \mathrm{r}(T)$. The Rank-Nullity Theorem 8.19 gives us $\dim V = \mathrm{n}(T) + \mathrm{r}(T) = 0 + \dim W$, so for any isomorphism, the dimension of the domain and codomain have to agree.
So to show that $T(v_1), \ldots, T(v_n)$ is a basis, it is enough to show that it is linearly independent, because we have the right number of vectors for "check one get one free on bases" (Proposition 6.13). Suppose

$$\lambda_1 T(v_1) + \cdots + \lambda_n T(v_n) = 0.$$

Then as T is linear, we have

$$T(\lambda_1 v_1 + \cdots \lambda_n v_n) = 0,$$

i.e. $\lambda_1 v_1 + \cdots \lambda_n v_n \in \mathrm{Ker}(T) = \{0\}$. So $\lambda_1 v_1 + \cdots \lambda_n v_n = 0$. But $v_1, \ldots, v_n$ is a basis, so it is linearly independent, so all $\lambda_j = 0$. So $T(v_1), \ldots, T(v_n)$ are linearly independent, and so form a basis of $W = \operatorname{Im} T$. □

Theorem 8.38: (Dimension n means iso to $\mathbb{R}^n$.)
Any real vector space of dimension n is isomorphic to $\mathbb{R}^n$. Any complex vector space of dimension n is isomorphic to $\mathbb{C}^n$.

PROOF. To do the proof of both statements at once, we write F to mean $\mathbb{R}$ or $\mathbb{C}$.

Let V be a vector space of dimension n. Then it has some basis $v_1, \dots, v_n$. Define a linear map $E\colon V \longrightarrow F^n$ by

$$E(v_1) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_1 \qquad E(v_2) = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_2$$

$$\vdots \qquad\qquad \vdots$$

$$E(v_{n-1}) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix} = e_{n-1} \qquad E(v_n) = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix} = e_n.$$

This extends to a unique linear map by Proposition 8.12. (We have $E(v) = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$, where $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ is the unique linear combination expressing v in terms of the given basis.) The inverse of E is

$$E^{-1}\left(\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}\right) = \lambda_1 v_1 + \cdots + \lambda_n v_n,$$

so E is indeed an isomorphism.

Exercise: Check that these two maps are indeed inverse to each other. □

Remark 8.39: Notice that the isomorphism from V to $\mathbb{R}^n$ given in the proof above gives each vector $v \in V$ as a coordinate vector $[v]_B$ with respect to the given basis. So "writing a vector as a coordinate vector with respect to a given basis" is really applying an isomorphism, and then working in $\mathbb{R}^n$ instead.

Corollary 8.40: (Iso ⇔ same dimension.)
Two vector spaces V and W are isomorphic if and only if they have the same dimension.

Proof. The rank nullity theorem already shows us that if V and W are isomorphic, then they have the same dimension. Conversely, if they have the same dimension n, then by Theorem 8.38, both spaces are isomorphic to F^n i.e. to $\mathbb{R}^n$ if they are both real vector spaces, or to $\mathbb{C}^n$ if they are both complex vector spaces, so there are isomorphisms $T\colon V \longrightarrow F^n$ and $S\colon W \longrightarrow F^n$. As the inverse of an isomophism is also an isomorphism (Corollary 8.34) and the composite of isomorphisms is an isomorphism (Proposition 8.23), we have an isomorphism $S^{-1} \circ T\colon V \longrightarrow W$. □

Study guide.

Concept review

- ⋄ Isomorphisms and inverses.
- ⋄ Relationship between injectivity and surjectivity for maps from a vector space to itself.
- ⋄ Uniqueness of inverses.
- ⋄ Inverse of a composite.
- ⋄ Conditions for vector spaces to be isomorphic.

Skills

- ⋄ Find inverses to isomorphisms.
- ⋄ Find an isomorphism between vector spaces of the same dimension.

E. Matrix representation of linear maps

Using uniqueness of basis respresetation (Theorem 5.13), we defined the concept of a *coordinate vector* in Chapter 5 (Linear Independence and Bases). We can use this idea to represent any linear map between finite-dimensional vector spaces as a matrix.

Definition 8.41: Let $T\colon V \longrightarrow W$ be a linear map between finite-dimensional spaces, and let $B = \{v_1, \dots, v_n\}$ be a basis for V and $C = \{w_1, \dots, w_m\}$ be a basis for W. Then the **matrix for T with respect to the bases B and C** is the matrix A whose columns are the images of the basis vectors of V, written as coordinate vectors with respect to the basis C of W.

$$[T]_{C,B} = {}_C[T]_B = A = \begin{pmatrix} [T(v_1)]_C & [T(v_2)]_C & \cdots & [T(v_n)]_C \end{pmatrix}$$

This matrix is therefore a $m \times n$ matrix.

Lemma 8.42: (Linear map as matrix transformation)
Given a linear map $T\colon V \longrightarrow W$ and bases $B = \{v_1, \dots, v_n\}$ for V and $C = \{w_1, \dots, w_m\}$ for W as above, then for any $v \in V$,

$$[T(v)]_C = {}_C[T]_B[v]_B.$$

In words: to work out the coordinate vector of the image of v, we multiply the matrix for T with the coordinate vector of v.
The order of the bases in the label is meant to show you that "the B cancels out the B and leaves the C".

Proof. Let $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$, so the coordinate vector of v with respect to basis B of V is

$$[v]_B = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Then by linearity, $T(v) = \lambda_1 T(v_1) + \cdots + \lambda_n T(v_n)$.
If $A = {}_C[T]_B$ is the matrix for T with respect to the given bases, then column k of this matrix is the vector $a_k = [T(v_k)]_C$. As taking coordinate vectors is a linear isomorphism (see proof of Theorem 8.38), the coordinate vector of $T(v)$ is

$$[T(v)]_C = \lambda_1[T(v_1)]_C + \cdots + \lambda_n[T(v_n)]_C = \lambda_1 a_1 + \cdots + \lambda_n a_n = A\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

□

Notice that the coordinate vectors of the basis of V are

$$[v_1]_B = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_1 \qquad [v_2]_B = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = e_2$$

$$\vdots \qquad\qquad \vdots$$

$$[v_{n-1}]_B = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{pmatrix} = e_{n-1} \qquad [v_n]_B = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{pmatrix} = e_n.$$

This can help you remember or work out what the columns of the matrix A are: since matrix multiplication gives

$$Ae_1 = A\begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = a_1 \qquad \text{the first column of } A,$$

and similarly $Ae_k = a_k$, the kth column of A, you see that the columns are the images of the basis vectors for V.

Another picture that might help you visualise what is going on:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "E"'] & W \arrow[d, "E'"] \\ \mathbb{R}^n \arrow[r, "A"'] & \mathbb{R}^m \end{tikzcd} \qquad \begin{tikzcd} v \arrow[r, mapsto, "T"] \arrow[d, mapsto, "E"'] & T(v) \arrow[d, mapsto, "E'"] \\ {[v]_B} \arrow[r, mapsto, "A"'] & A[v]_B = [T(v)]_C \end{tikzcd}$$

Remark 8.43: In the case when $T\colon V \longrightarrow V$ goes from a vector space to itself, we usually take the same basis for V "on both sides". We then just write $[T]_B$ instead of ${}_B[T]_B$.
However, there is going to be an exception to this later on, when we talk about base change matrices.

Examples 8.44: a) The identity map $\mathrm{id}\colon V \longrightarrow V$ is represented by the identity matrix with respect to any basis of V, as long as we choose the same basis for domain and codomain.

b) The zero map $0\colon V \longrightarrow W$ is represented by the zero matrix with respect to any bases for V and W.

c) The dilation $T\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ with $T(v) = rv$ has matrix

$$A = \begin{pmatrix} r & 0 & 0 & \cdots & 0 \\ 0 & r & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & r & 0 \\ 0 & \cdots & 0 & 0 & r \end{pmatrix} = rI$$

with r on the diagonal and 0 everywhere else. This is for any basis of $\mathbb{R}^n$, as long as we take the same basis for domain and codomain.

d) The projection $S\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ with $S\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\right) = x_1$ has matrix

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

with respect to the standard basis vectors e_1, e_2 of $\mathbb{R}^2$ and the basis $w_1 = 1$ for $\mathbb{R}$.

e) Consider $T\colon P_n \longrightarrow P_{n+1}$ with $T(p) = xp$. If we use the standard bases $1, x, \ldots, x^n$ and $1, x, \ldots, x^n, x^{n+1}$ for the two polynomial spaces, this has $(n+2) \times (n+1)$ matrix

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

This is because $T(1) = x$, so the image of the first basis vector is the second basis vector, and the first column of the matrix represents this. Similarly, $T(x^k) = x^{k+1}$, so the kth column must have 0s everywhere and a 1 in the $k+1$st row.

f) Consider $D\colon P_{n+1} \longrightarrow P_n$ given by $D(p) = p'$, the derivative of the polynomial p. With respect to the standard bases $1, x, \ldots, x^n, x^{n+1}$ and $1, x, \ldots, x^n$, this has $(n+1) \times (n+2)$ matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 0 & n+1 \end{pmatrix}$$

Again this is because

- $\diamond$ $D(1) = 0$, so the first column must be 0.

- $D(x) = 1$, so the second column must be the first basis vector, which is represented by $\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$.
- $D(x^k) = kx^{k-1}$, so the kth column has a k in the $k-1$st entry, and 0s everywhere else.

Working out the matrix for a linear map wrt given bases
Let $T\colon V \longrightarrow W$ be a linear map, $v_1, \ldots, v_n$ basis B for V and $w_1, \ldots, w_m$ basis C for W. This is how we work out the matrix for T with respect to these bases.

- Write down the definition
$$_C[T]_B = \begin{pmatrix} [T(v_1)]_C & [T(v_2)]_C & \cdots & [T(v_n)]_C \end{pmatrix},$$
ideally already substituting the given vectors $v_1, \ldots, v_n$ into it (rather than writing the symbol v_1). This helps because you can refer back to it without having to keep it in your mind. You can also get some marks already.
- Work out the images of the basis vectors:
$$T(v_1) = \ldots, \quad T(v_2) = \ldots, \quad \cdots, \quad T(v_n) = \ldots$$
I.e. you have to apply the given map to the given basis vectors of V.
- Work out the coordinate vectors of the images you have just worked out. I.e. you are looking for $\lambda_1, \ldots, \lambda_m$ such that $T(v_1) = \lambda_1 w_1 + \lambda_2 w_2 + \cdots + \lambda_m w_m$. You do this by solving an inhomogeneous linear system. (See if you can solve it just by looking: if not, use Gauss-Jordan algorithm.) You have to do this for each $T(v_i)$.
  TIP: if you are using Gauss-Jordan, you can do them all at the same time: write the vectors $w_1, \ldots, w_m$ into a matrix, and then write several augmentations after it: one for each image $T(v_i)$ that you have worked out.
- You should now have $[T(v_1)]_C = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_m \end{pmatrix}$, and similarly for all the images.
- Assemble all these coordinate vectors into the matrix
$$_C[T]_B = \begin{pmatrix} [T(v_1)]_C & [T(v_2)]_C & \cdots & [T(v_n)]_C \end{pmatrix}.$$

Examples 8.45:

- We saw that the projection $S\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ with $S\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\right) = x_1$ has matrix
$$A = \begin{pmatrix} 1 & 0 \end{pmatrix}$$
with respect to the standard basis vectors e_1, e_2 of $\mathbb{R}^2$ and the basis $w_1 = 1$ for $\mathbb{R}$.
  If we take basis $B = \{v_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\}$ for the domain instead, then we get matrix
$$C = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$
Why? As described in the yellow box, we work out $S(v_1) = 1$ and $S(v_2) = 1$, which is already given as coordinate vector in $\mathbb{R}$, so we write those into the two columns.
  Why does this still represent the same linear map? When we use matrix C, we have to use coordinate vectors with respect to the new basis. So given $v = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$, we have to write it in the new basis: $v = x_2 v_1 + (x_1 - x_2) v_2$. So the new coordinate vector for v is $[v]_B = \left(\begin{smallmatrix} x_2 \\ x_1 - x_2 \end{smallmatrix}\right)$, and then
$$C[v]_B = \begin{pmatrix} 1 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_1 - x_2 \end{pmatrix} = x_2 + (x_1 - x_2) = x_1.$$
So we do really get the same linear map. Going from $\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ to $\left(\begin{smallmatrix} x_2 \\ x_1 - x_2 \end{smallmatrix}\right)$ and from A to C is called **base change**, and we will see it in more detail next.
- Consider the map
$$T\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} x_1 + x_2 \\ -2x_1 + 4x_2 \end{smallmatrix}\right).$$

With respect to the standard basis, this has matrix

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

How do we work out the matrix with respect to the basis

$$B = \{v_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)\}?$$

We need to find

$$[T]_B = \left(\left[T\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)\right]_B \quad \left[T\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}\right)\right]_B \right).$$

We work out

$$T(v_1) = \left(\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right) = 2v_1 + 0v_2$$
$$T(v_2) = \left(\begin{smallmatrix} 3 \\ 6 \end{smallmatrix}\right) = 0v_1 + 3v_2$$

so the matrix is

$$C = [T]_B = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix}$$

This happens to be a very nice matrix, and it tells us something about the map that we could not see from the first matrix: it scales any multiple of v_1 by 2, and any multiple of v_2 by 3. We will see this in more detail in the next chapter about eigenvalues and eigenvectors.

◇ Consider the map

$$T\left(\left(\begin{smallmatrix} x_1 \\ x_2 \\ x_3 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} x_1+x_2 \\ x_1-x_3 \end{smallmatrix}\right).$$

This is a map $T\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$. With respect to the standard bases for both $\mathbb{R}^3$ and $\mathbb{R}^2$, this has matrix

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

Let's work out the matrix wrt basis $B = \left\{ v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \right\}$ for $\mathbb{R}^3$ and basis $C = \left\{ w_1 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, w_2 = \begin{pmatrix} 3 \\ 1 \end{pmatrix} \right\}$ for $\mathbb{R}^2$.

We want to find

$${}_C[T]_B = \left(\left[T\left(\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)\right)\right]_B \quad \left[T\left(\left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right)\right)\right]_B \quad \left[T\left(\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)\right)\right]_B \right).$$

So we first work out

$$T\left(\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right) \qquad T\left(\left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right) \qquad T\left(\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right).$$

Now we need to find the coordinate vectors of these with respect to basis C. We can solve 3 inhomogeneous systems at once:

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

So this tells us that

$$\begin{pmatrix}2\\0\end{pmatrix} = -\frac{2}{5}\begin{pmatrix}1\\2\end{pmatrix} + \frac{4}{5}\begin{pmatrix}3\\1\end{pmatrix} \qquad \text{so} \qquad \left[\begin{pmatrix}2\\0\end{pmatrix}\right]_C = \begin{pmatrix}-\frac{2}{5}\\ \frac{4}{5}\end{pmatrix}$$

$$\begin{pmatrix}2\\1\end{pmatrix} = \frac{1}{5}\begin{pmatrix}1\\2\end{pmatrix} + \frac{3}{5}\begin{pmatrix}3\\1\end{pmatrix} \qquad \text{so} \qquad \left[\begin{pmatrix}2\\1\end{pmatrix}\right]_C = \begin{pmatrix}\frac{1}{5}\\ \frac{3}{5}\end{pmatrix}$$

$$\begin{pmatrix}1\\1\end{pmatrix} = \frac{2}{5}\begin{pmatrix}1\\2\end{pmatrix} + \frac{1}{5}\begin{pmatrix}3\\1\end{pmatrix} \qquad \text{so} \qquad \left[\begin{pmatrix}1\\1\end{pmatrix}\right]_C = \begin{pmatrix}\frac{2}{5}\\ \frac{1}{5}\end{pmatrix}$$

So we put it together into a matrix:

$${}_C[T]_B = \begin{pmatrix}-\frac{2}{5} & \frac{1}{5} & \frac{2}{5}\\ \frac{4}{5} & \frac{3}{5} & \frac{1}{5}\end{pmatrix}$$

How do we check if the answer is correct? We cannot check it as easily as when we use the standard basis. The check happens in the three lines where we can work out if the vectors are really those given linear combinations of the basis C.

Example 8.46: Not lectured: this is an extra worked example for you, but we don't have time to go through everything in the lectures.

To see why we want to use the same basis on both sides when looking at maps $T\colon V \longrightarrow V$, consider the dilation map $T(v) = 3v$ on $\mathbb{R}^3$. The matrix for T in the standard basis is

$$A = \begin{pmatrix}3 & 0 & 0\\ 0 & 3 & 0\\ 0 & 0 & 3\end{pmatrix} = 3I,$$

which visibly looks like "multiplying by 3". If we take basis $B = \left\{v_1 = \left(\begin{smallmatrix}1\\1\\1\end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix}1\\1\\0\end{smallmatrix}\right), v_3 = \left(\begin{smallmatrix}1\\0\\0\end{smallmatrix}\right)\right\}$ for both sides, then T has matrix

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

again, because $T(v_1) = 3v_1$, etc. But if we look at the matrix for T using B for the domain (source) and the standard matrix for the codomain (target), then T has the matrix

$$C = \begin{pmatrix}3 & 3 & 3\\ 3 & 3 & 0\\ 3 & 0 & 0\end{pmatrix}$$

because $T(v_1) = \left(\begin{smallmatrix}3\\3\\3\end{smallmatrix}\right)$, $T(v_2) = \left(\begin{smallmatrix}3\\3\\0\end{smallmatrix}\right)$ and $T(v_3) = \left(\begin{smallmatrix}3\\0\\0\end{smallmatrix}\right)$. This matrix is not as useful, because we do not see the very important property that "T is just multiplyling by 3".

We can even make T have matrix

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

if we take the standard basis for the domain and the basis consisting of 3 times the standard vectors for the codomain. That does not tell us much about what the map does, other than that it is an isomorphism.

This is the reason that when we have a map $T\colon V \longrightarrow V$ from one vector space to itself, we (almost always) want to use the *same* basis on both sides.

Proposition 8.47: (Matrix of composite is product of matrices.)
If $T\colon U \longrightarrow V$ and $S\colon V \longrightarrow W$ are two linear maps, and B_1, B_2, B_3 are bases for U, V, W respectively, then
$$ {}_{B_3}[S{\circ}T]_{B_1} = {}_{B_3}[S]_{B_2}\ {}_{B_2}[T]_{B_1}. $$

In words: composition of maps corresponds to multiplying matrices. (We've chosen the placement of the bases as subscripts so that this composition comes out looking visually helpful.)

PROOF. We apply the definition of the matrix representing a linear map. Given as exercise in Lectures. You should either try to work it out yourself, to practice if you understand the definition (best learning opportunity), or follow the proof below, making sure you understand it (second best).
Suppose $B_1 = \{u_1, \ldots, u_l\}$, $B_2 = \{v_1, \ldots, v_m\}$ and $B_3 = \{w_1, \ldots, w_n\}$, and write $D = {}_{B_3}[S{\circ}T]_{B_1}$, $C = {}_{B_3}[S]_{B_2}$ and $A = {}_{B_2}[T]_{B_1}$. So we want to show $D = CA$. I'm deliberately trying to avoid having a matrix B, so as not to confuse it with the bases.
The kth column of the matrix $D = {}_{B_3}[S{\circ}T]_{B_1}$ is $d_k = [S(T(u_k))]_{B_3}$. In words: the image under $S{\circ}T$ of the kth basis vector of U, written in the basis of W. We know by how matrix multiplication works that $d_k = Ca_k$ C times the kth column of A (c.f. Definition of matrix multiplication, Chapter 1).
Using how linear maps work as matrix transformations (Lemma 8.42), we know that $[T(u_k)]_{B_2} = a_k$ in words, the coordinate vector of the image of the kth basis vector of U is the kth column of matrix A. And further, $Ca_k = {}_{B_3}[S]_{B_2}[T(u_k)]_{B_2} = [S(T(u_k))]_{B_3}$. So the image of u_k under the composite $S{\circ}T$ is indeed the kth column of the product CA, which means by definition $D = CA$. □

Proposition 8.48: (Isos have invertible matrices.)
Let $T\colon V \longrightarrow W$ be linear and B a basis of V, C a basis for W. Then T is an isomorphism if and only if the matrix for T with respect to B and C is invertible. In this case,
$$ {}_B[T^{-1}]_C = ({}_C[T]_B)^{-1}. $$

Given as exercise.

EXPLANATORY PROOF: We will fix some notation first. Let $B = \{v_1, v_2, \ldots, v_n\}$ and $C = \{w_1, w_2, \ldots, w_n\}$ be the bases (recall iso if and only if same dimension), and let $v \in V$ be $v = \lambda_1 v_1 + \cdots + \lambda_n v_n$ in terms of the basis vectors. So the coordinate vector for v is $[v]_B = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$.
Let $A = {}_C[T]_B$ be the matrix for T, and have columns $A = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \end{pmatrix}$.
The main idea of the proof is this: we view v as the vector $\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$ and T as multiplying by the matrix A. Then to "undo T", we have to multiply by the matrix A^{-1}.
Suppose T is an isomorphism, then it has an inverse $T^{-1}\colon W \longrightarrow V$, with $T{\circ}T^{-1} = \mathrm{id}_W$ and $T^{-1}{\circ}T = \mathrm{id}_V$. Let the inverse have matrix $D = {}_B[T^{-1}]_C$. Since the matrix of a composite is the product of the matrices (Prop. 8.47), and the identity map is represented by the identity matrix with respect to any basis, we have $DA = AD = I_n$. So $D = A^{-1}$, and A is invertible.
Conversely, suppose A is invertible. Then for any column vector $\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix}$, we can calculate the column vector $A^{-1}\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$. Then, if $w = \mu_1 w_1 + \cdots + \mu_n w_n$ in W, define $S\colon W \longrightarrow V$ as

$$ S(w) = \lambda_1 v_1 + \cdots + \lambda_n v_n, \qquad \text{i.e. } [S(w)]_B = A^{-1}[w]_C = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}. $$

We want to show that S is the inverse of T. To work out $S(T(v))$ and $T(S(w))$, we first work out what coordinate vectors they correspond to. We know

$$A^{-1}\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}=\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}, \text{ so}$$

$$A\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}=\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}.$$

So to work out $S(T(v))$, we have to find the coordinate vector of $T(v)$, apply A^{-1} to it, which gives us the coordinate vector of $S(T(v))$.
We have

$$[S(T(v))]_B = A^{-1}[T(v)]_C = A^{-1}A[v]_B = [v]_B.$$

Writing out the full column vectors:

$$[S(T(v))]_B = A^{-1}\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix} = A^{-1}A\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}=\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix}.$$

And conversely

$$[T(S(w))]_C = A[S(w)]_B = AA^{-1}[w]_C = [w]_C.$$

Writing out the full column vectors:

$$[T(S(w))]_C = A\begin{pmatrix}\lambda_1\\ \vdots\\ \lambda_n\end{pmatrix} = AA^{-1}\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}=\begin{pmatrix}\mu_1\\ \vdots\\ \mu_n\end{pmatrix}.$$

So if $v=\lambda_1v_1+\cdots+\lambda_nv_n$, then $S(T(v))=\lambda_1v_1+\cdots+\lambda_nv_n=v$, and if $w=\mu_1w_1+\cdots+\mu_nw_n$, then $T(S(w))=\mu_1v_1+\cdots+\mu_nv_n=w$. So $S=T^{-1}$ and T is an isomorphism. □

Higher-level proof: This is just turning the above proof into short-hand. It's really the same proof.
Recall the isomorphism $E_1\colon V\longrightarrow\mathbb{R}^n$ from "dimension n means iso to $\mathbb{R}^n$", Theorem 8.38. So we also have $E_2\colon W\longrightarrow\mathbb{R}^n$. If A is the matrix for T with respect to bases B and C and we write A also for the matrix transformation $A\colon\mathbb{R}^n\longrightarrow\mathbb{R}^n$, then $T=E_2^{-1}AE_1$. (To calculate T, turn a vector into its coordinate vector, multiply by A, and turn the resulting coordinate vector into an element of W.) This also gives $A=E_2TE_1^{-1}$.

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "E_1"'] & W \arrow[d, "E_2"] \\ \mathbb{R}^n \arrow[r, "A"'] & \mathbb{R}^n \end{tikzcd}$$

If T is an isomorphism, then T^{-1} also has a matrix C representing it, and we have $CA=E_1T^{-1}TE_1^{-1}=E_1\mathrm{id}_VE_1^{-1}=I_n$ and $AC=E_2TT^{-1}E_2^{-1}=E_2\mathrm{id}_WE_2^{-1}=I_n$, so $C=A^{-1}$.
Conversely, if A is invertible, then the inverse of T is $S=E_1^{-1}A^{-1}E_2$. (To work out S of some vector, turn the vector into a coordinate vector, multiply it by A^{-1}, and turn it into an element of V.)
We have

$$ST=E_1^{-1}A^{-1}E_2E_2^{-1}AE_1=E_1^{-1}A^{-1}AE_1=E_1^{-1}I_nE_1=\mathrm{id}_V,$$

and

$$TS=E_2^{-1}AE_1E_1^{-1}A^{-1}E_2=E_2^{-1}AA^{-1}E_2=E_2^{-1}I_nE_2=\mathrm{id}_W,$$

so indeed $S=T^{-1}$.

We can summarise the whole situation like this:

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, "E_1"'] \arrow[rr, bend left, "\mathrm{id}_V"] & W \arrow[r, "T^{-1}"] \arrow[d, "E_2"] & V \arrow[d, "E_1"] \\ \mathbb{R}^n \arrow[r, "A"'] \arrow[rr, bend right, "I_n"'] & \mathbb{R}^n \arrow[r, "A^{-1}"'] & \mathbb{R}^n \end{tikzcd} \quad \text{and} \quad \begin{tikzcd} W \arrow[r, "T^{-1}"] \arrow[d, "E_2"'] \arrow[rr, bend left, "\mathrm{id}_W"] & V \arrow[r, "T"] \arrow[d, "E_1"] & W \arrow[d, "E_2"] \\ \mathbb{R}^n \arrow[r, "A^{-1}"'] \arrow[rr, bend right, "I_n"'] & \mathbb{R}^n \arrow[r, "A"'] & \mathbb{R}^n \end{tikzcd}$$

□

Study guide.

Concept review

- ◇ Linear maps as matrix transformations.
- ◇ When to use different bases for domain and codomain, and when to use the same basis.
- ◇ Identity and dilation have the same matrix with respect to any basis.
- ◇ Matrix of composite map and matrix of inverse map.

Skills

- ◇ Find the matrix for a given linear map with respect to given bases.
- ◇ Find the coordinate vector of the image of some vector under a linear map.

F. Base change

We will now look at how to switch between coordinate vectors with respect to different bases:

Definition 8.49: Let V be a finite dimensional vector space, and B_1 and B_2 be bases of V. Then the matrix for the identity map $\mathrm{id}\colon V \longrightarrow V$ with respect to B_1 for the domain and B_2 for the codomain,

$$P = P_{B_1 \longrightarrow B_2} = {}_{B_2}[\mathrm{id}]_{B_1}$$

is called the **base change matrix** from basis B_1 to basis B_2.

Note that any base change matrix is automatically an invertible matrix, because the identity map id is an isomorphism (c.f. "Isos have invertible matrices", Proposition 8.48).
We define base change matrices this way because they satisfy:

Proposition 8.50: (Base change)
Let V be a finite dimensional vector space, and let B_1 and B_2 be bases of V. Then for any $v \in V$,

$$[v]_{B_2} = P_{B_1 \longrightarrow B_2}[v]_{B_1}.$$

In words: to change the coordinate vector for v with respect to basis B_1 into the coordinate vector of v with respect to basis B_2, we multiply it by the base change matrix.

Proof. Because of how the linear map $\mathrm{id}\colon V \longrightarrow V$ is viewed as a matrix transformation (c.f. Lemma 8.42), we have

$$[\mathrm{id}(v)]_{B_2} = {}_{B_2}[\mathrm{id}]_{B_1}[v]_{B_1},$$

which exactly says

$$[v]_{B_2} = P_{B_1 \longrightarrow B_2}[v]_{B_1}.$$

□

Example 8.51: Consider the standard basis $E = \{e_1 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right), e_2 = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)\}$ for $\mathbb{R}^2$, and the basis $B = \{v_1 = \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)\}$. Then the base change matrix $P_{B\longrightarrow E} = P$ has as first column the first basis vector of basis B, written in the basis E. So

$$P_{B\longrightarrow E} = \begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}$$

We saw earlier that if $[v]_E = \left(\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right)$, then $[v]_B = \left(\begin{smallmatrix}x_2\\x_1-x_2\end{smallmatrix}\right)$. Then

$$P[v]_B = \begin{pmatrix}1 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}x_2\\ x_1-x_2\end{pmatrix} = \begin{pmatrix}x_2+(x_1-x_2)\\ x_2\end{pmatrix} = \begin{pmatrix}x_1\\ x_2\end{pmatrix} = [v]_E.$$

Slogan: "Base change matrix from any 'other' basis *to* standard basis is easy to write down: the columns are just the 'other' basis vectors."

To work out the base change matrix the other way round, i.e. $P_{E\longrightarrow B}$, we have to write e_1 and e_2 in terms of v_1 and v_2:

$$e_1 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right) = v_2 \qquad \text{so} \qquad [e_1]_B = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$$
$$e_2 = \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right) = v_1 - v_2 \qquad \text{so} \qquad [e_2]_B = \left(\begin{smallmatrix}1\\-1\end{smallmatrix}\right)$$

and so the base change matrix is

$$P_{E\longrightarrow B} = \begin{pmatrix}0 & 1\\ 1 & -1\end{pmatrix},$$

and indeed

$$P_{E\longrightarrow B}[v]_E = \begin{pmatrix}0 & 1\\ 1 & -1\end{pmatrix}\begin{pmatrix}x_1\\ x_2\end{pmatrix} = \begin{pmatrix}x_2\\ x_1-x_2\end{pmatrix} = [v]_B.$$

Notice that $P_{E\longrightarrow B}P_{B\longrightarrow E} = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right) = I_2$ and $P_{B\longrightarrow E}P_{E\longrightarrow B} = I_2$, so $P_{E\longrightarrow B} = P^{-1}_{B\longrightarrow E}$. This is always the case.

Lemma 8.52: (Backwards base change has inverse base change matrix.)
Let V be a finite dimensional vector space, and let B_1 and B_2 be bases of V. Then

$$P_{B_2\longrightarrow B_1} = P^{-1}_{B_1\longrightarrow B_2}.$$

Proof. Let $\dim V = n$. Since the matrix of a composite is the product of the matrices (Prop. 8.47), we know that

$$P_{B_2\longrightarrow B_1}P_{B_1\longrightarrow B_2} = {}_{B_2}[\mathrm{id}]_{B_2}\ {}_{B_2}[\mathrm{id}]_{B_1} = {}_{B_1}[\mathrm{id}\circ\mathrm{id}]_{B_1} = I_n$$

and similarly

$$P_{B_1\longrightarrow B_2}P_{B_2\longrightarrow B_1} = I_n,$$

so $P_{B_2\longrightarrow B_1} = P^{-1}_{B_1\longrightarrow B_2}$ as claimed. □

This helps us to work out a base change matrix for more complicated bases.

Example 8.53: Consider the basis $B = \{v_1 = \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right)\}$ for $\mathbb{R}^2$. Calculating the base change matrix $P_{E\longrightarrow B}$ in this case is not completely straight-forward: we cannot just *see* what e_1 and e_2 are in terms of these new basis vectors v_1 and v_2. But we saw earlier that $P_{B\longrightarrow E}$ is easy:

$$P_{B\longrightarrow E} = \begin{pmatrix}1 & 1\\ 1 & 2\end{pmatrix}.$$

And so we know that the base change matrix the other way round is the inverse of this matrix, which we know how to calculate. So

$$P_{E\longrightarrow B} = P^{-1}_{B\longrightarrow E} = \begin{pmatrix}2 & -1\\ -1 & 1\end{pmatrix},$$

and indeed

$$\begin{aligned} e_1 &= \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right) = 2\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) - \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right) = 2v_1 - v_2 \\ e_2 &= \left(\begin{smallmatrix}0\\1\end{smallmatrix}\right) = -\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) + \left(\begin{smallmatrix}1\\2\end{smallmatrix}\right) = -v_1 + v_2. \end{aligned}$$

Proposition 8.54: (Base change for matrices)
Let $T\colon V \longrightarrow W$ be a linear map between finite-dimensional vector spaces, and let B_1, B_2 be two bases for V, C_1, C_2 two bases for W. Then

$$_{C_1}[T]_{B_1} = P^{-1}_{C_1\longrightarrow C_2}\ {}_{C_2}[T]_{B_2} P_{B_1\longrightarrow B_2}.$$

PROOF.

$$\begin{tikzcd} \text{basis } B_1 & V \arrow[r, "{}_{C_1}[T]_{B_1}"] \arrow[d, "P_{B_1\longrightarrow B_2}"'] & W \arrow[d, "P_{C_1\longrightarrow C_2}"] & \text{basis } C_1 \\ \text{basis } B_2 & V \arrow[r, "{}_{C_2}[T]_{B_2}"'] & W & \text{basis } C_2 \end{tikzcd}$$

We know that for $v \in V$, we have

$$_{C_1}[T]_{B_1}[v]_{B_1} = [T(v)]_{C_1} \qquad \text{and} \qquad {}_{C_2}[T]_{B_2}[v]_{B_2} = [T(v)]_{C_2}.$$

And using the base change matrices, we have

$$[v]_{B_2} = P_{B_1\longrightarrow B_2}[v]_{B_1} \qquad \text{and} \qquad [T(v)]_{C_2} = P_{C_1\longrightarrow C_2}[T(v)]_{C_1}.$$

Putting these together, we get two expressions for $[T(v)]_{C_1}$: first we get

$$\begin{aligned} {}_{C_2}[T]_{B_2}[v]_{B_2} &= {}_{C_2}[T]_{B_2} P_{B_1\longrightarrow B_2}[v]_{B_1} \\ &= [T(v)]_{C_2} = P_{C_1\longrightarrow C_2}[T(v)]_{C_1} \end{aligned}$$

which implies

$$[T(v)]_{C_1} = P^{-1}_{C_1\longrightarrow C_2}\ {}_{C_2}[T]_{B_2} P_{B_1\longrightarrow B_2}[v]_{B_1}.$$

But by definition we also have

$$[T(v)]_{C_1} = {}_{C_1}[T]_{B_1}[v]_{B_1},$$

so ${}_{C_1}[T]_{B_1} = P^{-1}_{C_1\longrightarrow C_2}\ {}_{C_2}[T]_{B_2} P_{B_1\longrightarrow B_2}$ as required. □

You need to check carefully that the order of matrices matches what bases are being used: ${}_{C_1}[T]_{B_1} = P^{-1}_{C_1\longrightarrow C_2}\ {}_{C_2}[T]_{B_2} P_{B_1\longrightarrow B_2}$ can be applied to a coordinate vector given in basis B_1. Then $P_{B_1\longrightarrow B_2}$ can act on it: it takes a coordinate vector in basis B_1 and turns it into a coordinate vector with basis B_2. Then ${}_{C_2}[T]_{B_2}$ can act on that one: it takes a coordinate vector with basis B_2 and gives you the image of that vector, written in basis C_2. Then $P^{-1}_{C_1\longrightarrow C_2}$ turns that into a coordinate vector with basis C_1. And on the other side of the equation, ${}_{C_1}[T]_{B_1}$ just does it all in one step: takes a coordinate vector with basis B_1 and gives you the image of that vector written in basis C_1.
As we in general have different spaces for domain and codomain, the base change formula looks like

$$\boxed{A_1 = Q^{-1} A_2 P.}$$

with different base change matrices P and Q, in the different vector spaces.

Example 8.55: We worked out matrices for the projection $S\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ with $S\left(\left(\begin{smallmatrix}x_1\\x_2\end{smallmatrix}\right)\right) = x_1$:

$A = \begin{pmatrix}1 & 0\end{pmatrix}$ with respect to the standard basis for $\mathbb{R}^2$ and standard basis 1 for $\mathbb{R}$,

$C = \begin{pmatrix}1 & 1\end{pmatrix}$ with respect to basis $B = \{v_1 = \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)\}$ for $\mathbb{R}^2$ and standard basis for $\mathbb{R}$.

We worked out the base change matrix

$$P = P_{E\longrightarrow B} = \begin{pmatrix} 0 & 1 \\ 1 & -1 \end{pmatrix},$$

and indeed,

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

(here $Q = 1$ is a 1×1 identity matrix, as we are not changing the basis in $\mathbb{R}$).

Corollary 8.56: (Base change for square matrices)
If $T : V \longrightarrow V$ and B_1, B_2 are bases for V, then

$$[T]_{B_1} = P^{-1}_{B_1\longrightarrow B_2}[T]_{B_2}P_{B_1\longrightarrow B_2}.$$

PROOF. It follows immediately from the previous result, we are just using the same bases for both sides of the map. □

In nicer notation, if the base change matrix is $P = P_{B_1\longrightarrow B_2}$, then

$$\boxed{A_1 = P^{-1}A_2P.}$$

Here we have the *same* base change in both domain and codomain.
This relationship between matrices is very important, and so we give it a name:

Definition 8.57: Two square matrices A and B are **similar** if they represent the same linear map with respect to different bases, or equivalently if there exists an invertible P such that $A = P^{-1}BP$.

Example 8.58: We saw that the map

$$T\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} x_1+x_2 \\ -2x_1+4x_2 \end{smallmatrix}\right).$$

has matrices

$$A = \begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix} \qquad \text{with respect to the standard basis,}$$

$$C = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \qquad \text{with respect to basis } B = \left\{v_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)\right\}.$$

We worked out the base change matrix

$$P = P_{E\longrightarrow B} = \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} \qquad \text{with} \qquad P^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}$$

and indeed,

$$A = P^{-1}CP$$
$$\begin{pmatrix} 1 & 1 \\ -2 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}.$$

So A and C are similar matrices.

Similar matrices have several properties in common.

Proposition 8.59: *Similar matrices have the same determinant.*

PROOF. If $A = P^{-1}BP$ with P invertible, then
$$\det A = \det(P^{-1}BP) = \det P^{-1} \det B \det P = \det(P^{-1}P) \det B = \det I_n \det B = \det B. \qquad \square$$

Example 8.60: From before, we have

$$\det A = \begin{vmatrix} 1 & 1 \\ -2 & 4 \end{vmatrix} = 6$$
$$\det C = \begin{vmatrix} 2 & 0 \\ 0 & 3 \end{vmatrix} = 6.$$

Proposition 8.61: *Similar matrices have the same trace.*

PROOF. Exercise. Use $\operatorname{tr}(AB) = \operatorname{tr}(BA)$, which you should prove first (by looking at elements). $\square$

So as any matrix which represents a linear map has the same determinant, we can define

Definition 8.62: Let $T: V \longrightarrow V$ be a linear map from a finite dimensional vector space to itself. Then the **determinant of** T is the determinant of any matrix which represents T with respect to some basis B. The **trace of** T is the trace of any matrix which represents T.

By the previous results, these is a well-defined concepts, as all these matrices have the same determinant and trace. Note that this only works for a linear map from one vector space *to itself*, in the same way as determinant and trace are only defined for *square* matrices.

Study guide.

Concept review

- ◇ Base change matrix, inverse of base change matrix.
- ◇ Base change from "other" basis to standard basis.
- ◇ Base change for matrices (both square and not square).
- ◇ Relationship between different matrices for the same linear map.
- ◇ Similar matrices.
- ◇ Determinant and trace of a linear map.

Skills

- ◇ Find a base change matrix from one basis to another (either directly, or using an inverse matrix).
- ◇ Find the correct equation for two matrices for the same linear map and the corresponding base change matrices.
- ◇ Find the determinant and trace of a linear map.

G. Linear Maps: Study guide collation

Just putting together all the study guides from the different sections.

Concept review.

- ◇ Linear map.
- ◇ Domain and codomain.
- ◇ Composition of maps.
- ◇ Zero map, identity map.
- ◇ Linear map is determined by values on a basis.
- ◇ Build up a repertoire of and intuition for linear maps.

- ◇ Kernel and image of linear map.
- ◇ Properties of kernel and image.
- ◇ Rank and nullity of linear map.
- ◇ Relationship between rank and nullity of a linear map.
- ◇ Injective and surjective functions.
- ◇ Behaviour of injectivity and surjectivity under composition.
- ◇ Isomorphisms and inverses.
- ◇ Relationship between injectivity and surjectivity for maps from a vector space to itself.
- ◇ Uniqueness of inverses.
- ◇ Inverse of a composite.
- ◇ Conditions for vector spaces to be isomorphic.
- ◇ Linear maps as matrix transformations.
- ◇ When to use different bases for domain and codomain, and when to use the same basis.
- ◇ Identity and dilation have the same matrix with respect to any basis.
- ◇ Matrix of composite map and matrix of inverse map.
- ◇ Base change matrix, inverse of base change matrix.
- ◇ Base change from "other" basis to standard basis.
- ◇ Base change for matrices (both square and not square).
- ◇ Relationship between different matrices for the same linear map.
- ◇ Similar matrices.
- ◇ Determinant and trace of a linear map.

Skills.

- ◇ Determine whether a function is a linear map.
- ◇ Find a formula for a linear map given values on a basis.
- ◇ Find the kernel and image of a linear map.
- ◇ Find bases for kernel and image of a linear map.
- ◇ Find rank and nullity of a linear map.
- ◇ Determine whether a map is injective or surjective.
- ◇ Find inverses to isomorphisms.
- ◇ Find an isomorphism between vector spaces of the same dimension.
- ◇ Find the matrix for a given linear map with respect to given bases.
- ◇ Find the coordinate vector of the image of some vector under a linear map.
- ◇ Find a base change matrix from one basis to another (either directly, or using an inverse matrix).
- ◇ Find the correct equation for two matrices for the same linear map and the corresponding base change matrices.
- ◇ Find the determinant and trace of a linear map.

CHAPTER 9

Eigenvectors and Eigenvalues

A. Definitions

We have seen that linear maps can be represented as matrices with respect to different bases. We will now look at finding a very nice basis for a linear map. We will start with a definition. In this whole chapter, we will only consider linear maps from a space to itself, corresponding to square matrices.

Definition 9.1: If $T\colon V \longrightarrow V$ is a linear map, then $v \neq 0 \in V$ is called an **eigenvector** with **eigenvalue** λ if $T(v) = \lambda v$.

So an eigenvector is a vector whose "direction" is not changed by the linear map. As T is linear, if v is an eigenvector, then any multiple of v is also an eigenvector with the same eigenvalue: $T(\mu v) = \mu T(v) = \mu\lambda v = \lambda(\mu v)$.

CAREFUL: We do not count $v = 0$ as an eigenvector, because $T(0) = 0 = \lambda 0$ for any λ. So it is important to remember that eigenvectors have to be **non-zero**.

Examples 9.2: a) For any scalar λ, the map $T(v) = \lambda v$ has any vector as an eigenvector, with eigenvalue λ.

b) We saw earlier the map

$$T\left(\left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} x_1+x_2 \\ -2x_1+4x_2 \end{smallmatrix}\right).$$

This satisfies

$$T\left(\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} 2 \\ 2 \end{smallmatrix}\right) = 2\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$$
$$T\left(\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)\right) = \left(\begin{smallmatrix} 3 \\ 6 \end{smallmatrix}\right) = 3\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$$

so we have eigenvector $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ with eigenvalue 2 and eigenvector $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ with eigenvalue 3.

c) [**Kernel means 0 as eigenvalue**] For any map $T\colon V \longrightarrow V$, if $v \in \mathrm{Ker}(T)$, then $T(v) = 0 = 0v$, so any non-zero vector in the kernel of T is an eigenvector with eigenvalue 0. This means that 0 is an eigenvalue of T if and only if T has non-trivial kernel, if and only if T is not injective.

d) Consider the map

$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} -2x_3 \\ x_1+2x_2+x_3 \\ x_1+3x_3 \end{pmatrix}$$

which has matrix

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

and let

$$v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}.$$

We work out

$$\begin{aligned} T(v_1) &= \begin{pmatrix} -2 \\ 0 \\ 2 \end{pmatrix} = 2v_1 \\ T(v_2) &= \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} = 2v_2 \\ T(v_3) &= \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} = v_3 \end{aligned}$$

so T has eigenvectors v_1 and v_2 with eigenvalue 2, and v_3 with eigenvalue 1. Note that any linear combination of v_1 and v_2 is also an eigenvector with eigenvalue 2:

$$T(av_1 + bv_2) = aT(v_1) + bT(v_2) = 2av_1 + 2bv_2 = 2(av_1 + bv_2).$$

This leads us to a general result:

Proposition 9.3: (Eigenspace)
Given a linear map $T: V \longrightarrow V$ with some eigenvalue λ, then the set of eigenvectors of λ, together with the 0 vector, form a subspace V_λ of V.

PROOF. Exercise: we have seen all the ingredients we need in the examples above. □

Definition 9.4: This subspace V_λ is called an **eigenspace** of T.

Careful! We have a separate eigenspace per eigenvalue, not one eigenspace for everything together! This is very important!

Examples 9.5: In the same examples as above, we have:

a) $T: V \longrightarrow V$ with $T(v) = \lambda v$ has $V_\lambda = V$: the whole space V is one eigenspace.

b) The map

$$T\left(\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}\right) = \begin{pmatrix} x_1+x_2 \\ -2x_1+4x_2 \end{pmatrix}$$

has eigenspaces

$$\begin{aligned} V_2 &= \langle \begin{pmatrix} 1 \\ 1 \end{pmatrix} \rangle = \text{Span}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ V_3 &= \langle \begin{pmatrix} 1 \\ 2 \end{pmatrix} \rangle = \text{Span}\begin{pmatrix} 1 \\ 2 \end{pmatrix}. \end{aligned}$$

c) [**Kernel is 0-eigenspace**] For any map $T: V \longrightarrow V$, we have $V_0 = \text{Ker}(T)$.

d) The map

$$T\left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}\right) = \begin{pmatrix} -2x_3 \\ x_1+2x_2+x_3 \\ x_1+3x_3 \end{pmatrix}$$

with

$$v_1 = \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, v_3 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}$$

has eigenspaces

$$\begin{aligned} V_2 &= \langle v_1, v_2 \rangle = \text{Span}(v_1, v_2) \\ V_1 &= \langle v_3 \rangle. \end{aligned}$$

Corollary 9.6: (Isos via eigenvalues) *$T: V \longrightarrow V$ is an isomorphism if and only if 0 is not an eigenvalue of T.*

PROOF. $V_0 = \text{Ker}(T)$, so $\text{Ker}(T) = 0$ if and only if 0 is not an eigenvalue. □

Study guide.

Concept review

- ◇ Eigenvalues, eigenvector.
- ◇ Eigenspace.
- ◇ Connection of kernel with eigenvalues/eigenspaces.
- ◇ Connection of isomorphisms and eigenvalues.

Skills

- ◇ (next section)

B. Finding eigenvalues: characteristic polynomial

So how do we find eigenvalues and eigenvectors? Recall

Lemma 9.7: (Determinant finds isos.)
If V is finite dimensional, a linear map $T: V \longrightarrow V$ is an isomorphism if and only if $\det(T) \neq 0$. Equivalently, if A is a matrix for T with respect to some basis, then T is an isomorphism if and only if $\det(A) \neq 0$.

Proof. Recall that $\det(T)$ is defined as $\det(A)$ for any matrix A representing T, so the above two statements really are the same. This is Theorem 3.55. □

Corollary 9.8: *The same T as above has $\operatorname{Ker}(T) = 0$ if and only if $\det(T) \neq 0$.*

Proof. By "check one get one free for isos" (Proposition 6.13), we know that T is injective if and only if it is surjective. And by "injectivity via kernels" (Proposition 8.24) we know that $\operatorname{Ker}(T) = 0$ if and only if T is injective. □

So now we can prove:

Theorem 9.9: (Eigenvalues are the roots of the characteristic polynomial.)
If $T: V \longrightarrow V$ is a linear map, then λ is an eigenvalue of T if and only if $\det(\lambda\mathrm{id} - T) = 0$. Equivalently, if A is a matrix for T with respect to some basis, then λ is an eigenvalue of T if and only if $\det(\lambda I - A) = 0$.

Proof. We have

$$\begin{array}{lll} & \exists v \neq 0 \text{ such that} & T(v) = \lambda v \\ \Leftrightarrow & \exists v \neq 0 \text{ such that} & \lambda v - T(v) = 0 \\ \Leftrightarrow & \exists v \neq 0 \text{ such that} & (\lambda\mathrm{id} - T)(v) = 0 \\ \Leftrightarrow & \exists v \neq 0 \text{ such that} & v \in \operatorname{Ker}(\lambda\mathrm{id} - T) \\ \Leftrightarrow & & \operatorname{Ker}(\lambda\mathrm{id} - T) \neq 0 \\ \Leftrightarrow & & \det(\lambda\mathrm{id} - T) = 0 \end{array}$$

To justify the second assertion, we have to show that it does not matter *which* matrix representation we choose. Suppose A and B are two matrices for T with respect to different bases. Then there is a base change matrix P (which is invertible) so that $A = P^{-1}BP$. Then

$$\begin{aligned} \det(\lambda I - A) &= \det(\lambda I - P^{-1}BP) \\ &= \det(P^{-1}(\lambda I - B)P) \\ &= \det P^{-1} \det(\lambda I - B) \det P \\ &= \det(P^{-1}P) \det(\lambda I - B) \\ &= \det(\lambda I - B). \end{aligned}$$

So this quantity is not dependent on which matrix representation of T we choose.

Really what is behind this is that determinant does not change under base change, and the identity map has the identity matrix with respect to *any* basis, so if we change the basis for T to go from A to B, we might as well change the basis for id as well, and we still get the identity matrix. $\square$

Definition 9.10: Given $T\colon V \longrightarrow V$, the **characteristic polynomial of** T is $\chi_T(t) = \det(t\mathrm{id} - T)$, or if A is a matrix representing T, then $\chi_T(t) = \chi_A(t) = \det(tI - A)$.

So the theorem above shows us that λ is an eigenvalue of T if and only if λ is a root of the characteristic polynomial of T. We also saw above that it does not matter which matrix representation of T we take.
So

Fact 9.11: The eigenvalues are *exactly* the roots of the characteristic polynomial.

Why is $\det(tI - A)$ a polynomial?

Example 9.12: If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is 2×2, then

$$\det(tI - A) = \begin{vmatrix} t-a & -b \\ -c & t-d \end{vmatrix} = (t-a)(t-d) - bc = t^2 - (a+d)t + ad - bc$$

is a polynomial of degree 2.
As we defined determinants "by induction" as linear combinations of determinants of smaller matrices, this inductively shows that $\det(tI - A)$ is a polynomial in t of degree n, where A is an $n \times n$ matrix.
So any $n \times n$ matrix has at most n eigenvalues.

Here are a few more examples for you of working out eigenvalues:

Examples 9.13: a) not lectured, extra worked example
If

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

then

$$\chi_A(t) = \det(tI - A) = \begin{vmatrix} t-1 & -1 \\ +2 & t-4 \end{vmatrix}$$
$$= (t-1)(t-4) + 2 = t^2 - 5t + 6 = (t-2)(t-3)$$

so the eigenvalues of A are 2 and 3, as we saw above.

b) Consider

$$R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

If we first think of this geometrically: if we view it as a map $\mathbb{R}^2 \longrightarrow \mathbb{R}^2$, then the first standard basis vector $e_1 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ is mapped to $\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) = e_2$, and e_2 is mapped to $-e_1$. So this is a rotation in $\mathbb{R}^2$ by 90° anti-clockwise. We should expect it to have no eigenvalues.
We work out the characteristic polynomial:

$$\chi_R(t) = \det(tI - R) = \begin{vmatrix} t & 1 \\ -1 & t \end{vmatrix} = t^2 + 1$$

This has no real roots, so indeed it has no real eigenvalues. (We will talk about complex eigenvalues later.

Warning: Don't forget to change the signs of the entries in the matrix when you write down the entries of $\det(tI - A)$.

For some matrices, we can just read off the eigenvalues from the diagonal.

Proposition 9.14: (Eigenvalues of diagonal and triangular matrices)
If A is

(i) a diagonal matrix, or
(ii) an upper triangular matrix, or
(iii) a lower triangular matrix,

then the eigenvalues of A are exactly the diagonal entries.

Proof. By Proposition 3.49, the determinant of an upper or lower triangular matrix is the product of diagonal entries. So if A has diagonal entries a_{ii}, then $tI - A$ has diagonal entries $(t - a_{ii})$, so $\chi_A(t) = \det(tI - A) = (t - a_{11})(t - a_{22}) \cdots (t - a_{nn})$, which has roots the diagonal entries of A. This applies to any of the three cases. □

Proposition 9.15: *Similar matrices have the same eigenvalues and characteristic polynomial.*

Proof. If A and B are similar, there exists an invertible P such that $A = P^{-1}BP$, so

$$\begin{aligned}
\chi_A(t) &= \det(tI - A) \\
&= \det(tI - P^{-1}BP) \\
&= \det(P^{-1}(tI - B)P) \\
&= \det P^{-1} \det(tI - B) \det P \\
&= \det(P^{-1}P) \det(tI - B) \\
&= \det I \det(tI - B) \\
&= \chi_B(t).
\end{aligned}$$

So they have the same characteristic polynomial. Since the eigenvalues are exactly the roots of the characteristic polynomial, A and B also have the same eigenvalues.
You can also think of it this way: A and B are similar if and only if they represent the same linear map, and characteristic polynomial and eigenvalues are properties of linear maps, so they don't change when we look at the same linear map. □

So now we want to know how to find eigenvectors. Using a similar argument as for eigenvalues, we see that $v \neq 0$ is an eigenvector for T with eigenvalue λ if and only if $(\lambda \mathrm{id} - T)(v) = 0$.

Lemma 9.16: (Eigenspace as kernel)
For a linear map $T\colon V \longrightarrow V$, the eigenspace $V_\lambda = \mathrm{Ker}(\lambda \mathrm{id} - T)$.

Proof. From the above, we see that $v \in V_\lambda$ if and only if $v \in \mathrm{Ker}(\lambda \mathrm{id} - T)$. □

So since we already know how to find elements of and/or bases for a kernel, we can now find elements of and/or bases for eigenspaces.
So, to summarise:

To **find eigenvalues and eigenvectors** of a matrix A, we

- ⋄ calculate the characteristic polynomial $\chi_A(t) = \det(tI - A)$;
- ⋄ factorise the characteristic polynomial to find its roots, which are the eigenvalues.
- ⋄ For each eigenvalue λ separately, we find a basis of $\mathrm{Ker}(\lambda I - A)$.

Examples 9.17: a) Consider

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

for which we worked out

$$\chi_A(t) = (t-2)(t-2)(t-1),$$

so the eigenvalues are 2 and 1.

Eigenvectors for $\lambda = 2$:

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

so a basis for V_2 is $v_1 = \left(\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right)$, $v_2 = \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right)$. This means that a general eigenvector for eigenvalue 2 is $\left(\begin{smallmatrix} s \\ t \\ -s \end{smallmatrix}\right)$.

Eigenvectors for $\lambda = 1$:

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

so a basis for V_1 is $v_3 = \left(\begin{smallmatrix} -2 \\ 1 \\ 1 \end{smallmatrix}\right)$.

b) Note that eigenvalues need not be integers - we just often have them as integers in exercises so that it's easier to enter in Numbas. In real life, in fact the matrices may not even have integer entries of course.

Let $A = \begin{pmatrix} 0 & 3 \\ 4 & 0 \end{pmatrix}$, then $\chi_A(t) = \begin{vmatrix} t & -3 \\ -4 & t \end{vmatrix} = t^2 - 12 = (t - 2\sqrt{3})(t + 2\sqrt{3})$, so the eigenvalues are $2\sqrt{3}$ and $-2\sqrt{3}$.

So we now work out the eigenvectors: for $\lambda = 2\sqrt{3}$, we find the kernel of

$$A - 2\sqrt{3}I = \begin{pmatrix} -2\sqrt{3} & 3 \\ 4 & -2\sqrt{3} \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -\frac{\sqrt{3}}{2} \\ 0 & 0 \end{pmatrix}$$

so the eigenvector is $\begin{pmatrix} \frac{\sqrt{3}}{2} \\ 1 \end{pmatrix}$, or if we want it without fractions, we can use $\begin{pmatrix} \sqrt{3} \\ 2 \end{pmatrix}$.

For $\lambda = -2\sqrt{3}$, we find the kernel of

$$A + 2\sqrt{3}I = \begin{pmatrix} 2\sqrt{3} & 3 \\ 4 & 2\sqrt{3} \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & \frac{\sqrt{3}}{2} \\ 0 & 0 \end{pmatrix}$$

so the eigenvector is $\begin{pmatrix} -\sqrt{3} \\ 2 \end{pmatrix}$.

Proposition 9.18: (Eigenvalues and eigenvectors of powers of a matrix)
If an $n \times n$ matrix A has eigenvector $v \neq 0$ with eigenvalue λ, then A^k has eigenvector v with eigenvalue λ^k, for any natural number k.

Proof. If $Av = \lambda v$, then $A^2 v = A(\lambda v) = \lambda^2 v$, so by induction $A^k v = A(\lambda^{k-1} v) = \lambda^k v$. □

Proposition 9.19: (Eigenvalues and eigenvectors of inverse)
If A is an invertible matrix, then A^{-1} has the same eigenvectors as A, with inverses of the corresponding eigenvalues.

PROOF. If $Av = \lambda v$ for $v \neq 0$, and A is invertible, then $\lambda \neq 0$ (c.f. isos via eigenvalues, Corollary 9.6). Then

$$\begin{aligned} && Av &= \lambda v \\ \Leftrightarrow && v &= \lambda A^{-1} v \\ \Leftrightarrow && \frac{1}{\lambda} v &= A^{-1} v. \end{aligned}$$ □

Remark 9.20: Notice the difference between taking powers, say squaring, and taking the inverse. Taking the inverse is a two-way operation, because $(A^{-1})^{-1} = A$. So, if A is invertible, then v is an eigenvector of A if and only if it is an eigenvector of A^{-1}.
But for squaring, it is only a one-directional statement: if v is an eigenvector of A, then v is also an eigenvector of A^2. But A^2 may have other eigenvectors, which A does not have.
For example: $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ has eigenvalues -1 and 1, with eigenvectors $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ for -1 and $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ for 1. So there are two separate one-dimensional eigenspaces for A.
But $A^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, which has eigenvalue 1, and every vector is an eigenvector. So for example $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ is an eigenvector for A^2, but it is not an eigenvector for A. The eigenspaces have been combined to be for the same eigenvalue, which gives a 2-dimensional eigenspace - this is the sum of the two previous eigenspaces (which is bigger than the union).
This corresponds to the fact that you can't undo squaring: you can't tell if the 1 came from $(-1)^2$ or $(+1)^2$.

Study guide.
Concept review

- ◇ Characteristic polynomial, and its link to eigenvalues.
- ◇ Eigenvalues of diagonal and triangular matrices.
- ◇ Relationship of eigenvalues of similar matrices.
- ◇ Eigenspace as kernel of a particular linear map.
- ◇ Eigenvalues and eigenvectors of powers or inverses of a map/matrix.

Skills

- ◇ Find the characteristic polynomial of a map/matrix.
- ◇ Find the eigenvalues of a map/matrix.
- ◇ Find the eigenvectors of a map/matrix.
- ◇ Find the eigenvalues of an inverse matrix.

C. Diagonalisation

We started by saying that we want to find a nice basis for a linear map. Eigenvectors help with that.

Proposition 9.21: (Basis of eigenvectors gives diagonal matrix.)
If $v_1, v_2, \ldots, v_n$ is a basis for V consisting of eigenvectors for $T \colon V \longrightarrow V$, then the matrix for T with respect to this basis is diagonal.

Proof. Suppose $T(v_k) = \lambda_k v_k$ (where the λ_k are not necessarily distinct). Then using the definition of a matrix representing a linear map, we get

$$A = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & \lambda_{n-1} & 0 \\ 0 & \cdots & 0 & 0 & \lambda_n \end{pmatrix}$$

which is diagonal. □

This is the nicest form we can expect from a linear map: it tells us exactly in which directions the linear map dilates (stretches/shrinks) vectors by exactly how much.
However, we can't always expect to get such a nice form. We've already had one example where there were no real eigenvalues. But that is not the only problem.

Example 9.22: Consider the map which has matrix

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

with respect to the standard basis. Then

$$\chi_A(t) = \begin{vmatrix} t-1 & 3 & -2 \\ 1 & t+5 & -6 \\ -2 & 2 & t \end{vmatrix} = t^2(t+4).$$

So 0 and -4 are eigenvalues. We work out eigenvectors, first for $\lambda = 0$:

$$\begin{pmatrix} 1 & -3 & 2 \\ -1 & -5 & 6 \\ 2 & -2 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -3 & 2 \\ 0 & -8 & 8 \\ 0 & 4 & -4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

(doing row operations to reduce to echelon form)

so a basis for V_0 is $\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)$.
For $\lambda = -4$: (swap first two rows for easier calculations)

$$A + 4I = \begin{pmatrix} 5 & -3 & 2 \\ -1 & -1 & 6 \\ 2 & -2 & 4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & -6 \\ 0 & -8 & 32 \\ 0 & -4 & 16 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 1 & -6 \\ 0 & 1 & -4 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & -4 \\ 0 & 0 & 0 \end{pmatrix}$$

so a basis for V_{-4} is $\left(\begin{smallmatrix} 2 \\ 4 \\ 1 \end{smallmatrix}\right)$.
We only have 2 linearly independent eigenvectors, so we cannot hope to make a basis for all of $\mathbb{R}^3$ out of these.

The aim now is to investigate this issue, and come up with some situations where we know we will be able to find a basis of eigenvectors.

Definition 9.23: Given an eigenvalue λ for T, the **algebraic multiplicity** of λ is the power of the linear factor $(t - \lambda)$ in the characteristic polynomial, and the **geometric multiplicity** of λ is the dimension of the eigenspace V_λ.

Example 9.24: For

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

eigenvalue 0 has algebraic multiplicity 2 and geometric multiplicity 1, and eigenvalue -4 has algebraic and geometric multiplicity 1.

We see immediately that

If λ is an eigenvalue, then
- ⋄ algebraic multiplicity $\geqslant 1$ and
- ⋄ geometric multiplicity $\geqslant 1$.

Otherwise λ would not even *be* an eigenvalue. We will see later that geometric multiplicity $\leqslant$ algebraic multiplicity.
We saw that similar matrices have the same eigenvalues and characteristic polynomials, so they have the same algebraic multiplicities. We can't expect them to have exactly the same eigenvectors, but they do have something in common about their eigenspaces.

Proposition 9.25: *Similar matrices have the same eigenspace dimensions, i.e. the same geometric multiplicities.*

PROOF. Let $A = P^{-1}BP$ for P invertible. We have already proved that A and B have the same eigenvalues. Let V_λ be the λ-eigenspace of A, and W_λ be the λ-eigenspace of B. Then

$$\begin{aligned} & Av = \lambda v \\ \Leftrightarrow \quad & P^{-1}BPv = \lambda v \\ \Leftrightarrow \quad & B(Pv) = \lambda(Pv). \end{aligned}$$

So v is an eigenvector for A with eigenvalue λ if and only if Pv is an eigenvector for B with eigenvalue λ. Given a basis $v_1, \ldots, v_k$ of V_λ, then $Pv_1, \ldots, Pv_k$ are all in W_λ, and as P is an invertible map, these vectors are still linearly independent. You can view this sentence as a "hidden exercise": is it clear to you why the conclusion of this sentence is true? If not, see if you can work through it.
So $\dim W_\lambda \geqslant \dim V_\lambda$. We can do the same with a basis $w_1, \ldots, w_l$ for W_λ and P^{-1}, giving $\dim V_\lambda \geqslant \dim W_\lambda$. So the dimensions are the same. □
What is really going on is that the restriction of P to V_λ gives an isomorphism $P' \colon V_\lambda \longrightarrow W_\lambda$.
If you're interested in making that more precise:
Consider the restriction of T_P to V_λ, i.e. $S \colon V_\lambda \longrightarrow V$ with $S(v) = Pv$, but we're only allowed to use v from V_λ, not all of V. We worked out above that then $\operatorname{Im} S \subseteq W_\lambda$, i.e. $S(v)$ lands in W_λ. Then you can work out that $\operatorname{Ker} S = \operatorname{Ker} P \cap V_\lambda$. This is always the case for any restriction. So $\operatorname{Ker} S = 0$, and S is an isomorphism from V_λ to its image. Then by rank-nullity, $\mathrm{r}(S) = \dim V_\lambda$, but we also know $\operatorname{Im} S \leqslant W_\lambda$, so $\dim V_\lambda \leqslant \dim W_\lambda$. Then you can either do it all again with A, B swapped, or you can show directly that all of W_λ is in the image of S by saying $w \in W_\lambda$ has $P^{-1}w \in V_\lambda$, so S is surjective onto W_λ.

Let's start with a definition of what we are aiming for.

Definition 9.26: A linear map $T \colon V \longrightarrow V$ is **diagonalisable** if there exists a basis for V such that the matrix for T with respect to this basis is a diagonal matrix.
An $n \times n$ matrix A is **diagonalisable** if it is similar to a diagonal matrix.

Recall that this means that there is some invertible P (a base change matrix) such that $A = P^{-1}DP$, where D is diagonal. Another way to put diagonalisability is

Theorem 9.27: *A linear map $T \colon V \longrightarrow V$ is diagonalisable if and only if there is a basis of V consisting of eigenvectors of T.*

PROOF. We already saw above (Prop. 9.21) that if we have a basis of eigenvectors then the corresponding matrix is diagonal. Conversely, if $v_1, \ldots, v_n$ is a basis which gives us a diagonal matrix with entries $\lambda_1, \dots, \lambda_k$ on the diagonal, then $T(v_k) = \lambda_k v_k$ by definition of a matrix representing a linear map, so each v_k is an eigenvector of T. □

Let us make a first step towards achieving this goal.

Proposition 9.28: (Evectors for distinct evalues)
For a linear map $T\colon V \longrightarrow V$, if $v_1, \dots, v_k$ are eigenvectors with distinct eigenvalues, then they are linearly independent.

Proof. We have $T(v_i) = \lambda_i v_i$, with all λ_i distinct. Suppose that $v_1, \dots, v_k$ are linearly dependent (so we are aiming for a contradiction). As the set of one vector v_1 is linearly independent (as $v_1 \neq 0$), there must be some largest m such that $v_1, \dots, v_m$ are linearly independent, but $v_1, \dots, v_m, v_{m+1}$ are linearly dependent. Because of our assumuption, $m+1 \leqslant k$. Then there exist scalars such that

$$\mu_1 v_1 + \cdots + \mu_m v_m + \mu_{m+1} v_{m+1} = 0.$$

Applying T to this vector gives

$$\mu_1 \lambda_1 v_1 + \cdots + \mu_m \lambda_m v_m + \mu_{m+1} \lambda_{m+1} v_{m+1} = 0.$$

If we now multiply the first equation by λ_{m+1} and subtract that from the second equation, we get

$$\mu_1(\lambda_1 - \lambda_{m+1})v_1 + \cdots + \mu_m(\lambda_m - \lambda_{m+1})v_m + 0v_{m+1} = 0.$$

But as all the λ_i are distinct, and $v_1, \dots, v_m$ are linearly independent, this implies that $\mu_1 = \cdots = \mu_m = 0$. Then we also get $\mu_{m+1} = 0$, which contradicts our assumption that $v_1, \dots, v_{m+1}$ are linearly dependent.
So in fact, $v_1, \dots, v_k$ are linearly independent. □
Summary: Suppose the vectors are linearly dependent, and look at the "switch": where does it go from lin indep set to lin dep set. Have a dependecy relation, and eliminate the last vector by combining "apply T" and "multiply by λ_{m+1}" and subtracting both. This gives contradiction.

Examples 9.29: a)

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

has eigenvalues 2 and 3, so the corresponding eigenvectors $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right)$ are linearly independent.

b)

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

has eigenvalues -1 with evector $\left(\begin{smallmatrix} 0 \\ 3 \\ -1 \end{smallmatrix}\right)$, 2 with evector $\left(\begin{smallmatrix} 1 \\ 0 \\ -1 \end{smallmatrix}\right)$ and 5 with evector $\left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)$. So the previous result tells us that the set of these three vectors is linearly independent. As it is a set of three vectors, this makes it a basis of $\mathbb{R}^3$ (by "check one get one free for bases", Prop. 6.13).

c)

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

has eigenvalue 2 with eigenvectors $v_1 = \left(\begin{smallmatrix} -1 \\ 0 \\ 1 \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right)$, and eigenvalue 1 with eigenvector $v_3 = \left(\begin{smallmatrix} -2 \\ 1 \\ 1 \end{smallmatrix}\right)$. So the previous result tells us that v_1, v_3 are linearly independent, and also v_2, v_3 are linearly independent. The result does not tell us anything about the eigenvectors for the same eigenvalue 2, but we chose them to be linearly independent.

Corollary 9.30: *If* $\dim V = n$ *and* $T\colon V \longrightarrow V$ *has* n *distinct eigenvalues, then* T *is diagonalisable.*

Examples 9.31: Examples a) and b) above have distinct eigenvalues, so they are diagonalisable. c) does not have distinct eigenvalues, so this result does not tell us whether it is diagonalisable or not.

But even when we have some repeated eigenvalues, we have a chance of diagonalisation, as we saw in an earlier example. When does it happen that we don't have enough linearly independent eigenvectors? Clearly then some of the eigenspaces are not big enough. But how big can they get?

Proposition 9.32: (geometric mult $\leqslant$ algebraic mult)
For a linear map $T\colon V \longrightarrow V$ *and an eigenvalue* λ *of* T*, the geometric multiplicity of* λ *is at most the algebraic multiplicity of* λ*.*

Proof. Let V_λ be the corresponding eigenspace, and let $v_1, \dots, v_k$ be a basis for V_λ. So k is the geometric multiplicity. Extend $v_1, \dots, v_k$ to a basis $v_1, \dots, v_k, v_{k+1}, \dots, v_n$ of V (possible by Theorem 6.15). Then let A be the matrix for T with respect to this particular basis. As the first k basis vectors are eigenvectors, the first k columns of A have only λ in the diagonal entry and the rest is 0. We don't know anything about the remaining columns of A.

$$A = \begin{pmatrix} \lambda & 0 & 0 & \cdots & 0 & * & \cdots & * \\ 0 & \lambda & 0 & \cdots & \vdots & * & \cdots & * \\ \vdots & & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & & & \ddots & 0 & \vdots & & \vdots \\ 0 & & \cdots & 0 & \lambda & * & \cdots & * \\ 0 & & \cdots & & 0 & * & \cdots & * \\ \vdots & & \cdots & & \vdots & * & \cdots & * \\ 0 & & \cdots & & 0 & * & \cdots & * \end{pmatrix} = \begin{pmatrix} \lambda I_k & * \\ 0 & * \end{pmatrix}$$

Through consecutively expanding in the first k columns, we see that

$$\chi_A(t) = (t-\lambda)^k p(t)$$

for some polynomial $p(t)$ of degree $n-k$. So λ has algebraic multiplicity of at least k. Since similar matrices have the same characteristic polynomials and hence the same algebraic multiplicities, this proves that

$$\text{geometric multiplicity of } \lambda \leqslant \text{ algebraic multiplicity of } \lambda. \qquad \square$$

Summary: "Do as best we can" by taking basis for eigenspace as part of the whole basis, then check what power $(t-\lambda)$ has in the characteristic polynomial for that matrix.

Example 9.33: You can check in all the examples we've had (also in the Workbook) that this is true.

So we see that to have a chance of getting a basis of eigenvectors, we would need the geometric multiplicities to be equal to the algebraic multiplicities, as the algebraic multiplicities add up to at most n.
We saw earlier that eigenvectors for distinct eigenvalues form a linearly independent set. Now we want to show that if we have a basis for each eigenspace and put them all together, then we still have a linearly independent set.

Proposition 9.34: (Combining eigenspace bases)
Let $T\colon V \longrightarrow V$ be a linear map. Let $B_1, \ldots, B_k$ be bases for distinct eigenspaces of T. Then the union of these forms a linearly independent set in V.

SKETCH OF PROOF: This proof is very similar to the proof of "evectors for distinct evalues", Prop. 9.28. We just have to make a few tweaks. So I'm going to just give you a sketch, you can fill in the details.
To fix some notation, let the ith eigenspace have eigenvalue λ_i, and let $B_i = \{v_1^i, v_2^i, \ldots, v_{l_i}^i\}$. So we put a superscript which tells us which basis the vector is in, and the subscript counts the vectors in a given basis, and B_i has l_i elements. Notice that no vector can appear in two bases at once, since all these vectors are eigenvectors and each eigenspace has a different eigenvalue.
Supppose the set of all vectors is linearly dependent. As before, this means there is some m such that the first m vectors on the list are linearly independent, but the first $m+1$ are linearly dependent. So we have some dependence relation between the vectors. If the last vector on the list is the first vector in a new eigenspace basis, we can use the same method as before to show that all coefficients must be 0. So let's assume the last vector on the list is somewhere in the middle of an eigenspace basis:

$$\mu_1 v_1^1 + \cdots + \mu_{m-j+1} v_1^i + \cdots + \mu_m v_j^i + \mu_{m+1} v_{j+1}^i = 0$$

Then applying T to this, and subtracting λ_i times the original equation, we get

$$\mu_1(\lambda_1 - \lambda_i)v_1^1 + \cdots + \mu_{m-j}(\lambda_{i-1} - \lambda_i) + \mu_{m-j+1}(\lambda_i - \lambda_i)v_1^i + \cdots + \mu_{m+1}(\lambda_i - \lambda_i)v_{j+1}^i = 0,$$

i.e. all the contributions from the same eigenspace as the very last vector disappear. But we know that the vectors that still have non-zero coefficients are linearly independent, so $\mu_1 = \cdots = \mu_{m-j} = 0$. This leaves us with

$$\mu_{m-j+1} v_1^i + \cdots + \mu_m v_j^i + \mu_{m+1} v_{j+1}^i = 0,$$

but now these vectors are all in the basis B_i, so they are also linearly independent, so all the remaining μs are also zero.
So the set of all bases of the distinct eigenspaces is linearly independent. $\square$

Example 9.35: Example c) we had before:

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

has eigenvalue 2 with eigenvectors $v_1 = \left(\begin{smallmatrix} -1 \\ 0 \\ 1 \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right)$, and eigenvalue 1 with eigenvector $v_3 = \left(\begin{smallmatrix} -2 \\ 1 \\ 1 \end{smallmatrix}\right)$. As we chose v_1 and v_2 to be a basis of the 2-eigenspace, this result tells us that v_1, v_2, v_3 is a linearly independent set (and so a basis for $\mathbb{R}^3$, by "check one get one free for bases", Prop. 6.13.)

Theorem 9.36: (Diagonalisability via geometric multiplicities)
A linear map $T\colon V \longrightarrow V$ with $\dim V = n$ is diagonalisable if and only if the sum of all geometric multiplicities of the distinct eigenvalues is n.

PROOF. If T is diagonalisable, we have a basis of eigenvectors. Each eigenvector is in some eigenspace basis, so the sum of geometric multiplicities of distinct eigenvalues is at least n. But since geometric multiplicities are at most the algebraic multiplicities, the sum of geometric multiplicities is also at most n, so it is equal to n.
Conversely, if the sum of geometric multiplicities is n, then the union of all the bases of all the distinct eigenspaces has size n, so (by Prop. 9.34) is a linearly independent set of size n, hence a basis by "check one get on free for bases", Proposition 6.13. $\square$

As well as knowing whether a matrix is diagonalisable, we might also want to know what the diagonal form is, and what the base change matrix is. We can in fact already infer that from the results we've discussed so far, but let us summarise it:

Theorem 9.37: (Diagonal form and eigenvector base change)
If $T\colon V \longrightarrow V$ is diagonalisable, then it's diagonal form D has all eigenvalues of T on the diagonal.
Furthermore, if A is the matrix for T in the standard basis (for V where this makes sense), and P is the matrix whose columns are the eigenvectors of T, corresponding in order to the diagonal entries of D, then $D = P^{-1}AP$.

Proof. Essentially the whole statement follows directly from Theorem 9.27: T is diagonalisable iff there is a basis of eigenvectors. To see that we have the base change the correct way round: if E is the standard basis and B the eigenvector basis, then $P = P_{B\longrightarrow E}$, so

$$[T]_B = P_{E\longrightarrow B}[T]_E P_{B\longrightarrow E}$$

i.e.

$$D = P^{-1}AP. \qquad \square$$

Examples 9.38: a) Revisiting

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

we have eigenvector base change matrix

$$P = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix},$$

and then

$$\begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} = D = P^{-1}AP.$$

b) Revisiting

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

we have

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

and then

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 5 \end{pmatrix} = D = P^{-1}AP.$$

c) Revisiting again

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

we have eigenvector base change matrix

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

and then

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} = D = P^{-1}AP.$$

has eigenvalue 2 with eigenvectors $v_1 = \left(\begin{smallmatrix} -1 \\ 0 \\ 1 \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} 0 \\ 1 \\ 0 \end{smallmatrix}\right)$, and eigenvalue 1 with eigenvector $v_3 = \left(\begin{smallmatrix} -2 \\ 1 \\ 1 \end{smallmatrix}\right)$.

If we are working with real numbers, this is the best we can hope for, as we don't know whether the map has "enough eigenvalues", i.e. whether the sum of algebraic multiplicities is n. We can do a little better if we work with complex numbers, as we will see in the next section.

Having diagonalised a matrix not only tells us more about the properties of the linear map, but also let's us calculate powers more easily.

Proposition 9.39: (Powers via diagonalisation)
If A is a square matrix which is diagonalisable as $A = PDP^{-1}$ with D diagonal, then $A^k = PD^kP^{-1}$, where D^k is diagonal with diagonal entries the kth powers of the diagonal entries of D.

PROOF. Exercise. □

Let's have a little summary from the previous few sections of properties that are invariant under base change, or properties that stay the same for similar matrices :

Properties of linear maps (i.e. Similarity invariants)

Property	Explanation
Rank	dimension of image
Nullity	dimension of kernel
Invertibility	or "being an isomorphism"
Determinant	originally defined for matrix, now shown $\det(P^{-1}AP) = \det A$.
Trace	originally defined for matrix, now shown $\mathrm{tr}(P^{-1}AP) = \mathrm{tr}A$.
Eigenvalues	
Characteristic polynomial	so also algebraic mult of evalues
Dimension of eigenspaces	i.e. geometric mult of evalues

Study guide.

Concept review

- ◇ Diagonalisation of a map/matrix.
- ◇ Algebraic and geometric multiplicities of eigenvalues, and their relationships.
- ◇ Different conditions for a map/matrix to be diagonalisable.
- ◇ Relationship of eigenvectors for distinct eigenvalues.

Skills

- ◇ Work out the algebraic and geometric multiplicities of eigenvalues.
- ◇ Determine whether a map/matrix is diagonalisable.
- ◇ If the map is diagonalisable, find the diagonal form.
- ◇ If the map is diagonalisable, find a basis which consists of eigenvectors.

D. Complex eigenvalues

All the theory we have developed about eigenvalues and eigenvectors of course works for linear maps between complex vector spaces as well. If we are working over $\mathbb{C}$, we should not expect eigenvalues to be real. However, we saw that even for *real* vector spaces, sometimes a linear map might not have real eigenvalues. But the main important fact about complex numbers is that

Very Important Fact 9.40: (Fundamental Theorem of Algebra)
Any polynomial can be factorised completely into linear factors over $\mathbb{C}$.

(Not proved in this course. A very nice proof uses complex analysis.)
This means that we always have "enough eigenvalues" over $\mathbb{C}$: any polynomial of degree n will have exactly n linear factors, so (counting with multiplicity) it will have exactly n roots.

Example 9.41: Recall the 90° rotation

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

with characteristic polynomial

$$\chi_R(t) = \det(tI - R) = \begin{vmatrix} t & 1 \\ -1 & t \end{vmatrix} = t^2 + 1$$

This has no real roots, but it does have complex roots $\pm i$. If we now think of $\mathbb{R}^2$ via the Argand plane as $\mathbb{C}$, which is a real vector space of dimension 2, then

$\binom{x}{y}$	corresponds to	$x + iy$,
$\binom{1}{0}$	corresponds to	$1 + 0i$ on the x-axis,
$\binom{0}{1}$	corresponds to	$0 + i$ on the y-axis,
$\binom{-1}{0}$	corresponds to	$-1 + 0i$ on the negative x-axis,

so under this correspondance $R(1) = i$ and $R(i) = -1$, which is exactly a 90° rotation of the Argand plane.
So we can view complex eigenvalues of real matrices as telling us something about the rotational properties of the matrix.
Similarly, the general rotation matrix

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

which rotates the plane anti-clockwise by θ has characteristic polynomial

$$\chi_{R_\theta}(t) = \begin{vmatrix} t - \cos\theta & \sin\theta \\ -\sin\theta & t - \cos\theta \end{vmatrix} = (t - \cos\theta)^2 + (\sin\theta)^2 = t^2 - 2\cos\theta t + 1$$
$$= t^2 - (e^{i\theta} + e^{-i\theta})t + 1 = (t - e^{i\theta})(t - e^{-i\theta})$$

so this has eigenvalues $e^{i\theta}$ and $e^{-i\theta}$, which correspond to the complex numbers we get when we rotate 1 by θ anti-clockwise or clockwise respectively.
Here we are using $e^{i\theta} = \cos\theta + i\sin\theta$, which gives $\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2}$. In lectures we'll draw an Argand plane here to demonstrate.

We see that there are two complex eigenvalues in these examples but it's only one rotation. We can look at the reasons for this from several different viewpoints. One is this:

Proposition 9.42: (Complex roots of real polynomials)
If p is a real polynomial, then any complex roots appear in complex conjugate pairs.

Proof. Let $p(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$ with all $a_i \in \mathbb{R}$. Suppose that $p(z) = 0$ for some particular $z \in \mathbb{C}$. Then we look at the complex conjugate $\overline{z}$:

$$\begin{aligned} p(\overline{z}) &= a_n \overline{z}^n + \cdots + a_1 \overline{z} + a_0 \\ &= \overline{a_n z^n + \cdots + a_1 z + a_0} && \text{because all } a_i \in \mathbb{R}, \text{ (so } \overline{a_i} = a_i). \\ &= \overline{p(z)} \\ &= 0. \end{aligned}$$

So if z is a root then so is it's complex conjugate. $\square$

This shows that the roots *have* to come in these complex conjugate pairs. How you can make sense of it geometrically:
Suppose you have $\mathbb{R}^2$ as the x, y-plane in $\mathbb{R}^3$. If you look at it "from above", so from the side which has positive z-values, then sending $\left(\begin{smallmatrix}1\\0\end{smallmatrix}\right)$ to $\left(\begin{smallmatrix}0\\1\end{smallmatrix}\right)$ looks like a 90° anti-clockwise rotation. But if you look at it *from below*, then the same mapping now looks like a clockwise rotation.
So when we say that we view $\mathbb{R}^2$ as $\mathbb{C}$, there are these two inbuilt "viewpoints", and the characteristic polynomial doesn't know which we've chosen, so it has both these roots.

Here is a totally different way of looking at complex eigenvalues of real matrices:

Example 9.43: The same matrix $R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ can also be viewed as a complex matrix, so as a map $\mathbb{C}^2 \longrightarrow \mathbb{C}^2$ (since $\mathbb{R}^2 \subseteq \mathbb{C}^2$). So given the eigenvalues $\pm i$, we can find complex eigenvectors:

$$R - iI = \begin{pmatrix} -i & -1 \\ 1 & -i \end{pmatrix} \xrightarrow{\text{R1 } \times i} \begin{pmatrix} 1 & -i \\ 1 & -i \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix}$$

so the eigenvector for i is $v_1 = \left(\begin{smallmatrix} i\\1\end{smallmatrix}\right)$.

$$R + iI = \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} \xrightarrow{\text{R1 } \times -i} \begin{pmatrix} 1 & i \\ 1 & i \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & i \\ 0 & 0 \end{pmatrix}$$

so the eigenvector for $-i$ is $v_2 = \left(\begin{smallmatrix} -i\\1\end{smallmatrix}\right)$.

If we view a real matrix as a complex matrix like this and find complex eigenvalues and eigenvectors, we have

Proposition 9.44: (Complex evalues and evectors of real matrices)
If a real matrix A has a complex evalue λ with complex evector v, then $\overline{\lambda}$ is also an evalue of A, with evector $\overline{v}$.

Proof. We already know from the previous result that roots of the (real) characteristic polynomial of A come in complex conjugate pairs. But even easier (recall that evector $v \neq 0$):

$$\begin{aligned} & Av = \lambda v \\ \Leftrightarrow \quad & \overline{Av} = \overline{\lambda v} \\ \Leftrightarrow \quad & A\overline{v} = \overline{\lambda} \cdot \overline{v} && \text{as } A \text{ is real.} \end{aligned}$$ $\square$

Examples 9.45: a) We already saw the 90° rotation R.

b) A general rotation

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

with evalues $\lambda = e^{i\theta} = \cos\theta + i\sin\theta$ and $e^{-i\theta} = \cos\theta - i\sin\theta = \overline{\lambda}$ has evectors $v_1 = \left(\begin{smallmatrix} i \\ 1 \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} -i \\ 1 \end{smallmatrix}\right)$.

Here is the working out, but not lectured: we need to use $e^{i\theta} = \cos\theta + i\sin\theta$, which gives $\cos\theta = \frac{e^{i\theta}+e^{-i\theta}}{2}$, and $\sin\theta = \frac{e^{i\theta}-e^{-i\theta}}{2i}$. Don't worry, you won't be asked to do something this involved in the exam. But you may well have to do it later in your life! Even possibly in your job.

$$R_\theta - e^{i\theta}I = \begin{pmatrix} \cos\theta - e^{i\theta} & -\sin\theta \\ \sin\theta & \cos\theta - e^{i\theta} \end{pmatrix} = \begin{pmatrix} \frac{-e^{i\theta}+e^{-i\theta}}{2} & -\frac{e^{i\theta}-e^{-i\theta}}{2i} \\ \frac{e^{i\theta}-e^{-i\theta}}{2i} & \frac{-e^{i\theta}+e^{-i\theta}}{2} \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} -(e^{i\theta}-e^{-i\theta}) & -(-i)(e^{i\theta}-e^{-i\theta}) \\ (-i)(e^{i\theta}-e^{-i\theta}) & -(e^{i\theta}-e^{-i\theta}) \end{pmatrix} \qquad \text{using } \frac{1}{i} = -i$$

$$\longrightarrow \begin{pmatrix} -(e^{i\theta}-e^{-i\theta}) & -(-i)(e^{i\theta}-e^{-i\theta}) \\ 0 & 0 \end{pmatrix} \qquad \text{using } i(-i) = 1$$

$$\longrightarrow \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix} \qquad \text{using } i(-i) = 1$$

c) The real matrix

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

has characteristic polynomial $(t-i)(t+i)$, and eigenvectors $v_1 = \begin{pmatrix} -\frac{2}{5} + \frac{1}{5}i \\ 1 \end{pmatrix}$ for i and $v_2 = \begin{pmatrix} -\frac{2}{5} - \frac{1}{5}i \\ 1 \end{pmatrix} = \overline{v_1}$ for $-i$.

The main result of this section is related to the fact that complex polynomials factor totally into linear factors.

Theorem 9.46: (Complex diagonalisable iff algebraic = geometric mult)
Let $T\colon V \longrightarrow V$ be a linear map on a complex *vector space V. Then T is diagonalisable (over $\mathbb{C}$) if and only if the geometric multiplicity of every eigenvalue agrees with the algebraic multiplicity.*

Proof. By the Fundamental Theorem of Algebra (Very Important Fact 9.40), the sum of algebraic multiplicities is $\dim V = n$. By "Diagonalisability via geometric multiplicities" (Theorem 9.36), we know that T is diagonalisable if and only if the geometric multiplicities add to n. As geometric $\leqslant$ algebraic multiplicities (Prop. 9.32) and the sum of algebraic multiplicities is n, this now happens if and only if geometric = algebraic multiplicity for every single eigenvalue. □

NOTE: This theorem is only true over the complex numbers! This means that the base change matrix has to be allowed to be complex.

Later on you might learn also about *minimal polynomials* and *Jordan normal form*, which are very good tools to know about a linear map and it's eigenvalues, eigenspaces and diagonalisability. But for us, this is all for now.

Study guide.

Concept review

- ⋄ Complex evalue of 2×2 matrix: shows rotational properties.
- ⋄ Complex roots of real polynomials.

◇ Complex evalues and evectors of real matrices.
◇ Fundamental Theorem of Algebra.

Skills

◇ Factorise polynomials into complex linear factors.
◇ Find complex evalues and evectors.
◇ Diagonalise a matrix over $\mathbb{C}$.

E. Eigenvectors and Eigenvalues: Study guide collation

Just putting together all the study guides from the different sections.

Concept review.

◇ Eigenvalues, eigenvector.
◇ Eigenspace.
◇ Connection of kernel with eigenvalues/eigenspaces.
◇ Connection of isomorphisms and eigenvalues.
◇ Characteristic polynomial, and its link to eigenvalues.
◇ Eigenvalues of diagonal and triangular matrices.
◇ Relationship of eigenvalues of similar matrices.
◇ Eigenspace as kernel of a particular linear map.
◇ Eigenvalues and eigenvectors of powers or inverses of a map/matrix.
◇ Diagonalisation of a map/matrix.
◇ Algebraic and geometric multiplicities of eigenvalues, and their relationships.
◇ Different conditions for a map/matrix to be diagonalisable.
◇ Relationship of eigenvectors for distinct eigenvalues.
◇ Complex evalue of 2×2 matrix: shows rotational properties.
◇ Complex roots of real polynomials.
◇ Complex evalues and evectors of real matrices.
◇ Fundamental Theorem of Algebra.

Skills.

◇ Find the characteristic polynomial of a map/matrix.
◇ Find the eigenvalues of a map/matrix.
◇ Find the eigenvectors of a map/matrix.
◇ Find the eigenvalues of an inverse matrix.
◇ Work out the algebraic and geometric multiplicities of eigenvalues.
◇ Determine whether a map/matrix is diagonalisable.
◇ If the map is diagonalisable, find the diagonal form.
◇ If the map is diagonalisable, find a basis which consists of eigenvectors.
◇ Factorise polynomials into complex linear factors.
◇ Find complex evalues and evectors.
◇ Diagonalise a matrix over $\mathbb{C}$.

CHAPTER 10

Inner Products

A. Standard real inner product

You might know from $\mathbb{R}^2$ or $\mathbb{R}^3$ the concept of a *dot product* and the *length* of a vector:

- ◇ If $x = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ and $y = \left(\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right)$, then $x \cdot y = x_1 y_1 + x_2 y_2$.
- ◇ The length is $\|x\| = \sqrt{x_1^2 + x_2^2}$ (using Pythagoras).

It is very useful to view the dot product as a matrix product of a row vector times a column vector:

$$x \cdot y = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = x_1 y_1 + x_2 y_2$$

by the general matrix multiplication rules.
These concepts are easily generalised to $\mathbb{R}^n$:

Definition 10.1: The **standard inner product** (also called **Euclidean inner product**) on $\mathbb{R}^n$ is a function $\langle -, - \rangle \colon \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ defined by

$$\begin{aligned} \langle v, w \rangle = v^T w &= \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} \\ &= v_1 w_1 + v_2 w_2 + \cdots + v_n w_n = \sum_{i=1}^{n} v_i w_i. \end{aligned}$$

The **norm** (or length or magnitude) of a vector in $\mathbb{R}^n$ is defined as

$$\|v\| = \sqrt{\langle v, v \rangle}.$$

Examples 10.2: a) If $v = \left(\begin{smallmatrix} 2 \\ -1 \\ 3 \\ -5 \end{smallmatrix}\right)$ then $\|v\| = \sqrt{2^2 + (-1)^2 + 3^2 + (-5)^2)} = \sqrt{39}$.

b) If v is as above and $w = \left(\begin{smallmatrix} -3 \\ -4 \\ 1 \\ 0 \end{smallmatrix}\right)$, then $\langle v, w \rangle = 2 \cdot (-3) + (-1) \cdot (-4) + 3 \cdot 1 + (-5) \cdot 0 = 2$.

c) It is up to you whether you prefer writing $\langle v, w \rangle$ or $v^T w$. HOWEVER be careful with the use of $v \cdot w$, the "dot product", as this can sometimes lead to confusion. It is MUCH SAFER to use $v^T w$, in the matrix multiplication form.

From the idea of length, we might expect this norm to satisfy certain properties, such as:

- ◇ The norm of any vector is non-negative.
- ◇ The zero vector is the only vector with norm 0.
- ◇ Multiplying a vector by a scalar multiplies the norm by the absolute value of that scalar.

We might also expect something that resembles a triangle inequality.

We will also see some important propties of the inner product, which are: The inner product

- ◇ is linear in each entry.
- ◇ is symmetric.

- ⋄ encodes orthogonality.

Let's start with the the inner product properties, since we defined the norm using the inner product.

Theorem 10.3: (Inner product properties)
The standard inner product $\langle -,-\rangle \colon \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ satisfies:

(i) $\langle \lambda u + \mu v, w\rangle = \lambda\langle u, w\rangle + \mu\langle v, w\rangle$ *and* $\langle v, \lambda u + \mu w\rangle = \lambda\langle v, u\rangle + \mu\langle v, w\rangle$ ***(bilinear)***
(ii) $\langle v, w\rangle = \langle w, v\rangle$ ***(symmetric)***
(iii) $\langle v, v\rangle \geqslant 0$, *and* $\langle v, v\rangle = 0 \Leftrightarrow v = 0$ ***(positive definite)***

Proof. (i) Matrix multiplication is linear (Prop. 1.67). This gives

$$\langle v, \lambda u + \mu w\rangle = v^T(\lambda u + \mu w) = \lambda v^T u + \mu v^T w = \lambda\langle v, u\rangle + \mu\langle v, w\rangle.$$

We will get the other statement once we also have symmetry.

(ii) $\langle v, w\rangle = v^T w = (v^T w)^T = w^T v = \langle w, v\rangle$. We are using: transpose of 1×1-matrix (i.e. a number) does not change it.

Or write it out explicitely as $\langle v, w\rangle = \sum\limits_{i=1}^{n} v_i w_i$.

(iii) $\langle v, v\rangle = \sum\limits_{i=1}^{n} v_i v_i = \sum\limits_{i=1}^{n} (v_i)^2$. Squares are always non-negative, and a sum of non-negative terms is non-negative.

Also $\langle v, v\rangle = v_1^2 + \cdots + v_n^2 = 0$ if and only if each summand is 0 separately, as they are all non-negative (so we can't cancel anything out). This is the case if and only if $v = 0$. □

Remark 10.4: The bilinear property is related to two linear maps we already know:

- ⋄ Given $v \in \mathbb{R}^n$, we can view v as a $(1 \times n)$-matrix v^T, and then the corresponding linear map is $T_{v^T} : \mathbb{R}^n \longrightarrow \mathbb{R}$ with $T_{v^T}(w) = v^T w = \langle v, w\rangle$. This corresponds to linearity in the second entry of the inner product.
- ⋄ The actual process of "viewing v as v^T" is in itself a linear map $(-)^T \colon \mathbb{R}^n \longrightarrow \mathscr{M}_{1,n}$, sending a vector to a matrix (and a matrix of course in itself also represents a linear map). (For interest, linking to the future:) You will later see this as a map

$$v \longmapsto \langle v, -\rangle \in V^* = \{ \text{ linear maps } V \longrightarrow \mathbb{R}\}.$$

Here V^* is called the **dual space** of V. This linear map corresponds to linearity in the first entry.

Example 10.5: Not lectured, explaining above remark.

a) Given $v = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, we have the linear map

$$T_{v^T} : \mathbb{R}^n \longrightarrow \mathbb{R} \qquad \text{with} \qquad T_{v^T}(w) = w_1 + 2w_2 + w_3.$$

b) Any vector $\begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ can be seen as the $(1 \times n)$-matrix $\begin{pmatrix} v_1 & v_2 & \dots & v_n \end{pmatrix}$. In some sense we are saying "taking the transpose (of a matrix) is a linear operation".

The inner product properties also immediately give us the properties for the norm:

Corollary 10.6: (Properties of norm)
The standard norm on $\mathbb{R}^n$ *satisfies:*

(i) $\|v\| \geqslant 0$.
(ii) $\|v\| = 0 \Leftrightarrow v = 0$.
(iii) $\|\lambda v\| = |\lambda|\|v\|$.

Proof. Exercise: use inner product properties. □

It is often useful to remember

$$\|v\|^2 = \langle v, v \rangle.$$

Now that we have something like length, we can talk about "making vectors a unit length".

Definition 10.7: A **unit vector** is a vector of norm 1. Given $v \in \mathbb{R}^n$, if $v \neq 0$, we can **normalise** v to the unit vector $\frac{1}{\|v\|}v$.

Examples 10.8: a) The standard basis vectors are all unit vectors.
b) If $v = \begin{pmatrix} 2 \\ -1 \\ 3 \\ -5 \end{pmatrix}$ then its normalisation is $u = \frac{1}{\|v\|}v = \frac{1}{\sqrt{39}}\begin{pmatrix} 2 \\ -1 \\ 3 \\ -5 \end{pmatrix}$.
c) If $v = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$, then its normalisation is $\frac{1}{\sqrt{n}}v$.

In $\mathbb{R}^2$ (and by extension in $\mathbb{R}^3$), we can talk about angles between vectors. Using a diagram in $\mathbb{R}^2$, we can see that, given two unit vectors $x = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$ and $y = \left(\begin{smallmatrix} y_1 \\ y_2 \end{smallmatrix}\right)$, we have

$$x_1y_1 + x_2y_2 = y_1 = \cos\theta.$$

If $x = \left(\begin{smallmatrix} x_1 \\ x_2 \end{smallmatrix}\right)$ is a general vector, then by rotating the plane we still get

$$x_1y_1 + x_2y_2 = \cos\theta,$$

where θ is the angle between the two vectors which satisfies $0 \leqslant \theta \leqslant 180°$.
Draw a diagram to illustrate, and let $x_1 = \cos\theta_1, x_2 = \sin\theta_1, y_1 = \cos\theta_2, y_2 = \sin\theta_2$. Then the angle θ between the vectors is $(\pm)(\theta_1 - \theta_2)$. And

$$x_1y_1 + x_2y_2 = \cos\theta_1\cos\theta_2 + \sin\theta_1\sin\theta_2 = \cos(\theta_1 - \theta_2) = \cos\theta.$$

In particular we see that

$$x_1y_1 + x_2y_2 = 0 \Leftrightarrow x \text{ and } y \text{ are perpendicular to each other.}$$

While it does not make entire sense to generalise angles to $\mathbb{R}^n$, we *are* very much interested in this generalisation of vectors being perpendicular.

Definition 10.9: Two vectors $v, w \in \mathbb{R}^n$ are called **orthogonal** exactly when $\langle v, w \rangle = 0$.

Examples 10.10: a) The standard basis vectors are pairwise orthogonal.
b) $v_1 = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ are orthogonal in $\mathbb{R}^2$:

$$\langle v_1, v_2 \rangle = 1 \cdot 1 + 1 \cdot (-1) = 0.$$

Of course in $\mathbb{R}^2$, there is only one vector which is orthogonal to a given vector (up to linear multiples). This is not true in $\mathbb{R}^n$:
c) The vector $\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right) \in \mathbb{R}^3$ is orthogonal to any vector which satisfies $x_1 + x_2 + x_3 = 0$, so for example

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, v_1 + v_2 = \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix}$$

and any other linear combination of v_1, v_2. You can think of this as *the plane orghogonal to* $\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)$.
d) Similarly in $\mathbb{R}^4$, the vector $\left(\begin{smallmatrix} 1 \\ 1 \\ 1 \\ 1 \end{smallmatrix}\right)$ is orthogonal to any vector satisfying $x_1 + x_2 + x_3 + x_4 = 0$, which you can think of as a **hyperplane** with basis

$$v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}, v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ -1 \end{pmatrix}, v_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}.$$

Here "hyper"plane means "space of dimension one less than ambient space". We call it hyper*plane* because our most comfortable space is three-dimensional, and in that case we get a plane.

e) This generalises to $\mathbb{R}^n$.

Exercise 10.11: Find the vectors which are orthogonal to $\begin{pmatrix} 1 \\ 3 \\ 2 \\ -1 \\ 4 \end{pmatrix} \in \mathbb{R}^5$. Hint: use an equation like we did above.

Using this inner product, we can also determine vectors.

Proposition 10.12: (Inner product detects 0)
For given $v \in \mathbb{R}^n$, $\langle v, w \rangle = 0$ for all $w \in \mathbb{R}^n$ if and only if $v = 0$.

Proof. Since we have $\langle v, w \rangle = 0$ *for all* $w \in \mathbb{R}^n$, also $\langle v, v \rangle = 0$, which gives $v = 0$. The other direction was already proved in "inner product properties" (Theorem 10.3). □

Example 10.13: a) Suppose you want to know "which vectors are orthogonal to everything in $\mathbb{R}^5$". We are looking for $v \in \mathbb{R}^5$ such that $\langle v, w \rangle = 0$ for all $w \in \mathbb{R}^5$. So the previous result tells us: only 0.

b) If we have v_1 and v_2 such that $\langle v_1, w \rangle = \langle v_2, w \rangle$ *for all* $w \in \mathbb{R}^n$, then we can conclude:

$$\langle v_1, w \rangle = \langle v_2, w \rangle \Rightarrow \langle v_1 - v_2, w \rangle = 0 \Rightarrow v_1 - v_2 = 0 \Rightarrow v_1 = v_2.$$

So if we ask: can the line $x_1 = x_2$ (i.e. the line generated by the vector $u = \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$) in $\mathbb{R}^2$ have several vectors which are orthogonal to it? We might know that $v_1 = \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right)$ is such a vector. Suppose v_2 is another. Then

- For $w = \lambda\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$, i.e. on the line, we have $\langle v_1, w \rangle = \langle v_2, w \rangle = 0$.
- We have $\langle v_1, v_1 \rangle = 2$, and $\langle v_2, v_1 \rangle = a$, some real number.
- We know that $\left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right)$ and v_1 form a basis of $\mathbb{R}^2$, so for any $v \in \mathbb{R}^2$, we can write $v = \mu_1 u + \mu_2 v_1$, and then

$$\begin{aligned} & \langle v_2, v \rangle = \mu_1 \langle v_2, u \rangle + \mu_2 \langle v_2, v_1 \rangle = 0 + a\mu_2 \\ \text{and} \quad & \langle v_1, v \rangle = \mu_1 \langle v_1, u \rangle + \mu_2 \langle v_1, v_1 \rangle = 0 + 2\mu_2. \end{aligned}$$

  Then $\langle \frac{a}{2} v_1, v \rangle = \langle v_2, v \rangle$ *for all* $v \in \mathbb{R}^2$. So $v_2 = \frac{a}{2} v_1$ must be a scalar multiple of v_1. So we have only "one direction" of vector which is orthogonal to the given line.
  Of course we could have found this out more easily as well. I just wanted to demonstrate a kind of way this could be used.

We will now see a very useful property of the inner product which implies the triangle inequality for the norm.

Theorem 10.14: (Cauchy-Schwarz inequality)
For $u, v \in \mathbb{R}^n$, we have

$$\langle u, v \rangle^2 \leqslant \|u\|^2 \|v\|^2$$

($\Rightarrow |\langle u, v \rangle| \leqslant \|u\|\|v\|$ and $\langle u, v \rangle \leqslant \|u\|\|v\|$).
Equality holds if and only if $u = \mu v$ for some $\mu \in \mathbb{R}$. (I.e. only when u, v are parallel.)

Proof. Consider $\langle u + tv, u + tv \rangle \geqslant 0$, which is a quadratic polynomial in t, and non-negative as the inner product is positive definite (see Theorem 10.3). This is the main idea and super trick of this proof! Give it the awe it deserves :-)
We have

$$\langle u + tv, u + tv \rangle = \langle u, u \rangle + t\langle u, v \rangle + t\langle v, u \rangle + t^2 \langle v, v \rangle = \|u\|^2 + 2t\langle u, v \rangle + \|v\|^2 t^2.$$

The roots of this quadratic are

$$\frac{-2\langle u, v\rangle \pm \sqrt{(2\langle u, v\rangle)^2 - 4\|u\|^2\|v\|^2}}{2\|v\|^2}.$$

Since we know the quadratic is non-negative, it has exactly one or zero roots, so the discriminant must be non-positive (either discriminant is 0: one real root, or discriminant negative: only complex roots, no real roots), i.e.

$$4\langle u, v\rangle^2 - 4\|u\|^2\|v\|^2 \leqslant 0$$

which gives

$$\langle u, v\rangle^2 \leqslant \|u\|^2\|v\|^2.$$

Then, taking square roots, we get $|\langle u, v\rangle| \leqslant \|u\|\|v\|$, but as $\langle u, v\rangle \leqslant |\langle u, v\rangle|$, we can also say

$$\langle u, v\rangle \leqslant \|u\|\|v\|.$$

To get equality in the discriminant, we need $\langle u + tv, u + tv\rangle = 0$, which happens if and only if $u + tv = 0$, i.e. $u = -tv$ for some t. □

Using this, we can prove:

Corollary 10.15: (Triangle inequality for norm)
Given $v, w \in \mathbb{R}^n$, then $\|v + w\| \leqslant \|v\| + \|w\|$.

Proof.

$$\begin{aligned}
\|v + w\|^2 &= \langle v + w, v + w\rangle \\
&= \|v\|^2 + 2\langle v, w\rangle + \|w\|^2 \\
&\leqslant \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 && \text{using Cauchy-Schwarz} \\
&= (\|v\| + \|w\|)^2
\end{aligned}$$

which implies

$$\|v + w\| \leqslant \|v\| + \|w\|$$

as both sides are non-negative. □

Study guide.

Concept review

- ◇ Standard (Euclidean) Inner Product, norm, and the connection between the two.
- ◇ Inner product properties: it's bilinear, symmetric, positive definite.
- ◇ Norm properties that we expect from a "length".
- ◇ Unit vectors, normalisation of vectors.
- ◇ Orthogonal vectors. Hyperplane orthogonal to a vector.
- ◇ Inner product detects 0.
- ◇ Cauchy-Schwarz inequality.
- ◇ Triangle inequality of norm.

Skills

- ◇ Work out norm of a given vector.
- ◇ Work out inner product of two given vectors.
- ◇ Use properties of inner product for expansion/manipulation.
- ◇ Normalise a vector to be a unit vector.
- ◇ Determine whether vectors are orthogonal.

B. General inner products

The standard (Euclidean) inner product is not the only one on $\mathbb{R}^n$. We can generalise this concept.

Definition 10.16: Let V be a real vector space. Then a **(real) inner product** on V is a function $\langle -, -\rangle\colon V \times V \longrightarrow \mathbb{R}$ satisfying:

(i) $\langle \lambda u + \mu v, w\rangle = \lambda\langle u, w\rangle + \mu\langle v, w\rangle$ and $\langle v, \lambda u + \mu w\rangle = \lambda\langle v, u\rangle + \mu\langle v, w\rangle$ **(bilinear)**
(ii) $\langle v, w\rangle = \langle w, v\rangle$ **(symmetric)**
(iii) $\langle v, v\rangle \geqslant 0$, and $\langle v, v\rangle = 0 \Leftrightarrow v = 0$ **(positive definite)**

A vector space V together with an inner product is called an **inner product space**.

So instead of *proving* these properties, we just *ask* them, and call any such function which satisfies them an inner product.

Examples 10.17: a) Obviously the Euclidean inner product is an example.

b) [**Weighted inner product**] Let D be a diagonal matrix with positive diagonal entries $d_1, d_2, \ldots, d_n > 0$. Then

$$\langle v, w\rangle = v^T Dw = d_1v_1w_1 + d_2v_2w_2 + \cdots + d_nv_nw_n$$

is also an inner product. The first two points are easy to check using $v^T Dw$ and matrix multiplication properties. To see it is positive definite:

$$\langle v, v\rangle = d_1v_1^2 + \cdots + d_nv_n^2 \geqslant 0$$

because the squares are $\geqslant 0$ and the $d_i > 0$. This sum is 0 if and only if all summands are 0, so as the d_i are *strictly* positive, this happens if and only if $v = 0$.

Such an inner product can be useful in a situation or problem where not all the different directions are "equally good" for something. For example, if you want to move something, moving it in one direction might be harder or more expensive or more time-consuming than in another direction. A weighted inner product can incorporate such situations.

c) A variation of this is to take a real symmetric matrix A with positive eigenvalues, and take $\langle v, w\rangle = v^T Aw$. (The proof of this relies on the fact that any real symmetric matrix is diagonalisable, which we haven't proved yet. C.f. Theorem 10.32.)

With the stated result, this is then really the same as a weighted inner product, just that the weightings are not necessariliy in the x and y directions (or equivalent), but in some other directions, given by the eigenvectors of the matrix used.

d) If V is the vector space of $n \times n$ matrices, then

$$\langle A, B\rangle = \mathrm{tr}(A^T B)$$

is an inner product.

Exercise: check the three properties in the definition of inner product in this example. Remember: $\mathrm{tr}(C^T) = \mathrm{tr}(C)$. For positive definiteness, you'll have to look at actual elements.

e) Let $V = P_n$. Then

$$\langle p, q\rangle = \int_{-1}^{1} p(x)q(x)dx$$

is an inner product.

(i) $\int_{-1}^{1}(ap_1(x) + bp_2(x))q(x)dx = a\int_{-1}^{1} p_1(x)q(x)dx + b\int_{-1}^{1} p_2(x)q(x)dx$, and similarly in the second entry.

(ii) $\int_{-1}^{1} p(x)q(x)dx = \int_{-1}^{1} q(x)p(x)dx$.

(iii) $\int_{-1}^{1} p(x)p(x)dx \geqslant 0$ because the integrand is non-negative on all of $[-1, 1]$. Why does $\int_{-1}^{1} p(x)p(x)dx = 0$ imply $p = 0$? As $p(x)^2 \geqslant 0$ on $[-1, 1]$, we can first conclude that $p(x) = 0$ on $[-1, 1]$. But why is it then the zero-polynomial? A polynomial of degree n has at most n roots, but we have found a whole interval $[-1, 1]$ where it is 0, so it must be the zero polynommial.

The integral inner product also works on the vector space of integrable (or continuous or differentiable or infinitely differentiable) functions defined on a given interval (if we make the integral go over the whole integral as well).

f) Let V be a suitable space of random variables. Then

$$\langle X, Y\rangle = \mathrm{E}(XY)$$

is an inner product.

Exercise: prove this, using properties of expectation from probability.

Not lectured: Taking covariance of two random variables also gives an inner product. Check it with covariance facts.

Such inner product spaces are very important and used in all sorts of areas, ranging from pure maths via numerical methods needed in mathematical modelling, to quantum physics.

Because of the way we proved everything in the previous section, only using the properties of bilinearity, symmetry and positive definiteness, all the results still hold. So we can

- ◇ define a **norm** corresponding to the inner product, which satisfies all the same properties;
- ◇ have **unit** vectors and normalise with respect to the inner product;
- ◇ have **orthogonal** vectors with respect to the inner product;
- ◇ use Cauchy-Schwarz with respect to the inner product.

Examples 10.18: a) Such different inner products give rise to *different lengths*. For example, for a weighted inner product on $\mathbb{R}^2$ with $D = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$, we have

$$\langle u, v \rangle = 3u_1v_1 + 2u_2v_2$$

and, for example,

$$\| \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) \| = \sqrt{3}, \qquad \| \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) \| = \sqrt{2}.$$

This can be useful if, for example, we want to include how difficult it is to travel or move something in a certain direction, and so want to weight the "length" or inner product so that two vectors with same norm represent travel with the same time or the same amount of energy, or something similar.

This gives rise to strange "unit circles" as well. In lectures, we'll draw one.

b) Cauchy-Schwarz still holds for these different kind of inner products, so for the integral one, we have

$$\int_{-1}^{1} p(x)q(x)dx \leqslant \sqrt{\int_{-1}^{1} p(x)^2 dx \int_{-1}^{1} q(x)^2 dx}.$$

and also

$$\mathrm{E}(XY) \leqslant \sqrt{\mathrm{E}(X^2)\mathrm{E}(Y^2)}.$$

Who would have thought!

Study guide.

Concept review

- ◇ General inner product.
- ◇ Examples of weighted inner product, matrix inner product, integral inner product.

Skills

- ◇ Work out norm of a given vector for different examples of (non-Euclidean) inner products.
- ◇ Work out inner product of two given vectors for different examples of inner products.
- ◇ Normalise a vector to be a unit vector with respect to different examples of inner products.
- ◇ Determine whether vectors are orthogonal with respect to different examples of inner products.

C. Orthonormal bases and Gram-Schmidt

With the concept of orthogonal vectors we have seen previously, we might be interested in having a basis where all the vectors are orthogonal to each other, as we are used to from the standard basis.

In this section, V is any real inner product space V with inner product $\langle -, - \rangle$.

Definition 10.19: A set of vectors is called **orthogonal** if the vectors are pairwise orthogonal. A set of vectors is called **orthonormal** if it is orthogonal and all vectors are unit vectors (i.e. have norm 1). A basis which is also an orthonormal set is called an **orthonormal basis**.

Examples 10.20: (Using the Standard Inner Product)
First we look at examples with the standard or Euclidean inner product.

a) The standard basis in $\mathbb{R}^n$ is an orthonormal basis.
b) $v_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ form an orthogonal set, but they both have length $\sqrt{2}$. So

$$u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}$$

is an orthonormal basis of $\mathbb{R}^2$.
c) Not lectured. Here is an orthogonal basis of $\mathbb{R}^n$ (which is not the standard basis):

$$v_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ 1 \end{pmatrix}, \quad v_2 = \begin{pmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad v_3 = \begin{pmatrix} 1 \\ 1 \\ -2 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad v_4 = \begin{pmatrix} 1 \\ 1 \\ 2 \\ -4 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad v_5 = \begin{pmatrix} 1 \\ 1 \\ 2 \\ 4 \\ -8 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix},$$

$$\ldots, \quad v_k = \begin{pmatrix} 1 \\ 2^0 \\ 2^1 \\ 2^2 \\ \vdots \\ 2^{k-3} \\ -2^{k-2} \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad v_{n-1} = \begin{pmatrix} 1 \\ 2^0 \\ 2^1 \\ 2^2 \\ 2^3 \\ 2^4 \\ \vdots \\ 2^{n-4} \\ -2^{n-3} \\ 0 \end{pmatrix}, \quad v_n = \begin{pmatrix} 1 \\ 2^0 \\ 2^1 \\ 2^2 \\ 2^3 \\ 2^4 \\ \vdots \\ 2^{n-4} \\ 2^{n-3} \\ -2^{n-2} \end{pmatrix}$$

Then we can normalise it to get an orthonormal basis (but we won't show it here).

How does this work: Construct v_2 so that it is orthogonal to v_1, but only uses the first two entries. This means that when you take $n = 2$, you can cut off the zeros, and you have an orthogonal set of size 2 for $\mathbb{R}^2$. So it needs $x_1 + x_2 = 0$ to be orthogonal to v_1.

Then construct v_3 to be orthogonal to v_1 and v_2, but only uses the first three entries, so that it works in $\mathbb{R}^3$. So we need $x_1 = x_2$ to be orthogonal to v_2, and $x_1 + x_2 + x_3 = 0$ to be orthogonal to v_1.

So you continue like this. For v_k to be orthogonal to $v_2, v_3, \ldots, v_{k-2}$, it needs the same first $k-2$ entries as v_{k-1}. Then to be orthogonal to v_{k-1}, the $(k-1)$th entry has to be the negative of the$(k-1)$th entry of v_{k-1}. And then the last non-zero entry (i.e. kth entry) is determined so that the vector is orthogonal to v_1, i.e. the sum of all entries has to be 0. So that last non-zero entry is $-$(sum of all previous entries)$= -(1 + 2^0 + 2^2 + \cdots + 2^{k-3}) = -2^{k-2}$.

Examples 10.21: (Using other inner products) Let's see a few examples with different inner products.

a) **[Weighted inner product]** If we use an inner product weighted by $D = \left(\begin{smallmatrix} 3 & 0 \\ 0 & 2 \end{smallmatrix}\right)$ as in a previous example, then the standard basis in $\mathbb{R}^2$ is not orthonormal any more.

$$\begin{aligned} \langle \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) \rangle &= 0 && \text{so they are still orthogonal} \\ \|\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)\| &= \sqrt{3} \\ \|\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)\| &= \sqrt{2} \end{aligned}$$

so

$$\begin{pmatrix} \frac{1}{\sqrt{3}} \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}$$

is an orthonormal basis in this inner product space.

Also

$$\langle \left(\begin{smallmatrix} 1 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right) \rangle = 3 - 2 = 1$$

so these two vectors are *not* orthogonal any more.

Not lectured, but for extra practice/understanding: Instead,

$$\langle \left(\begin{smallmatrix} 2 \\ 3 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right) \rangle = 6 - 6 = 0$$

or

$$\langle \left(\begin{smallmatrix} 2 \\ 1 \end{smallmatrix}\right), \left(\begin{smallmatrix} 1 \\ -3 \end{smallmatrix}\right) \rangle = 6 - 6 = 0$$

or some other such combination.

b) **[Integral inner product on polynomial space]** Here is a set of orthogonal polynomials in P_2, using the integral inner product:

$$p_0(x) = 1, \qquad p_1(x) = x, \qquad p_2(x) = \frac{1}{2}(3x^2 - 1).$$

They are the first three **Legendre Polynomials**. It's easy to see that $\langle 1, x \rangle = 0$; you can check the other conditions $\langle p_0, p_2 \rangle = 0$ and $\langle p_1, p_2 \rangle$ as well. Remember to integrate over $[-1, 1]$. You won't be asked this in the exam, but it's super useful for a wide range of topics in your mathematical future.

You see from above that for every n, $\mathbb{R}^n$ does have an orthonormal basis. We could ask the other way round: if a given set is orthogonal, does it have to be linearly indepdenent? The answer is yes:

Proposition 10.22: (Orthogonal sets are linearly independent.)
If $v_1, \ldots, v_k$ is an orthogonal set of non-zero vectors in the inner product space V, then it is linearly independent.

PROOF. Summary: take inner product with each vector.
Consider

$$\lambda_1 v_1 + \cdots \lambda_k v_k = 0$$

and now apply $\langle v_1, - \rangle$ to the equation:

$$\begin{aligned} \Rightarrow \qquad \lambda_1 \langle v_1, v_1 \rangle + \lambda_2 \langle v_1, v_2 \rangle + \cdots + \lambda_n \langle v_1, v_k \rangle &= 0 \\ \Rightarrow \qquad \lambda_1 \langle v_1, v_1 \rangle + 0 &= 0 \\ \Rightarrow \qquad \lambda_1 &= 0 \end{aligned}$$

because $\langle v_1, v_1 \rangle = \|v_1\|^2 \neq 0$.
Similarly applying $\langle v_i, - \rangle$ for all the $i = 2, \ldots, k$ gives all other $\lambda_i = 0$. So we have a linearly independent set. □

So if $\dim V = n$, then having n orthogonal (or orthonormal) vectors gives us a basis. The good thing about orthonormal bases is that the coordinates (i.e. the coefficients in front of each basis vector) can be calculated much more easily than for a general basis.

Proposition 10.23: (Coordinates for orthonormal basis)
If $u_1, \dots, u_n$ is an orthonormal basis, then for any v, we have

$$v = \langle v, u_1 \rangle u_1 + \langle v, u_2 \rangle u_2 + \cdots + \langle v, u_n \rangle u_n.$$

Proof. As $u_1, \dots, u_n$ is a basis, we know that there are some λ_i such that

$$v = \lambda_1 u_1 + \cdots + \lambda_n u_n.$$

Then

$$\begin{aligned}\langle v, u_i \rangle &= \langle \lambda_1 u_1 + \cdots + \lambda_n u_n, u_i \rangle \\ &= \lambda_1 \langle u_1, u_i \rangle + \cdots + \lambda_n \langle u_n, u_i \rangle \\ &= \lambda_1 0 + \cdots + 0 + \lambda_i \langle u_i, u_i \rangle + 0 + \cdots + \lambda_n 0 \\ &= \lambda_i\end{aligned}$$

since $\langle u_j, u_i \rangle = 0$ for $j \leqslant i$ and $\langle u_i, u_i \rangle = 1$ for the orthonormal basis. □

This is *very* different to how we find coordinates for a general basis (which is not orthonormal)! There we need to know all the basis vectors to be able to find the coordinates. Here, as long as we know the vector is part of some orthonormal basis, we can find the coordinate in that direction just using that single vector. (This is a super amazing fact.)

Example 10.24: Recall the orthonormal basis

$$u_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, u_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

of $\mathbb{R}^2$ we worked out above. To find the coordinates with respect to this basis for $v = \left(\begin{smallmatrix} -4 \\ 13 \end{smallmatrix}\right)$, we can calculate:

- $\lambda_1 = \langle v, u_1 \rangle = \frac{1}{\sqrt{2}}((-4) \cdot 1 + 13 \cdot 1) = \frac{9}{\sqrt{2}}$
- $\lambda_2 = \langle v, u_2 \rangle = \frac{1}{\sqrt{2}}(-4) \cdot 1 + 13 \cdot (-1)) = \frac{-17}{\sqrt{2}}$

And indeed

$$\frac{9}{\sqrt{2}} u_1 + \frac{-17}{\sqrt{2}} u_2 = \frac{9}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \frac{-17}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 9 - 17 \\ 9 + 17 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} -8 \\ 26 \end{pmatrix} = \begin{pmatrix} -4 \\ 13 \end{pmatrix}$$

Hurray!

We can also turn any basis into an orthonormal basis. This is much easier to understand on an example first, so we'll do an example and then the general result.

Example 10.25: (Applying Gram-Schmidt normalisation)
Let $v_1 = \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right), v_2 = \left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right), v_3 = \left(\begin{smallmatrix} 1 \\ 0 \\ 0 \end{smallmatrix}\right)$, which is a basis of $\mathbb{R}^3$. We will, step by step, turn it into an *orthonormal* basis of $\mathbb{R}^3$.

- Start with $\widetilde{u_1} = v_1 = \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right)$.
- Now, using the way we can determine coordinates for an orthonormal basis (Prop. 10.23), we want to take off the component of v_2 which is in direction $\widetilde{u_1}$:

$$\widetilde{u_2} = v_2 - \frac{\langle v_2, \widetilde{u_1} \rangle}{\|\widetilde{u_1}\|^2} \widetilde{u_1} = \left(\begin{smallmatrix} 1 \\ 1 \\ 0 \end{smallmatrix}\right) - \frac{2}{3} \left(\begin{smallmatrix} 1 \\ 1 \\ 1 \end{smallmatrix}\right) = \begin{pmatrix} \frac{1}{3} \\ \frac{1}{3} \\ -\frac{2}{3} \end{pmatrix} = \frac{1}{3} \left(\begin{smallmatrix} 1 \\ 1 \\ -2 \end{smallmatrix}\right)$$

(We need the $\frac{1}{\|\widetilde{u_1}\|^2}$ there because we have not normalised $\widetilde{u_1}$ yet.)

> "Be amazed" interlude: This works because of the amazing fact about coordinates for orthonormal bases. We don't even know the whole basis $\widetilde{u_1}, \widetilde{u_2}, \widetilde{u_3}$ yet, we're still constructing it. But *still* we can already work out the coordinate of v_2 in direction

$\widetilde{u_1}$. This is *not possible* for a "commonplace" basis!!! There you'd need to know *all* the basis vectors before you can work out any coordinates.

◇ For the third one, we take off the components in the directions of the two we have already made:

$$\widetilde{u_3} = v_3 - \frac{\langle v_3, \widetilde{u_1}\rangle}{\|\widetilde{u_1}\|^2}\widetilde{u_1} - \frac{\langle v_3, \widetilde{u_2}\rangle}{\|\widetilde{u_2}\|^2}\widetilde{u_2} = \begin{pmatrix}1\\0\\0\end{pmatrix} - \frac{1}{3}\begin{pmatrix}1\\1\\1\end{pmatrix} - \frac{1}{2}\begin{pmatrix}\frac{1}{3}\\\frac{1}{3}\\-\frac{2}{3}\end{pmatrix} = \begin{pmatrix}\frac{1}{2}\\-\frac{1}{2}\\0\end{pmatrix} = \frac{1}{2}\begin{pmatrix}1\\-1\\0\end{pmatrix}$$

Now we can check that

◇ $\langle \widetilde{u_1}, \widetilde{u_2}\rangle = 0$

◇ $\langle \widetilde{u_1}, \widetilde{u_3}\rangle = 0$

◇ $\langle \widetilde{u_2}, \widetilde{u_3}\rangle = 0$

(Compare with the orthogonal set of $\mathbb{R}^n$ in the example c after that definition.)

Geometric explanation:
Suppose you arrange a blackboard so that v_1 goes straight up in the blackboard surface, and v_2 goes at an angle to v_1 but is also in the blackboard surface. (I.e. we put the blackboard in the position of the surface spanned by v_1, v_2.)
Then the plane orthogonal to v_1 is the plane that is horizontal (to the floor) and 90° to the blackboard. Imagine taking a piece of paper to stick out in front of the blackboard. So since we want an orthogonal basis, we want all our other vectors to be in the plane of that piece of paper.
v_2 is somewhere in the blackboard plane, and we get $\widetilde{u_2}$ by "projecting v_2 onto the piece of paper". Imagine shining a light from exactly above the blackboard, and the shadow the vector v_2 makes on the paper is $\widetilde{u_2}$.
v_3 is lurking around somewhere in space, imagine it as an arrow which is sticking out at any old angle from the point where v_1, v_2 start. So we first want to project it onto the piece of paper (shine a light from above the blackboard, take the shadow of the arrow on the paper). This corresponds to "taking $-\frac{\langle v_3, \widetilde{u_2}\rangle}{\|\widetilde{u_1}\|^2}\widetilde{u_1}$".
Now look at the piece of paper which is sticking out of the blackboard from above. It has one vector $\widetilde{u_2}$ drawn on it, and another vector $w = v_3 - \frac{\langle v_3, \widetilde{u_1}\rangle}{\|\widetilde{u_1}\|^2}\widetilde{u_1}$ drawn on it (the projection of the arrow onto the piece of paper). So now draw the line on the paper which is perpendicular to $\widetilde{u_2}$, and project w onto that line: that corresponds to "taking $-\frac{\langle v_3, \widetilde{u_2}\rangle}{\|\widetilde{u_2}\|^2}\widetilde{u_2}$", and gives us $\widetilde{u_3}$.

So now we just have to normalise them all to get

$$\begin{aligned}
u_1 &= \frac{1}{\|\widetilde{u_1}\|}\widetilde{u_1} = \frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix},\\
u_2 &= \frac{1}{\|\widetilde{u_2}\|}\widetilde{u_2} = \frac{1}{-\frac{\sqrt{6}}{3}}\begin{pmatrix}\frac{1}{3}\\\frac{1}{3}\\-\frac{2}{3}\end{pmatrix} = \frac{1}{\sqrt{6}}\begin{pmatrix}1\\1\\-2\end{pmatrix},\\
u_3 &= \frac{1}{\|\widetilde{u_3}\|}\widetilde{u_3} = \frac{1}{\frac{1}{\sqrt{2}}}\begin{pmatrix}\frac{1}{2}\\-\frac{1}{2}\\0\end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\-1\\0\end{pmatrix}
\end{aligned}$$

and we have an orthonormal basis. Notice the special thing about this basis:

$$\mathrm{Span}(u_1) = \mathrm{Span}(v_1), \qquad \mathrm{Span}(u_1, u_2) = \mathrm{Span}(v_1, v_2), \qquad \mathrm{Span}(u_1, u_2, u_3) = \mathrm{Span}(v_1, v_2, v_3).$$

This is nice, because we may have chosen v_1 to be a special direction, e.g. an eigenvector direction or something. And we may have chose the "blackboard plane" to be special as well. So this algorithm preserves us these directions/planes.

So here it is in general.

Theorem 10.26: (Gram-Schmidt normalisation)
Given a basis $v_1, v_2, \ldots, v_n$, an orthonormal basis $u_1, \cdots, u_n$ can be constructed with the property that $\mathrm{Span}(v_1, \cdots, v_k) = \mathrm{Span}(u_1, \cdots, u_k)$ *for any $k = 1, \cdots, n$, by the following inductive steps:*

- $\widetilde{u_1} = v_1$
- $\widetilde{u_2} = v_2 - \frac{\langle v_2, \widetilde{u_1}\rangle}{\|\widetilde{u_1}\|^2}\widetilde{u_1}$ *(take off the component of v_2 which is in direction $\widetilde{u_1}$)*
- $\widetilde{u_3} = v_3 - \frac{\langle v_3, \widetilde{u_1}\rangle}{\|\widetilde{u_1}\|^2}\widetilde{u_1} - \frac{\langle v_3, \widetilde{u_2}\rangle}{\|\widetilde{u_2}\|^2}\widetilde{u_2}$ *(take off components in directions $\widetilde{u_1}$ and $\widetilde{u_2}$)*
- $\vdots$
- $\widetilde{u_n} = v_n - \sum_{i=1}^{n-1} \frac{\langle v_n, \widetilde{u_i}\rangle}{\|\widetilde{u_i}\|^2}\widetilde{u_i}$ *(take off the components in previous directions)*
- *For each k, $u_k = \frac{1}{\|\widetilde{u_k}\|}\widetilde{u_k}$.* *(normalise all the vectors)*

Proof. Not lectured, so won't be in exam, but is part of your education.
We first check that $\widetilde{u_k}, \widetilde{u_l}$ are orthogonal, for $k \neq l$. Let's start with 1 and 2:

$$\begin{aligned}
\langle \widetilde{u_1}, \widetilde{u_2}\rangle &= \langle \widetilde{u_1}, (v_2 - \frac{\langle v_2, \widetilde{u_1}\rangle}{\|\widetilde{u_1}\|^2}\widetilde{u_1})\rangle \\
&= \langle \widetilde{u_1}, v_2\rangle - \frac{\langle v_2, v_1\rangle}{\|v_1\|^2}\langle \widetilde{u_1}, \widetilde{u_1}\rangle && \text{(linear in second entry)} \\
&= \langle \widetilde{u_1}, v_2\rangle - \langle v_2, \widetilde{u_1}\rangle && \text{(def of norm)} \\
&= 0 && \text{(symmetric)}
\end{aligned}$$

Now suppose that for a given $k \leqslant n$, the vectors $\widetilde{u_1}, \ldots, \widetilde{u_k}$ are all orthogonal to each other, i.e. $\langle \widetilde{u_i}, \widetilde{u_j}\rangle = 0$ for all $i \neq j \leqslant k$. Then we look at $\widetilde{u_{k+1}}$:

$$\begin{aligned}
\langle \widetilde{u_l}, \widetilde{u_{k+1}}\rangle &= \left\langle \widetilde{u_l}, \left(v_{k+1} - \sum_{i=1}^{k} \frac{\langle v_{k+1}, \widetilde{u_i}\rangle}{\|\widetilde{u_i}\|^2}\widetilde{u_i}\right)\right\rangle \\
&= \langle \widetilde{u_l}, v_{k+1}\rangle - \sum_{i=1}^{k} \frac{\langle v_{k+1}, \widetilde{u_i}\rangle}{\|\widetilde{u_i}\|^2}\langle \widetilde{u_l}, \widetilde{u_i}\rangle && \text{(linear in second entry)} \\
&= \langle \widetilde{u_l}, v_{k+1}\rangle - \frac{\langle v_{k+1}, \widetilde{u_l}\rangle}{\|\widetilde{u_l}\|^2}\langle \widetilde{u_l}, \widetilde{u_l}\rangle && \text{(by assumption the others are 0)} \\
&= 0
\end{aligned}$$

So by induction, all the $\widetilde{u_i}$ are pairwise orthogonal, so their normalised version u_i form an orthonormal set.
It is clear by contruction that $\mathrm{Span}(v_1, \cdots, v_k) = \mathrm{Span}(u_1, \cdots, u_k)$ for any $k = 1, \cdots, n$. □

Examples 10.27: Not lectured, here to help you with understanding. You could view them as exercises to which you can check the answer.

a) Let's apply Gram-Schmidt to $v_1 = \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix}1\\-1\end{smallmatrix}\right)$, using the weighted inner product with $D = \left(\begin{smallmatrix}3&0\\0&2\end{smallmatrix}\right)$.

$$\begin{aligned}
\widetilde{u_1} &= v_1 = \left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) \\
\widetilde{u_2} &= v_2 - \frac{\langle v_2, \widetilde{u_1}\rangle}{\|\widetilde{u_1}\|^2}\widetilde{u_1} = \left(\begin{smallmatrix}1\\-1\end{smallmatrix}\right) - \frac{1}{5}\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) = \frac{1}{5}\left(\begin{smallmatrix}4\\-6\end{smallmatrix}\right)
\end{aligned}$$

We see that $\langle \widetilde{u_1}, \widetilde{u_2}\rangle = \frac{1}{5}(3 \cdot 1 \cdot 4 - 2 \cdot 1 \cdot 6) = 0$. So we just normalise:

$$\begin{aligned}
u_1 &= \frac{1}{\|\widetilde{u_1}\|}\widetilde{u_1} = \frac{1}{\sqrt{5}}\left(\begin{smallmatrix}1\\1\end{smallmatrix}\right) \\
u_2 &= \frac{1}{\|\widetilde{u_2}\|}\widetilde{u_2} = \frac{1}{\sqrt{30}}\left(\begin{smallmatrix}2\\-3\end{smallmatrix}\right).
\end{aligned}$$

b) If you apply Gram-Schmidt to the standard basis $1, x, x^2$ etc. of a polynomial space, then you get the Legendre polynomials.

Study guide.

Concept review

- ⋄ Othogonal set, orthonormal set of vectors; orthonormal basis.
- ⋄ Orthogonal set of vectors is linearly independent.
- ⋄ The special way of determining coordinates for an orthonormal basis.
- ⋄ Gram-Schmidt normalisation.

Skills

- ⋄ Determine whether a set is orthogonal or orthonormal.
- ⋄ Find coordinates of a vector with respect to an orthonormal basis.
- ⋄ Create an orthonormal basis out of a given basis, using Gram-Schmidt.

D. Complex inner product

We can generalise the inner product to complex numbers, but we have to be a little careful. What we want is to still have *real* numbers as the "length" of vectors. So we define

Definition 10.28: The **complex inner product** on $\mathbb{C}^n$ is a function $\langle -, - \rangle \colon \mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C}$ defined by

$$\begin{aligned}\langle v, w\rangle = v^T\overline{w} &= \begin{pmatrix} v_1 & v_2 & \cdots & v_n \end{pmatrix} \overline{\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}} \\ &= v_1\overline{w_1} + v_2\overline{w_2} + \cdots + v_n\overline{w_n} = \sum_{j=1}^{n} v_j\overline{w_j}.\end{aligned}$$

The **norm** (or length or magnitude) of a vector in $\mathbb{C}^n$ is defined as

$$\|v\| = \sqrt{\langle v, v\rangle} \in \mathbb{R}.$$

IMPORTANT: the norm of a vector is indeed in $\mathbb{R}$, because

$$\langle v, v\rangle = v_1\overline{v_1} + v_2\overline{v_2} \cdots v_n\overline{v_n} = |v_1|^2 + \cdots |v_n|^2 \in \mathbb{R},$$

using the modulus of complex numbers, which is a *real* number.

Example 10.29: a) If $v = \begin{pmatrix} 2 \\ -i \\ 3 \\ -5+2i \end{pmatrix}$ then

$$\|v\| = \sqrt{2^2 + (-i)i + 3^2 + (-5+2i)(-5-2i)} = \sqrt{4 + 1 + 9 + (5^2 + 2^2)} = \sqrt{43}.$$

b) If v is as above and $w = \begin{pmatrix} -3i \\ -4+i \\ -2 \\ 0 \end{pmatrix}$, then

$$\begin{aligned}\langle v, w\rangle &= 2 \cdot \overline{(-3i)} + (-i) \cdot \overline{(-4+i)} + 3 \cdot \overline{(-2)} + (-5+2i) \cdot 0 \\ &= 2 \cdot (3i) + (-i)(-4-i) + 3 \cdot (-2) + 0 \\ &= -7 + 10i.\end{aligned}$$

The complex inner product is not quite bilinear: because we have the complex conjugation on one of the entries, we only get something that is "nearly" linear in the second entry. We call the complex inner product "sesqui-linear". "Sesqui" means "one-and-a-half" in Latin, so it is "one-and-a-half linear": linear on one side, and linear but with complex conjugate scalars on the other side. You could call it "linear in the first entry, and conjugate-linear in the second entry".

Theorem 10.30: (Complex inner product properties)
The complex inner product $\langle -, - \rangle \colon \mathbb{C}^n \times \mathbb{C}^n \longrightarrow \mathbb{C}$ satisfies:

(i) $\langle \lambda u + \mu v, w \rangle = \lambda \langle u, w \rangle + \mu \langle v, w \rangle$
and $\langle v, \lambda u + \mu w \rangle = \overline{\lambda}\langle v, u \rangle + \overline{\mu}\langle v, w \rangle$ ***(sesqui-linear)***

(ii) $\langle v, w \rangle = \overline{\langle w, v \rangle}$ ***(conjugate-symmetric)***

(iii) $\langle v, v \rangle \geqslant 0$, *and* $\langle v, v \rangle = 0 \Leftrightarrow v = 0$ ***(positive definite)***

The first two properties together are also called **hermitian**.

Proof. (i) We have

$$\langle \lambda u + \mu v, w \rangle = (\lambda u + \mu v)^T \overline{w} = \lambda u^T \overline{w} + \mu v^T \overline{w}$$

and

$$\langle v, \lambda u + \mu w \rangle = v^T \overline{(\lambda u + \mu w)} = \overline{\lambda} v^T \overline{u} + \overline{\mu} v^T \overline{w} = \overline{\lambda}\langle v, u \rangle + \overline{\mu}\langle v, w \rangle.$$

(ii) $\langle v, w \rangle = v^T \overline{w} = (v^T \overline{w})^T = \overline{w}^T v = \overline{w^T \overline{v}} = \overline{\langle w, v \rangle}$.

Or do it explicitely with entries: $\langle v, w \rangle = \sum_{j=1}^n v_i j \overline{w_j} = \overline{\sum_{j=1}^n \overline{v_j} w_j} = \overline{\langle w, v \rangle}$.

(iii) $\langle v, v \rangle = \sum_{j=1}^n v_j \overline{v_j} = \sum_{j=1}^n |v_j|^2$ using the modulus of a complex number. As before, a sum of non-negative terms is non-negative, and 0 if and only if each summand is 0. □

IMPORTANT: The $\geqslant$ symbol in the last point only makes sense because in that case we are dealing with *real* numbers. We cannot "compare" complex numbers, meaning complex numbers cannot be "bigger" or "smaller" than another one.

And now we can repeat *all* the results from the previous section, they all carry over to complex norm and inner product. They will be listed hereafter for your convenience.

There is an important result about *real* matrices which needs the complex inner product for it's proof:

Theorem 10.31: (Real symmetric matrix has real evalues.)
Any real symmetric matrix only has real eigenvalues.

What we mean by this is that the characteristic polynomial factors into linear factors *over* $\mathbb{R}$, not just over $\mathbb{C}$.

Proof. View a real symmetric matrix A as a complex matrix, with (potentially complex) eigenvalue λ. Say $Av = \lambda v$ for $v \neq 0 \in \mathbb{C}^n$. Then, using the complex inner product, we have

$$\begin{aligned} v^T \overline{Av} &= v^T \overline{(\lambda v)} = \overline{\lambda} v^T \overline{v} && \text{but also, as } A \text{ real} \\ &= v^T A \overline{v} = (A^T v)^T \overline{v} = (Av)^T \overline{v} && \text{as } A \text{ symmetric} \\ &= \lambda v^T \overline{v}. \end{aligned}$$

So alltogether we get

$$\overline{\lambda} v^T \overline{v} = \lambda v^T \overline{v} \qquad \Leftrightarrow \qquad (\lambda - \overline{\lambda}) v^T \overline{v} = 0.$$

But $v \neq 0$, so $v^T \overline{v} \neq 0$, so $\lambda = \overline{\lambda}$ and so the eigenvalue λ is real. □

In fact, we can say more:

Theorem 10.32: (Real symmetric matrix is orthogonally diagonalisable.)
For any real symmetric matrix A, there is a basis of eigenvectors for A which is also an orthonormal basis.

Proof. Not examinable, but still part of your education. Important for second year. We will do this in steps:

Proposition 10.33: (Evectors for different evalues of a real symmetric matrix are orthogonal.)
If A is a real symmetric matrix with evalues $\lambda_1 \neq \lambda_2$, then eigenvectors v_1, v_2 for λ_1, λ_2 are orthogonal.

PROOF. Note that as a real symmetric matrix has real eigenvalues (Thm 10.31), it also has real eigenvectors. So we can use the real inner product here.
We have

$$\begin{aligned}\lambda_1\langle v_1, v_2\rangle &= \langle \lambda_1 v_1, v_2\rangle = \langle Av_1, v_2\rangle \\ &= (Av_1)^T v_2 = v_1^T A^T v_2 = v_1^T A v_2 && \text{as } A \text{ symmetric} \\ &= \langle v_1, Av_2\rangle = \langle v_1, \lambda_2 v_2\rangle \\ &= \lambda_2\langle v_1, v_2\rangle\end{aligned}$$

So $(\lambda_1 - \lambda_2)\langle v_1, v_2\rangle = 0$. As $(\lambda_1 - \lambda_2) \neq 0$, we must have $\langle v_1, v_2\rangle = 0$, i.e. the vectors are orthogonal. □

Then we need a result which splits off one vector (and its whole subspace):

Proposition 10.34: (Line and orthogonal complement)
Given $v \neq 0 \in V$, then $V = \langle v\rangle \oplus \langle v\rangle^\perp$.

PROOF. Here $\langle v\rangle$ is the one-dimensional subspace spanned by v, i.e. $\langle v\rangle = \{kv \mid k \in \mathbb{R}\}$. And $\langle v\rangle^\perp$ is the **orthogonal complement** of v: all vectors which are orthogonal to v. I.e. $\langle v\rangle^\perp = \{w \mid \langle v, w\rangle = 0\}$. We proved in a workbook question that this set of all vectors orthogonal to v is indeed a subspace of V.
To show this is a direct sum, we have to show that the intersection is zero:
Let $w \in \langle v\rangle \cap \langle v\rangle^\perp$. Then $w = kv$ for some k, and also $\langle v, w\rangle = 0$. So $\langle v, kv\rangle = k\langle v, v\rangle = 0$. But $\langle v, v\rangle = \|v\|^2 \neq 0$ as $v \neq 0$, so $k = 0$, so $w = 0$. So the intersection of these two spaces is only zero. $\langle v\rangle \cap \langle v\rangle^\perp = 0$.
So the sum is a direct sum. To show that we get all of V, we look at dimensions:
$\langle v\rangle^\perp$ is the kernel of the linear map $\langle v, -\rangle \colon \mathbb{R}^n \longrightarrow \mathbb{R}$. The image is all of $\mathbb{R}$ (as long as $v \neq 0$, which it is). This is because we can get any real number by using $\langle v, kv\rangle = k\langle v, v\rangle$. So by rank-nullity, we have

$$\dim(\langle v\rangle^\perp) + \dim(\mathbb{R}) = n$$

which gives $\dim\langle v\rangle^\perp = n - 1$. So the dimension of the direct sum is

$$\dim(\langle v\rangle \oplus \langle v\rangle^\perp) = \dim(\langle v\rangle) + \dim(\langle v\rangle^\perp) = 1 + n - 1 = n$$

and so this direct sum is all of V. □

Using this result, we can split off one eigenvector and use induction.

PROOF OF THEOREM 10.32. A is a real symmetric matrix, so it has real eigenvalues. Let λ be an eigenvalue with eigenvector v. Then $V = \langle v\rangle \oplus \langle v\rangle^\perp$, by Prop. 10.34.
From Prop. 10.33 we know that all other existing eigenvectors are in $\langle v\rangle^\perp$, so it makes sense to split off V in this fashion. We now look at how A works on $\langle v\rangle^\perp$, so that we can use induction (as this space has one dimension less than V).
Given $w \in \langle v\rangle^\perp$, we want to show that $Aw \in \langle v\rangle^\perp$ as well. As A is symmetric, we have $A = A^T$. So

$$\langle v, Aw\rangle = v^T Aw = v^T A^T w = (Av)^T w = \langle Av, w\rangle = \langle \lambda v, w\rangle = \lambda\langle v, w\rangle = 0.$$

So $Aw \in \langle v\rangle^\perp$. This means we can view A as a linear map $A \colon \langle v\rangle^\perp \longrightarrow \langle v\rangle^\perp$, on a space of dimension $n-1$. By induction hypothesis, $\langle v\rangle^\perp$ has an orthonormal basis of eigenvectors of A. So if we add v (normalised to length 1), then we have an orthonormal basis of eigenvectors for A as a basis of V. Note that v is orthogonal to all the other vectors we have in this basis, since they are in the orthogonal complement of v.

The induction base case is when $\dim(V) = 1$: then we have just one eigenvector, and normalising it gives an orthonormal basis of the one-dimensional space.
Hence every real symmetric matrix is diagonalisable, with an orthonormal basis of eigenvectors. $\square$

Examples 10.35: Here are two examples. (We did these in the live lecture rather than on pre-recorded videos.)

- Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 2 & 2 \\ 3 & 2 & 1 \end{pmatrix}$. This is a real symmetric matrix, so we know it is diagonalisable without having to do any calculations.

  It has characteristic polynomial

$$\begin{aligned} \chi(t) &= \begin{vmatrix} t-1 & -2 & -3 \\ -2 & t-2 & -2 \\ -3 & -2 & t-1 \end{vmatrix} \\ &= (t-1)(t-2)(t-1) - 12 - 12 - 9(t-2) - 4(t-1) - 4(t-1) \\ &= (t^2 - 2t + 1)(t-2) - 24 - 17t + 26 \\ &= t^3 - 4t^2 - 12t \\ &= t(t-6)(t+2) \end{aligned}$$

  So the eigenvalues are 0, 6 and -2. (Notice: even if we did not know from A being symmetric, now we definitely know that A is diagonalisable, because it has 3 distinct evalues.)

  To work out the evector for $\lambda = 6$: we need to take the matrix $A - 6I$ to RREF, i.e.

$$\begin{pmatrix} -5 & 2 & 3 \\ 2 & -4 & 2 \\ 3 & 2 & -5 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -\frac{2}{5} & -\frac{3}{5} \\ 0 & -\frac{16}{5} & \frac{16}{5} \\ 0 & \frac{16}{5} & -\frac{16}{5} \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & -\frac{2}{5} & -\frac{3}{5} \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix}$$

  So the eigenvector is $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$.

  The eigenvector for $\lambda = -2$ is $\begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix}$, and for $\lambda = 0$ it is $\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$.

  We see that these three eigenvectors form an orthogonal set. But they are not an orthonormal set, as they don't have length 1. The "orthogonal" comes automatically here, but we have to normalise them ourselves to get the orthonormal basis of eigenvectors mentioned in Theorem 10.32.

- Let $A = \begin{pmatrix} \frac{7}{3} & 0 & -\frac{\sqrt{2}}{3} \\ 0 & 2 & 0 \\ -\frac{\sqrt{2}}{3} & 0 & \frac{8}{3} \end{pmatrix}$. This is also real symmetric, so we know it is diagonalisable.

  The characteristic polynomial is $\chi(t) = (t-2)^2(t-3)$. So we have a repeated eigenvalue: in the usual case, when we don't have the additional information of it being a symmetric matrix, we would have to work out the geometric multiplicities to decide whether this matrix is diagonalisable. So here the result about real symmetric matrices really gives us new information.

  We can work out that the eigenvectors are:

  for $\lambda = 2$ we have $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} \sqrt{2} \\ 0 \\ 1 \end{pmatrix}$;

for $\lambda = 3$ we have $\begin{pmatrix} -1 \\ 0 \\ \sqrt{2} \end{pmatrix}$.

But as for $\lambda = 2$, we have a two-dimensional eigenspace, we could also have chosen:

for $\lambda = 2$ we take $v_1 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $v_2 = \begin{pmatrix} \sqrt{2} \\ 1 \\ 1 \end{pmatrix}$;

for $\lambda = 3$ we take $v_3 = \begin{pmatrix} -1 \\ 0 \\ \sqrt{2} \end{pmatrix}$.

The first option gives an orthogonal set (which we can normalise), but the second option is not orthogonal. While v_1 and v_3 are orthogonal and v_2 and v_3 are orthogonal, because they come from different eigenvalues, we see that v_1 and v_2 are not orthogonal. Because we have a two-dimensional eigenspace, we have to *choose* the basis for this space to be orthogonal; that part is not automatic.

But we can make such a choice, so we can get an orthonormal basis of eigenvectors, as mentioned in Theorem 10.32.

Notice this is actually not unique: here is another orthogonal set of eigenvectors (which we can normalise):

for $\lambda = 2$ we take $\begin{pmatrix} \sqrt{2} \\ \sqrt{3} \\ 1 \end{pmatrix}$ and $\begin{pmatrix} \sqrt{2} \\ -\sqrt{3} \\ 1 \end{pmatrix}$;

for $\lambda = 3$ we take $v_3 = \begin{pmatrix} -1 \\ 0 \\ \sqrt{2} \end{pmatrix}$.

For a two-dimensional space, we can choose different orthogonal bases.

(Study guide at end)
For your convenience, a list of results for the complex inner product:

Corollary 10.36: ("Properties of (complex) norm")
The complex norm on $\mathbb{C}^n$ satisfies:

(i) $\|v\| \geqslant 0$.
(ii) $\|v\| = 0 \Leftrightarrow v = 0$.
(iii) $\|\lambda v\| = |\lambda|\|v\|$.

PROOF. Exercise: use complex inner product properties. □

It is often useful to remember

$$\boxed{\|v\|^2 = \langle v, v\rangle.}$$

Definition 10.37: A **unit vector** is a vector of norm 1. Given $v \in \mathbb{C}^n$, if $v \neq 0$, we can **normalise** v to the unit vector $\frac{1}{\|v\|}v$.

Examples 10.38: a) The standard basis vectors of $\mathbb{C}^n$ are all unit vectors.
b) If $v = \begin{pmatrix} 2 \\ -i \\ 3 \\ -5+2i \end{pmatrix}$ as above, then its normalisation is $u = \frac{1}{\|v\|}v = \frac{1}{\sqrt{43}}\begin{pmatrix} 2 \\ -i \\ 3 \\ -5+2i \end{pmatrix}$.

c) If $v = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{C}^n$, then its normalisation is still $\frac{1}{\sqrt{n}} v$.

d) If $w = \begin{pmatrix} i \\ i \\ \vdots \\ i \end{pmatrix} \in \mathbb{C}^n$, then $w = iv$ for the v in the previous example, and so $\|w\| = |i|\|v\| = \|v\|$, so the normalisation is $\frac{1}{\sqrt{n}} w$.

Definition 10.39: Two vectors $v, w \in \mathbb{C}^n$ are called **orthogonal** exactly when $\langle v, w \rangle = 0$.

Examples 10.40: a) The standard basis vectors are pairwise orthogonal.
b) $v_1 = \left(\begin{smallmatrix} 1 \\ i \end{smallmatrix}\right)$ and $v_2 = \left(\begin{smallmatrix} i \\ 1 \end{smallmatrix}\right)$ are orthogonal in $\mathbb{C}^2$:

$$\langle v_1, v_2 \rangle = 1 \cdot \bar{i} + i \cdot \bar{1} = -i + i = 0.$$

As in $\mathbb{R}$, we can use the inner product to determine vectors:

Proposition 10.41: (Inner product detects 0)
For given $v \in \mathbb{C}^n$, $\langle v, w \rangle = 0$ for all $w \in \mathbb{C}^n$ if and only if $v = 0$.

Proof. Same as for $\mathbb{R}$. □

We still have Cauchy-Schwarz:

Theorem 10.42: (Cauchy-Schwarz inequality)
For $u, v \in \mathbb{C}^n$, we have

$$|\langle u, v \rangle|^2 \leqslant \|u\|^2 \|v\|^2$$

($\Rightarrow |\langle u, v \rangle| \leqslant \|u\|\|v\|$)

Proof. Same proof as for $\mathbb{R}$, because all the quantities above are *real* numbers. Notice we can't go to $\langle u, v \rangle \leqslant \|u\|\|v\|$ without modulus, because that is (potentially) a complex number and can't be compared. □

And again we get:

Corollary 10.43: (Triangle inequality for norm)
Given $v, w \in \mathbb{R}^n$, then $\|v + w\| \leqslant \|v\| + \|w\|$.

Proof. Same as for $\mathbb{R}$. □

We also still have orthogonal and orthonormal bases, and Gram-Schmidt normalisation.

Study guide.

Concept review

- ◇ Complex inner product and norm, and relationship between the two.
- ◇ Complex inner product properties: it's sesqui-linear, conjugate-symmetric, positive definite.
- ◇ Real symmetric matrix has real eigenvalues.
- ◇ Real symmetric matrix has orthonormal basis of eigenvectors (without proof).
- ◇ Properties of inner product and norm from before, now for complex inner product.

Skills

- ◇ Work out norm of a complex vector.
- ◇ Work out inner product of two complex vectors.
- ◇ Use properties of complex inner product for expansion/manipulation.
- ◇ Normalise a complex vector to be a unit vector.
- ◇ Determine whether complex vectors are orthogonal.

E. Study guide collation

Just putting together all the study guides from the different sections.

Concept review.

◇ Standard (Euclidean) Inner Product, norm, and the connection between the two.
◇ Inner product properties: it's bilinear, symmetric, positive definite.
◇ Norm properties that we expect from a "length".
◇ Unit vectors, normalisation of vectors.
◇ Orthogonal vectors. Hyperplane orthogonal to a vector.
◇ Inner product detects 0.
◇ Cauchy-Schwarz inequality.
◇ Triangle inequality of norm.
◇ General inner product.
◇ Examples of weighted inner product, matrix inner product, integral inner product.
◇ Othogonal set, orthonormal set of vectors; orthonormal basis.
◇ Orthogonal set of vectors is linearly independent.
◇ The special way of determining coordinates for an orthonormal basis.
◇ Gram-Schmidt normalisation.
◇ Complex inner product and norm, and relationship between the two.
◇ Complex inner product properties: it's sesqui-linear, conjugate-symmetric, positive definite.
◇ Real symmetric matrix has real eigenvalues.
◇ Real symmetric matrix has orthonormal basis of eigenvectors (without proof).
◇ Properties of inner product and norm from before, now for complex inner product.

Skills.

◇ Work out norm of a given vector.
◇ Work out inner product of two given vectors.
◇ Use properties of inner product for expansion/manipulation.
◇ Normalise a vector to be a unit vector.
◇ Determine whether vectors are orthogonal.
◇ Work out norm of a given vector for different examples of (non-Euclidean) inner products.
◇ Work out inner product of two given vectors for different examples of inner products.
◇ Normalise a vector to be a unit vector with respect to different examples of inner products.
◇ Determine whether vectors are orthogonal with respect to different examples of inner products.
◇ Determine whether a set is orthogonal or orthonormal.
◇ Find coordinates of a vector with respect to an orthonormal basis.
◇ Create an orthonormal basis out of a given basis, using Gram-Schmidt.
◇ Work out norm of a complex vector.
◇ Work out inner product of two complex vectors.
◇ Use properties of complex inner product for expansion/manipulation.
◇ Normalise a complex vector to be a unit vector.
◇ Determine whether complex vectors are orthogonal.