Linear Algebra

Jonny Evans

May 18, 2021

Contents

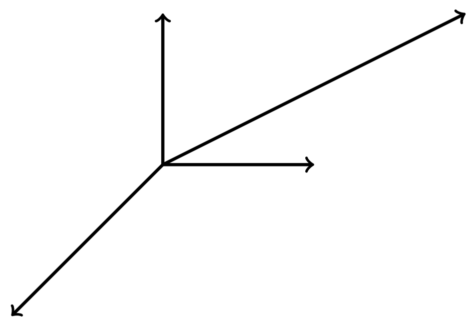
1 Week 1, Session 1: Matrices and transformations

1.1 Matrices

1.1.1 Vectors in the plane

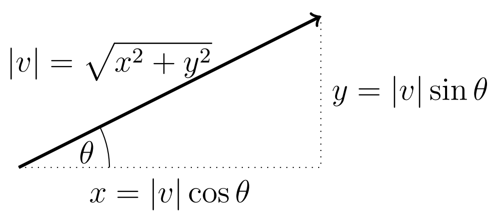
A vector is an arrow in the plane (later we'll deal with vectors in higher-dimensional spaces). We encode this arrow as a pair of numbers $\begin{pmatrix} x \\ y \end{pmatrix}$. The number x tells us how far to the right the arrow points; the number y tells us how far upwards it points. If the arrow points to the *left* then x is negative; if it points downwards then y is negative.

Can you match up the vectors $\begin{pmatrix} x \\ y \end{pmatrix}$ with those in the diagram? (Some of the vectors are not depicted).



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

A lot of this module will focus on the interplay between algebra (like column vectors) and geometry (like arrows in the plane).



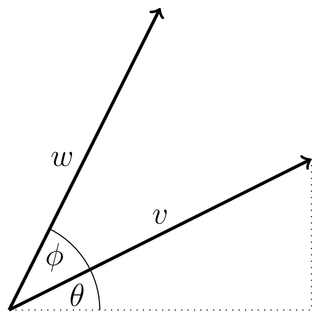
Suppose $v = \begin{pmatrix} x \\ y \end{pmatrix}$ is a vector. What is it's length? By Pythagoras's theorem, it's $\sqrt{x^2 + y^2}$. I'll write this as $|v|$, which you can read out loud as "norm v". The angle that v makes with the horizontal is $\theta = \arctan(y/x)$ (by trigonometry). If we want to write x and y in terms of $|v|$ and θ, we get (again, using trigonometry):

$$x = |v| \cos\theta, \qquad y = |v| \sin\theta.$$

So $v = \begin{pmatrix} |v| \cos\theta \\ |v| \sin\theta \end{pmatrix}$.

1.1.2 2-by-2 matrices

What happens if I rotate v by an angle ϕ anticlockwise? We get a new vector w, which we can express in terms of v and ϕ.



Rotation preserves lengths, so $|w| = |v|$.

The angle that w makes with the horizontal is $\theta + \phi$. Therefore

$$w = \begin{pmatrix} |v|\cos(\theta+\phi) \\ |v|\sin(\theta+\phi) \end{pmatrix}.$$

We can expand this using the trigonometric addition formulae:

$$w = \begin{pmatrix} |v|\cos(\theta+\phi) \\ |v|\sin(\theta+\phi) \end{pmatrix} = \begin{pmatrix} |v|\cos\theta\cos\phi - |v|\sin\theta\sin\phi \\ |v|\sin\theta\cos\phi + |v|\cos\theta\sin\phi \end{pmatrix}.$$

Using $x = |v|\cos\theta$ and $y = |v|\sin\theta$, we get

$$w = \begin{pmatrix} x\cos\phi - y\sin\phi \\ x\sin\phi + y\cos\phi \end{pmatrix}.$$

This expresses w in terms of the original vector $v = \begin{pmatrix} x \\ y \end{pmatrix}$ and the angle ϕ of rotation. We now invent a piece of notation which separates out the dependence of w on v from its dependence on ϕ: we write

$$w = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

You can just think of this as a shorthand for $\begin{pmatrix} x\cos\phi - y\sin\phi \\ x\sin\phi + y\cos\phi \end{pmatrix}$, keeping track of where all the coefficients sit.

More generally, given a 2-by-2 array of numbers $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and a vector $v = \begin{pmatrix} x \\ y \end{pmatrix}$, we define the product

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

This defines the *action of a matrix on a vector*. This notation completely separates out the rotation (M) from the vector we started with (v).

Now we don't have to limit ourselves to rotations: any matrix $\begin{pmatrix} a & b \\ x & d \end{pmatrix}$ defines a geometric transformation of the plane. This is the transformation

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

We'll see lots of examples in the next video (rotations, reflections, shears,...).

1.1.3 Mnemonic

How do you remember the formula for a matrix acting on a vector? The mnemonic I like is as follows. To get the first entry of Av, you "multiply the top row of A into v", that is you perform the multiplications ax and by (working across the top row of A and down the column of v) and sum them.

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

To get the second entry, you multiply the second row of A into v.

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

In the next video, we'll see lots of examples of transformations of the plane coming from 2-by-2 matrices.

1.2 Matrices: examples

1.2.1 Vectors in the plane

In the last video, we saw that a 2-by-2 matrix of numbers $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ defines a geometric transformation of the plane $\mathbf{R}^2$:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}.$$

Remark 1.1. Recall that $\mathbf{R}$ denotes the real number line. $\mathbf{R}^2$ denotes the 2-dimensional plane of all column vectors of height 2 (i.e. $\begin{pmatrix} x \\ y \end{pmatrix}$); $\mathbf{R}^3$ denotes the 3-dimensional space of all column vectors of height 3 (i.e. $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$; and more generally, $\mathbf{R}^n$ denotes the n-dimensional space of all column vectors of height n (i.e. $\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$).

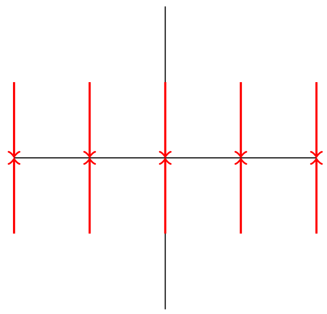
Just as the coordinates $\begin{pmatrix} x \\ y \end{pmatrix}$ encode points in the plane, we should think of the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ as encoding a transformation of the plane. In this lecture, we will take a range of examples and see what the corresponding transformation looks like.

1.2.2 Example 1

Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. If I apply A to $v = \begin{pmatrix} x \\ y \end{pmatrix}$ then I get

$$Av = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ 0 \end{pmatrix}.$$

This takes v to the point on the x-axis with the same x-coordinate, so A represents a vertical projection map to the x-axis.



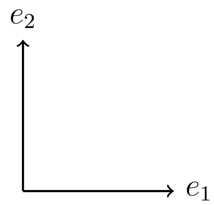
1.2.3 Example 2

Consider the action of $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. This sends $\begin{pmatrix} x \\ y \end{pmatrix}$ to $\begin{pmatrix} x \\ y \end{pmatrix}$; this transformation leaves everything as it was: it is called the *identity transformation*. We call this matrix the *identity matrix*, and we often write this matrix as I; it plays the role of the number 1 in the algebra of matrices.

1.2.4 Useful lemma

Lemma 1.2. *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, *let* $e_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ *and* $e_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. *Then*

- Ae_1 *is the first column of* A, *i.e.* $\begin{pmatrix} a \\ c \end{pmatrix}$.
- Ae_2 *is the second column of* A, *i.e.* $\begin{pmatrix} b \\ d \end{pmatrix}$.



We'll call e_1, e_2 basis vectors, *which basically means that any other vector can be written as a combination of* e_1 *and* e_2 *in a unique way. More on this in MATH220.*

Proof. We'll just check it for Ae_1:

$$Ae_1 = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} a+0 \\ c+0 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix}.$$

The calculation for Ae_2 is similar.

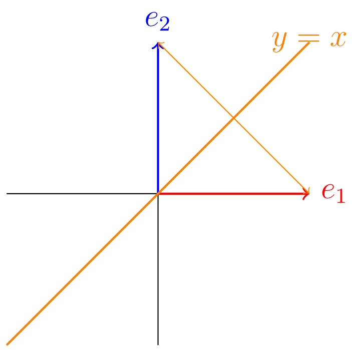
□

1.2.5 Example 3

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

- Where does e_1 go? It goes to the first column of A, which is e_2.
- Where does e_2 go? It goes to the second column of A, which is e_1.

So e_1 and e_2 get switched. This corresponds to a reflection in the line $y = x$:



Let's check that the line $y = x$ is indeed fixed by the action of A. The vectors $\begin{pmatrix} x \\ x \end{pmatrix}$ (and only these ones) lie on this line, so let's compute:

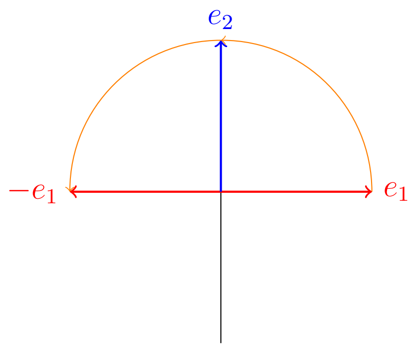
$$A \begin{pmatrix} x \\ x \end{pmatrix} = \begin{pmatrix} x \\ x \end{pmatrix},$$

which indeed tells us that the points on the line $y = x$ are fixed.

1.2.6 Example 4

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

- Where does e_1 go? It goes to the first column of A, which is e_2.
- Where does e_2 go? It goes to the second column of A, which is $-e_1$.



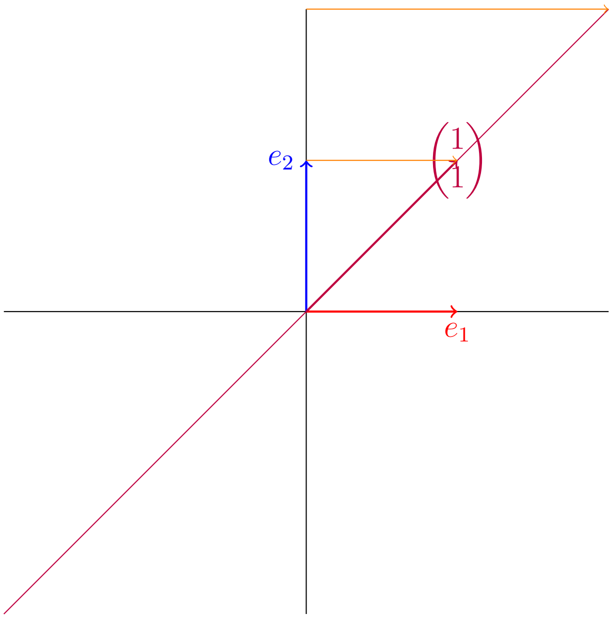
We see that this looks like a 90 degree ($\pi/2$ radian) rotation. This makes sense, because the matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ for a rotation by an angle θ specialises to A when $\theta = \pi/2$, because $\cos(\pi/2) = 0$ and $\sin(\pi/2) = 1$.

1.2.7 Example 5

Take $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. We have

- $e_1 \mapsto e_1$,
- $e_2 \mapsto \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

So e_1 is fixed, but e_2 is slanted over in the x-direction. In fact, the whole y-axis gets slanted in the x-direction, for example if we compute $A\begin{pmatrix} 0 \\ 2 \end{pmatrix}$ we get $\begin{pmatrix} 2 \\ 2 \end{pmatrix}$



1.2.8 Example 6

As one final example, we'll take $A = \begin{pmatrix} -3 & 16 \\ -1 & 5 \end{pmatrix}$. What on earth does this correspond to? I claim that it corresponds to a shear in a different direction. How can we find the fixed direction?

If $v = \begin{pmatrix} x \\ y \end{pmatrix}$ points in the direction fixed by A then $v = Av$ (that's what it means to be fixed). Therefore

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -3 & 16 \\ -1 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

In other words, the first entries of v and Av must coincide, and so must the second entries. This gives us a pair of linear simultaneous equations:

$$x = -3x + 16y, \qquad y = -x + 5y.$$

These are both equivalent to $y = x/4$, so the line $y = x/4$ is fixed.

Remark 1.3. Not all matrices have fixed directions, but if they do then this method will find it.

1.2.9 Outlook

In the next video, we will take a look at bigger matrices and higher-dimensional spaces.

1.3 Bigger matrices

1.3.1 Bigger matrices

Just as a 2-by-2 matrix defines a transformation of the plane, an m-by-n matrix defines a transformation $\mathbf{R}^n \to \mathbf{R}^m$. An m-by-n matrix is a rectangular array of numbers with m rows and n columns.

Example 1.4. $\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$ is a 2-by-3 matrix.

The transformation $\mathbf{R}^n \to \mathbf{R}^m$ associated to an m-by-n matrix A is the map $v \mapsto Av$ where:

- $v = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$
- $A = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ A_{21} & A_{22} & \cdots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \cdots & A_{mn} \end{pmatrix}$

- Av is the vector whose jth entry is obtained by multiplying the jth row of A into the column vector v, that is

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

This vector Av has height m because there are m rows of A to multiply into the vector v.

For example,

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} ax + by + cz \\ dx + ey + fz \\ gx + hy + iz \end{pmatrix}$$

shows how a 3-by-3 matrix eats a vector of height 3 and outputs a vector of height 3.

Example 1.5. Take $A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$. We get

$$Av = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \\ z \end{pmatrix}.$$

We see that this is a rotation of 3-dimensional space which fixes the z-axis and rotates by θ in the xy-plane. We call it a *rotation by θ about the z-axis.*

Example 1.6. Take $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. We need to feed A a vector of height 3; it will output a vector of height 2. In other words, A defines a transformation $\mathbf{R}^3 \to \mathbf{R}^2$. What is the transformation?

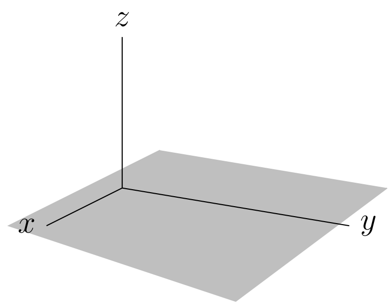
$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

This is the projection to the xy-plane (which squishes the z-axis to the origin).

Example 1.7. Take $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}$. This gives a map $\mathbf{R}^2 \to \mathbf{R^3}$:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$$

This is the *inclusion* map of the 2-dimensional xy-plane into 3-dimensional space (putting it at height zero).



These rectangular (nonsquare) matrices change the dimension of the space we're working with, e.g. map from a lower to a higher dimensional space or vice versa. You might wonder why we matrices which are bigger than 3-by-3, given that we live in a 3-dimensional universe. In fact:

- the theory of special relativity treats space and time on an equal footing, and the *Lorentz transformations*, which describe all the weird relativistic effects like time dilation and length contraction, mix up space and time, and are given by 4-by-4 matrices.
- in statistics, data is often represented as a vector of samples; the more samples you have, the bigger the dimension of the vector you need to encode them.

1.3.2 More examples

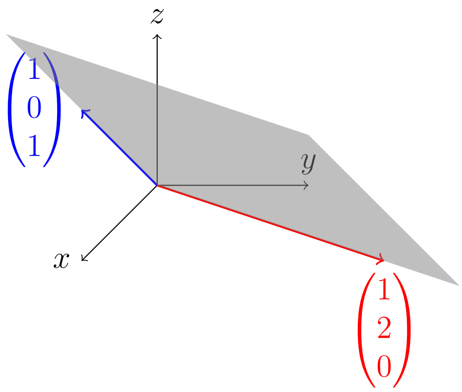
Example 1.8. Take $A = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}$. This defines a map $\mathbf{R}^2 \to \mathbf{R}^3$:

$$A \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ 2x \\ y \end{pmatrix}.$$

What does this map "look like"? Its *image* (the set of points in 3d which have the form Av for some $v \in \mathbf{R}^2$) is a plane. To visualise the plane, we and draw the images of the x- and y-axes in $\mathbf{R}^2$:

- The x-axis (vectors of the form $\begin{pmatrix} x \\ 0 \end{pmatrix}$) goes to the set of vectors $\begin{pmatrix} x \\ 2x \\ 0 \end{pmatrix}$.
- The y-axis (vectors of the form $\begin{pmatrix} 0 \\ y \end{pmatrix}$) goes to the set of vectors $\begin{pmatrix} y \\ 0 \\ y \end{pmatrix}$.

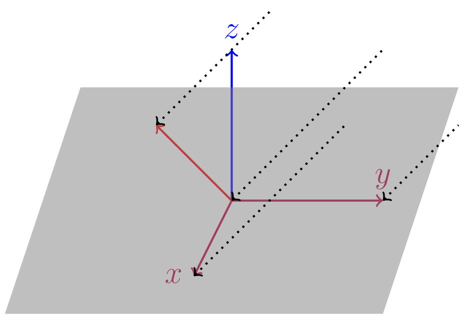
The image of A is the unique plane containing these two lines.



Example 1.9. Take $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$. This defines a map $\mathbf{R}^3 \to \mathbf{R}^2$:

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x - z \\ y - z \end{pmatrix}.$$

What does this map look like? Let's imagine it's projecting from 3-dimensional space onto the xy-plane (by including $\mathbf{R}^2$ into $\mathbf{R}^3$ as the xy-plane). The points $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ on the xy-plane go to $\begin{pmatrix} x \\ y \end{pmatrix}$ (i.e. they stay where they are). The point $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ on the z-axis goes to $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$. This means that everything is being projected onto the xy-plane; the projection is along straight line rays which point in the $\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$-direction (because to get from $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} -1 \\ -1 \\ 0 \end{pmatrix}$ you have to go backwards 1 in each of the x, y and z directions.



This line along which we're projecting has a name: it's called the *kernel* of A. More on this later.

2 Week 1, Session 2: Matrix algebra

2.1 Matrix multiplication, 1

2.1.1 Composing transformations

Recall that a 2-by-2 matrix defines for us a transformation of the plane. Suppose we are given two matrices

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}.$$

They each define a transformation of the plane. What happens if we *first* do the transformation associated to B, and *then* do the transformation associated to A?

We get a new transformation associated to a new matrix, which we call AB.

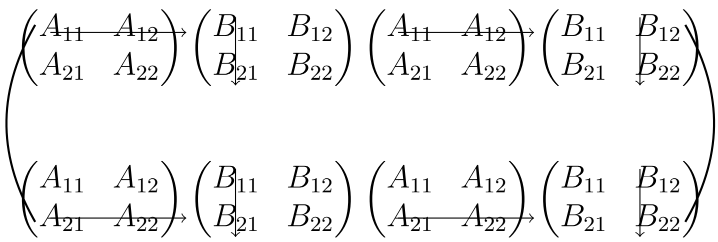
$$\begin{aligned} A(B(v)) &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11}x + B_{12}y \\ B_{21}x + B_{22}y \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11}x + A_{11}B_{12}y + A_{12}B_{21}x + A_{12}B_{22}y \\ A_{21}B_{11}x + A_{21}B_{12}y + A_{22}B_{21}x + A_{22}B_{22}y \end{pmatrix} \\ &= \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &=: (AB)v \end{aligned}$$

Definition 2.1. Given matrices $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, and $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$, we define the matrix product

$$AB = \begin{pmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{pmatrix}.$$

2.1.2 Mnemonic

How on earth can we remember this formula? Here is a mnemonic. Just like when we act on a vector using a matrix, we can think of the entries of AB as "multiplying a row of A into a column of B". More specifically, to get the ijth entry of AB (i.e. ith row and jth column) we multiply the ith row of A into the jth column of B:



2.2 Matrix multiplication, 2

2.2.1 Examples

We're going to do some examples of matrix multiplication.

Example 2.2. Consider the 90 degree rotation matrix $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. We have

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

This makes sense: two 90 degree rotations compose to give a 180 degree rotation, which sends every point $\begin{pmatrix} x \\ y \end{pmatrix}$ to its opposite point $\begin{pmatrix} -x \\ -y \end{pmatrix}$.

Example 2.3. More generally, if

$$R_\alpha = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \qquad R_\beta = \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix}$$

are two rotations then the composite is

$$\begin{aligned} R_\alpha R_\beta &= \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{pmatrix} \\ &= \begin{pmatrix} \cos\alpha\cos\beta - \sin\alpha\sin\beta & -\cos\alpha\sin\beta - \sin\alpha\cos\beta \\ \sin\alpha\cos\beta + \cos\alpha\sin\beta & -\sin\alpha\sin\beta + \cos\alpha\cos\beta \end{pmatrix} \\ &= \begin{pmatrix} \cos(\alpha+\beta) & -\sin(\alpha+\beta) \\ \sin(\alpha+\beta) & \cos(\alpha+\beta) \end{pmatrix} \\ &= R_{\alpha+\beta}. \end{aligned}$$

(using trigonometric addition formulas). This is what we expect, of course: rotating by β and then α amounts to rotating by $\alpha + \beta$.

Example 2.4. Let $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ be the identity matrix and A be any matrix. Then

$$\begin{aligned} IA &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\ &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \\ &= A. \end{aligned}$$

Similarly, $AI = A$. As you can see, the identity matrix really plays the role of the number 1 here.

Example 2.5. Let $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ and $B = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then

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

but

$$BA = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

This shows that the *order* in which we multiply matrices matters: $AB \neq BA$. So matrix multiplication is not commutative!

As an exercise, can you think of a matrix B which does not commute with $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$?

2.3 Matrix multiplication, 3

2.3.1 Multiplying bigger matrices

Suppose A is an m-by-n matrix (m rows and n columns) and B is an n-by-p matrix (n rows and p columns). To save our sanity, suppose $m = 2$, $n = 3$, $p = 4$:

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{pmatrix}, \qquad B = \begin{pmatrix} B_{11} & B_{12} & B_{13} & B_{14} \\ B_{21} & B_{22} & B_{23} & B_{24} \\ B_{31} & B_{32} & B_{33} & B_{34} \end{pmatrix}.$$

I claim that there's an obvious way to define AB given everything we've seen so far.

- To get the top left entry of AB, we multiply the top row of A into the first column of B, giving $A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31}$.
- To get the next entry along, we multiply the top row of A into the second column of B.
- We keep going: to get the entry of AB in the ith row and the jth column, we multiply the ith row of A into the jth column of B.

This means we end up with 2 rows (same number as A) and 4 columns (same number as B).

You might ask: what happens if A is m-by-n and B is k-by-p but $n \neq k$? For example:

$$\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \end{pmatrix}$$

doesn't make any sense: the rows of A have length 2 and the columns of B have height 1, so we can't multiply rows into columns.

This makes sense: A defines a transformation $\mathbf{R}^2 \to \mathbf{R}^2$ and B defines a transformation $\mathbf{R}^2 \to \mathbf{R}$, so while you can define $BA \colon \mathbf{R}^2 \to \mathbf{R}^2 \to \mathbf{R}$, you have no way of composing the transformations as AB (the domain of A is not the target of B).

As an exercise, do the following multiplications:

$$\begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

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

$$\begin{pmatrix} 1 & -1 & 1 & -1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

$$\begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 & -1 \end{pmatrix}$$

(see the video for solutions).

2.4 Index notation

2.4.1 Index notation

I can write the entries of an m-by-n matrix A as

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} & \cdots & A_{1n} \\ A_{21} & A_{22} & A_{23} & \cdots & A_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ A_{m1} & A_{m2} & A_{m3} & \cdots & A_{mn} \end{pmatrix}.$$

Here A_{ij} denotes the entry sitting in the ith row and the jth column.

One advantage of writing matrices like this is that it gives a compact formula for operations like matrix multiplication: rather than writing out the full matrix AB, we can just write a formula for the ijth entry $(AB)_{ij}$. Suppose that A is m-by-n and B is n-by-p.

To get $(AB)_{ij}$, we need to multiply the ith row of A into the jth column of B, in other words:

$$\begin{aligned} (AB)_{ij} &= \begin{pmatrix} A_{i1} & A_{i2} & \cdots & A_{in} \end{pmatrix} \begin{pmatrix} B_{1j} \\ B_{2j} \\ \vdots \\ B_{nj} \end{pmatrix} \\ &= A_{i1}B_{1j} + A_{i2}B_{2j} + \cdots + A_{in}B_{nj} \\ &= \sum_{k=1}^{n} A_{ik}B_{kj}. \end{aligned}$$

In this last step, we just introduced a "dummy index" k to keep track of the terms in the sum. We now have a nice compact formula for the ijth entry of AB: it's $\sum_{k=1}^{n} A_{ik}B_{kj}$.

2.4.2 Associativity of matrix multiplication

To demonstrate how useful index notation is, let's prove that matrix multiplication is associative, that is:

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

We can just write out the formula for the ijth entry on each side and check they give the same answer. For the left-hand side:

$$\begin{aligned}((AB)C)_{ij} &= \sum_k (AB)_{ik} C_{kj} \\ &= \sum_k \sum_\ell A_{i\ell} B_{\ell k} C_{kj}\end{aligned}$$

where we've used the formula for matrix multiplication twice (using all sorts of different letters). Note that the second time I use the formula, I can't use the letter k for my dummy index because k already means something in the expression; that's why I introduced ℓ.

For the right-hand side:

$$\begin{aligned}(A(BC))_{ij} &= \sum_k A_{ik} (BC)_{kj} \\ &= \sum_k A_{ik} \sum_\ell B_{k\ell} C_{\ell j}\end{aligned}$$

We can take the factor A_{ik} inside the sum (just by multiplying out the whole expression), which gives:

$$(A(BC))_{ij} = \sum_k \sum_\ell A_{ik} B_{k\ell} C_{\ell j}.$$

This looks very similar to the formula for the left-hand side, but the indices k and ℓ have been swapped. That doesn't matter: k and ℓ are dummy indices, so we can just rename them. We'll relabel k as ℓ and ℓ as k:

$$(A(BC))_{ij} = \sum_\ell \sum_k A_{i\ell} B_{\ell k} C_{kj}.$$

Finally, we can switch the order of the sums without worrying because they're finite sums. This gives exactly the same formula that we had on the left-hand side.

Index notation is very heavily used in subjects like general relativity. For example, the Riemann curvature tensor $R^i{}_{jk\ell}$ is an object with four indices, some up and some down!

2.5 Other operations

2.5.1 Matrix addition

In this video, we'll define some further operations you can do to produce new matrices. The first is *matrix addition* If we have two m-by-n matrices A and B with entries A_{ij} and B_{ij}, we can form a new matrix $A + B$ with

$$(A + B)_{ij} = A_{ij} + B_{ij}.$$

In other words, you take the ijth entries of both matrices and add them.

Example 2.6.

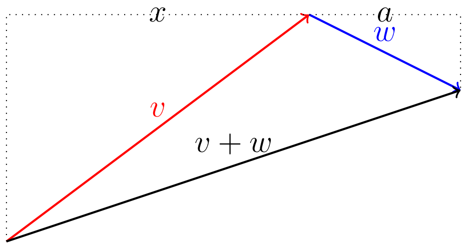
$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} + \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 0 \end{pmatrix}.$$

2.5.2 Special case: vector addition

This is most useful when A and B are both column vectors, i.e. m-by-1 matrices. Let's see what it means in for vectors in $\mathbf{R}^2$. The formula is

$$\begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} x+a \\ y+b \end{pmatrix}.$$

Geometrically, we add two vectors $v = \begin{pmatrix} x \\ y \end{pmatrix}$ and $w = \begin{pmatrix} a \\ b \end{pmatrix}$ by translating w to the tip of v and drawing the arrow from the tail of v to the tip of w. One can see from the picture that the x- (respectively y-) coordinate of this arrow is the sum of the x- (respectively y-) coordinates of v and w.



2.5.3 Rescaling

Given a number λ and a matrix A, you can form the matrix λA whose entries are λ times the entries of A.

Example 2.7. $2\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$.

2.5.4 Matrix exponentiation

The exponential of a number x is defined by the Taylor series of exp:

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

We can use the same definition to define the exponential of a matrix:

$$\exp(A) = \sum_{n=0}^{\infty} \frac{1}{n!} A^n.$$

Here, A^0 is understood to mean the identity matrix I (the analogue for matrices of the number 1).

Example 2.8. Consider $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Since $A^2 = 0$, all the higher powers of A vanish (the name for this is *nilpotence*: some power of A is zero), so the matrix exponential becomes

$$\exp(A) = I + A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

So we get the matrix for a shear as the exponential of a nilpotent matrix.

In fact,

$$\exp \begin{pmatrix} 0 & t \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix},$$

so we get a whole family of matrices which shear further and further to the right as t varies.

Example 2.9. Take $A = \begin{pmatrix} 0 & -t \\ t & 0 \end{pmatrix}$. We have

$$\begin{aligned} A &= t \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ A^2 &= \begin{pmatrix} -t^2 & 0 \\ 0 & -t^2 \end{pmatrix} = -t^2 I \\ A^3 &= -t^3 \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ A^4 &= -t^4 I \\ &\vdots \end{aligned}$$

and in the end we get

$$\begin{aligned} \exp(A) = I & \qquad +t \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ -\frac{t^2}{2!} I & \qquad -\frac{t^3}{3!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ +\frac{t^4}{4!} I & \qquad +\frac{t^5}{5!} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \\ \vdots & \qquad \vdots \end{aligned}$$

The coefficient of I is the Taylor series for $\cos t$; the coefficent of $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ is the Taylor series for $\sin t$, so overall we get

$$\exp(A) = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}.$$

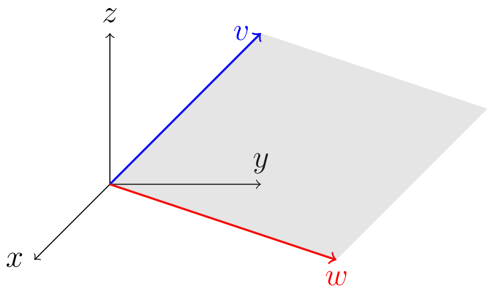
So we get a general rotation matrix in 2-d by exponentiating this very simple matrix.

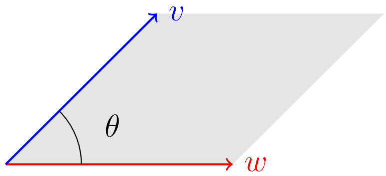
3 Week 2, Session 1: Dot products and rotations

3.1 Dot product

Given two vectors $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ and $w = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ in $\mathbf{R}^n$, what is the angle between them?

To define angles in $\mathbf{R}^n$, note that the two vectors v and w are contained in a unique 2-plane, and we mean the usual angle between v and w inside that plane.





To compute that angle, we introduce the *dot product* $v \cdot w$.

Definition 3.1. The *dot product* $v \cdot w$ is the number $v_1 w_1 + \cdots + v_n w_n$.

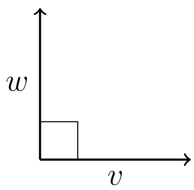
Theorem 3.2. *If θ is the angle between v and w then*

$$v \cdot w = |v||w| \cos\theta.$$

Recall that $|v|$ means the length of v.

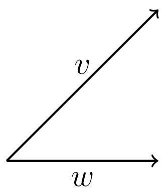
We will prove this in due course, but first we'll explore it a little.

Example 3.3. Let $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $w = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. We have $v \cdot w = 1 \times 0 + 0 \times 1 = 0$.



This makes sense: the angle between v and w is $\pi/2$ radians, and $\cos(\pi/2) = 0$. In this case, we say the vectors are *orthogonal to one another* (equivalent to "perpendicular" or "at right angles").

Example 3.4. Let $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. We can see the angle between them should be 45 degrees ($\pi/4$ radians).

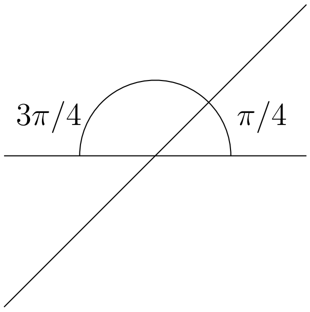


Let's confirm this: we have $v \cdot w = 1 \times 1 + 0 \times 1 = 1$. We also have $|v| = \sqrt{1+1} = \sqrt{2}$ by Pythagoras and $|w| = 1$, so

$$1 = v \cdot w = |v||w| \cos\theta = \sqrt{2}\cos\theta,$$

so $\cos\theta$ should be $1/\sqrt{2}$. Indeed, $\cos(\pi/4) = 1/\sqrt{2}$.

Even if you didn't know the angle, you could figure it out as $\arccos(1/\sqrt{2})$. You might object that arccos is multivalued, for example $\cos(\pi/4) = \cos(3\pi/4)$. This just corresponds to the fact that there are different ways of picking "the" angle between v and w (e.g. clockwise or anticlockwise).



3.2 Transposition

You may have noticed that the definition of the dot product looks a lot like matrix multiplication. In fact, it is a special case of matrix multiplication:

$$v_1w_1 + \cdots + v_nw_n = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}.$$

Technically, the matrix product gives a 1-by-1 matrix whose unique entry is the dot product, but let's not be too pedantic.

Here, we took the column vector $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ and turned it on its side to get a row vector which we call the *transpose* of v, written:

$$v^T = \begin{pmatrix} v_1 & \cdots & v_n \end{pmatrix}$$

.

More generally, you can transpose a matrix:

Definition 3.5. Given an m-by-n matrix A with entries A_{ij}, we get an n-by-m matrix A^T whose ijth entry is A_{ji}, i.e.

$$(A^T)_{ij} = A_{ji}$$

Example 3.6. $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}.$

Example 3.7. $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$.

So the rows of A become the columns of A^T.

With all this in place, we observe that the dot product $v \cdot w$ is $v^T w$.

Lemma 3.8. $(AB)^T = B^T A^T$.

Proof. Writing out the ijth entry of $(AB)^T$ using index notation, we get:

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

Similarly expanding $B^T A^T$ we get

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

The two expressions differ only by the order of the factors A_{jk} and B_{ki}.

The order of these factors doesn't matter: A_{jk} and B_{ki} are just numbers (entries of A and B), so they commute. This is one reason index notation is so convenient: it converts expressions involving noncommuting objects like matrices into expressions involving commuting quantities (numbers).

□

3.3 Orthogonal matrices

Definition 3.9. An n-by-n matrix A is *orthogonal* if $A^T A = I$.

This is equivalent to saying that:

- the columns of A, considered as column vectors, are orthogonal to one another, and
- each column has length one when considered as a vector.

Example 3.10. The matrix $A = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is orthogonal:

$$\begin{aligned} A^T A &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\theta + \sin^2\theta & -\sin\theta\cos\theta + \cos\theta\sin\theta \\ \sin\theta\cos\theta - \cos\theta\sin\theta & \cos^2\theta + \sin^2\theta \end{pmatrix} \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I. \end{aligned}$$

Lemma 3.11. *If A is orthogonal then the geometric transformation of $\mathbf{R}^n$ defined by A preserves dot products, i.e. $(Av) \cdot (Aw) = v \cdot w$.*

Proof.

$$(Av) \cdot (Aw) = v^T A^T A w = v^T I w = v^T w = v \cdot w.$$

□

In particular, orthogonal matrices preserve lengths of vectors. This is because $|v| == \sqrt{v_1^2 + \cdots + v_n^2} = \sqrt{v \cdot v}$ and orthogonal matrices preserve dot products.

The sorts of transformations you should have in mind when you think of orthogonal matrices are rotations, reflections and combinations of rotations and reflections.

We now turn to the proof of the formula

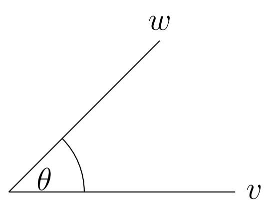
$$v \cdot w = |v||w| \cos \theta$$

where θ is the angle between v and w.

Proof. The "proof" I'll give is more of an intuitive argument. I'll point out where the dodgy bits are as we go on.

The angle between v and w is defined by taking the plane spanned by v and w and taking the angle between them inside that plane. Therefore without loss of generality, we can assume v and w are in the standard $\mathbf{R}^2$. (This is where the dodgy bit is: technically, we need to check that the plane spanned by v and w is the same, geometrically (isometric to) the standard $\mathbf{R}^2$; this would require us to appeal to something like Gram-Schmidt orthogonalisation, which you'll see in a later course on linear algebra.)

Further, we can rotate using our 2-by-2 rotation matrix so that v points in the positive x-direction. We can do this without changing the angle θ (because rotations preserve angles) and without changing the dot product (because rotations are orthogonal, so don't change dot products).



Finally, compute

$$v \cdot w = v_1 w_1 + v_2 w_2 = v_1 w_1$$

because $v_2 = 0$. We have $v_1 = |v|$ because v points along the positive x-axis. We also have $w_1 = |w| \cos \theta$. So overall we have $v \cdot w = |v||w| \cos \theta$.

□

If you were disturbed by my claim that rotations preserve angles without justification, you should have been even more disturbed by the fact that I didn't actually define angles in the plane at all. This is not intended as a complete rigorous derivation from axioms, but as an appeal to your geometric intuition.

3.4 Rotations

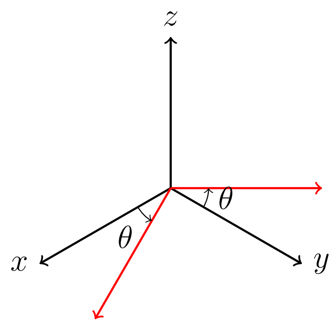
We'll analyse some examples of 3-by-3 rotation matrices, and then see to figure out the axis and angle of rotation for a general 3-by-3 rotation matrix.

3.4.1 Example 1

Let $A = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$. This is an example of a 3-by-3 rotation matrix. The top-left 2-by-2 block rotates the xy-plane, and the 1 in the bottom-right tells us that the z-axis is fixed. More precisely:

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x\cos\theta - y\sin\theta \\ x\sin\theta + y\cos\theta \\ z \end{pmatrix},$$

so the vectors $\begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ in the xy-plane get rotated by the 2-by-2 rotation matrix $\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, and the vector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ which points along the z-axis is fixed.



A key point here is that the axis of rotation is fixed, i.e. any vector u pointing in the z-direction satisfies $u = Au$.

3.4.2 Example 2

Let $B = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$. This is another example of a 3-by-3 rotation matrix. What is the axis of rotation? We need to find a vector $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ such that $u = Bu$. In other words,

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} z \\ y \\ -x \end{pmatrix}.$$

This is three equations (one for each component):

$$x = z, \qquad y = y, \qquad -x = z.$$

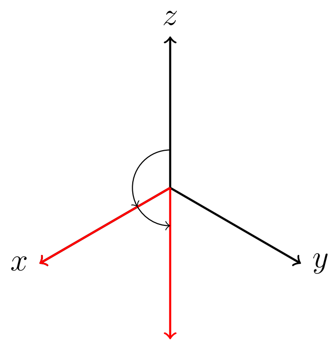
The equation $y = y$ is trivially satisfied. The other two equations imply $x = z = 0$. Therefore $u = \begin{pmatrix} 0 \\ y \\ 0 \end{pmatrix}$, and u must point along the y-axis.

By comparing with Example 1, the angle of rotation can be found by:

- taking a vector v which lives in the plane orthogonal to the axis,
- applying B to get Bv,
- computing the angle between v and Bv using dot products.

For example, we could take $v = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ (as $v \cdot u = 0$ so v is orthogonal to the axis). Then $Bv = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$, so $v \cdot Bv = 0$. If θ is the angle between v and Bv then this implies $\cos\theta = 0$, so $\theta = \pm 90$ degrees.

In fact, we can understand exactly what this rotation is doing by drawing the images of the basis vectors e_1, e_2, e_3 under B (i.e. the three columns of B). The vector e_2 is fixed, e_3 goes to e_1 and e_1 goes to $-e_3$, so this rotates by 90 degrees about the y-axis, sending the positive z-axis to the positive x-axis.



3.4.3 Example 3

Take $C = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$. This is another 3-by-3 rotation matrix; we'll find the axis and angle of rotation.

Remark 3.12. These examples are carefully chosen to be rotation matrices. Note that if I gave you a random 3-by-3 matrix, it probably wouldn't be a rotation matrix, and isn't guaranteed to have any fixed vectors at all.

To find the axis $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, we need to solve $u = Cu$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} z \\ x \\ y \end{pmatrix}.$$

The first two equations imply $x = y = z$, so the third is redundant, and any vector of the form $u = \begin{pmatrix} x \\ x \\ x \end{pmatrix}$ is fixed. In other words, the axis points in the $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$-direction.

To find the angle, pick a vector v orthogonal to u. For example, $v = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ satisfies $u \cdot v = 0$ so is orthogonal to u. We compute

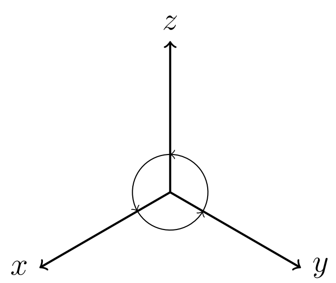
$$v \cdot Cv = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix} = -1.$$

We also know that if θ is the angle between v and Cv then $-1 = v \cdot Cv = |v||Cv| \cos\theta$. Since $|v| = |Cv| = \sqrt{2}$, we get

$$\cos\theta = -1/2.$$

This tells us that θ is something like $2\pi/3$ (or any of the other values that have $\cos\theta = -1/2$). Let's draw a picture to convince ourselves it's really $2\pi/3$ (120 degrees).

The axis of rotation points out of the screen:



4 Week 2, Session 2: Simultaneous equations

4.1 Simultaneous equations and row operations

Thus far in the course, we have focused on the geometric aspects of matrices and the transformations they determine. Now we'll approach the subject from the point of view of simultaneous equations.

4.1.1 Simultaneous equations

A system of simultaneous linear equations, for example

$$x - y = -1, \qquad x + y = 3,$$

conmprises a finite list of linear equations in some number of variables. Linear means that each term is either constant of else linear in the variables (i.e. the terms are things like $3x$ or $-4y$, not x^2 or xye^x).

A system of simultaneous linear equations is a matrix equation in disguise. For example, the system above can be written as

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

Indeed, if we multiply this expression out, we get

$$\begin{pmatrix} x - y \\ x + y \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix},$$

and these two vectors are equal if and only if their components are equal, which is equivalent to the two equations we started with.

We will abbreviate such a matrix equation by writing a so-called *augmented matrix*: we write the matrix of coefficients, then a vertical bar, then the column of constants:

$$\begin{pmatrix} 1 & -1 & | & -1 \\ 1 & 1 & | & 3 \end{pmatrix}$$

This is really just a shorthand for the equation

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

4.1.2 Solving these equations

To solve this system, we will manipulate the equations one at a time. We will see what happens to the augmented matrix as we perform these manipulations. We start with:

$$x - y = -1, \qquad x + y = 3, \qquad \begin{pmatrix} 1 & -1 & | & -1 \\ 1 & 1 & | & 3 \end{pmatrix}.$$

We can eliminate x from the second equation by subtracting equation 1 from equation 2:

$$x - y = -1, \qquad 2y = 4, \qquad \begin{pmatrix} 1 & -1 & | & -1 \\ 0 & 2 & | & 4 \end{pmatrix}.$$

This has had the effect on the augmented matrix of subtracting the first row from the second. We will write this operation as

$$R_2 \mapsto R_2 - R_1.$$

Next, we can divide equation 2 by 2:

$$x - y = -1, \qquad y = 2, \qquad \begin{pmatrix} 1 & -1 & | & -1 \\ 0 & 1 & | & 2 \end{pmatrix}.$$

This has the effect of dividing row 2 by 2, which we write as

$$R_2 \mapsto \frac{1}{2} R_2.$$

Finally, we add equation 2 to equation 1:

$$x = 1, \qquad y = 2, \qquad \begin{pmatrix} 1 & 0 & | & 1 \\ 0 & 1 & | & 2 \end{pmatrix}.$$

This is the row operation

$$R_1 \mapsto R_1 + R_2.$$

We have now solved our system of equations.

4.1.3 Row operations

This process of solving simultaneous equations can therefore be understood as performing a sequence of row operations on the augmented matrix.

Definition 4.1.

- Type I row operations: Replace row i by row i plus a multiple (say λ) of row j. We write this as $R_i \mapsto R_i + \lambda R_j$. In terms of equations, this means we're adding/subtracting a multiple of equation j to equation i.
- Type II row operations: Replace row i by a nonzero multiple (say λ) of row i. We write this as $R_i \mapsto \lambda R_i$. In terms of equations, this means we're multiplying an equation by a nonzero constant.
- Type III row operations: Swap row i and row j. This corresponds to reordering your equations.

Remark 4.2. We don't allow ourselves to multiply an equation by zero: this will change our system of equations by effectively ignoring some of them.

In our example, we "solved" the equation when we reached $x = 1, y = 2$. This meant that the augmented matrix had the identity matrix on the left of the vertical bar. So the aim of the row operations is to put the augmented matrix in the form $(I|b)$ where I is the identity matrix and b is a column vector (of "constants"). Of course, this will sometimes fail:

Example 4.3. Consider the system $x + y = 1$ (one equation, two variables). The augmented matrix is now

$$\begin{pmatrix} 1 & 1 & | & 1 \end{pmatrix}.$$

No matter what you do, you can't put this into the form $(I|b)$ because the matrix on the left hand side of the vertical bar isn't a square matrix (so can't be turned into the identity matrix by row operations). Nonetheless, the equation is easy to solve: we have $x = 1 - y$, so given any y we get a solution $\begin{pmatrix} 1-y \\ y \end{pmatrix}$. So we don't *need* to get to the identity matrix so say our system is "solved". In the next few videos, we'll see the correct thing to aim for is to put the matrix on the left of the bar into *reduced echelon form*.

In the previous video, we saw how a system of simultaneous linear equations can be encoded as an augmented matrix and how manipulations of the equations corresponded to performing row operations on the augmented matrix. In the example we worked through, we started with:

$$x - y = -1, \qquad x + y = 3, \qquad \begin{pmatrix} 1 & -1 & | & -1 \\ 1 & 1 & | & 3 \end{pmatrix}$$

and ended up with

$$x = 1, \qquad y = 2, \qquad \begin{pmatrix} 1 & 0 & | & 1 \\ 0 & 1 & | & 2 \end{pmatrix}.$$

Because the block on the left of the vertical bar in the augmented matrix is the identity, this tells us that the variables x and y are completely determined. It's not always possible to make a matrix equal to the identity by performing row operations. It is, however, always possible to put it into a special form called *reduced echelon form*. We'll see how this works in the next three videos.

4.2 Echelon form

Definition 4.4. Suppose we have an m-by-n (augmented) matrix like:

$$\begin{pmatrix} 0 & 5 & 0 & 3 & | & 2 \\ 0 & 0 & 0 & 0 & | & 1 \\ 7 & 1 & 0 & 3 & | & 4 \end{pmatrix}$$

For each nonzero row, the *leading entry* is the leftmost nonzero entry (left of the bar). In some row, if there are no nonzero entries left of the bar, then there is no leading entry in this row. In the example above, the leading entries are:

- for the first row: 5,
- for the second row: there is no leading entry (nothing nonzero to the left of the bar),
- for the third row: 7.

Example 4.5. The n-by-n identity matrix has leading entries $1, 1, 1, \ldots, 1$. These move to the right as you work down the rows.

Definition 4.6. A matrix is in *echelon form* if all the zero-rows are at the bottom and the leading entries move strictly to the right as you go down the rows.

Example 4.7. Our earlier example is not in echelon form: there's a row of zeros in the middle, which should be at the bottom. Even if that weren't a problem, its leading entries (5 and 7) move to the left as you go down the rows.

Example 4.8. The matrix $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is in echelon form. All zero-rows are at the bottom. The leading entries are (1 in the top row and 2 in the second) move to the right, so this is in echelon form. The "staircase of zeros" in the bottom left of the matrix is the reason that echelon form is called echelon form: the French word *échelle* means ladder.

Example 4.9. The matrix $\begin{pmatrix} 3 & 1 & 0 \\ 4 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$ is not in echelon form: the leading indices don't move strictly to the right (3 and 4 are in the same column instead of moving to the right.

Which of the following matrices are in echelon form?

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

Next time, we'll look at *reduced echelon form*.

4.3 Reduced echelon form

4.3.1 Definition

Definition 4.10. A matrix is in *reduced echelon form* if:

- it's in echelon form,
- all its leading entries are equal to 1,
- (*) in any column containing a leading entry, there are no other nonzero entries.

Consider the following matrices:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 1 \end{pmatrix}, \qquad C = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & -1 \end{pmatrix}$$
$$D = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 2 & 0 \end{pmatrix}, \qquad E = \begin{pmatrix} 3 & 0 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad F = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

- A is in echelon form an its leading entries are all 1. However, condition (*) doesn't hold: for example column 2 contains a leading entry (1) and another nonzero entry (the 2 above it).
- B is in echelon form, its leading entry is 1 and, in the column containing the leading entry, there are no other nonzero entries, so B is in reduced echelon form.
- C is in reduced echelon form.
- D and F are not even in echelon form.
- E is in echelon form, but its leading entries are not 1, so it is not in reduced echelon form.

4.3.2 Reason for the definition

Consider the augmented matrix

$$\begin{pmatrix} 1 & 0 & 1 & 1 & | & 5 \\ 0 & 1 & 2 & -1 & | & 13 \end{pmatrix}.$$

This corresponds to the system of simultaneous equations:

$$w + y + z = 5, \qquad x + 2y - z = 13$$

. These are two equations in four variables, so the best we can hope for is to use the equations to express some of the variables (dependent variables) in terms of the rest (free variables). Then, for any value of the free variables, we get a solution (where the dependent variables are determined by the free variables and the equations).

We get $w = 5 - y - z$ and $x = 13 - 2y + z$, using w, x as dependent variables and y, z as free variables. Why did I pick w, x as the dependent variables? If we had tried to use the first equation to express y in terms of w and z, since y appears in the second equation, we would need to substitute $y = 5 - w - z$ into the second equation to proceed. We didn't need to do that for w or x because w only appeared in equation 1 and x only appeared in equation 2. That is a consequence of condition (*): if we take our dependent variables to be the ones whose coefficients are the leading entries then they only appear in one of the equations.

4.3.3 Another example

Example 4.11. Consider the augmented matrix $\begin{pmatrix} 1 & 2 & 0 & 1 & | & 1 \\ 0 & 0 & 1 & 8 & | & 2 \\ 0 & 0 & 0 & 0 & | & 0 \end{pmatrix}$. This corresponds to the equations:

$$w + 2x + z = 1, \qquad y + 8z = 2, \qquad 0 = 0.$$

The third equation here is trivially satisfied, so we can ignore it. The dependent variables are w and y, as they have the leading entries as their coefficients. We get

$$w = 1 - 2x - z, \qquad y = 2 - 8z,$$

with x, z as free variables.

Note that x also appears only in the first equation, so we could use x as a dependent variable, but that's just luck: we're only guaranteed that the variables whose coefficients are leading entries will appear in precisely one equation.

Example 4.12. Consider the augmented matrix $\begin{pmatrix} 1 & 2 & 0 & 1 & | & 1 \\ 0 & 0 & 1 & 8 & | & 2 \\ 0 & 0 & 0 & 0 & | & 3 \end{pmatrix}$ which differs from the previous one only in the very last entry. That corresponds to the system:

$$w + 2x + z = 1, \qquad y + 8z = 2, \qquad 0 = 3.$$

The third equation here doesn't hold, which means there is *no solution*: even if you can find w, x, y, z satisfying the first two equations, you'll never be able to solve the third.

4.3.4 Strategy

Our strategy (which we'll see implemented in some examples next time) is as follows:

- given a system of equations, write down an augmented matrix;
- perform row operations on the augmented matrix until the part left of the bar is in reduced echelon form;
- read off from this whether solutions exist (they do unless one of the rows is $\begin{pmatrix} 0 & \cdots & 0 & | & b \end{pmatrix}$ for $b \neq 0$);
- if solutions exist, write down the solution by finding the dependent variables in terms of the free variables.

4.4 Echelon examples

4.4.1 Example 1

Take the system

$$\begin{aligned} x + 2y + z &= 5 \\ -x + y + 2z &= 1 \\ x - z &= 1. \end{aligned}$$

The augmented matrix is

$$\begin{pmatrix} 1 & 2 & 1 & | & 5 \\ -1 & 1 & 2 & | & 1 \\ 1 & 0 & -1 & | & 1 \end{pmatrix}$$

To clear the first column, we do $R_2 \mapsto R_2 + R_1$ and $R_3 \mapsto R_3 - R_1$:

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

Now we do $R_2 \mapsto \frac{1}{3}R_2$ and $R_3 \mapsto -\frac{1}{2}R_3$:

$$\begin{pmatrix} 1 & 2 & 1 & | & 5 \\ 0 & 1 & 1 & | & 2 \\ 0 & 1 & 1 & | & 2 \end{pmatrix}$$

Finally, clear column 2 with $R_3 \mapsto R_3 - R_2$ and $R_1 \mapsto R_1 - 2R_2$:

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

This is now in reduced echelon form. It corresponds to the equations

$$\begin{aligned} x - z &= 1 \\ y + z &= 2 \\ 0 &= 0. \end{aligned}$$

The third equation is trivially satisfied: it is just a check that our equation really has a solution. The other two can be used to express the dependent variables x, y in terms of the free variable z. Note that x, y are chosen as dependent variables because they are the variables whose coefficients are leading entries.

4.4.2 Example 2

Take the system

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

The augmented matrix is

$$\begin{pmatrix} 2 & 2 & 0 & | & 2 \\ 1 & 1 & 1 & | & 8 \\ 1 & 1 & -1 & | & -5 \end{pmatrix}.$$

We do $R_1 \mapsto \frac{1}{2}R_1$:

$$\begin{pmatrix} 1 & 1 & 0 & | & 1 \\ 1 & 1 & 1 & | & 8 \\ 1 & 1 & -1 & | & -5 \end{pmatrix}.$$

Now we clear column 1 with $R_2 \mapsto R_2 - R_1$ and $R_3 \mapsto R_3 - R_1$:

$$\begin{pmatrix} 1 & 1 & 0 & | & 1 \\ 0 & 0 & 1 & | & 7 \\ 0 & 0 & -1 & | & -6 \end{pmatrix}.$$

Finally, $R_3 \mapsto R_3 + R_2$ yields:

$$\begin{pmatrix} 1 & 1 & 0 & | & 1 \\ 0 & 0 & 1 & | & 7 \\ 0 & 0 & 0 & | & 1 \end{pmatrix}.$$

The left-hand block is now in reduced echelon form.

Translating this back into equations gives

$$x + y = 1, \qquad z = 7, \qquad 0 = 1.$$

This means there is no solution, because $0 \neq 1$.

4.4.3 Example 3

Take the very similar system

$$\begin{aligned} 2x + 2y &= 2 \\ x + y + z &= 8 \\ x + y - z &= b. \end{aligned}$$

Here, b is a parameter; we'll perform the same analysis and figure out what value b needs to take in order for the system to have a solution. The augmented matrix is

$$\begin{pmatrix} 2 & 2 & 0 & | & 2 \\ 1 & 1 & 1 & | & 8 \\ 1 & 1 & -1 & | & b \end{pmatrix}.$$

The same sequence of row operations yield

$$\begin{pmatrix} 1 & 1 & 0 & | & 1 \\ 0 & 0 & 1 & | & 7 \\ 0 & 0 & 0 & | & b+6 \end{pmatrix}.$$

We see this only has a solution if $b = -6$. In this case, the solution is $x = 1 - y$, $z = 7$ (y is free).

The moral of this is: for an augmented matrix $(A|v)$ with A in reduced echelon form,

- any row of zeros in A gives a constraint on the vector v which has to be satisfied if a solution is to exist,
- any other row gives you a way of expressing a single dependent variable in terms of free variables.

5 Week 3, Session 1: Geometry of simultaneous equations and inverses

5.1 Geometric viewpoint on simultaneous equations, 1

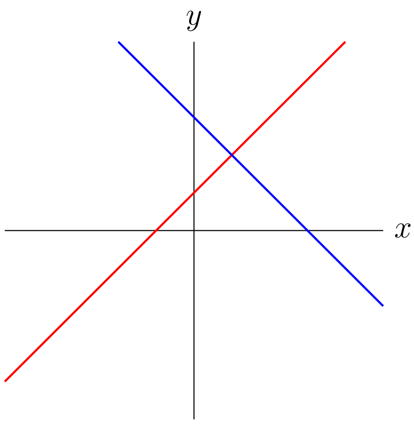
In this video and the next, we'll develop a geometric point of view on simultaneous equations.

5.1.1 2 variables

Example 5.1. We've repeatedly studied the system of equations

$$x - y = -1, \qquad x + y = 3.$$

We saw this has the unique solution $x = 1$, $y = 2$. Geometrically, $x - y = -1$ and $x + y = 3$ define lines in the xy-plane: the first (red) is $y = x + 1$ (slope 1, y-intercept 1), the second (blue) is $y = -x + 3$ (slope -1, y-intercept 3):



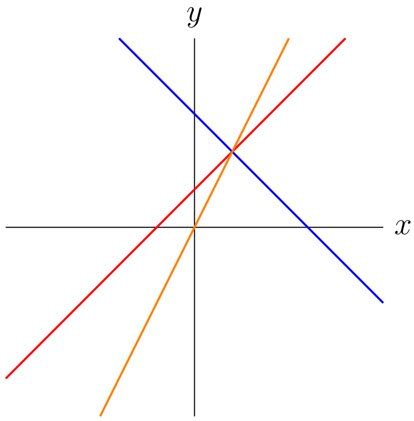
They intersect at exactly one point, namely $(1, 2)$. This is precisely the solution of the equations ($x = 1$, $y = 2$). That makes sense: the first line is the set of points whose x and y coordinates satisfy $x - y = -1$, the second line is the set of points whose coordinates satisfy $x + y = 3$; the intersection is the set of points whose coordinates satisfy both equations.

So solving simultaneous linear equations in two variables is the same thing as intersecting lines.

Example 5.2. Let's add another equation:

$$x - y = -1, \qquad x + y = 3, \qquad 2x - y = 0.$$

This adds another line (orange) into our diagram.

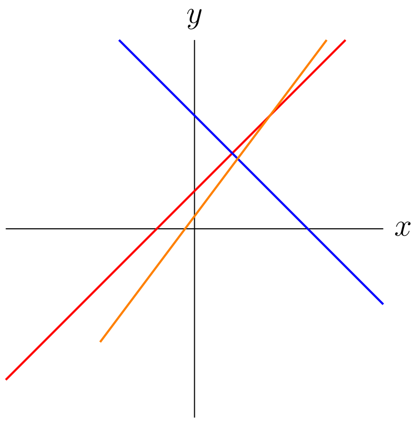


Since these lines have a common triple intersection point $(1, 2)$, this is still a solution to this system of equations.

If we had picked a slightly third different equation, defining a line which didn't pass through $(1, 2)$, there would be no solutions. For example:

$$x - y = -1, \qquad x + y = 3, \qquad 2x - y = 1,$$

defines three lines which have no triple intersection point, so there is no common solution.



5.1.2 Lines in the plane

Each equation $a_1x + a_2y = b$ defines a line in the plane. A solution of the system of equations is a common intersection point between these lines.

The line $L = \{(x, y) \in \mathbf{R}^2 \ : \ a_1x + a_2y = b$ is equivalent to $y = -\frac{a_1}{a_2}x + \frac{b}{a_2}$ (if $a_2 \neq 0$, otherwise it's vertical), so the slope is $-a_1/a_2$ (infinity if $a_2 = 0$).

Lemma 5.3. *The vector* $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ *is orthogonal to* L*. (It's also called the "normal vector to* L*").*

Proof. If I pick two points $\begin{pmatrix} x \\ y \end{pmatrix}$ and $\begin{pmatrix} x' \\ y' \end{pmatrix}$ on L then their difference $\begin{pmatrix} x - x' \\ y - y' \end{pmatrix}$ points along L, so it's sufficient to check that

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \cdot \begin{pmatrix} x - x' \\ y - y' \end{pmatrix} = 0.$$

We have $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \cdot \begin{pmatrix} x - x' \\ y - y' \end{pmatrix} = a_1x + a_2y - a_1x' - a_2y' = 0 - 0$, since $a_1x + a_2y = a_1x' + a_2y' = b$ (both points lie on L).

□

5.2 Geometric viewpoint on simultaneous equations, 2

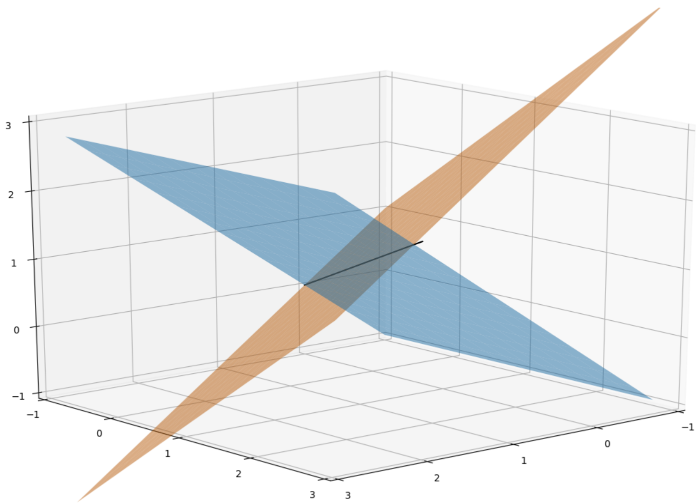
5.2.1 3 variables

Now let's look at higher dimensional systems.

Example 5.4. Consider the equations

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

The set of points whose coordinates satisfy these equations is a subset of 3-dimensional space $\mathbf{R}^3$. Each equation cuts out a plane; their intersection (the set of simultaneous solutions) is a line.



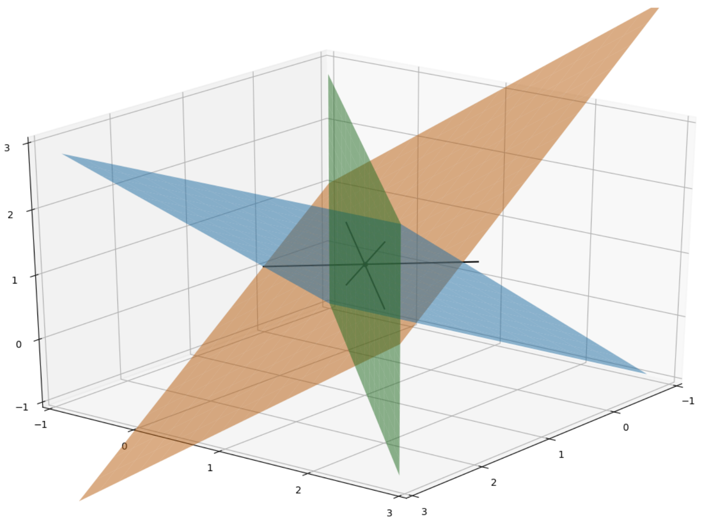
The solution of the equations is $z = x$, $y = 2x - 1$ (x is a free variable); this gives us a parametrisation of the line of intersection: $\begin{pmatrix} x \\ 2x - 1 \\ x \end{pmatrix}$.

Each equation $a_1x + a_2y + a_3z = b$ gives us a plane. This plane is orthogonal to the vector $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$. The z-intercept is where $x = y = 0$, so $z = b/a_3$.

Example 5.5. Let's consider the system

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

The first two planes are the same as before; the third equation gives a third plane. The set of solutions is the set of triple intersections between these planes. Solving, we get $x = y = z$ and $z = y = 2z - 1$, so $x = y = z = 1$. Therefore there is a unique triple intersection point at $(1, 1, 1)$.



5.2.2 Higher dimensions

An equation $a_1x_1 + \cdots + a_nx_n = b$ in n variables defines a *hyperplane* in n-dimensional space $\mathbf{R}^n$. This is the hyperplane orthogonal to the vector $\begin{pmatrix} a_1 \\ \cdots \\ a_n \end{pmatrix}$.

If $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ is a vector of variables and $Av = b$ is a system of m simultaneous equations encoded with an m-by-n matrix$A$ then each row of A defines a hyperplane $A_{i1}x_1 + \cdots + A_{in}x_n = b_i$ in $\mathbf{R}^n$. The set of solutions is a *subspace* of $\mathbf{R}^n$. Subspace is a catch-all name, which includes points, lines, planes, hyperplanes, and everything in between (for which we don't have everyday words).

Example 5.6. Consider the equations $w + x + y + z = 0$ and $x - y = 1$. This is a pair of equations in four variables, so defines a subspace of solutions in $\mathbf{R}^4$. Each equation gives a 3-dimensional hyperplane in $\mathbf{R}^4$. Their intersection is a 2-dimensional subspace (plane) living in $\mathbf{R}^4$. Solving the equations: $w = -2y - z$, $x = 1 - y$. We see that y, z are free variables and w, x are dependent variables. The fact that there are two free variables is another way of saying that the subspace of solutions is 2-dimensional: each free variable is a coordinate on the subspace of solutions.

Remark 5.7. More generally, the dimension of the subspace of solutions equals the number of free variables: we can think of the free variables as parameters for describing points in the subspace of solutions.

5.3 Subspaces

In this video we will explain in more detail what a *subspace* is. Recall from the last video that subspaces are generalisations of lines/planes and the set of solutions of a system of simultaneous equations in n variables forms a subspace of $\mathbf{R}^n$.

5.3.1 Linear subspaces

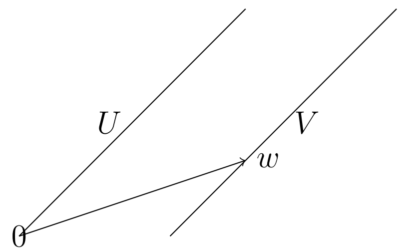
Definition 5.8. A subset $V \subset \mathbf{R}^n$ is a *linear subspace* if:

- for all $v, w \in V$, $v + w \in V$ (closed under addition),
- for all $v \in V$ and $\lambda \in \mathbf{R}$, $\lambda v \in V$ (closed under rescaling).

Note that nonempty linear subspaces always contain the origin: if $v \in V$ then $0v \in V$ by the second axiom, and $0v = 0$ is the zero-vector (the origin). Note that there are systems of simultaneous equations for which $(0, \ldots, 0)$ is not a solution; likewise, there are lines and planes that do not pass through the origin. So for our purposes, we need something more general than linear subspaces.

5.3.2 Affine subspaces

Definition 5.9. $V \subset \mathbf{R}^n$ is called an *affine subspace* if there exist a vector $w \in \mathbf{R}^n$ and a linear subspace $U \subset \mathbf{R}^n$ such that $V = w + U := \{w + u \ : \ u \in U\}$; in other words, V is obtained by translating U by the vector w.



Lemma 5.10. *Given a system of simultaneous equations in matrix form $Av = b$, the set of solutions $v = \begin{pmatrix} x_1 \\ \vdots \\ v_n \end{pmatrix}$ form an affine subspace of $\mathbf{R}^n$, (n is the number of variables). It is a linear subspace if and only if $b = 0$.*

Proof. First, let's assume $b = 0$. We'll prove that the set of solutions satisfies the axioms for being a linear subspace. If v, v' are solutions then $Av = Av' = 0$, so $A(v + v') = 0 + 0 = 0$, and if $\lambda \in \mathbf{R}$ then $A\lambda v = \lambda A v = \lambda 0 = 0$. Therefore both $v + v'$ and λv are solutions and the set of solutions is a linear subspace. It's also nonempty because $v = 0$ is always a solution to $Av = 0$.

Now suppose $b \neq 0$. If there are no solutions then the set of solutions is empty, and the empty set is an empty subspace. So suppose there is at least one solution w. Let U be the set of solutions to $Au = 0$. Let S be the set of solutions to $Av = b$. We will show that $S = w + U$:

- We first prove that $w + U \subset S$. This is because $A(w + u) = Aw + Au = b + 0 = b$, so $w + u \in S$.
- Then we prove that $S \subset w + U$. This is because $Av = b$ implies $A(v - w) = Av - Aw = b - b = 0$, so $v - w \in U$ and $w \in w + U$.

□

5.3.3 More properties

You'll see a lot more about subspaces in future courses on linear algebra, but I'll just give you a couple more nice facts you can see from the definitions and which are related to what we've been doing.

Lemma 5.11. *If $V, W \subset \mathbf{R}^n$ are linear subspaces then $V \cap W$ is also a linear subspace.*

Proof. We'll check that the two axioms hold.

Suppose $a, b \in V \cap W$. Then $a \in V$ and $b \in V$, so $a + b \in V$ because V is a linear subspace. Similarly $a \in W, b \in W$ implies $a + b \in W$. Therefore $a + b \in V \cap W$.

Similarly, $a \in V \cap W$ implies $a \in V$ and $a \in W$, so $\lambda a \in V$ and $\lambda a \in W$, so $\lambda a \in V \cap W$.

□

Another fact which is good to know but which I won't prove is:

Lemma 5.12. *A nonempty affine subspace is linear if and only if it contains the origin.*

You now have the beginnings of the language you need for talking about lines, planes and their generalisations in higher dimensions.

5.4 Inverses

5.4.1 When can you "divide" by a matrix?

In the next few videos, we're going to answer the question: *can you divide by a matrix?*

Theorem 5.13. *If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is a 2-by-2 matrix with $ad - bc \neq 0$ then the matrix $A^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ satisfies $A^{-1}A = AA^{-1} = I$. We say that A^{-1} is the* inverse *of A.*

This is the analogue of the reciprocal $x^{-1} = 1/x$ of a number: the equation $A^{-1}A = I$ is the analogue of $x^{-1}x = 1$.

Proof. Let's just check $A^{-1}A = I$ (check the other equality for yourself).

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

□

We can use this to "divide" by A: if we have a matrix equation $AB = C$ then we can multiply both sides on the left by A^{-1} to get $A^{-1}AB = A^{-1}C$, and since $A^{-1}A = I$ this means $B = A^{-1}C$.

With great power comes great responsibility: you should never write $\frac{B}{A}$ for matrices A, B! It's not clear if you're doing $A^{-1}B$ or BA^{-1} (and these are different because A^{-1} and B might not commute).

Example 5.14. We can use inverses to solve simultaneous equations. For example:

$$x - y = -1, \qquad x + y = 3$$

is equivalent to $Av = b$ where $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$ and $b = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$. We can compute $A^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, so

$$v = A^{-1}b = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} -1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix},$$

and this coincides with the solution $x = 1$, $y = 2$ we found earlier.

5.4.2 Bigger matrices

We'd like to generalise the notion of inverse to bigger matrices.

Definition 5.15. Let A be an n-by-n matrix. We say that A is invertible if there exists a matrix B such that $AB = BA = I$. (Here I is the n-by-n identity matrix). If such a B exists, then it's unique, so we're justified in calling it *the* inverse of A and writing it as A^{-1}.

To see that the inverse is unique when it exists, suppose you had two inverses B, C for A:

$$AB = AC = I, \qquad BA = CA = I.$$

Now evaluate BAC in two ways:

- $BAC = (BA)C = IC = C$
- $BAC = B(AC) = BI = B$

so $C = BAC = B$.

Lemma 5.16. *If A, B are invertible matrices then AB is invertible with inverse $(AB)^{-1} = B^{-1}A^{-1}$.*

Proof. we have

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

and

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

so $B^{-1}A^{-1}$ is an inverse for AB.

□

Had we tried to use $A^{-1}B^{-1}$ instead, we would have obtained $ABA^{-1}B^{-1}$, and we couldn't have cancelled anything because the various terms don't commute.

We saw for 2-by-2 matrices that a matrix is invertible if $ad - bc \neq 0$. Even more basically, a 1-by-1 matrix (x) is invertible if and only if $x \neq 0$. We'll see in a few videos' time that there is an analogous quantity (the *determinant*) $\det(A)$ which is nonzero precisely when A is invertible. First though, we're going to explain how to calculate inverses of n-by-n matrices.

5.5 Inverses: examples

5.5.1 Method for finding the inverse

Theorem 5.17. *Given an n-by-n matrix A, form the augmented matrix $(A|I)$ (i.e. you have the n-by-n identity matrix on the right hand side of the bar). Apply row operations to this augmented matrix until you obtain an augmented matrix $(B|C)$ with B in reduced echelon form.*

- *If $B = I$ then A is invertible with inverse C.*
- *If $B \neq I$ then A is not invertible.*

So if A is invertible, you start with $(A|I)$ and you end up with $(I|A^{-1})$. We're going to postpone the proof for now, and work out some examples before developing the theory necessary for the proof.

5.5.2 Examples

Example 5.18. Let $A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$. The augmented matrix is

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

To put the left-hand block into reduced echelon form, we do $R_2 \mapsto R_2 - R_1$:

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

then we do $R_2 \mapsto \frac{1}{2}R_2$:

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

Next we do $R_1 \mapsto R_1 + R_2$:

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

Now the left-hand block is in reduced echelon form; it is the identity matrix, so A is invertible and its inverse is the right-hand block:

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

You can check this agrees with the 2-by-2 inverse we talked about last time.

Example 5.19. Here's a messy example. Let's invert $A = \begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 4 \\ 1 & 5 & 0 \end{pmatrix}$. We have

$$(A|I) = \begin{pmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 0 & 1 & 4 & | & 0 & 1 & 0 \\ 1 & 5 & 0 & | & 0 & 0 & 1 \end{pmatrix}.$$

Perform $R_3 \mapsto R_3 - R_1$ to clear the first column:

$$\begin{pmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 0 & 1 & 4 & | & 0 & 1 & 0 \\ 0 & 3 & -3 & | & -1 & 0 & 1 \end{pmatrix}.$$

Next, $R_3 \mapsto R_3 - 3R_2$ gets us into echelon form:

$$\begin{pmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 0 & 1 & 4 & | & 0 & 1 & 0 \\ 0 & 0 & -15 & | & -1 & -3 & 1 \end{pmatrix}.$$

Now $R_3 \mapsto -R_3/15$:

$$\begin{pmatrix} 1 & 2 & 3 & | & 1 & 0 & 0 \\ 0 & 1 & 4 & | & 0 & 1 & 0 \\ 0 & 0 & 1 & | & 1/15 & 1/5 & -1/15 \end{pmatrix}.$$

Clear the last column with $R_2 \mapsto R_2 - 4R_3$ and $R_1 \mapsto R_1 - 3R_3$:

$$\begin{pmatrix} 1 & 2 & 0 & | & 4/5 & -3/5 & 1/5 \\ 0 & 1 & 0 & | & -4/15 & 1/5 & 4/15 \\ 0 & 0 & 1 & | & 1/15 & 1/5 & -1/15 \end{pmatrix}.$$

Finally, $R_1 \mapsto R_1 - 2R_2$ yields:

$$\begin{pmatrix} 1 & 0 & 0 & | & 4/3 & -1 & -1/3 \\ 0 & 1 & 0 & | & -4/15 & 1/5 & 4/15 \\ 0 & 0 & 1 & | & 1/15 & 1/5 & -1/15 \end{pmatrix}.$$

The left-hand block is the identity matrix, so A is invertible and the right-hand block is A^{-1}.

This means $A^{-1} = \begin{pmatrix} 4/3 & -1 & -1/3 \\ -4/15 & 1/5 & 4/15 \\ 1/15 & 1/5 & -1/15 \end{pmatrix}$. We can check this by computing $A^{-1}A$: we get I.

6 Week 3, Session 2: Inverses and determinants

6.1 Elementary matrices, 1

We introduce elementary matrices which have the following property: multiplying A on the left by an elementary matrix has the effect of performing a row operation to A.

6.1.1 Elementary matrices of type I

Definition 6.1. Let i, j be integers between 1 and n inclusive, and assume $i \neq j$. Let $\lambda \in \mathbf{R}$. Then $E_{ij}(\lambda)$ is the matrix which has:

- 1s on the diagonal,
- 0s everywhere else except in the ijth position, where there is a λ.

Such a matrix is an *elementary matrix of type I.*

For example, if $n = 3$ you get things like:

$$E_{12}(4) = \begin{pmatrix} 1 & 4 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad E_{32}(-5) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -5 & 1 \end{pmatrix}.$$

If $n = 4$, you get things like

$$E_{23}(7) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 7 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Lemma 6.2. *Let A and $E_{ij}(\lambda)$ be n-by-n matrices. Then $E_{ij}(\lambda)A$ is the matrix obtained from A by doing $R_i \mapsto R_i + \lambda R_j$.*

Proof. Let's consider the case $i < j$ (the other case is similar so we omit it). Consider the product

$$\begin{pmatrix} 1 & & \text{col } i & & \text{col } j & & \\ & \ddots & \downarrow & & \downarrow & & \\ \text{row } i & \rightarrow & 1 & & \lambda & & \\ & & & \ddots & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{pmatrix} A_{11} & \cdots & \cdots & \cdots & \cdots & A_{1n} \\ \vdots & & & & & \vdots \\ A_{i1} & & & & & A_{in} \\ \vdots & & & & & \vdots \\ A_{j1} & & & & & A_{jn} \\ \vdots & & & & & \vdots \\ A_{n1} & \cdots & \cdots & \cdots & \cdots & A_{nn} \end{pmatrix}$$

The only difference the λ makes is when we multiply the ith row into a column of A (say the kth column). Instead of just picking up $1 \times A_{ik}$, we get $1 \times A_{ik} + \lambda \times A_{jk}$.

In other words, the result $E_{ij}(\lambda)A$ is obtained from A by adding λ times row j to row i.

$\square$

6.1.2 Elementary matrices of type II

Definition 6.3. Let i be an integer between 1 and n inclusive and let $\lambda \neq 0$ be a real number. Define $E_i(\lambda)$ to be the matrix which has 1s on the diagonal everywhere except in the iith position, where it has a λ. Such a matrix is an *elementary matrix of type II*.

Example 6.4. If $n = 3$,

$$E_2(\lambda) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad E_3(-1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Lemma 6.5. *$E_i(\lambda)A$ is obtained from A by performing the row operation $R_i \mapsto \lambda R_i$.*

Proof. In the product,

$$\begin{pmatrix} 1 & 0 & \cdots & & 0 \\ 0 & \ddots & & & \\ \vdots & & \lambda & & \vdots \\ & & & \ddots & 0 \\ 0 & & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} A_{11} & \cdots & A_{1n} \\ \vdots & & \vdots \\ A_{i1} & \cdots & A_{in} \\ \vdots & & \vdots \\ A_{n1} & \cdots & A_{nn} \end{pmatrix}$$

most of the entries are multiplied by 1, except those in the ith row of A, which are rescaled by λ.

□

6.2 Elementary matrices, 2

Lemma 6.6. *$E_{ij}(\lambda)$ is invertible with inverse $E_{ij}(-\lambda)$ and $E_i(\lambda)$ is invertible with inverse $E_i(1/\lambda)$ (recall that $\lambda \neq 0$ for type II elementary matrices).*

Proof. We'll prove it for 2-by-2 matrices to save ourselves from all the dots in the last few proofs. For the first:

$$E_{12}(\lambda)E_{12}(-\lambda) = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -\lambda \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & \lambda - \lambda \\ 0 & 1 \end{pmatrix} = I.$$

For the second:

$$E_1(\lambda)E_1(1/\lambda) = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1/\lambda & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \lambda/\lambda & 0 \\ 0 & 1 \end{pmatrix} = I.$$

□

We now proceed to the proof of the theorem:

Theorem 6.7. *If A is a matrix and you use row operations on $(A|I)$ until you get $(B|C)$ with B in reduced echelon form then:*

- *if B is I then A is invertible with inverse C;*
- *if $B \neq I$ then A is not invertible.*

Proof. Put A into reduced echelon form by performing row operations. Each row operation corresponds to multiplying A on the left by an elementary matrix. Therefore $B = M_k \cdots M_1 A$ for some sequence of elementary matrices M_i.

We're simultaneously performing row operations on both sides of the bar in the augmented matrix, so $C = M_k \cdots M_1 I = M_k \cdots M_1$, so C is a product of elementary matrices. Moreover, $B = M_k \cdots M_1 A = CA$.

If $B = I$ then $CA = I$ and A is invertible with $A^{-1} = C$.

If $B \neq I$, it needs to have a row of zeros. Indeed, suppose B is an n-by-n matrix in echelon form. If all rows are nonzero then there are n leading entries, which have to move to the right as you go down, so they have to live on the diagonal, which forces your matrix to be the identity.

Therefore there is a free variable in the solution to $Bv = 0$ (equivalently $Av = 0$). Therefore there is a whole line's worth of solutions, so there is at least one $v \neq 0$ such that $Av = 0$. If A were invertible then $v = A^{-1}0 = 0$, which is a contradiction. Therefore A is not invertible.

□

Corollary 6.8. *A product of elementary matrices is invertible; conversely, any invertible matrix is a product of elementary matrices.*

Proof. Recall that elementary matrices are invertible and their inverses are again elementary matrices.

If $M_k \cdots M_1$ is a product of elementary matrices then $(M_k \cdots M_1)^{-1} = M_1^{-1} \cdots M_k^{-1}$, so it is invertible. Conversely, if A is invertible then its inverse has the form $M_k \cdots M_1$ for some sequence of elementary matrices (as we just saw in the previous proof). Therefore $A = M_1^{-1} \cdots M_k^{-1}$ is a product of elementary matrices.

□

6.3 Determinants

6.3.1 Definition

We saw that a 2-by-2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has an inverse if and only if $ad - bc \neq 0$. This condition generalises to bigger matrices as follows. To each n-by-n matrix A, we will define a number $\det(A)$ (the *determinant* of A) such that A has an inverse if and only if $\det(A) \neq 0$.

Definition 6.9. Let A be an n-by-n matrix. The number $\det(A)$ is defined as follows.

- Pick an entry from each row in such a way that no two of your choices are in the same column. (If A were 2-by-2 then you could either pick a and d or b and c).

- Multiply your chosen entries together (if A were 2-by-2 this would give us ad or bc).
- Do the same for every other choice you could have made, and add up the results with signs (I'll explain the signs later). (If A were 2-by-2, we would get $\pm ad \pm bc$).

Before we explain the signs, let's see how many choices there would be if $n = 3$, so

$$A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}.$$

If you pick A_{11} from the top row, you can't pick anything in the first column for your other choices, so the only possibilities are A_{22}, A_{33} or A_{23}, A_{32} from the second and third rows. In this way, the possible choices give:

$$A_{11}A_{22}A_{33}, \qquad A_{11}A_{23}A_{32}, \qquad A_{12}A_{21}A_{33}, \qquad A_{12}A_{23}A_{31}, \qquad A_{13}A_{21}A_{32}, \qquad A_{13}A_{22}A_{31},$$

so there are six possibilities.

More generally:

Lemma 6.10. *Let A be an n-by-n matrix. The number of ways of choosing something from each row so that no two are in the same column is $n!$.*

Proof. For each choice, write down the column numbers in the order you chose them. For example:

- $A_{11}A_{22}A_{33}$ corresponds to choosing entry 1 from row 1, entry 2 from row 2 and entry 3 from row 3, so you'd write down 123;
- $A_{11}A_{23}A_{32}$ corresponds to choosing entry 1 from row 1, entry 3 from row 2 and entry 2 from row 3, so you'd write down 132.

This gives a permutation of $1, 2, \ldots, n$. Every permutation arises and a choice is determined by the permutation you wrote down, so the number of choices is the same as the number of permutations, which is $n!$.

□

6.3.2 Signs

Having made a choice as above and written down the corresponding permutation of $1, 2, \ldots, n$, the sign you assign to this choice is:

- $+1$ if your permutation is *even* (i.e. can be written as an even number of transpositions);
- -1 if your permutation is *odd* (i.e. can be written as an odd number of transpositions).

Example 6.11. Suppose you picked $A_{11}A_{22}A_{33}$, corresponding to the permutation 123. To get from 123 back to 123, I don't have to do anything (i.e. 0 transpositions) and 0 is an even number so this is an even permutation, and $A_{11}A_{22}A_{33}$ contributes with a plus sign to the determinant.

Example 6.12. Suppose you picked $A_{11}A_{23}A_{32}$ corresponding to the permutation 132. To get from 132 back to 123 you need to transpose 2 and 3, so that's one transposition, and the permutation is odd. Therefore $A_{11}A_{23}A_{32}$ contributes with a minus sign.

Here are the permutations and signs for the other choices when $n = 3$:

$A_{11}A_{22}A_{33}$	123	$+$
$A_{11}A_{23}A_{32}$	132	$-$
$A_{12}A_{21}A_{33}$	213	$-$
$A_{12}A_{23}A_{31}$	231	$+$
$A_{13}A_{21}A_{32}$	312	$+$
$A_{13}A_{22}A_{31}$	321	$-$

so

$$\det(A) = A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{13}A_{22}A_{31}.$$

Example 6.13. If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then our two choices are ad (permutation 12, even so plus sign) and bc (permutation 21, odd so minus sign), giving

$$\det(A)ad - bc.$$

6.3.3 Index notation

We have a nice formula in index notation:

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}.$$

Here:

- the sum is over all permutations σ of $1, 2, \ldots, n$
- $\operatorname{sgn}(\sigma)$ means the sign ($+1$ if σ is even, -1 if σ is odd).
- to understand the product, note that $A_{1\sigma(1)}$ is the choice from the first row (which is why the first index is 1) and it lives in column $\sigma(1)$ (remember σ is the permutation which tells us which column our choices live in). Similarly, $A_{2\sigma(2)}$ is the choice from the second row, etc.

6.3.4 Examples

Example 6.14. Let $A = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$. There's only one valid choice which doesn't involve any zeros, namely $A_{31}A_{22}A_{13} = 1 \times 1 \times 1$. This corresponds to the permutation 321. To get back to 123 we'd have to transpose 1 and 3, which is an odd number of transpositions, so it has a minus sign. Therefore $\det(A) = -1$.

Remark 6.15. To get back from 321 to 123, we could also have flipped 2 and 3 (getting to 231) then 1 and 3 (getting to 213) then 1 and 2 (getting to 123). This would have been three transpositions, which is still an odd number. The reason the definition of odd vs even permutations makes sense is that *any permutation can be written as either an odd number of transpositions or an even number, but never as both.* We won't prove that here.

Example 6.16. Let $B = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Again, there's only one valid choice which has no zeros, and this choice corresponds to the permutation 123, so the only contribution to the determinant is $+5 \times 1 \times 3 = 15$. Therefore $\det(B) = 15$.

This applies to any diagonal matrix, i.e. a matrix whose off-diagonal entries are all zero: the determinant is the product of diagonal entries (with no extra signs).

Example 6.17. Consider an *upper triangular matrix* (where all entries strictly below the diagonal are zero) like

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

From the last row you have to pick the last entry. Therefore from the penultimate row you have to pick the penultimate entry, and so on, and you end up being forced to pick only the diagonal entries. So you get the product of diagonal entries with no extra sign (in this example, $\det(C) = 1$).

Example 6.18. Since the elementary matrix $E_{ij}(\lambda) = \begin{pmatrix} 1 & 0 & \cdots & & \\ 0 & \ddots & & \lambda & \\ \vdots & & 1 & & \vdots \\ & & & \ddots & 0 \\ & & \cdots & 0 & 1 \end{pmatrix}$ is upper triangular (if $i < j$, or lower triangular if $i > j$), we see that $\det(E_{ij}(\lambda)) = 1$.

Example 6.19. Since the elementary matrix $E_i(\lambda) = \begin{pmatrix} 1 & 0 & \cdots & & \\ 0 & \ddots & & & \\ \vdots & & \lambda & & \vdots \\ & & & \ddots & 0 \\ & & \cdots & 0 & 1 \end{pmatrix}$ is a diagonal matrix, we get $\det(E_i(\lambda)) = \lambda$.

6.4 Properties of determinants

Recall that

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \cdots A_{n\sigma(n)}.$$

We'll use this formula to prove some nice properties of the determinant.

Lemma 6.20. *If two rows coincide then* $\det(A) = 0$. *For example* $\det \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 2 & 4 & 5 \end{pmatrix} = 0$.

Proof. If row i and row j coincide then $A_{ik} = A_{jk}$ for all k. Consider one of the terms

$$\mathrm{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{i\sigma(i)} \cdots A_{j\sigma(j)} \cdots A_{n\sigma(n)}$$

in the determinant. Since $A_{ik} = A_{jk}$ for all k, we have

$$A_{i\sigma(i)} = A_{j\sigma(i)}, \qquad A_{j\sigma(j)} = A_{i\sigma(j)}.$$

Therefore

$$\mathrm{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{i\sigma(i)} \cdots A_{j\sigma(j)} \cdots A_{n\sigma(n)} = \mathrm{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{i\sigma(j)} \cdots A_{j\sigma(i)} \cdots A_{n\sigma(n)}$$

This is another term in our determinant, corresponding to the permutation σ', which is defined by

$$\sigma'(i) = \sigma(j), \qquad \sigma'(j) = \sigma(i), \qquad \sigma'(k) = \sigma(k) \text{ for all } k \neq i, j.$$

In other words:

$$A_{1\sigma(1)} \cdots A_{i\sigma(j)} \cdots A_{j\sigma(i)} \cdots A_{n\sigma(n)} = A_{1\sigma'(1)} \cdots A_{i\sigma'(i)} \cdots A_{j\sigma'(j)} \cdots A_{n\sigma'(n)}$$

Since σ' differs from σ by a single transposition, $\mathrm{sgn}(\sigma') = -\mathrm{sgn}(\sigma)$. Therefore these two terms contribute with opposite signs to $\det(A)$, and cancel out:

$$\mathrm{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{i\sigma(j)} \cdots A_{j\sigma(i)} \cdots A_{n\sigma(n)} + \mathrm{sgn}(\sigma') A_{1\sigma'(1)} \cdots A_{i\sigma'(j)} \cdots A_{j\sigma'(i)} \cdots A_{n\sigma'(n)} = 0$$

Therefore all terms cancel in pairs, and the determinant is zero.

□

6.4.1 Type I row operations

Lemma 6.21. *If* A' *is obtained from* A *by the row operation* $R_i \mapsto R_i + \lambda R_j$ *then* $\det(A') = \det(A)$.

Proof. We have

$$\begin{aligned}
\det(A') &= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) A'_{1\sigma(1)} \cdots A'_{i\sigma(i)} \cdots A'_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) A_{1\sigma(1)} \cdots (A_{i\sigma(i)} + \lambda A_{j\sigma(i)}) \cdots A_{n\sigma(n)} \\
&= \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{i\sigma(i)} \cdots A_{n\sigma(n)} + \lambda \sum_{\sigma \in S_n} \mathrm{sgn}(\sigma) A_{1\sigma(1)} \cdots A_{j\sigma(i)} \cdots A_{n\sigma(n)}
\end{aligned}$$

The first term here is $\det(A)$. The second term is the determinant of the matrix obtained from A by replacing the ith row with the jth row; in other words, the jth row now appears twice. By the previous lemma, this second term vanishes, so $\det(A') = \det(A)$.

□

We can put any matrix into echelon form using type I row operations *without changing the determinant*. Once you're in echelon form (since you're then upper triangular), the determinant is just the product of diagonal entries.

6.4.2 Type III row operations

Lemma 6.22. *If A' is obtained from A by switching two rows (say $R_i \leftrightarrow R_j$ then* $\det(A') = -\det(A)$.

Proof. Every term in $\det(A)$ has the form

$$\mathrm{sgn}(\sigma)A'_{1\sigma(1)}\cdots A'_{i\sigma(i)}\cdots A'_{j\sigma(j)}\cdots A'_{n\sigma(n)}.$$

Since A' and A are related by switching rows i and j, this is just

$$\mathrm{sgn}(\sigma)A_{1\sigma(1)}\cdots A_{i\sigma(j)}\cdots A_{j\sigma(i)}\cdots A_{n\sigma(n)},$$

which is equal to

$$\mathrm{sgn}(\sigma)A_{1\sigma'(1)}\cdots A_{i\sigma'(i)}\cdots A_{j\sigma'(j)}\cdots A_{n\sigma'(n)},$$

where σ' is (as before):

$$\sigma'(i) = \sigma(j), \qquad \sigma'(j) = \sigma(i), \qquad \sigma'(k) = \sigma(k) \text{ for all } k \neq i, j.$$

Since $\mathrm{sgn}(\sigma') = -\mathrm{sgn}(\sigma)$, this is equal to

$$-\mathrm{sgn}(\sigma')A_{1\sigma'(1)}\cdots A_{i\sigma'(i)}\cdots A_{j\sigma'(j)}\cdots A_{n\sigma'(n)},$$

so for every term in $\det(A')$ there's a corresponding term in $\det(A)$ which differs only by a sign. Therefore $\det(A') = -\det(A)$.

□

6.4.3 Type II row operations

Lemma 6.23. *If A' is obtained from A by $R_i \mapsto \lambda R_i$ then* $\det(A') = \lambda\det(A)$.

Proof. Exercise.

□

6.5 Determinants: examples

Example 6.24. Let $A = \begin{pmatrix} 1 & 4 & -4 \\ -2 & -2 & -4 \\ 3 & -3 & 3 \end{pmatrix}$. Rather than applying our formula (sum over permutations), we're going to apply type I row operations to put this matrix into echelon form; this won't change the determinant, so we can take the determinant afterwards (at which point it will be the product of diagonal entries).

To clear the first column, we do $R_2 \mapsto R_2 + 2R_1$ and $R_3 \mapsto R_3 - 3R_1$:

$$\begin{pmatrix} 1 & 4 & -4 \\ 0 & 6 & -12 \\ 0 & -15 & 15 \end{pmatrix}$$

To clear the -15 below the diagonal, we do $R_3 \mapsto R_3 + \frac{5}{2}R_2$:

$$\begin{pmatrix} 1 & 4 & -4 \\ 0 & 6 & -12 \\ 0 & 0 & -15 \end{pmatrix}.$$

This is in echelon form. Therefore $\det(A) = 1 \times 6 \times (-15) = -90$.

Example 6.25. Let $B = \begin{pmatrix} 2 & -3 & -1 & 4 \\ 2 & -3 & 2 & 4 \\ 2 & -1 & -4 & -3 \\ 2 & -3 & 4 & 2 \end{pmatrix}$. Let's clear the first column with $R_2 \mapsto R_2 - R_1$, $R_3 \mapsto R_3 - R_1$, $R_4 \mapsto R_4 - R_1$:

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

Now let's do $R_2 \leftrightarrow R_3$:

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

The determinant just changed sign! We'll need to remember to stick an extra minus sign in at the end. To clear the final 5 below the diagonal, we do $R_4 \mapsto R_4 - \frac{5}{3}R_3$:

$$\begin{pmatrix} 2 & -3 & -1 & 4 \\ 0 & 2 & -3 & \\ 0 & 0 & 3 & 0-7 \\ 0 & 0 & 0 & -2 \end{pmatrix}$$

which is now in echelon form.

The determinant of B is therefore $-(2 \times 2 \times 3 \times -2) = 24$.

7 Week 4, Session 1: More about determinants

7.1 Further properties of determinants

Recall that $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if its determinant $ad - bc$ is nonzero. We'll prove the analogue of this for n-by-n matrices.

Theorem 7.1. *Let A be an n-by-n matrix. Then A is invertible if and only if* $\det(A) \neq 0$.

We'll also prove the following:

Theorem 7.2. $\det(AB) = \det(A)\det(B)$.

7.1.1 Proof of the invertibility criterion

Recall that A is invertible if and only if its reduced echelon form (say A_{red}) is the identity matrix. To get from A to A_{red}, we used some row operations. The effect of these on the determinant is:

- (type I) determinant unchanged,
- (type II) determinant rescaled by a nonzero factor,
- (type III) determinant switches sign.

Therefore $\det(A)$ differs from $\det(A_{red})$ by a nonzero factor. In particular, $\det(A) \neq 0$ if and only if $\det(A_{red}) \neq 0$.

If A is invertible, $\det(A_{red}) = \det(I) = 1$, so $\det(A) \neq 0$.

If A is not invertible, A_{red} has a row of zeros (A_{red} is a square matrix in reduced echelon form, so if every row is nonzero then the leading entries have to go along the diagonal, and you get $A_{red} = I$; but we're assuming A is not invertible). Since A_{red} has a row of zeros, $\det(A_{red}) = 0$, so $\det(A) = 0$ too.

7.1.2 Proof of multiplicativity of the determinant

We first prove $\det(AB) = \det(A)\det(B)$ in the case where A is an elementary matrix.

- If $A = E_{ij}(\lambda)$ then AB is obtained from B by $R_i \mapsto R_i + \lambda R_j$, so $\det(AB) = \det(B)$ (such a row operation doesn't change the determinant). We also have $\det(A) = 1$, so we can verify the formula $\det(AB) = \det(A)\det(B)$ in this case just by calculating both sides.
- If $A = E_i(\lambda)$ then AB is obtained from B by $R_i \mapsto \lambda R_i$, so $\det(AB) = \lambda\det(B)$. We also have $\det(A) = \lambda$, so again we can verify the formula just by computing both sides.

This shows that $\det(AB) = \det(A)\det(B)$ whenever A is an elementary matrix.

By induction, we can now show that $\det(AB) = \det(A)\det(B)$ whenever A is a product of elementary matrices, in other words whenever A is an invertible matrix.

If A is not invertible, then $\det(A) = 0$ (by the previous theorem) so to verify the formula in this case, we need to prove that $\det(AB) = 0$. Suppose for a contradiction that $\det(AB) \neq 0$. Then AB is invertible. But if AB is invertible then $B(AB)^{-1}$ is an inverse for A:

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

This gives a contradiction, so we deduce that $\det(AB) = 0$.

We have now checked the formula for all possible matrices A, which proves the theorem.

7.2 Cofactor expansion

Recall that the determinant of a 3-by-3 matrix $A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$ is

$$A_{11}A_{22}A_{33} - A_{11}A_{23}A_{32} - A_{12}A_{21}A_{33} + A_{12}A_{23}A_{31} + A_{13}A_{21}A_{32} - A_{13}A_{22}A_{31}.$$

We can group these into three sets of terms according to the factor A_{1k}: those which contain a factor of A_{11}, those with an A_{12} and those with an A_{13}:

$$A_{11}(A_{22}A_{33} - A_{23}A_{32}) - A_{12}(A_{21}A_{33} - A_{23}A_{31}) + A_{13}(A_{21}A_{32} - A_{22}A_{31}).$$

Note that the terms in brackets are themselves determinants of 2-by-2 submatrices:

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

When we group our determinant according to the choice of the factor A_{1k}, the expansion we get is:

$$\det(A) = A_{11} \det(C_{11}) - A_{12} \det(C_{12}) + A_{13} \det(C_{13}) - \cdots \pm A_{1n} \det(C_{1n}).$$

Here:

- signs alternate; in particular, the last $\pm$ is $+$ if n is odd and $-$ if n is even.
- C_{ij} is the submatrix of A obtained by removing the ith row and the jth column. For example, $C_{22} = \begin{pmatrix} A_{11} & X & A_{13} \\ X & X & X \\ A_{31} & X & A_{33} \end{pmatrix} = \begin{pmatrix} A_{11} & A_{13} \\ A_{31} & A_{33} \end{pmatrix}$ (where the Xs show which columns/rows we have removed).

This makes sense: if you pick A_{ij} then you can't pick anything else from row i or from column j.

Example 7.3. For the 3-by-3 matrix A as above, $C_{12} = \begin{pmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{pmatrix}$.

More generally, we could group terms according to the choice of factor A_{2k}, or A_{3k}, or any other row. The result of expanding about row i would have been:

$$\det(A) = \pm A_{i1} \det(C_{i1}) \mp A_{i2} \det(C_{i2} \pm \cdots$$

where the signs alternate (I'll talk more about exactly which signs occur below).

Or we could group terms according to the choice of entry in column 1, or column 2, etc. The result of expanding down column j would have been

$$\det(A) = \pm A_{1j}\det(C_{1j}) \mp A_{2j}\det(C_{2j} \pm \cdots$$

where the signs alternate.

7.2.1 Signs

The sign you put in front of $A_{ij}\det(C_{ij})$ is the ijth entry of this matrix:

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

For example, expanding a 3-by-3 matrix down column 2 would give

$$-A_{12}\det(C_{12}) + A_{22}\det(C_{22}) - A_{32}\det(C_{32}).$$

7.2.2 Examples

Rather than prove any of these formulae, which would be painful for everybody, let's do some examples.

Example 7.4. Let's compute the determinant of

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

using the cofactor expansion along the first row:

$$\begin{aligned} \det(A) &= 1 \times \det\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} - 2\det\begin{pmatrix} 4 & 6 \\ 7 & 9 \end{pmatrix} + 3\det\begin{pmatrix} 4 & 5 \\ 7 & 8 \end{pmatrix} \\ &= 5 \times 9 - 6 \times 8 - 2(4 \times 9 - 6 \times 7) + 3(4 \times 8 - 5 \times 7) \\ &= 45 - 48 - 2(36 - 42) + 3(32 - 35) \\ &= -3 + 12 - 9 \\ &= 0. \end{aligned}$$

Example 7.5. Just for comparison, let's expand down the first column (I said row in the video):

$$\begin{aligned} \det(A) &= 1\det\begin{pmatrix} 5 & 6 \\ 8 & 9 \end{pmatrix} - 4\det\begin{pmatrix} 2 & 3 \\ 8 & 9 \end{pmatrix} + 7\det\begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix} \\ &= -3 - 4(18 - 24) + 7(12 - 15) \\ &= -3 + 24 - 21 \\ &= 0. \end{aligned}$$

Which, reassuringly, gives the same answer.

Example 7.6. Let's calculate the determinant of

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

Expanding along the top row would give us four terms to calculate. If instead we expand around the first or second columns or the second or fourth rows, then we only have two terms to calculate. Let's do the second column (because the signs will be more fun that way).

We get:

$$\det(B) = -1 \det \begin{pmatrix} 0 & 4 & 5 \\ -1 & 1 & 1 \\ 0 & 2 & 3 \end{pmatrix} - 2 \det \begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 2 & 3 \end{pmatrix}$$

(care with signs!). For the first 3-by-3 submatrix (C_{12}) we have (expanding down the first column):

$$\det \begin{pmatrix} 0 & 4 & 5 \\ -1 & 1 & 1 \\ 0 & 2 & 3 \end{pmatrix} = -(-1)(4 \times 3 - 5 \times 2) = 2$$

For the second, we get (expanding down the first column):

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

so overall we get

$$\det(B) = -1 \times 2 - 2 \times 2 = -6.$$

Example 7.7. As a final sanity check, let's compute this again using row operations. If we do $R_3 \mapsto R_3 + R_1$, then B becomes

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

Now we switch $R_2 \leftrightarrow R_3$ and then $R_3 \leftrightarrow R_4$ (so the original row 2 ends up at the bottom):

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

Finally $R_4 \mapsto R_4 - 2R_3$:

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

This is in echelon form. We did two row-switches, so the determinant changed sign twice (i.e. didn't change sign overall), so that $\det(B)$ is just the product of the diagonal entries in this echelon matrix. In other words,

$$\det(B) = 1 \times 3 \times 2 \times (-1) = -6,$$

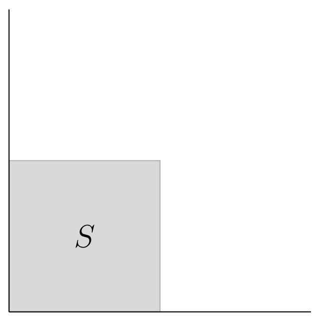
which agrees with the previous example.

7.3 Geometry of determinants, 1

7.3.1 2-by-2 determinants and area

In this video, we will reconnect with geometry.

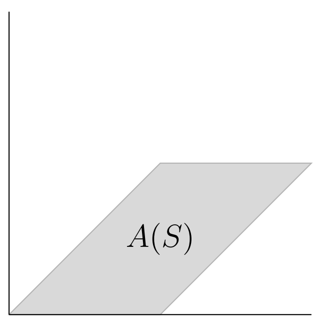
Theorem 7.8. *Let* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *be a 2-by-2 matrix and* $S \subset \mathbf{R}^2$ *be the unit square:*



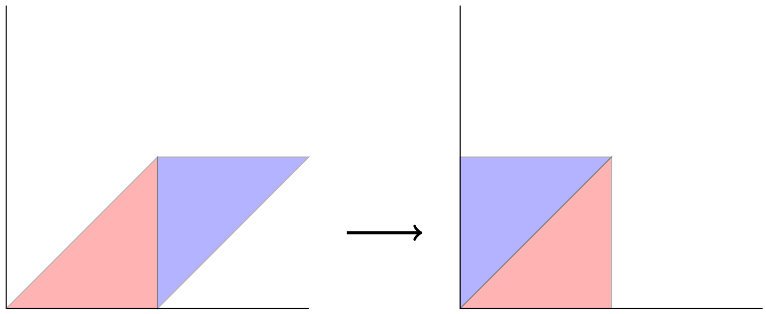
Let $A(S)$ *be the image of* S *under the geometric transformation* A *(*$A(S)$ *is a parallelogram. Then*

$$|\det(A)| = \text{area}(A(S)).$$

Example 7.9. If $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ then $A(S)$ looks like this (parallelogram with vertices at $(0,0), (1,0), (2,1), (1,1)$):

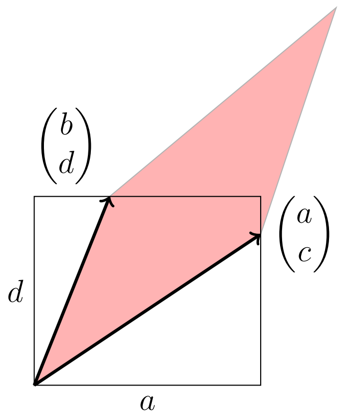


We can find the area of this parallelogram by decomposing it into two triangles, slicing it vertically between $(1,0)$ and $(1,1)$, and then shifting the right-hand triangle to the left until they form a rectangle (actually a square) with side lengths 1 and 1. Since this rearrangement didn't change the area, we see that the area of $A(S)$ is 1, which is indeed the determinant of A.



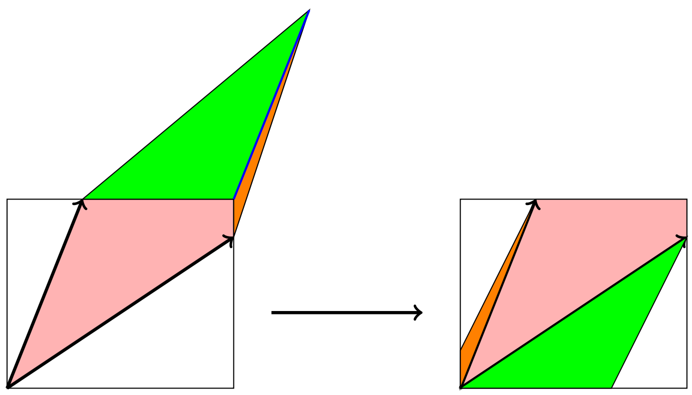
7.3.2 Proof of theorem

Take the parallelogram $A(S)$ with edges $A\begin{pmatrix}1\\0\end{pmatrix} = \begin{pmatrix}a\\c\end{pmatrix}$ and $A\begin{pmatrix}0\\1\end{pmatrix} = \begin{pmatrix}b\\d\end{pmatrix}$. We will draw a rectangle with its bottom-left corner at the origin, height d and base a. This passes through the vertices $\begin{pmatrix}a\\c\end{pmatrix}$ and $\begin{pmatrix}b\\d\end{pmatrix}$.

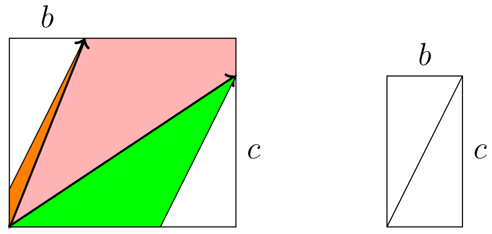


The area of this rectangle is ad. Let's subdivide $A(S)$ into pieces, rearrange them so they live inside this rectangle, and observe that there's still a bit missing, which has area bc, which will prove that $\text{area}(A(S)) = ad - bc$.

We draw a line connecting the top right corner of the rectangle to the top right corner of the parallelogram. Together with the rectangle, this cuts our parallelogram into three pieces. We shift the parts outside the rectangle until they sit inside the rectangle, and so that the edges which used to be opposite parallel edges match up:



What's missing? There are two (white) triangles still missing. These triangles share a parallel edge, so can be rearranged to form a rectangle. They have base b and height c, so the total missing area is bc. Therefore the area of $A(S)$ equals the area of the big rectangle (ad) minus the total area of these two triangles (bc).



7.4 Geometry of determinants, 2

7.4.1 Higher dimensions

Last time we saw a geometric interpretation of the determinant of a 2-by-2 matrix A as the area of a parallelogram obtained by applying A to a square. This result holds in all dimensions, though we will not prove it:

Theorem 7.10. *If A is an n-by-n matrix and $S \subset \mathbf{R}^n$ is the unit n-dimensional cube then $|\det(A)| = \mathrm{vol}(A(S))$.*

Remark 7.11. The absolute value signs around the determinant need to be there because areas and volumes are positive but determinants can be negative. In the proof of the 2-by-2 case, I hid the issue in the picture I drew: if I had switched the vectors $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$ in the picture, we would have ended up with $bc - ad$, i.e. $-\det(A)$.

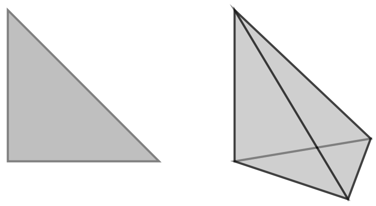
Remark 7.12. In higher dimensions, $A(S)$ is no longer a parallelogram, but a *parallelopiped* ("parallel-legged"): a shape whose faces are parallel pairs of congruent parallelograms.

7.4.2 Tetrahedra/simplices

Another useful characterisation of the determinant (which we won't prove) is the following.

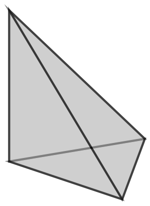
Theorem 7.13. *Let $a_1, \ldots, a_n$ be vectors in $\mathbf{R}^n$. Consider the* simplex *with vertices at $0, a_1, \ldots, a_n$. Then the volume of this simplex is $\frac{1}{n!}\det(A)$ where A is the matrix with columns $a_1, \ldots, a_n$.*

Remark 7.14. A simplex is the simplest convex shape you can draw with these vertices: you take the vertices, connect them by lines, and then fill in all the space in between. (More formally, this is called taking the convex hull of the vertices). In 2 dimensions this gives a triangle (3 vertices). In 3 dimensions you get a tetrahedron.

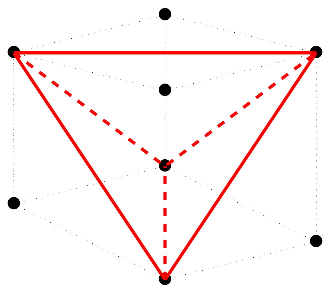


These two theorems are closely related: you can decompose a cube into $n!$ simplices. Rather than proving either of them, I want to use this second theorem to calculate the volumes of some tetrahedra in $\mathbf{R}^3$.

Example 7.15. Take $a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $a_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $a_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$. The corresponding tetrahedron has volume $\frac{1}{6} \det \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} = \frac{1}{6}$.



Example 7.16. Take a regular tetrahedron (d4 for the roleplayers out there) with vertices at $a_0 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, $a_1 = \begin{pmatrix} 1 \\ -1 \\ -1 \end{pmatrix}$, $a_2 = \begin{pmatrix} -1 \\ 1 \\ -1 \end{pmatrix}$ and $a_3 = \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix}$. (These vertices are half of the vertices of a regular cube centred at the origin with sidelength 2.)



Translating by $-a_0$, we get the vertices $0, a_1 - a_0, a_2 - a_0, a_3 - a_0$, so we can apply our theorem with

$$A = \begin{pmatrix} a_1 - a_0 & a_2 - a_0 & a_3 - a_0 \end{pmatrix} = \begin{pmatrix} 0 & -2 & -2 \\ -2 & 0 & -2 \\ -2 & -2 & 0 \end{pmatrix}.$$

We compute the determinant using the cofactor expansion along the top row:

$$\begin{aligned}
\text{vol} &= \left| \frac{1}{6} \det \begin{pmatrix} 0 & -2 & -2 \\ -2 & 0 & -2 \\ -2 & -2 & 0 \end{pmatrix} \right| \\
&= \frac{1}{6} \left| (-(-2) \det \begin{pmatrix} -2 & -2 \\ -2 & 0 \end{pmatrix} - 2 \det \begin{pmatrix} -2 & 0 \\ -2 & -2 \end{pmatrix} \right| \\
&= \frac{2}{6} \left| (-4 - 4) \right| \\
&= \frac{16}{6}
\end{aligned}$$

7.4.3 Determinant as scale factor for volumes

$\det(A)$ is the scale factor for volumes of cubes under the transformation A. Since any volume is defined by subdividing/approximating by smaller and smaller cubes, $\det(A)$ is the scale factor for any volume under the transformation A. From this point of view, the formula $\det(AB) = \det(A)\det(B)$ is just telling us that if we apply the transformation B and then the transformation A, the volumes rescale by $\det(B)$ and then by $\det(A)$, giving a total scale factor of $\det(A)\det(B)$ for the composite transformation.

8 Week 4, Session 2: Eigenvectors and eigenvalues

8.1 Eigenvectors and eigenvalues

8.1.1 Eigenvectors and eigenvalues

Given a complicated matrix, it can be difficult to determine salient information just by looking at the array of numbers. For example, the matrix:

$$A = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} + \frac{1}{\sqrt{3}} & -\frac{1}{3} + \frac{1}{\sqrt{3}} \\ \frac{1}{3} - \frac{1}{\sqrt{3}} & \frac{1}{3} & -\frac{1}{3} - \frac{1}{\sqrt{3}} \\ -\frac{1}{3} - \frac{1}{\sqrt{3}} & -\frac{1}{3} + \frac{1}{\sqrt{3}} & \frac{1}{3} \end{pmatrix}$$

describes a 90 degree rotation around $\begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$, but you wouldn't know it just by looking at it.

We saw that, if you have a rotation matrix, then, to determine the axis of rotation, you just need to solve $Av = v$ for v. We now introduce a generalisation of this equation.

Definition 8.1. Suppose A is an n-by-n matrix, $v \in \mathbf{C}^n$ is a complex vector and $\lambda \in \mathbf{C}$ is a complex number. If $Av = \lambda v$ then we say that v is an eigenvector of A with eigenvalue λ.

Remark 8.2. We'll usually require $v \neq 0$: this is because $v = 0$ is *always* a solution to $Av = \lambda v$, so we don't really care about it (it tells us nothing interesting about A). We'll often abbreviate "eigenvector" to "evec" and "eigenvalue" to "eval". The prefix "eigen-" comes from German, and means "self": v is sent to itself by A, up to a scale factor λ. You can use the prefix to construct many fun words, including eigenline (line of eigenvectors), eigendirection (direction in which eigenvector points), eigenspace (space consisting of eigenvectors for some particular eigenvalue), eigenfunctions (for example, $e^{\lambda x}$ is an eigenfunction of d/dx because $d(e^{\lambda x})/dx = \lambda e^{\lambda x}$.

It's hard to overemphasise the importance of eigenvalues and eigenvectors in mathematics and its applications. For example, in quantum mechanics, Schrödinger's equation is really just the statement that, if you have a quantum system, then the possible values that energy can take on are the eigenvalues of an operator called the Hamiltonian. This allows you predict the absorption/emission spectrum of hydrogen to great accuracy, which is one of the reasons people were so convinced by quantum mechanics early on. We'll focus on applications within mathematics, but this eigenstuff we're doing is really central in maths and in science.

8.1.2 The eigenvector equation

For each λ, $Av = \lambda v$ is an equation for v: you fix an eigenvalue and you look for eigenvectors associated to that eigenvalue. You always find the solution $v = 0$, but very rarely do you find anything else. The λs for which $Av = \lambda v$ has a nonzero solution are called the *eigenvalues* of A. Not every number occurs as an eigenvalue of A. In the next video we'll see how to figure out which numbers are eigenvalues for a given matrix.

In this video, we'll answer the question: if someone gives you an eigenvalue of A, how do you find the eigenvectors?

Example 8.3. If someone gives you $A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ and tells you that $\lambda = 1$ is an eigenvalue of A, you can find the eigenvectors by solving $Av = \lambda v = v$:

$$\begin{aligned} Av &= \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} 2x - y \\ x \end{pmatrix}, \end{aligned}$$

so $Av = v$ implies $2x - y = x$ and $x = y$. The second equation $x = y$ implies the first, so the eigenvectors are just the vectors of the form $\begin{pmatrix} x \\ x \end{pmatrix}$.

All we did here was to solve a system of simultaneous equations for v. So if someone gives you λ, finding v is something we know how to do. If we write this system in matrix form, we get $(A - \lambda I)v = 0$ (i.e. the constants are all zero, not λv).

What would happen if I picked something other than $\lambda = 1$?

Example 8.4. If I keep the same A, pick $\lambda = 2$ and try to solve $Av = 2v$ for v, I get:

$$2x - y = 2x, \qquad x = 2y,$$

which implies $y = 0$ and $x = 2y = 0$, so $v = 0$. This means that 2 is not an eigenvalue of A.

Currently we're just guessing λ, and there are infinitely many possible λs, so we need a better way of figuring out which λs are eigenvalues of A. That's what we'll do in the next video.

8.2 Finding eigenvalues

8.2.1 Characteristic polynomial

In the last video, we introduced the equation $Av = \lambda v$. For each λ, this gives us an equation for v. The question we will now answer is: for which $\lambda \in \mathbf{C}$ does this equation have a nonzero solution v?

Theorem 8.5. *$Av = \lambda v$ has a nonzero solution if and only if λ is a root of the* characteristic polynomial *of A:*

$$\det(A - tI) = 0.$$

Here, t is just a dummy variable we've introduced (not one of the components of v or anything like that). The characteristic polynomial is a polynomial in t of degree n. We'll do some examples, then prove the theorem.

8.2.2 Examples

Example 8.6. Suppose $A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$. The characteristic polynomial is:

$$\begin{aligned} \det(A - tI) &= \det\left(\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} - t\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \\ &= \det\begin{pmatrix} 2-t & -1 \\ 1 & -t \end{pmatrix} \\ &= -t(2-t) + 1 = t^2 - 2t + 1. \end{aligned}$$

The roots of this quadratic are $\frac{2\pm\sqrt{4-4}}{2} = 1$, i.e. this is a double root (the discriminant of the quadratic equation is zero). This means that the only eigenvalue of this matrix is 1: had we picked any other value for λ, we would not have been able to solve $Av = \lambda v$.

You will come to love the formula for solving quadratics; it lets you find the eigenvalues of any 2-by-2 matrix. By contrast, a 3-by-3 matrix will have a cubic characteristic polynomial. Whilst there is a formula for solving cubics, it's not nice. For 4-by-4 matrices, it gets still worse. For 5-by-5 and bigger matrices, the characteristic polynomial is a quintic or higher degree polynomial, and there's (provably) no general formula for the solution of a general quintic in terms of taking kth roots etc.

Example 8.7. Let $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then

$$\det(A - tI) = \det\begin{pmatrix} -t & -1 \\ 1 & -t \end{pmatrix} = t^2 + 1.$$

The roots (eigenvalues) are $\pm i$. This is why, even though our matrix is real, we may need to deal with complex numbers when we start working with eigenvalues and eigenvectors.

Let's figure out the eigenvectors. For $\lambda = i$, we need to solve $Av = iv$:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ix \\ iy \end{pmatrix}.$$

Multiplying this out gives:

$$\begin{pmatrix} -y \\ x \end{pmatrix} = \begin{pmatrix} ix \\ iy \end{pmatrix},$$

which implies $y = -ix$ and $x = iy$. The second equation follows from the first if you multiply by i. The eigenvectors for $\lambda = i$ are those of the form $\begin{pmatrix} x \\ -ix \end{pmatrix}$.

For $\lambda = -i$, we need to solve $Av = -iv$, which gives $y = ix$, and the eigenvectors are those of the form $\begin{pmatrix} x \\ ix \end{pmatrix}$.

8.2.3 Proof of theorem

If there exists a nonzero solution v to $Av = \lambda v$ then $(A - \lambda I)v = 0$. This implies that $A - \lambda I$ is not invertible; otherwise we get $v = (A - \lambda I)^{-1}0 = 0$. Therefore $\det(A - \lambda I) = 0$, so λ is a root of $\det(A - tI)$.

In fact, these are all "if and only if" statements. The only nonobvious one is to see that if $A - \lambda I$ is not invertible then there exists a nonzero v such that $(A - \lambda I)v = 0$ (you might like to think about why that's true).

8.3 Eigenexamples

Example 8.8. Let $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. The characteristic polynomial is

$$\det(A - tI) = \det \begin{pmatrix} 2-t & 1 \\ 1 & 1-t \end{pmatrix} = (2-t)(1-t) - 1 = t^2 - 3t + 1.$$

The roots of this polynomial are $\frac{3 \pm \sqrt{9-4}}{2} = \frac{3 \pm \sqrt{5}}{2}$.

Rather than doing the $\pm$ cases separately, I'll just leave $\pm$ in the notation; at the end, the answer will end up with a $\pm$ sign in and you pick the sign corresponding to the eigenvalue you want. The eigenvectors for the eigenvalue $\frac{3 \pm \sqrt{5}}{2}$ are the solutions to $Av = \lambda v$, i.e.

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \frac{3 \pm \sqrt{5}}{2} \begin{pmatrix} x \\ y \end{pmatrix}.$$

Multiplying out, this gives the equations

$$2x + y = \frac{3 \pm \sqrt{5}}{2} x, \qquad x + y = \frac{3 \pm \sqrt{5}}{2} y.$$

Rearranging, we get

$$y = \frac{-1 \pm \sqrt{5}}{2} x, \qquad x = \frac{1 \pm \sqrt{5}}{2} y.$$

As usual, the second equation follows from the first (check it!), so our eigenvectors are those of the form $\begin{pmatrix} x \\ \frac{-1 \pm \sqrt{5}}{2} x \end{pmatrix}$.

What this notation means is that the eigenvectors for $\frac{3+\sqrt{5}}{2}$ are $\begin{pmatrix} x \\ \frac{-1+\sqrt{5}}{2} x \end{pmatrix}$ and the eigenvectors for $\frac{3-\sqrt{5}}{2}$ are $\begin{pmatrix} x \\ \frac{-1-\sqrt{5}}{2} x \end{pmatrix}$. Another useful piece of notation is $\mp$ (which means $-\pm$).

Example 8.9. Let $A = \begin{pmatrix} \frac{3}{2} & \frac{5}{2} & 3 \\ -\frac{1}{2} & -\frac{3}{2} & -3 \\ 1 & 1 & 2 \end{pmatrix}$. The characteristic polynomial is:

$$\begin{aligned}
\det(A - tI) &= \det \begin{pmatrix} \frac{3}{2} - t & \frac{5}{2} & 3 \\ -\frac{1}{2} & -\frac{3}{2} - t & -3 \\ 1 & 1 & 2 - t \end{pmatrix} \\
&= \left(\frac{3}{2} - t\right) \det \begin{pmatrix} -\frac{3}{2} - t & -3 \\ 1 & 2 - t \end{pmatrix} - \frac{5}{2} \det \begin{pmatrix} -\frac{1}{2} & -3 \\ 1 & 2 - t \end{pmatrix} \\
&\qquad + 3 \det \begin{pmatrix} -\frac{1}{2} & -\frac{3}{2} - t \\ 1 & 1 \end{pmatrix} \\
&= \left(\frac{3}{2} - t\right) \left(-\left(\frac{3}{2} + t\right)(2 - t) + 3\right) - \frac{5}{2}\left(-\frac{1}{2}(2 - t) + 3\right) \\
&\qquad + 3\left(-\frac{1}{2} + \frac{3}{2} + t\right) \\
&= \left(\frac{3}{2} - t\right)(t^2 - t/2) - \frac{5}{2}(t/2 + 2) + 3t + 3 \\
&= -t^3 + 2t^2 + t - 2.
\end{aligned}$$

To find the roots of this cubic, we use a standard trick: we guess one of them! Let's guess $t = 1$. Amazingly this is a root (funny how often that happens in carefully designed examples) because $-1^3 + 2(1^2) + 1 - 2 = 0$. Once we have this, we know that our cubic factors as $(t - 1) \times$ a quadratic. Now we get the quadratic by doing polynomial long division. I don't know a good way of typesetting this, so you'll have to look at the video for the working! Note the similarities between long division and row reduction! These are both special cases of a general technique called "Groebner bases".

The result is:

$$-t^2 + t + 2.$$

This has roots $\frac{-1 \pm \sqrt{1+8}}{-2}$, i.e. -1 or 2.

This means that the eigenvalues are $-1, 1, 2$. What are the eigenvectors? I'll just work out the eigenvectors for $\lambda = 1$ and leave the others as a fun exercise.

We want to solve $Av = v$, which means

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

The first two rows are both equivalent to $x + 5y + 6z = 0$, the third means $z = -x - y$, so overall we get

$$y = -5x, \qquad z = 4x.$$

In other words, the eigenvectors have the form

$$\begin{pmatrix} x \\ -5x \\ 4x \end{pmatrix}.$$

8.4 Eigenspaces

In all the examples we've seen so far, the eigenvectors have all had a free variable in them. For example, in the last video, we found the eigenvectors for the matrix $A = \begin{pmatrix} \frac{3}{2} & \frac{5}{2} & 3 \\ -\frac{1}{2} & -\frac{3}{2} & -3 \\ 1 & 1 & 2 \end{pmatrix}$ to be:

- for $\lambda = 1$, $\begin{pmatrix} x \\ -5x \\ 4x \end{pmatrix}$,
- for $\lambda = 2$, $\begin{pmatrix} x \\ -x \\ 0 \end{pmatrix}$,
- for $\lambda = -1$, $\begin{pmatrix} x \\ -x \\ x \end{pmatrix}$

For the matrix $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ we found the eigenvalue $\frac{3\pm\sqrt{5}}{2}$ had eigenvectors $\begin{pmatrix} x \\ \frac{-1\pm\sqrt{5}}{2}x \end{pmatrix}$. All of these have the free parameter x.

This is a general fact:

Lemma 8.10. *If v is an eigenvector of A with eigenvalue λ ("a λ-eigenvector") then so is kv for any $k \in \mathbf{C}$.*

Proof. $A(kv) = kAv = k\lambda v = \lambda(kv)$.

□

So for example, the vectors $\begin{pmatrix} x \\ -x \\ x \end{pmatrix}$ are all just rescalings of $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$. Indeed, people often say things like "the eigenvector is $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$", when they mean "the eigenvectors are all the rescalings of $\begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}$". If you write this kind of thing in your answers, that's fine.

Example 8.11. Suppose we have the matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. The characteristic polynomial is $\det(I - \lambda I) = \det \begin{pmatrix} 1-\lambda & 0 \\ 0 & 1-\lambda \end{pmatrix} = (1-\lambda)^2$, so $\lambda = 1$ is the only eigenvalue. Any vector v satisfies $Iv = v$, so any vector $\begin{pmatrix} x \\ y \end{pmatrix}$ is a 1-eigenvector. This has two free parameters, so it is an *eigenplane*, not just an *eigenline*: there is a whole plane of eigenvectors for the same eigenvalue.

Theorem 8.12. *The set of eigenvectors with eigenvalue λ form a (complex) subspace of $\mathbf{C}^n$ (i.e. closed under complex rescalings and under addition).*

Proof. Let V_λ be the set of λ-eigenvectors of A. If $v \in V_\lambda$ then $kv \in V_\lambda$ (as we saw above). If $v_1, v_2 \in V_\lambda$ then

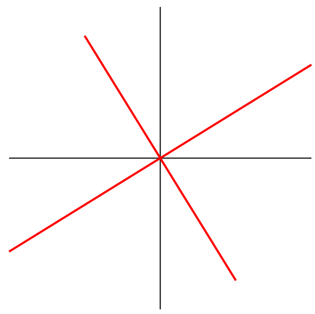
$$A(v_1 + v_2) = Av_1 + Av_2 = \lambda v_1 + \lambda v_2 = \lambda(v_1 + v_2),$$

so $v_1 + v_2$ is also a λ-eigenvector.

□

We call this subspace the λ-eigenspace. In all the examples we saw earlier (except I), the eigenspaces were 1-dimensional eigenlines (one free variable). So a matrix gives you a collection of preferred directions or subspaces (its eigenspaces), which tell you something about the matrix (e.g. if it's a rotation matrix, its axis will be one of these subspaces).

Example 8.13. For the example $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ we found the eigenvalues $\frac{3\pm\sqrt{5}}{2}$ and eigenvectors $\begin{pmatrix} x \\ \frac{-1\pm\sqrt{5}}{2}x \end{pmatrix}$. We now draw these two eigenlines (in red).



Note that these eigenlines look orthogonal; indeed, you can check that they are! You do this by taking the dot product of the eigenvectors (it's zero). This is true more generally for symmetric matrices (i.e. matrices A such that $A = A^T$).

9 Week 5, Session 1: Eigenapplications

9.1 Differential equations

9.1.1 Sketch of the idea

This is the first of three applications of eigenvectors and eigenvalues. Let $x(t), y(t)$ be functions of t, and consider the simultaneous linear ordinary differential equations

$$\begin{aligned} \dot{x} &= ax + by \\ \dot{y} &= cx + dy, \end{aligned}$$

where $\dot{x}$ denotes dx/dt and a, b, c, d are constants.

These equations are *coupled*: each equation involves x and y. Our goal is to decouple them, turning them into a pair of equations, each involving only one unknown function. We first rewrite the equations as

$$\dot{v} = Av$$

using:

$$v(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}, \qquad \dot{v} = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix}, \qquad A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Let λ_1, λ_2 be the eigenvalues of A and assume they are distinct; let u_1 and u_2 be eigenvectors for these eigenvalues. Write v as $\alpha(t)u_1 + \beta(t)u_2$. In other words, α and β are the components of v when in the u_1- and u_2-directions (compare with $v = x\begin{pmatrix} 1 \\ 0 \end{pmatrix} + y\begin{pmatrix} 0 \\ 1 \end{pmatrix}$).

Substituting $v = \alpha u_1 + \beta u_2$ into the left-hand side of $\dot{v} = Av$ gives:

$$\dot{v} = \dot{\alpha}u_1 + \dot{\beta}u_2$$

(because u_1, u_2 are constant).

Substituting it into the right-hand side gives

$$Av = \alpha A u_1 + \beta A u_2 = \lambda_1 \alpha u_1 + \lambda_2 \beta u_2$$

(because u_k is a λ_k-eigenvector for $k = 1, 2$).

Comparing the components in the u_1-direction, we get

$$\dot{\alpha} = \lambda_1 \alpha.$$

The u_2-components tells us that

$$\dot{\beta} = \lambda_2 \beta.$$

These equations now involve only α and β separately. We have decoupled the equations.

9.1.2 An example in detail

Consider the equations:

$$\dot{x} = y$$
$$\dot{y} = -x.$$

This is equivalent to the *simple harmonic oscillator equation* $\ddot{x} = -x$. To see this, differentiate $\dot{x} = y$ to get $\ddot{x} = \dot{y} = -x$.

In general, if you have a second-order equation like this, you can define $y = \dot{x}$ (which gives you one equation) and then use the second-order equation to express $\dot{y}$ in terms of x and y: this is a good trick for converting second-order equations into pairs of first order equations.

Remark 9.1. The equation $\ddot{x} = -x$ describes physical situations like a particle on a spring:

- one end of the spring is fixed at the origin;
- the particle sits at a distance x from the origin;
- Newton's law tells us that if m is the mass of the particle then $m\ddot{x}$ equals the force experience by the particle, which is $-kx$ by Hooke's law (for some constant k), so the equation of motion is $m\ddot{x} = -kx$;
- if we work in units where $m = k = 1$ then this gives us back our equation $\ddot{x} = -x$.

Let's solve this equation. We have $A = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. This has characteristic polynomial

$$\det \begin{pmatrix} -t & 1 \\ -1 & -t \end{pmatrix} = t^2 + 1,$$

which has roots $\lambda_1 = i$ and $\lambda_2 = -i$.

The eigenvectors are:

- for $\lambda_1 = i$, $\begin{pmatrix} 1 \\ i \end{pmatrix}$,
- for $\lambda_2 = -i$, $\begin{pmatrix} 1 \\ -i \end{pmatrix}$.

(We're just picking particular eigenvectors, not writing down the general eigenvector).

We write $\begin{pmatrix} x \\ y \end{pmatrix}$ in the form

$$\alpha u_1 + \beta u_2 = \alpha \begin{pmatrix} 1 \\ i \end{pmatrix} + \beta \begin{pmatrix} 1 \\ -i \end{pmatrix} = \begin{pmatrix} \alpha + \beta \\ i(\alpha - \beta) \end{pmatrix}.$$

Thus $\alpha = (x - iy)/2$ and $\beta = (x + iy)/2$.

Our equations are now:

$$\dot{\alpha} = \lambda_1 \alpha = i\alpha, \qquad \dot{\beta} = \lambda_2 \beta = -i\beta.$$

Dividing the first equation by α gives

$$\frac{\dot{\alpha}}{\alpha} = \frac{d}{dt}(\log \alpha) = i,$$

so $\log \alpha = it + \text{const}$, so $\alpha(t) = C_1 e^{it}$. Similarly, $\beta(t) = C_2 e^{-it}$.

Therefore

$$\begin{pmatrix} x \\ y \end{pmatrix} = C_1 e^{it} \begin{pmatrix} 1 \\ i \end{pmatrix} + C_2 e^{-it} \begin{pmatrix} 1 \\ -i \end{pmatrix},$$

so

$$\begin{aligned} x &= C_1 e^{it} + C_2 e^{-it} \\ y &= i(C_1 e^{it} - C_2 e^{-it}). \end{aligned}$$

This is the general solution to $\dot{x} = y$, $\dot{y} = -x$. You may be worried about the fact that there are is here: this is supposed to describe the motion of a particle on a spring, so x and y should be real numbers. The is will all cancel out if we pick appropriate initial conditions.

Suppose $x(0) = 1$ and $y(0) = 0$ (particle at rest at distance 1 from the origin). Substituting $t = 0$ into our general solution, we get

$$x(0) = 1 = C_1 + C_2, \qquad y(0) = 0 = i(C_1 - C_2).$$

This implies $C_1 = C_2 = 1/2$. Therefore

$$x(t) = \frac{1}{2}(e^{it} + e^{-it}), \qquad y(t) = \frac{i}{2}(e^{it} - e^{-it}).$$

This means that

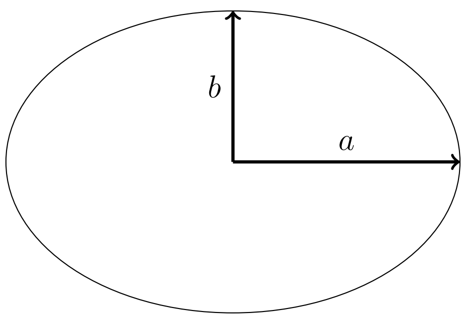
$$x(t) = \cos(t), \qquad y(t) = -\sin(t)$$

using the formulae for trigonometric functions in terms of complex exponentials. The minus sign in y is because our particle starts moving towards the origin as time increases.

The moral of this story is that you can use eigenvectors to decouple systems of linear differential equations.

9.2 Ellipses

An ellipse is a curve in the plane: it looks like a squashed circle:



The one in the diagram above has been squashed/stretched vertically/horizontally, and has equation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$. Here, a is the biggest value that x can take (because if $x = a$ then y has to be zero) and b is the biggest value that y can take. Suppose $a > b$. If you look at diameters (chords of the ellipse passing through the origin) then the longest will have length $2a$ (pointing in the x-direction) and the shortest will have length $2b$ (pointing vertically). We call a the *semimajor axis* and b the *semiminor* axis.

Suppose someone gives you an ellipse that has been squashed in some other direction. What is the equation for it? Conversely, if someone gives you the equation of an ellipse, how do you figure out the semimajor and semiminor axes?

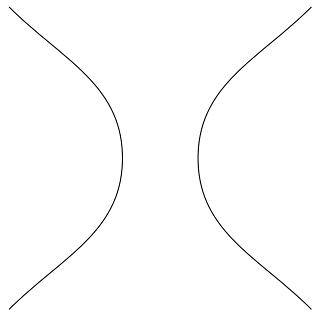
9.2.1 General equation of an ellipse

What's the general equation of an ellipse? Assuming its centre of mass is at the origin, the general equation has the form

$$Ax^2 + Bxy + Cy^2 = 1.$$

If I wanted the centre of mass to be elsewhere, I could add terms like $Dx + Ey$.

The constants A, B, C can't be just anything. For example, if we take $A = 1$, $B = 0$, $C = -1$, then we get $x^2 - y^2 = 1$, which is the equation of a hyperbola:



The condition on A, B, C we need to get an ellipse is *positive definiteness*:

Definition 9.2. The quadratic form $Ax^2 + Bxy + Cy^2$ is *positive definite* if it's positive whenever we substitute in real numbers x, y other than $0, 0$.

Definition 9.3. An ellipse is a subset of $\mathbf{R}^2$ cut out by the equation $Ax^2+Bxy+Cy^2 = 1$ where A, B, C are constants making the left-hand side positive definite.

9.2.2 Normal form for ellipses

Theorem 9.4. *Consider the matrix* $M := \begin{pmatrix} A & B/2 \\ B/2 & C \end{pmatrix}$. *Let u_1, u_2 be unit length eigenvectors of M (with eigenvalues λ_1, λ_2). Pick coordinates so that the new x- and y-axes point along the eigenvectors u_1, u_2 (and so that u_1 sits at $(1, 0)$ and u_2 sits at $(0, 1)$). In these new coordinates, the equation of the ellipse becomes*

$$\lambda_1 x^2 + \lambda_2 y^2 = 1.$$

This change of coordinates will actually be a rotation of the usual coordinates.

The matrix M arises as follows. Let $v = \begin{pmatrix} x \\ y \end{pmatrix}$. Then

$$Ax^2 + Bxy + Cy^2 = v^T M v.$$

Lemma 9.5. *If $\lambda_1 \neq \lambda_2$ then the eigenvectors u_1 and u_2 are orthogonal to one another. This works for any matrix M for which $M^T = M$.*

Proof. We have $Mu_1 = \lambda_1 u_1$ and $Mu_2 = \lambda_2 u_2$. Consider $u_1^T M u_2$:

$$\begin{aligned} u_1^T M u_2 &= u_1^T \lambda_2 u_2 \\ &= \lambda_2 u_1 \cdot u_2. \end{aligned}$$

We also have $u_1^T M u_2 = u_1^T M^T u_2$ (because $M^T = M$), so

$$\begin{aligned} u_1^T M u_2 &= u_1^T M^T u_2 \\ &= (Mu_1)^T u_2 \\ &= \lambda_1 u_1^T u_2 \\ &= \lambda_1 u_1 \cdot u_2. \end{aligned}$$

Therefore $(\lambda_1 - \lambda_2) u_1 \cdot u_2 = 0$. Since $\lambda_1 \neq \lambda_2$, we can divide by $\lambda_1 - \lambda_2$ and get $u_1 \cdot u_2 = 0$.

□

This is why the change of coordinates in the theorem is just a rotation: your eigenvectors are orthogonal, so just rotate your x and y-directions until they point in these directions.

9.2.3 Proof of theorem

In the new coordinates (which I'm still calling x, y), we have $v = xu_1 + yu_2$, so:

$$\begin{aligned} v^T M v &= (xu_1 + yu_2)^T M (xu_1 + yu_2) \\ &= (xu_1 + yu_2)^T (x\lambda_1 u_1 + y\lambda_2 u_2) \\ &= x^2 \lambda_1 u_1 \cdot u_1 + y^2 \lambda_2 u_2 \cdot u_2, \end{aligned}$$

where we have used $u_1 \cdot u_1 = u_2 \cdot u_2 = 1$ and $u_1 \cdot u_2 = 0$. This proves the theorem.

9.2.4 Semimajor and semiminor axes

The theorem tells us that the semimajor and semiminor axes point along the eigenvectors of M. Comparing the equations, we see that the semimajor and semiminor axes are $a = \frac{1}{\sqrt{\lambda_1}}$ and $b = \frac{1}{\sqrt{\lambda_2}}$.

9.2.5 Example

Consider the ellipse

$$\frac{3}{2}(x^2 + y^2) - xy = 1.$$

The matrix M is

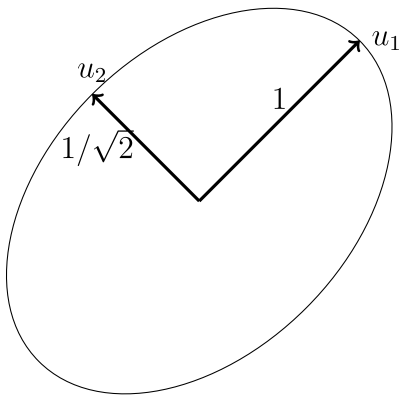
$$M = \begin{pmatrix} \frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{3}{2} \end{pmatrix}.$$

This has characteristic polynomial $\det \begin{pmatrix} 3/2 - t & -1/2 \\ -1/2 & 3/2 - t \end{pmatrix} = t^2 - 3t + 2$, which has roots $\frac{3 \pm \sqrt{9-8}}{2}$, i.e. $\lambda_1 = 1$ and $\lambda_2 = 2$.

The *unit* eigenvectors are:

- for $\lambda_1 = 1$, $u_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$,
- for $\lambda_2 = 2$, $u_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$,

What does this tell us? The semimajor and semiminor axes point in the u_1- and u_2-directions: rotated by 45 degrees from the usual axes. The lengths are $a = \frac{1}{\sqrt{\lambda_1}} = 1$ and $b = \frac{1}{\sqrt{\lambda_2}} = \frac{1}{\sqrt{2}}$.



9.2.6 Ellipsoids

Exactly the same thing works in higher dimensions: an ellipsoid is given by $Q(x_1, \ldots, x_n) = 1$ where Q is a positive definite quadratic form, $Q = v^T M v$ for some symmetric matrix M, and the ellipsoid is related to the standard ellipsoid

$$\sum \frac{x_k^2}{a_k^2} = 1$$

by rotating so that the $x_1, \ldots, x_n$ axes point along the eigendirections of M. The coefficients a_k are given by $\frac{1}{\sqrt{\lambda_k}}$ where λ_k are the eigenvalues.

9.3 Dynamics

We now turn to dynamics. Let v be a vector and A be a matrix. Consider the sequence $v, Av, A^2v, A^3v, \ldots$. We'll investigate what happens to this sequence $A^n v$ as $n \to \infty$.

9.3.1 Example: Fibonacci sequence

Let $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ and $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. We get

$$Av = \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \qquad A^2v = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \qquad A^3v = \begin{pmatrix} 3 \\ 5 \end{pmatrix}, \qquad A^4v = \begin{pmatrix} 5 \\ 8 \end{pmatrix},$$

and, more generally, $A^n v = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}$ where $F_0 = 1$ $F_1 = 1$, $F_2 = 2$, $F_3 = 3$, $F_4 = 5$, $F_5 = 8$, $F_6 = 13$ is the Fibonacci sequence.

Why are we getting the Fibonacci numbers? Suppose the formula $A^n v = \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}$ is true for some value of n; we'll prove it's true for all values of n by induction:

$$A^{n+1}v = AA^n v = A \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix} = \begin{pmatrix} F_{n+1} \\ F_n + F_{n+1} \end{pmatrix} = \begin{pmatrix} F_{n+1} \\ F_{n+2} \end{pmatrix},$$

where we used the recursive formula $F_{n+2} = F_{n+1} + F_n$ which defines the Fibonacci sequence.

As $n \to \infty$ both entries of the vector tends to infinity, but they do so in a particular way:

Theorem 9.6. *We have*

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \frac{1+\sqrt{5}}{2}.$$

This expression is the "golden ratio" $1.618\cdots$.

Proof. Write $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as $\alpha u_1 + \beta u_2$ where u_1 and u_2 are the λ_1- and λ_2-eigenvectors of $A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$. We'll figure out what these eigenvectors and eigenvalues are later.

Now $A^n v = A^n(\alpha u_1 + \beta u_2) = \alpha A^n u_1 + \beta A^n u_2$. We have $Au_1 = \lambda_1 u_1$, $A^2 u_1 = \lambda_1 A u_1 = \lambda_1^2 u_1$, and by induction we get

$$A^n u_1 = \lambda_1^n u_1, \qquad A^n u_2 = \lambda_2^n u_2.$$

Therefore $\begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix} = A^n v = \alpha\lambda_1^n u_1 + \beta\lambda_2^n u_2$.

I claim that $\lambda_1 = \frac{1+\sqrt{5}}{2} \approx 1.618\cdots$ and $\lambda_2 = \frac{1-\sqrt{5}}{2} \approx -0.618\cdots$. Therefore:

- $\lambda_1 > 1$, so $\lambda_1^n \to \infty$,

- $\lambda_2^n \to 0$ as $n \to \infty$. Note that λ_2 is negative, so its powers keep switching sign, but its absolute value is less than 1, so the absolute value of its powers get smaller and smaller as $n \to \infty$.

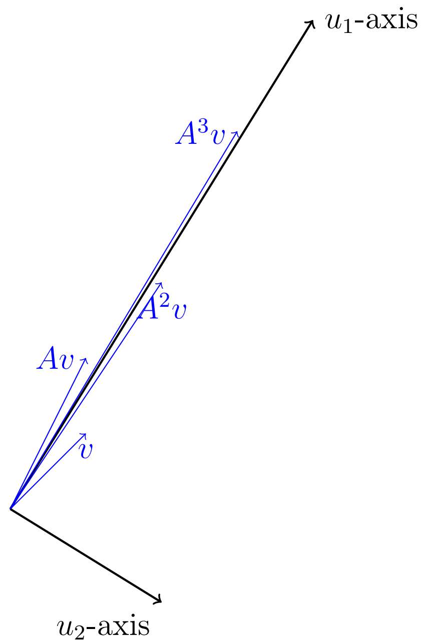
Therefore $\lim_{n\to\infty} \frac{F_{n+1}}{F_n}$ is the limit of the slopes of the vectors $\alpha\lambda_1^n u_1 + \beta\lambda_2^n u_2$, and the λ_2^n term is going to zero, so in the limit we just get the slope of the vector $\alpha\lambda_1^n u_1$, which is just a rescaling of u_1. Since rescaling doesn't change the slope, we get

$$\lim_{n\to\infty} \frac{F_{n+1}}{F_n} = \text{slope of } u_1.$$

We therefore need to figure out the slope of u_1 (and verify the claim about eigenvalues). The characteristic polynomial of A is $\det \begin{pmatrix} -t & 1 \\ 1 & 1-t \end{pmatrix} = t^2 - t + 1$, whose roots are $\frac{1\pm\sqrt{5}}{2}$ as required. The eigenvectors are $\begin{pmatrix} 1 \\ \frac{1\pm\sqrt{5}}{2} \end{pmatrix}$, so u_1 (corresponding to the plus sign) has slope $\frac{1+\sqrt{5}}{2}$, as required.

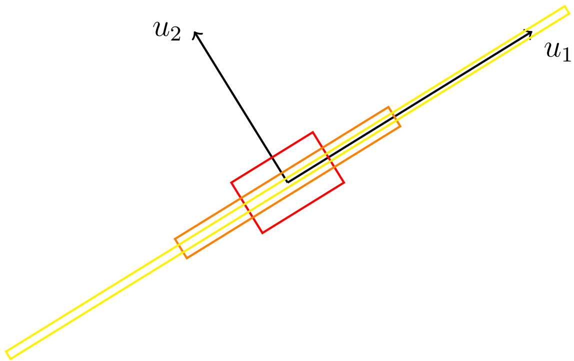
□

Here's a picture of the eigenlines (orthogonal to one another because the matrix A is symmetric) and the positions of $v, Av, A^2v, \ldots$. You can see that these vectors get closer and closer to the u_1-eigenline (and stretched out in the u_1-direction). They move from side to side of this axis because the sign of λ_2 is negative. So $A^n v$ gets more and more parallel to u_1 as $n \to \infty$.



9.3.2 Arnold cat map

Here's another nice example, due to Vladimir Arnold. Consider $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. This has eigenvalues $\frac{3\pm\sqrt{5}}{2}$: one of these is bigger than 1, the other is positive but less than 1. Here are the eigenlines, with a square S drawn on (whose sides are parallel to the eigenlines). We also draw $A(S)$ and $A^2(S)$. We can see that it gets stretched in the u_1-direction and squashed in the u_2-direction (because $\lambda_1 > 1$ and $\lambda_2 < 1$). In the limit, $A^n(S)$ gets thinner and thinner and closer to the u_1-eigenline.



This is called the Arnold cat map because of the following strange phenomenon. Take an infinite grid of squares in $\mathbf{R}^2$, take a picture of a cat, and put it into every square. Apply A to this grid of cats. The cats will get stretched and squashed in the eigendirections. Pick one of our original squares and look at what's there. We see a bunch of cats all chopped up and stretched and squished back into that square in some way. Now repeat, and repeat. What we see in our square is absolute carnage for a long time. But, amazingly, at some point, we our cat reappears almost identically to how it looked to begin with. This is not because of any periodicity: A^n is not the identity for any $n > 0$. This is instead an instance of "Poincaré recurrence": a phenomenon in dynamical systems which goes way beyond anything we're discussing in this course.

10 Week 5, Session 2: Linear maps and vector spaces

10.1 Linear maps

10.1.1 Two definitions of linearity

At the outset of this course, we talked about the geometric transformations coming from matrices (rotations, reflections, shears etc). These geometric transformations have a name: they are called *linear maps*. In this video we'll give two definitions of linear maps and show they're equivalent. The first definition encapsulates how we've been dealing with linear maps so far:

Definition 10.1. A map $f\colon \mathbf{R}^n \to \mathbf{R}^m$ is linear if there exists an m-by-n matrix A such that $f(v) = Av$ for all $v \in \mathbf{R}^n$.

However, some linear maps are more natural to describe in another way, without giving the matrix A.

Example 10.2. Fix n. Consider the space of polynomials $P(x)$ of degree at most n. Differentiation gives a map $\frac{d}{dx}\colon P \mapsto dP/dx$ from this space to itself. This map is linear. To understand why, we need to understand polynomials as vectors. We encode a polynomial

$$P(x) = a_n x^n + \cdots + a_1 x + a_0$$

as its vector of coefficients $\begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}$. Then

$$\frac{dP}{dx} = n a_{n-1} x^{n-1} + \cdots + 2a_2 x + a_1,$$

which corresponds to the vector of coefficients $\begin{pmatrix} a_1 \\ 2a_2 \\ \vdots \\ na_n \\ 0 \end{pmatrix}$. This is the same as

$$\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 3 & 0 & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & n \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix}$$

so differentiation is linear (given by a matrix). In other words, if v_P is the vector of coefficients of the polynomial P then $Dv_P = v_{dP/dx}$, where D is this matrix.

This way of encoding polynomials as vectors is a bit artificial. For example, I could have chosen to write the vector with a_n at the top and a_0 at the bottom, and the matrix D

would have ended up looking quite different. The fact that differentiation of polynomials is a linear map is an intrinsic fact about differentiation, and our proof above obscures that. So here's an equivalent definition of linearity which is more intrinsic.

Definition 10.3. A map f is linear if:

- $f(v+w) = f(v) + f(w)$ for all v, w
- $f(\lambda v) = \lambda f(v)$ for all v and for all $\lambda \in \mathbf{R}$.

I haven't specified the domain and target of f because I want to be intentionally vague: this definition makes sense whenever the domain and target of f admit operations of addition and rescaling (e.g. spaces of polynomials or functions as well as just $\mathbf{R}^n$). In the final video of the course, we'll see that the natural setting for this definition is the setting of *vector spaces*.

Example 10.4. Differentiation of polynomials is linear because

$$\frac{d}{dx}(P+Q) = \frac{dP}{dx} + \frac{dQ}{dx}$$

and

$$\frac{d(\lambda P)}{dx} = \lambda \frac{dP}{dx}$$

for any constant λ and polynomials $P(x), Q(x)$.

Example 10.5. The function $f\colon \mathbf{R} \to \mathbf{R}$ which converts metres to feet is linear. Since 1 metre is ≈ 3.281 feet, $f(x) = 3.281x$. If you double the number of metres, you double the number of feet. If you take two distances x metres and y metres you can add them and then convert to feet ($f(x+y)$) or you can convert and then add ($f(x)+f(y)$) and you get the same answer. So f is linear.

Example 10.6. The function $f\colon \mathbf{R} \to \mathbf{R}$ which converts Celsius to Kelvin is not linear. Recall that $f(0) \approx 273$. Any linear map satisfies $f(0) = 0$, because $f(0) = f(00) = 0f(0) = 0$ (some of those 0s are numbers, some are vectors!).

Example 10.7. I'm told the way they used to mark exams in Oxford was to take the marks from each question, square them and add them up. For example, if there were two questions and you got marks x and y then your final score would be x^2+y^2. This rewards those who do very well on a couple of questions (instead of scatter-gunning a little bit of knowledge over all questions). This function $f(x,y) = x^2+y^2$ is not a linear map! For example, if you score 1 and 0 then you get 1 in total, but if you double your score for x then you quadruple your total. Sadly for those taking the exam, $f(0,0) = 0$.

10.1.2 Equivalence

Lemma 10.8. *These two definitions of linearity are equivalent. In other words, the conditions*

- $f(v+w) = f(v) + f(w)$ *and* $f(\lambda v) = \lambda f(v)$

imply there exists a matrix A such that $f(v) = Av$, and any map of this form satisfies these conditions.

Proof. If $f(v) = Av$ for some matrix A then

$$f(v + w) = A(v + w) = Av + Aw = f(v) + f(w)$$

and

$$f(\lambda v) = A(\lambda v) = \lambda Av = \lambda f(v).$$

Conversely, consider the basis vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \qquad e_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \qquad \cdots, \qquad e_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}.$$

Let A be the matrix whose columns are $f(e_1), f(e_2), \ldots, f(e_n)$.

Then

$$f(v) = f\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = f(v_1 e_1 + \cdots + v_n e_n) = v_1 f(e_1) + \cdots + v_n f(e_n).$$

Since $f(e_k)$ is the kth column of A, it agrees with Ae_k (also the kth column of A). Therefore

$$f(v) = v_1 Ae_1 + \cdots + v_n Ae_n = A(v_1 e_1 + \cdots + v_n e_n) = Av,$$

which shows that f is linear in the sense that it has the form $f(v) = Av$ for some matrix A.

□

10.2 Kernels

10.2.1 Definition of the kernel

Recall the following definitions:

Definition 10.9. f is a linear map if

- $f(v + w) = f(v) + f(w)$ for all v, w
- $f(\lambda v) = \lambda f(v)$ for all v and all $\lambda \in \mathbf{R}$.

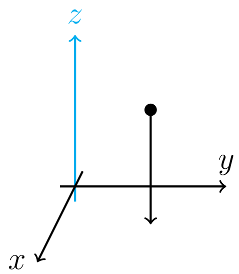
Definition 10.10. V is a linear subspace if

- $v, w \in V$ implies $v + w \in V$
- $v \in V$ implies $\lambda v \in V$ for all $\lambda \in \mathbf{R}$.

These two definitions are very similar. We will exploit this in the next two videos: given a linear map $f\colon \mathbf{R}^n \to \mathbf{R}^m$, we will associate to it two subspaces $\ker(f) \subset \mathbf{R}^n$ (the *kernel* of f) and $\mathrm{im}(f)$ (the *image of* f).

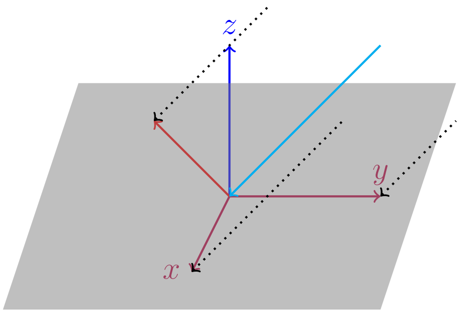
Definition 10.11. The kernel of f is the set of $v \in \mathbf{R}^n$ such that $f(v) = 0$. (If $m = n$ then this is just the 0-eigenspace of f).

Example 10.12. Let $f\colon \mathbf{R}^3 \to \mathbf{R}^3$ be the map $f(v) = Av$ for $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. Note that $f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$. This is the vertical projection to the xy-plane.



The kernel of f is the z-axis (blue in the figure; these are the points which project vertically down to the origin). That is $\ker(f) = \left\{ \begin{pmatrix} 0 \\ 0 \\ z \end{pmatrix} : z \in \mathbf{R} \right\}$.

Example 10.13. Recall the example $A = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$ (going from $\mathbf{R}^3$ to $\mathbf{R}^2$) from Section 1.3. This projects vectors into the plane; if we think of $\mathbf{R}^2$ as the xy-plane then we can visualise this map as the projection of vectors in the $\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$-direction until they live in the xy-plane.



We described this as projecting light rays in the $\begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}$ direction. In this case, the kernel of A is precisely the light ray which hits the origin, which is the line $\left\{ \begin{pmatrix} x \\ x \\ x \end{pmatrix} : x \in \mathbf{R} \right\}$ (light blue in the picture).

10.2.2 Kernel is a subspace

Lemma 10.14. *The kernel is a subspace.*

Proof. Given $v, w \in \ker(f)$, we need to show that $v + w \in \ker(f)$. Since $v, w \in \ker(f)$, we know that $f(v) = f(w) = 0$. Therefore $f(v + w) = f(v) + f(w)$ (since f is linear) $= 0 + 0 = 0$, so $v + w \in \ker(f)$. Similarly, $f(\lambda v) = \lambda f(v) = \lambda 0 = 0$.

□

10.2.3 Remarks

- $0 \in \ker(f)$ for any linear map f because $f(0) = 0$.
- If f is invertible then $\ker(f) = \{0\}$: if $v \in \ker(f)$ then $v = f^{-1}(0) = 0$.
- The "kernel" in a nut is the little bit in the middle that's left when you strip away the husk. If $f(v) = Av$ then we can think of $\ker(f)$ as the space of solutions to the simultaneous equations $Av = 0$, which is the intersection of the hyperplanes

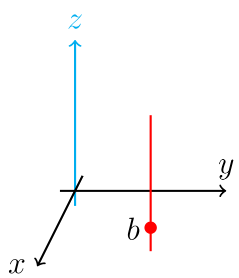
$$\begin{gathered} A_{11}v_1 + \cdots + A_{1n}v_n = 0 \\ \vdots \\ A_{m1}v_1 + \cdots + A_{mn}v_n = 0. \end{gathered}$$

  In other words, it's the little bit left over when you've intersected all these hyperplanes.

10.2.4 Simultaneous equations revisited

Lemma 10.15. *Consider the simultaneous equations $Av = b$ (A is an m-by-n matrix and $b \in \mathbf{R}^m$). Let $f(v) = Av$. The space of solutions to $Av = b$, if nonempty, is an affine translate of* $\ker(f)$.

Example 10.16. If $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ (so f is vertical projection to the xy-plane) then $Av = b$ has a solution only if b is in the xy-plane, and in that case it has a whole vertical line of solutions sitting above b.



This vertical line of solutions is parallel to the kernel of f (the z-axis), i.e. it is a translate of the kernel.

Proof. We saw this lemma earlier in a different guise in Section 5.3. Namely, we saw that if v_0 is a solution to $Av = b$ then the set of all solutions is the affine subspace $v_0 + U$ where U is the space of solutions to $Av = 0$. In other words, $U = \ker(f)$.

$\square$

In particular, we see that if $Av = b$ has a solution then it has a k-dimensional space of solutions, where k is the dimension of $\ker(f)$.

Remember that the space of solutions has dimension equal to the number of free variables when we put A into reduced echelon form. For example, $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is in reduced echelon form with two leading entries and one free variable, which is why we get 1-dimensional solution spaces.

Definition 10.17. The *nullity* of A (or of f) is the dimension of $\ker(f)$ (i.e. the number of free variables of A when put into reduced echelon form).

Our goal for the next video is to prove the *rank-nullity theorem* which gives us a nice formula relating the nullity to another important number called the rank.

10.3 Images

10.3.1 Definition

Definition 10.18. The *image* of a linear map $f\colon \mathbf{R}^n \to \mathbf{R}^m$ is the set of vectors $b \in \mathbf{R}^m$ such that $b = f(v)$ for some $v \in \mathbf{R}^n$. It is written as im.

If you think of applying a map as "following light rays" (like in some earlier examples), you can think of the image as the shadow your map casts.

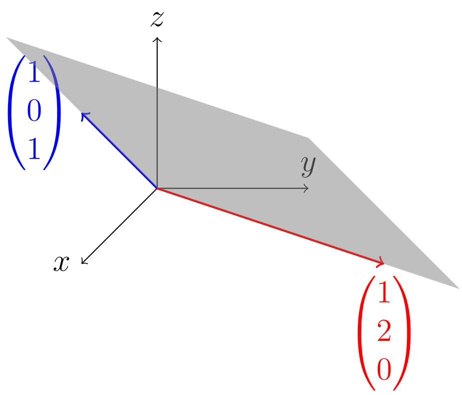
Example 10.19. If the map f is the vertical projection $f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x \\ y \\ 0 \end{pmatrix}$ then the image of f is the xy-plane. That is

$$\mathrm{im}(f) = \left\{ \begin{pmatrix} x \\ y \\ 0 \end{pmatrix} \ : \ x, y \in \mathbf{R} \right\}.$$

Example 10.20. Consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}$. The image of the corresponding linear map is the set of all vectors of the form

$$A\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+y \\ 2x \\ y \end{pmatrix}.$$

We studied this example earlier and even drew a picture of its image: it is the grey plane in the figure below. (There's a slight "videographic typo" (i.e. "mistake") in the video, see if you can spot it).



10.3.2 Remarks

- $0 \in \text{im}(f)$ because $0 = f(0)$.
- If f is invertible then $\text{im}(f) = \mathbf{R}^m$. This is because if $b \in \mathbf{R}^m$ then $b = f(f^{-1}(b))$, so $b \in \text{im}(f)$.

10.3.3 Image is a subspace

Lemma 10.21. *The image of f is a subspace.*

Proof. If $b_1, b_2 \in \text{im}(f)$ then so is $b_1 + b_2$. Since $b_1, b_2 \in \text{im}(f)$, $b_1 = f(v_1)$ and $b_2 = f(v_2)$ for some v_1, v_2. This means that $b_1 + b_2 = f(v_1) + f(v_2) = f(v_1 + v_2)$ (since f is linear), so $b_1 + b_2 \in \text{im}(f)$.

Similarly, $\lambda b_1 = \lambda f(v_1) = f(\lambda v_1)$ (since f is linear), so $\lambda b_1 \in \text{im}(f)$.

□

10.3.4 Relation with simultaneous equations

Lemma 10.22. *$Av = b$ has a solution if and only if $b \in \text{im}(f)$ where $f(v) = Av$.*

Proof. This is a tautology from the definition of image! $Av = b$ has a solution if and only if there is a v such that $f(v) = Av = b$.

□

So putting this together with the last lecture, we see that $Av = b$ has a solution if and only if $b \in \text{im}(f)$ and, if it has a solution, then the space of solutions is a translate of $\ker(f)$.

10.3.5 Rank

Definition 10.23. The *rank* of a linear map/matrix is the dimension of its image.

Theorem 10.24. (Rank-nullity theorem) *If A is an m-by-n matrix (or $f\colon \mathbf{R}^n \to \mathbf{R}^m$ is a linear map) then*

$$\mathrm{rank}(A) + \mathrm{nullity}(A) = n.$$

Here n is the number of columns or A (or the dimension of the target of f).

Example 10.25. The matrix $A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ sends everything to zero, so its image is a single point, which has dimension zero, so $\mathrm{rank}(A) = 0$. The kernel is the set of things which map to zero, and since everything maps to zero the kernel is $\mathbf{R}^3$. Therefore the nullity (dimension of the kernel) is three. Note that $0 + 3 = 3$ and $n = 3$ (A is a 3-by-3 matrix) so the rank-nullity theorem works.

The matrix $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ sends $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ to $\begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix}$, so its image is the x-axis. Therefore the rank (dimension of the image) is 1. The nullity is the number of free variables (B is in reduced echelon form already) which is 2 (there is one leading entry). Again, $1+2 = 3$, which is good. We can see that as the rank increases, the nullity goes down (as required by the rank-nullity theorem).

The matrix $C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ is the vertical projection to the xy-plane, so its rank is 2 (image is the xy-plane). Its nullity is 1 (one free variable). Again, $2 + 1 = 3$.

The identity matrix $I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ has rank 3 (for any v we have $v = Iv$ so every vector is in the image) and the nullity is 0 (only the origin maps to the origin). Again, $3+0 = 3$.

The rank-nullity theorem is basically saying that the map f is taking $\mathbf{R}^n$, crushing down some of the dimensions (those in the kernel), and mapping the rest faithfully onto the image (so the n dimensions of $\mathbf{R}^n$ either contribute to the kernel or to the image).

10.3.6 Proof of rank-nullity theorem

The nullity of A is the number of free variables of A when you put it into reduced echelon form. If we can show that the rank is the number of dependent variables then we're done: there are n variables which are either free (contributing to kernel) or dependent (contributing to rank). Recall that the dependent variables correspond to the columns with leading entries in reduced echelon form.

So we need to show that the rank is the number of leading entries of A in reduced echelon form.

First step: we prove that the rank doesn't change when we do a row operation. Suppose we start with a matrix A, do a row operation to get a matrix A'. We know there is an elementary matrix E such that $A' = EA$. This tells us immediately that $\mathrm{im}(A)$ and $\mathrm{im}(A')$ have the same dimension: $b \mapsto Eb$ gives us an "isomorphism" (invertible linear map) from the image of A to the image of A'.

As the rank doesn't change under row operations, we may assume without loss of generality that A is in reduced echelon form.

Second step: if A is in reduced echelon form then it has k nonzero rows (for some k) followed by $m - k$ zero rows. Now:

- The number k is the number of leading entries (because each nonzero row has a leading entry and each zero row doesn't).
- Recall that $Av = b$ has a solution if and only if $b_{k+1} = b_{k+2} = \cdots = b_m = 0$: these are the necessary and sufficient conditions for solving the simultaneous equations. If A has a zero row then b has to have a zero in that row, and if all these higher bs are zero then the other rows of A just give us equations which determine the dependent variables.

Since the image of A is the set of b for which $Av = b$ has a solution, this means that $\mathrm{im}(A)$ is the set of b for which $b_{k+1} = \cdots = b_m = 0$, i.e. those b of the form $\begin{pmatrix} b_1 \\ \vdots \\ b_k \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. This is a k-dimensional space, so we see that the rank equals k, the number of leading entries.

This completes the proof of the rank-nullity theorem.

10.4 Vector spaces

10.4.1 Definition

This lecture is intended as a foretaste of things to come. We will develop extra layers of abstraction, and this will allow us to apply the ideas of linear algebra in new contexts.

They key operations we've used on vectors are additon and rescaling:

$$v + w, \qquad \lambda v.$$

These operations are all we need to state the definitions of "linear map" and "linear subspace". But we can add and rescale things much more general than column vectors in $\mathbf{R}^n$.

Definition 10.26. A (real) vector space is a set V with operations

$$\begin{aligned} \text{add}\colon V \times V \to V & \qquad \text{written add}(v, w) = v + w \\ \text{rescale}\colon \mathbf{R} \times V \to V & \qquad \text{written rescale}(\lambda, v) = \lambda v \end{aligned}$$

The elements of V don't have to be elements of $\mathbf{R}^n$, and add and rescale don't have to be addition and rescaling of vectors in $\mathbf{R}^n$, they could be any other objects and operations which behave in a similar manner, by which I mean that the following conditions hold (for all $u, v, w \in V$ and $\lambda, \mu \in \mathbf{R}$):

- There is a distinguished element $0 \in V$.

- $u + (v + w) = (u + v) + w$
- $v + w = w + v$
- $v = 0 + v = v + 0$
- $v + (-v) = 0$
- $1v = v$
- $\lambda(\mu v) = (\lambda\mu)v$
- $\lambda + \mu)v = \lambda v + \mu v$
- $\lambda(v + w) = \lambda v + \lambda w$.

So a vector space is a set with operations that satisfy these conditions. In particular, $\mathbf{R}^n$ equipped with vector addition and rescaling is an example of a vector space. Indeed, it is a theorem that any finite-dimensional vector space is equivalent to $\mathbf{R}^n$ for some n. But if you allow yourself to consider infinite-dimensional vector spaces there are more interesting examples.

10.4.2 Function spaces

Example 10.27. Let V be the set of all functions from $\mathbf{R}$ to $\mathbf{R}$. Given functions f, g, define $f + g$ to be the function whose value at $x \in \mathbf{R}$ is $f(x) + g(x)$, and given a scalar $\lambda \in \mathbf{R}$ we define λf to be the function whose value at $x \in \mathbf{R}$ is $\lambda f(x)$. This gives V the structure of a vector space. It's not the same as $\mathbf{R}^n$ for any n: it is actually an infinite-dimensional vector space.

Example 10.28. Inside our previous example we can find some linear subspaces. For example, the set of continuous functions $\mathcal{C}^0(\mathbf{R})$ is a subspace. Inside that we have another subspace: the space of once continuously-differentiable functions $\mathcal{C}^1(\mathbf{R})$. Inside that we have the subspace of twice continuously-differentiable functions $\mathcal{C}^2(\mathbf{R})$. And so on. Inside all of these we have the infinitely-differentiable functions $\mathcal{C}^\infty(\mathbf{R})$, and, inside that, that space of analytic functions $\mathcal{C}^\omega(\mathbf{R})$ (infinitely-differentiable functions whose Taylor series converges on a neighbourhood of the origin). This gives us an infinite nested sequence of subspaces:

$$\mathcal{C}^\omega(\mathbf{R}) \subset \mathcal{C}^\infty(\mathbf{R}) \subset \cdots \subset \mathcal{C}^2(\mathbf{R}) \subset \mathcal{C}^1(\mathbf{R}) \subset \mathcal{C}^0(\mathbf{R}) \subset V$$

Inside the space of analytic functions, we have the space of polynomials (which we already met).

Example 10.29. Differentiation defines a linear map

$$\frac{d}{dx} \colon \mathcal{C}^1(\mathbf{R}) \to \mathcal{C}^0(\mathbf{R}), \qquad f \mapsto \frac{df}{dx}.$$

In other words, you start with a once continuously-differentiable function and differentiate it to get a continuous function). To show that it's linear, all we need to do is check that

$$\frac{d}{dx}(f + g) = \frac{df}{dx} + \frac{dg}{dx}, \qquad \frac{d}{dx}(\lambda f) = \lambda \frac{df}{dx}, \text{ for } \lambda \in \mathbf{R}.$$

What is the kernel of $\frac{d}{dx}$? It consists of functions whose derivative is zero, in other words constant functions. That is, $\ker(d/dx) \subset \mathcal{C}^1(\mathbf{R})$ is the 1-dimensional subspace of constant functions.

The eigenfunctions of d/dx (with eigenvalue λ) will be a function f such that $df/dx = \lambda f$. This is a differential equation for f; its solution is $f(x) = Ce^{\lambda x}$. This is why the exponential functions is so important: it's an eigenfunction of differentiation. Similarly the eigenfunctions of d^2/dx^2 are the solutions of $d^2f/dx^2 = \lambda f$, that is $A\cos(x\sqrt{-\lambda}) + B\sin(x\sqrt{-\lambda})$. This is a 2-dimensional eigenspace.

10.4.3 Algebraic numbers

Example 10.30. Here is an example from number theory. We have come across the idea that your vectors can have complex coefficients or real coefficients, but we can work much more generally by requiring our coefficients to live in some *coefficient field* K. In this example, we will take $K = \mathbf{Q}$ (the rational numbers), but you could imagine all sorts of things (the integers modulo 5, the 17-adic numbers, and goodness knows what else). The only difference this makes in the definition of a vector space is that rescaling can only be done by elements of K, that is the rescaling map is $\mathrm{rescale}\colon K \times V \to V$.

A number $x \in \mathbf{C}$ is *algebraic* if there exists a polynomial $P(z) = a_n z^n + \cdots + a_0$ with $a_n, \ldots, a_0 \in \mathbf{Q}$ for which $P(x) = 0$.

- For example, $\sqrt{2}$ is an algebraic number because it's a root of $z^2 = 2$.
- i is algebraic because it's a root of $z^2 = -1$.
- π and e are not algebraic (i.e. they're *transcendental*)

The set of all algebraic numbers is called $\overline{\mathbf{Q}}$.

Lemma 10.31. *$\overline{\mathbf{Q}}$ is a $\mathbf{Q}$-vector space.*

Proof. We need to show that if x, y are algebraic numbers and λ is a rational number then $x + y$ and λx are algebraic numbers.

To see that $\lambda x \in \overline{\mathbf{Q}}$, note that there is a polynomial $P(z) = a_n z^n + \cdots + a_0$ with $a_k \in \mathbf{Q}$ and $P(x) = 0$. Now $R(z) = \frac{a_n}{\lambda^n} z^n + \cdots + \frac{a_1}{\lambda} z + a_0$ satisfies $R(\lambda z) = P(z)$, so $R(\lambda x) = P(x) = 0$, showing that $\lambda x \in \overline{\mathbf{Q}}$. Note that we really need $\lambda \in \mathbf{Q}$ (or else the coefficients a_k/λ^k are not rational), so this is only a $\mathbf{Q}$-vector space (not a $\mathbf{R}$-vector space).

To show that $x + y$ is algebraic is much harder: we can't prove it here. In the words of Pierre Samuel (Algebraic Theory of Numbers), "The reader will have to exert himself to show that $\sqrt[3]{5} + \sqrt[3]{7}$ is an algebraic integer, and will be convinced that the steps which lead to a proof that this number is algebraic may not be easily generalised." The nicest proof uses the theory of modules over rings.

□

In fact, slightly more is true: the product of two algebraic numbers is also an algebraic number, so $\overline{\mathbf{Q}}$ is a *field*. One of the most complicated objects in mathematics is the group

$Gal(\overline{\mathbf{Q}} : \mathbf{Q})$, the *Galois group of* $\mathbf{Q}$. This is the set of invertible $\mathbf{Q}$-linear maps $f : \overline{\mathbf{Q}} \to \overline{\mathbf{Q}}$ which preserve the product ($f(ab) = f(a)f(b)$).

The elements of $Gal(\overline{\mathbf{Q}} : \mathbf{Q})$ are like infinitely large matrices with rational entries (because $\overline{\mathbf{Q}}$ is infinite-dimensional over $\mathbf{Q}$). One way people study this enormously complicated group is using *Galois representations*: associating to each $g \in Gal(\overline{\mathbf{Q}} : \mathbf{Q})$ a finite-dimensional matrix $M(g)$ such that $M(g_1 g_2) = M(g_1)M(g_2)$. Constructing Galois representations is a very difficult task, but the payoff can be enormous. Galois representations played an important role in Wiles's proof of Fermat's last theorem, and continue to play an important role in modern number theory.