# Linear Algebra

Jonny Evans

March 27, 2019

## Contents

The intention is that each section corresponds to a single lecture. The course comprises 19 lectures plus an in-class test in lecture 20. I have therefore left two hours leeway, which will get filled up with more examples if it becomes clear it is not needed.

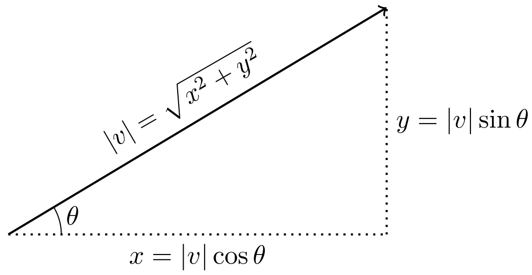
Sections marked with an asterisk I consider to be nonexaminable.

Workshops are scheduled to happen after lectures 2, 6, 10, 14, 18 and the first four will have assessed questions.

# 1 Matrices and transformations

## 1.1 Angles and rotations

A vector $v = \begin{pmatrix} x \\ y \end{pmatrix}$ in the plane is an arrow pointing $x$ units to the right and $y$ units up. By Pythagoras's theorem, the length of $v$ is $|v| = \sqrt{x^2 + y^2}$. If it makes an angle $\theta$ with the horizontal then $x = |v|\cos\theta$ and $y = |v|\sin\theta$.




**Theorem 1.1.** *Let $v = \begin{pmatrix} x \\ y \end{pmatrix}$ be a vector and let $w$ be the vector obtained by rotating $v$ an angle $\phi$ around its basepoint. Then*

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

*Proof.* We know that $v = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} |v|\cos\theta \\ |v|\sin\theta \end{pmatrix}$ where $\theta$ is the angle $v$ makes with the horizontal. After rotation, we know the following things about $w$:

- its length agrees with the length of $v$, i.e. $|w| = |v|$.
- the angle $w$ makes with the horizontal is $\theta + \phi$.

Therefore

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

## 1.2 Matrix notation

Inspired by Theorem 1.1, we introduce a new piece of notation which allows us to separate out the dependence of a rotated vector $w$ on the initial vector $v$ and on the rotation angle $\phi$.

**Definition 1.2.** A *2-by-2 matrix* is a 2-by-2 array of numbers, like $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Given a matrix and a vector $v = \begin{pmatrix} x \\ y \end{pmatrix}$, we define $Av$ to be the new vector

$$Av = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} := \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix}. \tag{1}$$

We say that $Av$ is obtained from $v$ by the action of $A$, in other words that matrices *act on* vectors.

In the context of Theorem 1.1, the matrix of the rotation by angle $\phi$ is

$$A = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix} \tag{2}$$

and the rotated vector is $w = Av$.

*Remark* 1.3. Vector notation lets us think of arrows as pairs of numbers. Matrix notation lets us think of transformations (rotations, reflections, etc) as grids of numbers.

*Remark* 1.4. How do you remember a formula like Eq. (1)? 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 $v$.

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

## 1.3 Linear maps

Now, given any 2-by-2 array of numbers, we get a geometric transformation of the plane $\mathbb{R}^2 \to \mathbb{R}^2$, $v \mapsto Av$. We call such a transformation arising from a matrix a *linear map*[1].

**Example 1.5.** The matrix $I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ represents the *identity transformation*, that is the map which sends the vector $\begin{pmatrix} x \\ y \end{pmatrix}$ to itself. For this reason, this matrix is usually called the identity matrix. It plays the same role in the theory of matrices that the number 1 plays in usual arithmetic, so sometimes I may end up writing 1 instead of $I$ (in my research I always write 1).

**Example 1.6.** The matrix $A = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ defines a reflection in the $y$-axis: the vector $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$, pointing along the $y$-axis, is fixed; the vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ pointing along the $x$-axis goes to $\begin{pmatrix} -1 \\ 0 \end{pmatrix}$.

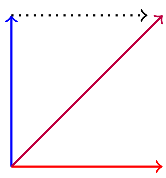
**Example 1.7.** Rotation by $\pi/2$ radians (90 degrees) is represented by the matrix in Eq. (2) with $\phi = \pi/2$, that is $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

**Example 1.8.** The matrix $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ represents a *shear* in the $x$-direction. For example, vectors along the $x$-axis are fixed:

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

vectors at height $y$ shear $y$ units to the right:

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

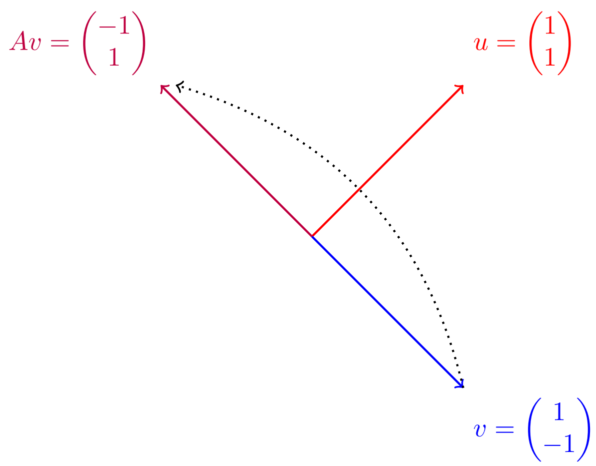


[1]We will see an equivalent definition of linear maps later, which makes no reference to matrices.

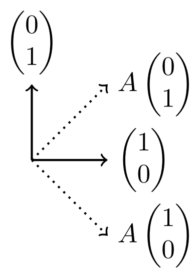
**Example 1.9.** Consider the matrix $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. This has a fixed vector $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ such that $Av = v$ (indeed, if $v = \begin{pmatrix} x \\ y \end{pmatrix}$ and $Av = v$ then

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

so any vector with $x = y$ is fixed. Moreover, $Aw = -w$ where $w = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$, which is orthogonal to $v$. Therefore $A$ represents a reflection in the line containing $v$.

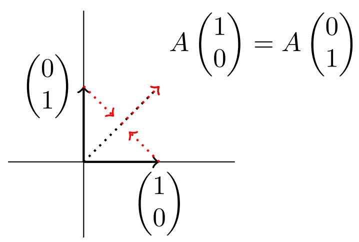



**Example 1.10.** If $A = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$ then we see $A\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $A\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Plotting these vectors, we can see that $A$ represents a rotation by $-\pi/4$ radians followed by a rescaling by $\sqrt{2}$.




**Example 1.11.** The matrix $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$. This gives the map $\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} x \\ 0 \end{pmatrix}$, which projects the plane vertically down to the $x$-axis.

**Example 1.12.** The matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ sends both $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to the vector $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. This means that the transformation defined by $A$ is an orthogonal projection to the $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$-line, followed by a rescaling by a factor of $2\sqrt{2}$.



## 1.4 Bigger matrices

Everything we've said so far generalises to higher dimensions.

**Definition 1.13.** An $n$-vector is a column of $n$ numbers. We write $\mathbb{R}^n$ for the set of all $n$-vectors[2]. An $m$-by-$n$ matrix is a rectangular array of numbers with $m$ rows and $n$ columns. Given an $n$-vector $v$ and an $m$-by-$n$ matrix $A$, we get an $m$-vector $Av$, whose $i$th entry is the result of multiplying the $i$th row of $A$ into the column vector $v$.

**Example 1.14.** A 3-by-3 matrix $\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ defines a linear map $\mathbb{R}^3 \to \mathbb{R}^3$, which takes the 3-vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ to

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

Again, the action of the matrix on the vector is defined by multiplying the rows of the matrix into the column vector. For example, the matrix $\begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$ defines a rotation by $\phi$ around the $z$-axis.

**Example 1.15.** A 2-by-3 matrix $\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}$ defines a linear map $\mathbb{R}^3 \to \mathbb{R}^2$:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} ax + by + cz \\ dx + ey + fz \end{pmatrix}.$$

For example, the matrix $\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ represents the map $\begin{pmatrix} x \\ y \\ z \end{pmatrix} \mapsto \begin{pmatrix} x \\ y \end{pmatrix}$, which is the projection from 3-dimensional space onto the $xy$-plane.

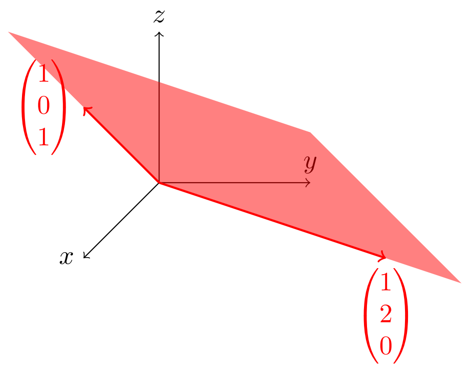
**Example 1.16.** In practice, matrices can be bigger than this. In special relativity, the maps which change from one spacetime reference frame to another are given by 4-by-4 matrices called *Lorentz matrices*; in statistics, in linear regression models, you work with matrices which have one row for each sample, so that could be very large.

**Example 1.17.** The 3-by-2 matrix $\begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}$ defines a linear map from $\mathbb{R}^2$ to $\mathbb{R}^3$ (an embedding from the plane into 3-dimensional space):

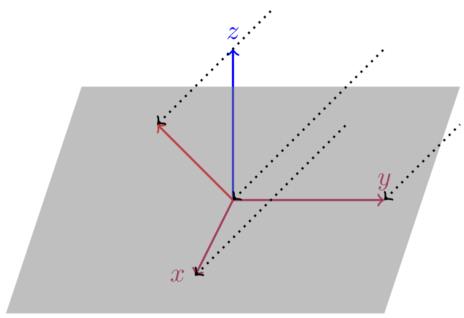
$$\begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 + v_2 \\ 2v_1 \\ v_2 \end{pmatrix}.$$

This sends the vector $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ to $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ to $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$. In the picture below we can see the image of the plane under this linear map.

[2] We could also work with vectors of complex numbers, in which case we'd write $\mathbb{C}^n$, or vectors of rational numbers, in which case we'd write $\mathbb{Q}^n$, or something else entirely.




**Example 1.18.** The matrix $\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix}$ defines a linear map from $\mathbb{R}^3$ to $\mathbb{R}^2$ (a projection from 3-dimensional space to the plane) which sends the basis vectors $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ to the vectors $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} -1 \\ -1 \end{pmatrix}$ respectively. Try to represent this projection in the picture below. The blue vectors point along the coordinate axes in 3-d. The red vectors are the images of the blue vectors under the projection (in two cases, the projection does nothing, so the blue and red vectors coincide; we draw them as purple). The dotted lines are the lines along which we're projecting. The grey shaded region is the plane onto which we're projecting.




**Example 1.19.** The matrix $A = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ defines a linear map $\mathbb{R}^3 \to \mathbb{R}^3$. We can see that

$$A \begin{pmatrix} 0 \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ y \\ z \end{pmatrix}, \qquad A \begin{pmatrix} x \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -x \\ 0 \\ 0 \end{pmatrix}.$$

This means that $A$ can be interpreted as a *reflection* in the $yz$-plane.

**Example 1.20.** The matrix $A = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ defines a linear map $\mathbb{R}^3 \to \mathbb{R}^3$ which fixes the vector $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ and effects a 90 degree rotation in the $xy$-plane. Similarly, the matrix $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}$ fixes the vector $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and effects a 90 degree rotation in the $yz$-plane, and the matrix $C = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$ fixes the vector $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and effects a 90 degree rotation in the $xz$-plane.

It is much harder (though still possible) to write down a general rotation matrix in three dimensions. We will revisit some three-dimensonal rotation matrices later in the week.

# 2 Matrix algebra

## 2.1 Matrix multiplication

Suppose we are given two matrices $A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$, and $B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$. 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 $AB$.

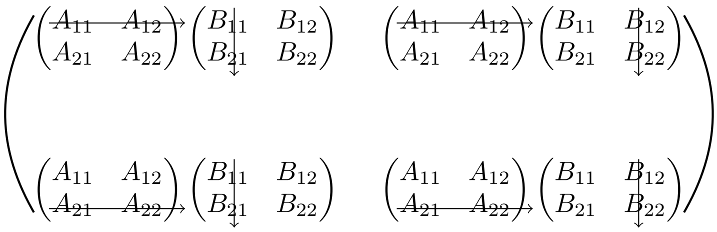
$$\begin{aligned} A(B(v)) &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \\ &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11}x + B_{12}y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

**Definition 2.1** (Matrix multiplication)**.** 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

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

How on earth can we remember this formula? Here are two mnemonics.

- 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 $ij$th entry of $AB$ (i.e. $i$th row and $j$th column) we multiply the $i$th row of $A$ into the $j$th column of $B$:



- We can also write a formula for the $ij$th entry:

$$(AB)_{ij} = \sum_{k=1}^{2} A_{ik}B_{kj}.$$

  For example, when $i = 1$, $j = 2$, this equation gives the entry of the product $AB$ in the first row and second column as

$$(AB)_{12} = A_{11}B_{12} + A_{12}B_{22}.$$

**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

$$R_{\theta_1} = \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix} \qquad R_{\theta_2} = \begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix}$$

are two rotations then the composite is

$$\begin{aligned} R_{\theta_1}R_{\theta_2} &= \begin{pmatrix} \cos\theta_1 & -\sin\theta_1 \\ \sin\theta_1 & \cos\theta_1 \end{pmatrix}\begin{pmatrix} \cos\theta_2 & -\sin\theta_2 \\ \sin\theta_2 & \cos\theta_2 \end{pmatrix} \\ &= \begin{pmatrix} \cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 & -\cos\theta_1\sin\theta_2 - \sin\theta_1\cos\theta_2 \\ \sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2 & -\sin\theta_1\sin\theta_2 + \cos\theta_1\cos\theta_2 \end{pmatrix} \\ &= \begin{pmatrix} \cos(\theta_1+\theta_2) & -\sin(\theta_1+\theta_2) \\ \sin(\theta_1+\theta_2) & \cos(\theta_1+\theta_2) \end{pmatrix} \\ &= R_{\theta_1+\theta_2}. \end{aligned}$$

(using trigonometric addition formulas). This is what we expect, of course: rotating by $\theta_2$ and then $\theta_1$ amounts to rotating by $\theta_1+\theta_2$.

**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.

## 2.2 Noncommutativity

*Remark* 2.5. You might be confused about why we write $AB$ for the transformation which *first* applies $B$ and *then* applies $A$. This actually makes perfect sense if you think of $A$ and $B$ as functions acting on vectors: remember that $f(g(x))$ means "apply $f$ to the result of first applying $g$ to $x$".

*Remark* 2.6. Order matters: most of the time, $AB$ is *not equal to* $BA$. In other words, *matrix multiplication is not commutative.* This makes matrices significantly more interesting algebraic objects than numbers.

## 2.3 Bigger matrices

From these examples, and what we've seen for 2-by-2 matrices, hopefully you can guess the definition of matrix multiplication.

**Definition 2.7.** If $A$ is an $m$-by-$n$ matrix and $B$ is a $n$-by-$p$ matrix then $AB$ is the $m$-by-$p$ matrix whose $ij$th entry is

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj}.$$

In other words, the entry in the $i$th row and $j$th column is obtained by multiplying the $i$th row of $A$ into the $j$th column of $B$. Because $A$ has $n$ columns and $B$ has $n$ rows, this multiplication makes sense.

*Remark* 2.8. This kind of notation where you see entries of the matrix written out with subscripts and sums all over the place is called *index notation.* It is extremely useful for when you would otherwise run out of letters to write your matrices (for example, if your matrix was $n$-by-$n$ and you didn't know what $n$ was). If you want to see it being used to great effect, open any textbook on general relativity, and glory in the "débauches des indices".

**Example 2.9.** Here are some examples of matrix multiplications:

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

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

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

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

## 2.4 Other operations

We have now seen that you can define the product of two matrices. In fact, you can do lots of other operations.

**Definition 2.10.** Given two $m$-by-$n$ matrices $A$ and $B$, we define their *sum* $A + B$ to be the $m$-by-$n$ matrix $A + B$ whose $ij$th entry is

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

That is, the entry in the $i$th row and $j$th column of $A + B$ is the sum of the corresponding entries for $A$ and $B$.

*Remark* 2.11. Matrix addition is kind of boring in comparison to matrix multiplication. Nonetheless, it plays an important role.

**Definition 2.12** (Scaling)**.** Given a matrix $A$ and a number $\lambda$, we define $\lambda A$ to be the matrix

$$\lambda A$$

whose $ij$th entry is $(\lambda A)_{ij} = \lambda A_{ij}$. If $A$ corresponds to some geometric transformation then $\lambda A$ corresponds to the same geometric transformation followed by a rescaling by a factor of $\lambda$.

More interestingly, we can define exponentials of matrices.

**Definition 2.13.** Given an $n$-by-$n$ matrix $A$, we define its *exponential* $\exp(A)$ to be the infinite sum

$$\exp(A) = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \cdots = \sum_{n \geq 0} \frac{1}{n!} A^n,$$

where we define $A^0 = I$.

**Example 2.14.** Let $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. Then

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

so $0 = A^3 = A^4 = \cdots$ and the infinite sum reduces to:

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

**Example 2.15.** Let $A = \begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}$. Then

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

Therefore

$$\begin{aligned} A^3 &= -\theta^2 IA = -\theta^2 A \\ A^4 &= -\theta^2 A^2 = (-\theta^2)^2 I = \theta^4 I \\ A^5 &= \theta^4 IA = \theta^4 A \\ A^6 &= \theta^4 A^2 = -\theta^6 I. \end{aligned}$$

Following this pattern, we get $A^{2n} = (-1)^{2n}\theta^{2n}$ and $A^{2n} = (-1)^{2n}\theta^{2n}A$. This means

$$\begin{aligned} \exp(A) &= \left( I + \frac{1}{2}A^2 + \frac{1}{4!}A^4 + \cdots \right) + \left( A + \frac{1}{3!}A^3 + \cdots \right) \\ &= \left( 1 - \frac{\theta^2}{2} + \frac{\theta^4}{4!} + \cdots \right) I + \left( \theta - \frac{\theta^3}{3!} + \cdots \right) A \\ &= \cos\theta I + \sin\theta A \\ &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \end{aligned}$$

This is, remarkably, the formula for a rotation matrix by an angle $\theta$. Starting from a very simple matrix $\begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix}$ and using the exponential function, we have ended up with the general formula for a rotation matrix in the plane. You can probably imagine that this becomes even more useful as a way of encoding rotations in 3-dimensions.

# 3 Dot products and orthogonal matrices

## 3.1 Dot product

Given two vectors $v, w \in \mathbb{R}^n$, how do you figure out the angle between them?

**Definition 3.1** (Dot product)**.** Given two vectors $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ and $w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}$, we define the *dot product* of $v$ and $w$ to be the number

$$v \cdot w := v_1 w_1 + \cdots + v_n w_n.$$

**Theorem 3.2.** *If $v$ and $w$ are separated by an angle $\phi$ then $v \cdot w = |v||w| \cos \phi$.*

We will prove the theorem momentarily. Let us first explore it a little.

**Example 3.3.** The vectors $v = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $w = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ satisfy $v \cdot w = 1 \times 0 + 0 \times 1 = 0$. Indeed, they are *orthogonal* to one another (i.e. at right-angles), so are separated by an angle $\pi/2$ radians, and $\cos(\pi/2) = 0$.

**Example 3.4.** The vectors $v = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $w = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ satisfy $v \cdot w = 1$, $|v| = \sqrt{2}$, $|w| = 1$, so if $\phi$ is the angle separating them then

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

so $\cos \phi = \frac{1}{\sqrt{2}}$, and $\phi = \pi/4$ radians.

*Remark* 3.5. You may be worried that $\cos \phi$ doesn't determine $\phi$ completely, for example $\cos(\pi/2) = \cos(3\pi/2) = 0$. However, the ambiguity is precisely whether you are measuring the angle from $v$ to $w$ clockwise or anticlockwise, so don't worry unless that distinction is important to you.

We now move in the direction of proving Theorem 3.2. Notice that the definition of dot product looks a lot like matrix multiplication. In fact,

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

In other words, we have turned one of our column vectors on its side to make it into a row vector. This operation is called *transposition*.

**Definition 3.6.** Given an $m$-by-$n$ matrix $A$ with entries $A_{ij}$, its *transpose* $A^T$ is defined to be the $n$-by-$m$ matrix with entries $A_{ji}$. For example

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

In other words, we can write dot product as $v \cdot w = v^T w$.

**Lemma 3.7.** *We have $(AB)^T = B^T A^T$.*

*Proof.* Since $A^T_{kj} = A_{jk}$ and $B^T_{ik} = B_{ki}$, we have

$$\begin{aligned}
(AB)^T_{ij} &= (AB)_{ji} \\
&= \sum_k A_{jk} B_{ki} \\
&= \sum_k A^T_{kj} B^T_{ik} \\
&= \sum_k B^T_{ik} A^T_{kj} \\
&= (B^T A^T)_{ij}.
\end{aligned}$$

Therefore $(AB)^T = B^T A^T$, because all the entries agree. □

*Remark* 3.8. You may complain that matrix multiplication is not commutative, so the step where we switch $A^T_{kj} B^T_{ik} = B^T_{ik} A^T_{kj}$ is not valid. Fortunately your objection is invalid: $A^T_{kj}$ and $B^T_{ik}$ are matrix *entries* (i.e. numbers!) not matrices themselves.

## 3.2 Orthogonal matrices

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

**Example 3.10.** The rotation matrix $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ is orthogonal. To see this, note that $R^T_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = R_{-\theta}$, so $R^T_\theta R_\theta = R_{\theta-\theta} = I$. In general, you should think of an orthogonal matrix as giving a higher-dimensional version of a rotation or reflection.

**Lemma 3.11.** *If $A$ is an orthogonal matrix then $(Av)\cdot(Aw) = v \cdot w$. In particular, the action of an orthogonal matrix doesn't change the lengths of vectors.*

*Proof.* We have

$$\begin{aligned}
(Av)\cdot(Aw) &= (Av)^T Aw \\
&= v^T A^T Aw \\
&= v^T I w \\
&= v^T w \\
&= v \cdot w.
\end{aligned}$$

The length of a vector $v$ is $\sqrt{v\cdot v} = \sqrt{v_1^2 + \cdots + v_n^2}$ by Pythagoras's theorem, so $|Av| = \sqrt{(Av)\cdot(Av)} = \sqrt{v \cdot v} = |v|$. □

*Proof of Theorem 3.2.* Because we're only interested in the two vectors $v$ and $w$, we can look at the plane which contains them, and we reduce to the case where $v$ and $w$ are 2-dimensional. Moreover, we can rotate so that $v$ points in the positive $x$-direction. Rotation is given by the action of an orthogonal matrix, so $v \cdot w$ is unchanged by this. If $v$ points in the positive $x$-direction then $v = \begin{pmatrix} |v| & 0 \end{pmatrix}$ and $v \cdot w = |v| w_1$, where $w = \begin{pmatrix} w_1 \\ w_2 \end{pmatrix}$. Since $w$ makes an angle $\phi$ with $v$, $w = \begin{pmatrix} |w|\cos\phi \\ |w|\sin\phi \end{pmatrix}$, so the formula follows. □

# 4 3-dimensional rotations

## 4.1 3-dimensional rotations

Armed with our newfound understanding of angles, let's take a look at some 3-by-3 rotation matrices and figure out what rotation is being represented.

**Example 4.1.** The matrix $A = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is a rotation matrix; just by looking at it, we can see that the $z$-axis is fixed:

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

and the $xy$-plane gets rotated by $\phi$; for example, the unit vector $v = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ goes to the unit vector $w = \begin{pmatrix} \cos\phi \\ \sin\phi \\ 0 \end{pmatrix}$, and $v \cdot w = \cos\phi$, so $v$ gets rotated by an angle $\phi$.

**Example 4.2.** In Example 1.20, I claimed that $C = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$ is a rotation matrix for $\mathbb{R}^3$. That means there's a fixed vector (the axis) and the plane orthogonal to the axis is rotated by some angle. Let's figure out what the axis is and what the angle is.

If $u = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is a fixed vector then $u = Cu$, which in this case means

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

This implies $-x = z = x$, so $x = z = 0$, and we see that the $y$-axis is fixed.

The $xz$-plane is orthogonal to the $y$-axis, so the next task is to find by what angle it is rotated. Let us pick a vector (say $v = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$) in that plane and act using $C$ to get a new vector $Cv = \begin{pmatrix} 0 \\ 0 \\ -1 \end{pmatrix}$. We note that $v \cdot Cv = 0$, so in this case the rotation must be through 90 degrees.

**Example 4.3.** Here is a more involved example. The matrix $D = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$ defines a rotation in 3 dimensions. To find the axis $u$ we need to solve $u = Du$:

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

which means $x = y = z$, so the axis points in the direction of $u = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. Now pick $v = \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}$ orthogonal to $u$ ($u \cdot v = 1 - 1 = 0$). Compute $Av = \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}$, and

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

Now $|v| = |Av| = \sqrt{2}$, so $\cos\phi = \frac{v\cdot Av}{\sqrt{2}\sqrt{2}} = -\frac{1}{2}$, so $\cos(\phi) = -1/2$ and $\phi = 2\pi/3$.

*Remark* 4.4. How do I recognise when a matrix is a rotation matrix? It turns out that the rotations are precisely the orthogonal matrices with determinant one (we will define the determinant of a matrix later).

## 4.2 Logarithms of rotations*

We saw earlier that $\exp\begin{pmatrix} 0 & -\theta \\ \theta & 0 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. This is a special case of a beautiful general fact.

**Definition 4.5.** We say that a matrix is symmetric (respectively antisymmetric) if $A^T = A$ (respectively $A^T = -A$).

**Theorem 4.6.** *If $A$ is an antisymmetric matrix then $\exp(tA)$ is orthogonal for all $t$. Conversely, if $\exp(tA)$ is orthogonal for all $t$ then $A$ is antisymmetric.*

*Proof.* If $A$ is antisymmetric then $\exp(tA)^T = \exp(tA^T) = \exp(-tA)$. We will see below that

$$\exp(-B)\exp(B) = I$$

for any matrix $B$, so this shows that $\exp(tA)$ is orthogonal for all $t$.

For the converse, you can differentiate the expression $\exp(tA)$ with respect to $t$. This is nothing scary: $\exp(tA)$ is just a matrix whose coefficients are functions of $t$, and differentiation just means differentiating the entries. Here are some properties of the matrix exponential which we need:

- $\frac{d}{dt}\exp(tA) = A\exp(tA)$
- $\exp(tA)^T = \exp(tA^T)$
- $\frac{d}{dt}(M(t)N(t)) = \frac{dM(t)}{dt}N + M(t)\frac{dN(t)}{dt}$ (Leibniz rule).

Assuming these properties, we have

$$0 = \left.\frac{d}{dt}\right|_{t=0} I = \left.\frac{d}{dt}\right|_{t=0} (\exp(tA)\exp(tA)^T) = A + A^T,$$

so $A$ is antisymmetric. □

Let's prove all the properties we wanted. I'll just assume we don't have to worry about convergence issues for the power series defining exp (it's one of the nicest power series around and you can always rely on it behaving the way you want it to).

**Lemma 4.7.** *If $B$ is a matrix then $\exp(B)\exp(-B) = I$.*

*Proof.* We have

$$\begin{aligned}
\exp(B)\exp(-B) &= \sum_{m\geq 0}\sum_{n\geq 0}\frac{1}{m!n!}B^m(-B)^n \\
&= \sum_{p\geq 0}\sum_{m=0}^{p}\frac{1}{m!(p-m)!}B^m(-B)^n \\
&= \sum_{p\geq 0}\frac{1}{p!}\sum_{m=0}^{p}\frac{p!}{m!(p-m)!}B^m(-B)^n \\
&= \sum_{p\geq 0}\frac{1}{p!}(B-B)^p \\
&= \exp(0) = I
\end{aligned}$$

where we substituted $p = m + n$ and rearranged the infinite sum on line 2, multiplied by $p!/p!$ on line 3, and used the binomial theorem on line 4. □

**Lemma 4.8.**

$$\frac{d}{dt}\exp(tA) = A\exp(tA).$$

*Proof.*

$$\begin{aligned}\frac{d}{dt}\exp(tA) &= \frac{d}{dt}\sum_{n\geq 0}\frac{t^n}{n!}A^n \\ &= \sum_{n\geq 1}\frac{t^{n-1}}{(n-1)!}A^n \\ &= A\sum_{m\geq 0}\frac{t^m}{m!}A^m \text{ relabelling } m = n-1.\end{aligned}$$ □

**Lemma 4.9.**

$$\exp(B)^T = \exp(B^T).$$

*Proof.* Clearly we have $(B_1+B_2)^T = B_1^T + B_2^T$, and we also have $(B^n)^T = (B^T)^n$ (using $(AB)^T = B^T A^T$ and induction). Therefore

$$\exp(B)^T = \left(\sum_{n\geq 0}\frac{1}{n!}B^n\right)^T = \sum_{n\geq 0}\frac{1}{n!}(B^n)^T = \sum_{n\geq 0}\frac{1}{n!}(B^T)^n = \exp(B^T).$$ □

**Lemma 4.10.**

$$\frac{d}{dt}(M(t)N(t)) = \frac{dM(t)}{dt}N + M(t)\frac{dN(t)}{dt}.$$

*Proof.* Let's use index notation. The $ij$ entry of $M(t)N(t)$ is $\sum_k M_{ik}(t)N_{kj}(t)$, so

$$\begin{aligned}\frac{d}{dt}(M(t)N(t))_{ij} &= \frac{d}{dt}\left(\sum_k M_{ik}(t)N_{kj}(t)\right) \\ &= \sum_k \frac{dM_{ik}(t)}{dt}N_{kj}(t) + \sum_k M_{ik}(t)\frac{dN_{kj}(t)}{dt} \\ &\qquad \text{using the usual Leibniz rule} \\ &= \left(\frac{dM(t)}{dt}N + M(t)\frac{dN(t)}{dt}\right)_{ij}.\end{aligned}$$ □

**Example 4.11.** The general 3-d rotation matrix is therefore $\exp\begin{pmatrix} 0 & \alpha & \gamma \\ -\alpha & 0 & \beta \\ -\gamma & -\beta & 0\end{pmatrix}$.

# 5 Simultaneous equations

## 5.1 Simultaneous equations

A system of simultaneous linear equations, like

$$\begin{aligned} x - y &= -1 \\ x + y &= 3 \end{aligned}$$

can be written as a single matrix equation $Av = b$, like

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

In fact, we often omit the $x$s and $y$s completely, and write instead the *augmented matrix*

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

## 5.2 Row operations

When we try to solve a system of equations like this, there are a bunch of operations we perform, like "add the second equation to the first" or "multiply the first equation by 5", and we can interpret these in terms of matrices. We illustrate this using the above example.

Start with:

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

Write the *augmented matrix*

$$\left(\begin{array}{cc|c} 1 & -1 & -1 \\ 1 & 1 & 3 \end{array}\right).$$

Subtract eq. 1 from eq. 2:

$$\begin{aligned} x - y &= -1 \\ 2y &= 4. \end{aligned}$$

Subtract row 1 from row 2:

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

Halve eq. 2

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

Halve row 2:

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

Add eq. 2 to eq. 1:

$$\begin{aligned} x &= 1 \\ y &= 2, \end{aligned}$$

Add row 2 to row 1:

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

and we're done.

i.e.

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

or $x = 1$, $y = 2$.

**Definition 5.1** (Row operations)**.** Given a matrix (possibly augmented with a vertical bar somewhere), we define the *row operations*:

- (Type I) $R_i \mapsto R_i + \lambda R_j$: "add $\lambda$ times the $j$th row to the $i$th row".

- (Type II) $R_i \mapsto \lambda R_i$: "multiply the $i$th row by $\lambda$".

So the sequence of row operations used in the above example was: $R_2 \mapsto R_2 - R_1$, $R_2 \mapsto \frac{1}{2}R_2$, $R_1 \mapsto R_1 + R_2$.

## 5.3 Echelon forms

The dream goal of solving simultaneous equations is to reduce to a system of the form

$$x = \text{something}, \quad y = \text{something else}, \ldots$$

If you can achieve this (which is not always possible, for example if your system has no solutions, or has many solutions) then, in terms of matrices, you have reduced the left-block of your augmented matrix to the identity matrix. In general, the best we can hope for is to reduce our matrix to so-called *reduced echelon form*.

**Definition 5.2** (Echelon forms)**.** Given a nonzero row $R$ of a matrix $M$, we define its *leading entry* to be the leftmost nonzero entry.

- We say that $M$ is in *echelon form* if, for every nonzero row $R_i$, the row $R_{i-1}$ immediately above it is nonzero and the leading entry of $R_{i-1}$ sits to the left of the leading entry of $R_i$. In other words, the bottom-left chunk of $M$ consists of zeros sitting in a configuration like a set of steps[3].
- We say that $M$ is in *reduced echelon form* if it is in echelon form, every leading entry is a 1 and every other entry in a column containing a leading entry vanishes. If the $i$th row of $M$ has leading entry $M_{ij} = 1$ then we will call $j$ the $i$th *leading index*. We call the other indices *free* and write $F$ for the set of free indices.

**Example 5.3.** Consider the following matrices

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

$$D = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \qquad E = \begin{pmatrix} 2 & 0 & 1 \\ 0 & 3 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad F = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

$$G = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \qquad H = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \qquad J = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

$A, B, C, E, G, H$ are in echelon form. $B, C, G, H$ are in reduced echelon form. $D, F, J$ are in neither. For the matrices in reduced echelon form:

- $B$ has one leading index, 1, and one free index 2.
- the leading indices of $C$ are $1, 2$; the free indices are $3, 4$.
- the leading indices of $G$ are $1, 3$; the free indices are $2, 4$.
- the leading indices of $H$ are $2, 3, 5$; the free indices are $1, 4$.

## 5.4 Echelon form and simultaneous equations

If $M$ is in reduced echelon form then it is very easy to understand the corresponding system of simultaneous equations $Mv = b$. Here are some illustrative examples.

[3] The word "echelon" comes from the French word "échelle" meaning "ladder".

**Example 5.4.** Suppose that $M = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 2 & -1 \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$. The system of simultaneous equations $Mv = b$ we get is

$$\begin{aligned} x_1 + x_3 + x_4 &= b_1 \\ x_2 + 2x_3 - x_4 &= b_2. \end{aligned}$$

We can rearrange:

$$\begin{aligned} x_1 &= b_1 - x_3 - x_4 \\ x_2 &= b_2 - 2x_3 + x_4. \end{aligned}$$

In other words, for every value of the variables $x_3, x_4$, we get a solution $v = \begin{pmatrix} b_1 - x_3 - x_4 \\ b_2 - 2x_3 + x_4 \\ x_3 \\ x_4 \end{pmatrix}$. The "free variables" $x_3, x_4$ are associated with free indices $3, 4$ and the "dependent variables" $x_1, x_2$ are associated with leading indices. Here, *dependent* means that the values of $x_1, x_2$ are determined by $x_3, x_4$ via the equations.

**Example 5.5.** Suppose that $M = \begin{pmatrix} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & 8 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$ and $b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix}$. The system of simultaneous equations $Mv = b$ we get is

$$\begin{aligned} x_1 + 2x_2 + x_4 &= b_1 \\ x_3 + 8x_4 &= b_2 \\ 0 &= b_3 \\ 0 &= b_4 \\ 0 &= b_5. \end{aligned}$$

This has solutions if and only if $b_3 = b_4 = b_5 = 0$. In the case when this condition holds, there are free variables $x_2, x_4$ (for the free indices) and dependent variables $x_1 = b_1 - 2x_2 - x_4$, $x_3 = b_2 - 8x_4$. The general solution is then $v = \begin{pmatrix} b_1 - 2x_2 - x_4 \\ x_2 \\ b_2 - 8x_4 \\ x_4 \end{pmatrix}$ (provided $b_3 = b_4 = b_5 = 0$).

More generally, the same reasoning shows:

**Theorem 5.6.** *Suppose that:*

- *$M$ is an $m$-by-$n$ matrix in reduced echelon form,*
- *the first $k \leq m$ rows of $M$ are non-zero and the final $m - k$ rows are zero,*
- *the leading entry in row $i \leq k$ is in column $j_i$ (so the leading indices are $j_1, \ldots, j_k$).*

*Then the general solution $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ exists if and only if $b_{k+1} = \cdots = b_m = 0$ and has free variables $x_p$ (where $p$ runs over the set $F$ of free indices), dependent variables $x_{j_i} = b_i - \sum_{p \in F} M_{ip} x_p$.*

*Remark* 5.7. In particular, the space of solutions has dimension equal to the number of free indices.

**Example 5.8.** Consider the matrix $A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \end{pmatrix}$. This is in reduced echelon form. If it is used to form a system of equations $Av = b$ then these equations have the form

$$\begin{aligned} x_1 + 2x_3 &= b_1 \\ x_2 + x_3 &= b_2 \end{aligned}$$

which can be solved immediately:

$$x_1 = b_1 - 2x_3, \qquad x_2 = b_2 - x_3.$$

In other words, for each $x_3$, we get a solution $v = \begin{pmatrix} b_1 - 2x_3 \\ b_2 - x_3 \\ x_3 \end{pmatrix}$.

**Example 5.9.** Consider the matrix $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}$. This is in reduced echelon form. If it is used to form a system of equations $Av = b$ then these equations have the form

$$\begin{aligned} x_1 &= b_1 \\ x_2 &= b_2 \\ 0 &= b_3 \end{aligned}$$

This system can be solved if and only if $b_3 = 0$, in which case it has a solution $v = \begin{pmatrix} b_1 \\ b_2 \\ x_3 \end{pmatrix}$ for every possible value of $x_3$.

**Example 5.10.** Consider the $n$-by-$n$ identity matrix $I$. This is in reduced echelon form. If it is used to form a system of equations $Iv = b$ then these equations have the unique solution $v = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$. (Duh[4].)

**Example 5.11.** Consider the $n$-by-$n$ zero matrix. This is in reduced echelon form. If it is used to form a system of equations $0v = b$ then these equations have solutions if and only if $b = 0$; if $b = 0$ then any $v$ is a solution.

In other words, once a matrix is in reduced echelon form, it becomes very transparent how to solve the corresponding system of simultaneous equations.

[4]This is a colloquial form of the Latin QED.

# 6 Echelon form theorems

## 6.1 Putting a matrix into echelon form

We will soon see that any matrix can be put into echelon form by row operations of type I, and further into reduced echelon form by row operations of types I and II. Let's see some examples.

**Example 6.1.** Consider the matrix

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

Clear column 1, row 3 using $R_3 \mapsto R_3 - R_1$

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

Clear column 1, row 4 using $R_4 \mapsto R_4 - \frac{1}{2}R_1$

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

Clear column 2, row 4 using $R_4 \mapsto R_4 - R_2$

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

Clear column 3, row 4 using $R_4 \mapsto R_4 + \frac{1}{3}R_3$

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

This is now in echelon form. We can go further to reduced echelon form.

Make leading entries in rows 1 and 3 equal to 1 using $R_1 \mapsto \frac{1}{2}R_1$ and $R_3 \mapsto \frac{1}{3}R_3$.

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

Clear column 3 using $R_1 \mapsto R_1 - R_3$ and $R_2 \mapsto R_2 - R_3$

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

Clear column 4 using $R_2 \mapsto R_2 - R_4$

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

This means that for any $b$ there is a unique solution to $Mv = b$ (no free variables and no constraints on $b$).

**Example 6.2.** In this example, we'll keep track of the augmented column. We start with the matrix

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & b_1 \\ 1 & 1 & -1 & b_2 \\ 4 & 0 & -2 & b_3 \\ 0 & 2 & -1 & b_4 \end{array}\right)$$

Clear column 1, using $R_2 \mapsto R_2 - R_1$ and $R_3 \mapsto R_3 - 4R_1$

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & b_1 \\ 0 & 2 & -1 & b_2 - b_1 \\ 0 & 4 & -2 & b_3 - 4b_1 \\ 0 & 2 & -1 & b_4 \end{array}\right)$$

Clear column 2, using $R_3 \mapsto R_3 - 2R_2$ and $R_4 \mapsto R_4 - R_2$

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & b_1 \\ 0 & 2 & -1 & b_2 - b_1 \\ 0 & 0 & 0 & b_3 - 4b_1 - 2(b_2 - b_1) \\ 0 & 0 & 0 & b_4 - (b_2 - b_1) \end{array}\right)$$

Make leading entry in row 2 equal to 1 using $R_2 \mapsto \frac{1}{2}R_2$.

$$\left(\begin{array}{ccc|c} 1 & -1 & 0 & b_1 \\ 0 & 1 & -1/2 & (b_2 - b_1)/2 \\ 0 & 0 & 0 & b_3 - 2b_1 - 2b_2 \\ 0 & 0 & 0 & b_4 + b_1 - b_2 \end{array}\right)$$

Clear column 2 using $R_1 \mapsto R_1 + R_2$

$$\left(\begin{array}{ccc|c} 1 & 0 & -1/2 & (b_1 + b_2)/2 \\ 0 & 1 & -1/2 & (b_2 - b_1)/2 \\ 0 & 0 & 0 & b_3 - 2b_1 - 2b_2 \\ 0 & 0 & 0 & b_4 + b_1 - b_2 \end{array}\right)$$

We see that the general solution exists if $b_4 + b_1 - b_2 = 0$ and $b_3 - 2b_1 - 2b_2 = 0$, in which case there is one free variable $x_3$ and two dependent variables

$$x_1 = (b_1 + b_2 + x_3)/2, \qquad x_2 = (b_2 - b_1 + x_3)/2.$$

For example, if $b = \begin{pmatrix} -3 \\ 0 \\ -6 \\ 3 \end{pmatrix}$ then $b_4 + b_1 - b_2 = 3 - 3 - 0 = 0$ and $b_3 - 2b_1 - 2b_2 = -6 + 6 - 0 = 0$, and we get the general solution $\begin{pmatrix} (x_3 - 3)/2 \\ (x_3 + 3)/2 \\ x_3 \end{pmatrix}$.

**Example 6.3.** Again, we'll keep track of the augmented column. We start with the matrix

$$\left(\begin{array}{cccc|c} 3 & -2 & -1 & -5 & b_1 \\ -5 & 3 & 2 & -3 & b_2 \\ 0 & -2 & -1 & 1 & b_3 \end{array}\right)$$

Clear column 1, row 2 using $R_2 \mapsto R_2 + \frac{5}{3}R_1$

$$\left(\begin{array}{cccc|c} 3 & -2 & -1 & -5 & b_1 \\ 0 & -1/3 & 1/3 & -34/3 & \frac{5}{3}b_1 + b_2 \\ 0 & -2 & -1 & 1 & b_3 \end{array}\right)$$

Clear column 2, row 3 using $R_3 \mapsto R_3 - 6R_2$

$$\left(\begin{array}{cccc|c} 3 & -2 & -1 & -5 & b_1 \\ 0 & -1/3 & 1/3 & -34/3 & \frac{5}{3}b_1 + b_2 \\ 0 & 0 & -3 & 69 & -10b_1 - 6b_2 + b_3 \end{array}\right)$$

Make leading entries in rows 1 and 2 equal to 1 using $R_1 \mapsto \frac{1}{3}R_1$ and $R_2 \mapsto -3R_2$.

$$\left(\begin{array}{cccc|c} 1 & -2/3 & -1/3 & -5/3 & \frac{1}{3}b_1 \\ 0 & 1 & -1 & 34 & -5b_1 - 3b_2 \\ 0 & 0 & -3 & 69 & -10b_1 - 6b_2 + b_3 \end{array}\right)$$

Clear column 2 using $R_1 \mapsto R_1 + \frac{2}{3}R_2$

$$\left(\begin{array}{cccc|c} 1 & 0 & -1 & 21 & -3b_1 - 2b_2 \\ 0 & 1 & -1 & 34 & -5b_1 - 3b_2 \\ 0 & 0 & -3 & 69 & -10b_1 - 6b_2 + b_3 \end{array}\right)$$

Make leading entry in row 3 equal to 1 using $R_3 \mapsto -\frac{1}{3}R_3$.

$$\left(\begin{array}{cccc|c} 1 & 0 & -1 & 21 & -3b_1 - 2b_2 \\ 0 & 1 & -1 & 34 & -5b_1 - 3b_2 \\ 0 & 0 & 1 & -23 & \frac{10}{3}b_1 + 2b_2 - \frac{1}{3}b_3 \end{array}\right)$$

Clear column 3 using $R_1 \mapsto R_1 + R_3$ and $R_2 \mapsto R_2 + R_3$

$$\left(\begin{array}{cccc|c} 1 & 0 & 0 & -2 & \frac{1}{3}(b_1 - b_3) \\ 0 & 1 & 0 & 11 & -\frac{1}{3}(5b_1 + 3b_2 + b_3) \\ 0 & 0 & 1 & -23 & \frac{10}{3}b_1 + 2b_2 - \frac{1}{3}b_3 \end{array}\right)$$

We see that this always has a solution, and the general solution is

$$\begin{pmatrix} \frac{1}{3}(b_1 - b_3) + 2x_4 \\ -\frac{1}{3}(5b_1 + 3b_2 + b_3) - 11x_4 \\ \frac{10}{3}b_1 + 2b_2 - \frac{1}{3}b_3 + 23x_4 \\ x_4 \end{pmatrix}$$

with one free variable $x_4$.

## 6.2 Echelon form theorems

**Theorem 6.4** (Echelon form)**.** *Every $m$-by-$n$ matrix $A$ can be put into echelon form using only the row operations $R_i \mapsto R_i + \lambda R_j$.*

*Proof.* We will prove the theorem by induction on the size of the matrix. Suppose we have proved the theorem for all $m'$-by-$n$ matrices with $m' < m$. The base case for induction is then $m = 1$ but if there is only one row then the matrix is automatically in echelon form, which proves the base case. Now for the induction step.

If your matrix is zero then it's already in echelon form, so without loss of generality, assume that there is a nonzero row.

- Of all the nonzero rows, pick the row $R_i$ whose leading entry $A_{ij}$ is furthest to the left (i.e. $j$ is minimal); if there are several such rows, pick the topmost (i.e. with $i$ minimal).
- If $i \neq 1$ (i.e. if $R_i$ is not the top row) then apply the row operation $R_1 \mapsto R_1 + R_i$ so that the top row also has leading entry $A_{ij}$.
- For $k = 2, \ldots, m$, apply the row operation $R_k \mapsto R_k - \frac{A_{kj}}{A_{ij}} R_1$. This ensures that the leading entries of all nonzero rows below the top are to the right of the leading entry of the top row.

Now consider the $(m-1)$-by-$n$ submatrix $A'$ you get by erasing the top row $R_1$. By induction, we can put this into echelon form using only row operations $R'_i \mapsto R'_i + \lambda R'_j$. Such operations don't introduce any leading entries in column $j$ or to the left of it because our submatrix $A'$ has zero entries in all these columns. Therefore, if we pop $R_1$ back on top of $A'$, the result in in echelon form. Since the row operations didn't affect $R_1$, we can think of them as row operations on $A$, so we have put $A$ into echelon form using only row operations of the specified type. □

**Theorem 6.5** (Reduced echelon form)**.** *Every $m$-by-$n$ matrix $A$ can be put into reduced echelon form by a sequence of row operations $R_i \mapsto R_i + \lambda R_j$ and $R_k \mapsto \lambda R_k$ ($\lambda \neq 0$).*

*Proof.* First, use Theorem 6.4 to put $A$ into echelon form. Now, for each nonzero row $R_i$, with leading entry $A_{ij}$, perform the row operation $R_i \mapsto \frac{1}{A_{ij}} R_i$ to make the leading entry equal to 1. Finally, for every nonzero row $R_i$ and every row $R_k$ with $k \neq i$, perform the row operation $R_k \mapsto R_k - A_{kj} R_i$. This clears out the nonzero entries in columns above and below the leading entry $A_{ij}$ of $R_i$. The result is in reduced echelon form. □

# 7 Inverses

## 7.1 Definition and basic properties

We've seen how to multiply and even exponentiate matrices. Can we "divide" by a matrix?

**Theorem 7.1.** *If* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *is a 2-by-2 matrix with* $ad - bc \neq 0$ *then the matrix*

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

*is an* inverse *for* $A$ *in the sense that* $AA^{-1} = A^{-1}A = I$.

*Proof.* We'll just check $A^{-1}A = I$.

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

□

*Remark* 7.2. This is great. However, you should never write $A^{-1}$ as $\frac{1}{A}$. The reason is that $\frac{B}{A}$ could mean $A^{-1}B$ or $BA^{-1}$ and these are in general different matrices (because matrix multiplication is not commutative).

We want to generalise this idea to $n$-by-$n$ matrices.

**Definition 7.3.** Let $A$ be an $n$-by-$n$ (square!) matrix. We say that $A$ is *invertible* if there exists a matrix $A^{-1}$ such that $A^{-1}A = AA^{-1} = I$.

*Remark* 7.4. Note that if an inverse exists, it is unique because if $B, C$ are two inverses for $A$ then $AB = AC = I$ and so $B = BI = BAB = BAC = IC = C$.

**Lemma 7.5.** *If* $A$ *and* $B$ *are invertible with inverses* $A^{-1}$ *and* $B^{-1}$ *then* $AB$ *is invertible with inverse* $B^{-1}A^{-1}$ *(note the order is reversed!)*

*Proof.* We have

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

Similarly, one can show $(B^{-1}A^{-1})(AB) = I$. □

In this section, we will see an algorithm to test if a matrix is invertible and, if it is, compute its inverse. We will later introduce a quantity called the *determinant* of a square matrix which is the analogue of $ad - bc$ for 2-by-2 matrices in the sense that a matrix is invertible if and only if its determinant is nonzero (and there's a formula for the inverse in terms of determinants).

## 7.2 Inverse matrices and reduced echelon form

Observe that finding $A^{-1}$ is equivalent to solving the simultaneous equations associated to $Av = b$. Indeed, if $A$ is invertible then $v = A^{-1}b$ is a solution to $Av = b$. Since we know how to solve simultaneous equations, we also know how to find inverses! In fact, we were secretly doing this already in the chapter on simultaneous equations.

The following theorem makes this precise.

**Theorem 7.6.** *Given an n-by-n matrix A, form the* augmented matrix $(A|I_n)$ *(where $I_n$ is the n-by-n identity matrix). Use row operations on the augmented matrix to put A into reduced echelon form. Then A is invertible if and only if the reduced echelon form of A is $I_n$, and in this case, the result of putting $(A|I_n)$ into reduced echelon form is $(I_n|A^{-1})$.*

We will first use the theorem to compute some examples of inverse matrices, then we will develop a little more theory and prove the theorem.

## 7.3 Examples

**Example 7.7.** Let's invert the matrix $\begin{pmatrix} -3 & -2 & -4 \\ 2 & 3 & 3 \\ -1 & 4 & -4 \end{pmatrix}$. We start by writing the augmented matrix

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

Clear column 1, row 2 using $R_2 \mapsto R_2 + (2/3)R_1$

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

Clear column 1, row 3 using $R_3 \mapsto R_3 + (-1/3)R_1$

$$\left(\begin{array}{ccc|ccc} -3 & -2 & -4 & 1 & 0 & 0 \\ 0 & 5/3 & 1/3 & 2/3 & 1 & 0 \\ 0 & 14/3 & -8/3 & -1/3 & 0 & 1 \end{array}\right)$$

Clear column 2, row 3 using $R_3 \mapsto R_3 + (-14/5)R_2$

$$\left(\begin{array}{ccc|ccc} -3 & -2 & -4 & 1 & 0 & 0 \\ 0 & 5/3 & 1/3 & 2/3 & 1 & 0 \\ 0 & 0 & -18/5 & -11/5 & -14/5 & 1 \end{array}\right)$$

Make leading entry in row 1 equal to 1 using $R_1 \mapsto (-1/3)R_1$.

$$\left(\begin{array}{ccc|ccc} 1 & 2/3 & 4/3 & -1/3 & 0 & 0 \\ 0 & 5/3 & 1/3 & 2/3 & 1 & 0 \\ 0 & 0 & -18/5 & -11/5 & -14/5 & 1 \end{array}\right)$$

Make leading entry in row 2 equal to 1 using $R_2 \mapsto (3/5)R_2$.

$$\left(\begin{array}{ccc|ccc} 1 & 2/3 & 4/3 & -1/3 & 0 & 0 \\ 0 & 1 & 1/5 & 2/5 & 3/5 & 0 \\ 0 & 0 & -18/5 & -11/5 & -14/5 & 1 \end{array}\right)$$

Clear column 2 using $R_1 \mapsto R_1 + (-2/3)R_2$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 6/5 & -3/5 & -2/5 & 0 \\ 0 & 1 & 1/5 & 2/5 & 3/5 & 0 \\ 0 & 0 & -18/5 & -11/5 & -14/5 & 1 \end{array}\right)$$

Make leading entry in row 3 equal to 1 using $R_3 \mapsto (-5/18)R_3$.

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 6/5 & -3/5 & -2/5 & 0 \\ 0 & 1 & 1/5 & 2/5 & 3/5 & 0 \\ 0 & 0 & 1 & 11/18 & 7/9 & -5/18 \end{array}\right)$$

Clear column 3 using $R_1 \mapsto R_1 + (-6/5)R_3$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -4/3 & -4/3 & 1/3 \\ 0 & 1 & 1/5 & 2/5 & 3/5 & 0 \\ 0 & 0 & 1 & 11/18 & 7/9 & -5/18 \end{array}\right)$$

Clear column 3 using $R_2 \mapsto R_2 + (-1/5)R_3$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -4/3 & -4/3 & 1/3 \\ 0 & 1 & 0 & 5/18 & 4/9 & 1/18 \\ 0 & 0 & 1 & 11/18 & 7/9 & -5/18 \end{array}\right)$$

Now the right-hand block is the inverse of the matrix we started with.

**Example 7.8.** We start with the matrix

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

Clear column 1, row 2 using $R_2 \mapsto R_2 + (1)R_1$

$$\left(\begin{array}{cccc|cccc} 1 & -1 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 3 & 1 & 1 & 0 & 0 \\ -1 & 1 & 1 & -3 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 7 & 0 & 0 & 0 & 1 \end{array}\right)$$

Clear column 1, row 3 using $R_3 \mapsto R_3 + (1)R_1$

$$\left(\begin{array}{cccc|cccc} 1 & -1 & 0 & 3 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 3 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 7 & 0 & 0 & 0 & 1 \end{array}\right)$$

Clear column 1, row 4 using $R_4 \mapsto R_4 + (-1)R_1$

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

Clear column 2, row 4 using $R_4 \mapsto R_4 + (-1)R_2$

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

Clear column 2 using $R_1 \mapsto R_1 + (1)R_2$

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 1 & 6 & 2 & 1 & 0 & 0 \\ 0 & 1 & 1 & 3 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & -1 & 0 & 1 \end{array}\right)$$

Clear column 3 using $R_1 \mapsto R_1 + (-1)R_3$

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 6 & 1 & 1 & -1 & 0 \\ 0 & 1 & 1 & 3 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & -1 & 0 & 1 \end{array}\right)$$

Clear column 3 using $R_2 \mapsto R_2 + (-1)R_3$

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 6 & 1 & 1 & -1 & 0 \\ 0 & 1 & 0 & 3 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & -1 & 0 & 1 \end{array}\right)$$

Clear column 4 using $R_1 \mapsto R_1 + (-6)R_4$

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 13 & 7 & -1 & -6 \\ 0 & 1 & 0 & 3 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & -1 & 0 & 1 \end{array}\right)$$

Clear column 4 using $R_2 \mapsto R_2 + (-3)R_4$

$$\left(\begin{array}{cccc|cccc} 1 & 0 & 0 & 0 & 13 & 7 & -1 & -6 \\ 0 & 1 & 0 & 0 & 6 & 4 & -1 & -3 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & -2 & -1 & 0 & 1 \end{array}\right)$$

Now the right-hand block is the inverse of the matrix we started with.

# 8 Inverses from echelon form

## 8.1 Elementary matrices

**Definition 8.1** (Elementary matrices I)**.** If $i \neq j$, we write $E_{ij}(\lambda)$ for the matrix with ones on the diagonal and zeros elsewhere, except for a $\lambda$ in position $ij$ ($i$th row, $j$th column). For example, if we're working with 3-by-3 matrices then

$$E_{12}(2) = \begin{pmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad E_{32}(7) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 7 & 1 \end{pmatrix}, \quad E_{13}(t) = \begin{pmatrix} 1 & 0 & t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

**Lemma 8.2.** *If $A$ and $E_{ij}(\lambda)$ are $n$-by-$n$ matrices then $E_{ij}(\lambda)A$ is the matrix obtained from $A$ by the row operation $R_i \mapsto R_i + \lambda R_j$.*

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

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

The only difference the $\lambda$ makes is when we multiply the $i$th row into a column of $A$ (say the $k$th 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 $\lambda$ times row $j$ to row $i$. □

**Definition 8.3** (Elementary matrices II)**.** We define the elementary matrix $E_i(\lambda)$ to be the matrix with 1s on the diagonal and zeros elsewhere, except that the $ii$ entry is $\lambda$. For example, if we're working with 4-by-4 matrices then

$$E_1(5) = \begin{pmatrix} 5 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, E_3(F) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & F & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

**Lemma 8.4.** *If $A$ and $E_i(\lambda)$ are $n$-by-$n$ matrices then $E_i(\lambda)A$ is the matrix obtained from $A$ by the row operation $R_i \mapsto \lambda R_i$.*

*Proof.* The only difference between multiplying by the identity and multiplying by $E_i(\lambda)$ is that when you multiply the $i$th row of $E_i(\lambda)$ into the $j$th column of $A$, you pick up a factor of $\lambda$. Therefore $E_i(\lambda)A$ differs from $A$ only in that every element on the $i$th row is multiplied by $\lambda$. For example, if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then

$$E_1(\lambda)A = \begin{pmatrix} \lambda & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda a & \lambda b \\ c & d \end{pmatrix}.$$ □

**Lemma 8.5.** *An elementary matrix $E_{ij}(\lambda)$ is invertible with inverse $E_{ij}(-\lambda)$. An elementary matrix $E_i(\lambda)$ is invertible if $\lambda \neq 0$, in which case its inverse is $E_i(1/\lambda)$.*

*Proof.* Consider the product

$$\begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & \lambda & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & -\lambda & \\ & & & \ddots & & \\ & & & & 1 & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

The only difference between this and $II = I$ is when you multiply row $i$ into column $j$, when you get $1 \times (-\lambda) + \lambda \times 1 = 0$. Therefore this product equals $I$.

Consider the product

$$\begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & \lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix} \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1/\lambda & & \\ & & & \ddots & \\ & & & & 1 \end{pmatrix}$$

The only difference between this product and $II = I$ is when you multiply row $i$ into column $i$, at which point you get $\lambda \times (1/\lambda) = 1$, so this product equals $I$. □

## 8.2 Proof of Theorem 7.6

*Proof of Theorem 7.6.* Suppose we have put $A$ into reduced echelon form using a sequence of row operations $r_1, \ldots, r_k$. Each row operation is equivalent to multiplying (on the left) by some elementary matrix $M_1, \ldots, M_k$. Therefore the reduced echelon form of $A$ is

$$C := M_k M_{k-1} \cdots M_1 A.$$

If $C$ is the identity then $M_k \cdots M_1 A = I$, so $M_k \cdots M_1 = A^{-1}$. If we perform the same row operations to the identity matrix (sitting on the right hand side of the augmented matrix $(A|I_n)$) then we get $M_k \cdots M_1 I = A^{-1}$.

If $C$ is not the identity matrix, then, since $C$ is a square matrix in reduced echelon form, there must be a row of $C$ which vanishes. Say this is the $i$th row. If $v$ is the vector with zeros everywhere except a 1 in the $i$th row then $Cv = 0$. Now $Cv = M_k \cdots M_1 Av = 0$, and $M_1, \ldots, M_k$ are invertible, so $Av = M_1^{-1} \cdots M_k^{-1} 0 = 0$. Therefore $A$ has nontrivial kernel. If $A$ were invertible then the only solution to $Av = 0$ is $v = A^{-1}0 = 0$, so the kernel would be trivial. Therefore $A$ is only invertible if its reduced echelon form is the identity matrix. □

**Corollary 8.6.** *A product of elementary matrices is invertible and, conversely, any invertible matrix is a product of elementary matrices.*

*Proof.* Each elementary matrix is invertible, so in for a product $M_k \cdots M_1$ of elementary matrices, the inverse is $M_1^{-1} \cdots M_k^{-1}$. Conversely, if $A$ is invertible then its reduced echelon form is the identity and its inverse is a product of elementary matrices $M_k \cdots M_1$ by Theorem 7.6. The inverse of an elementary matrix is again elementary, therefore $A = M_1^{-1} \cdots M_k^{-1}$ is a product of elementary matrices. □

# 9 Determinants

## 9.1 Definition and basic examples

We have seen that a 2-by-2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. We would like a similarly nice characterisation of invertibility for $n$-by-$n$ matrices. We will see that **a matrix is invertible if and only if its *determinant* is nonzero**.

**Definition 9.1** (Determinant)**.** If $A$ is an $n$-by-$n$ matrix with entries $A_{ij}$ then we define the *determinant* $\det(A)$ to be the number obtained as follows.

- Pick $n$ entries of $A$ with no two in the same row and no two in the same column. If we write the entry from the $i$th row as $A_{i\sigma(i)}$ (i.e. it's in the $\sigma(i)$th column) then this means that the map $i \mapsto \sigma(i)$ is a permutation of $\{1, \ldots, n\}$; there are $n!$ ways of making such a choice.
- Multiply these entries together to get the number $\pm A_{1\sigma(1)} \cdots A_{n\sigma(n)}$. The sign in this expression is taken to be $-1$ if your permutation is "odd" (i.e. if it involves an odd number of swaps) and $+1$ if your permutation is "even" (involves an even number of swaps). We will write $sgn(\sigma)$ for this sign.
- Repeat this for every possible choice $\sigma$ and sum the numbers that you get.

In brief:

$$\det(A) = \sum_{\sigma} sgn(\sigma) A_{1\sigma(1)} \cdots A_{n\sigma(n)},$$

where the sum is taken over all permutations $\sigma$.

**Example 9.2.** If $n = 2$ then there are $n! = 2$ choices:

- $\sigma$ could be the identity permutation $1 \mapsto 1$, $2 \mapsto 2$. This is an even permutation (it involves zero swaps and zero is even) so we get $A_{1\sigma(1)}A_{2\sigma(2)} = A_{11}A_{22}$.
- $\sigma$ could be the swap $1 \leftrightarrow 2$. This is an odd permutation (it involves one swap and one is odd) so we get $-A_{1\sigma(1)}A_{2\sigma(2)} = -A_{12}A_{21}$.

If $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then this translates into the two terms $ad$ and $-bc$, which we sum to get $\det(A) = ad - bc$.

**Example 9.3.** If $n = 3$ and $A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$ then we get $n! = 6$ choices:

| $\sigma$ | identity | $1 \leftrightarrow 2$ | $1 \leftrightarrow 3$ | $2 \leftrightarrow 3$ | cyclic (123) | cyclic (132) |
|---|---|---|---|---|---|---|
| contribution | $aef$ | $-bdi$ | $-ceg$ | $-afh$ | $bfg$ | $cdh$ |

so

$$\det(A) = aei + bfg + cdh - bdi - ceg - afh.$$

**Example 9.4** (Diagonal matrices)**.** If $D$ is a diagonal matrix with entries $\lambda_1, \ldots, \lambda_n$:

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

then there is only one way to pick a nonzero entry from each row, which gives $\det(D) = \lambda_1 \cdots \lambda_n$.

**Example 9.5** (Upper triangular matrices)**.** Suppose that $T$ is an *upper triangular matrix*, in other words all the entries below the diagonal are zero:

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

Then we need to pick something from the first column, which has to be $A_{11}$, then something from the second column but this may not be on the first row as we already picked something from the first row, so this must be $A_{22}$, then something from the third column, but this cannot be on the first or second rows, so it must be $A_{33}$, and so on, so we see that $\det(A) = A_{11} \cdots A_{nn}$ (i.e. $\det(A)$ is the product of the diagonal entries). Similarly if $A$ is lower-triangular.

**Example 9.6.** If $E_{ij}(\lambda)$ is an elementary matrix with ones on the diagonal and zeros elsewhere except for a $\lambda$ in position $ij$ then $\det(E_{ij}(\lambda)) = 1$. This is because $E_{ij}(\lambda)$ is upper (respectively lower) triangular (when $i < j$ or $i > j$ respectively).

**Example 9.7.** If $E_i(\lambda)$ is the elementary matrix with ones on the diagonal and zeros elsewhere except for a $\lambda$ in position $ii$, then $E_i(\lambda)$ is diagonal and its determinant is $\lambda$.

## 9.2 Some properties of the determinant

**Lemma 9.8.** *If two rows of $A$ coincide (that is, for some $i \neq j$, we have $A_{ik} = A_{jk}$ for all $k$) then* $\det(A) = 0$.

*Proof.* If two rows coincide then each term

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

cancels with the term

$$sgn(\sigma')(\cdots)A_{i\sigma'(i)}(\cdots)A_{j\sigma'(j)}(\cdots)$$

where $\sigma'$ is the permutation obtained by performing $\sigma$ and then switching $i \leftrightarrow j$. The point is that this doesn't change the value of the product (because $A_{i\sigma(i)} = A_{j\sigma(i)}$ and $A_{i\sigma(j)} = A_{j\sigma(j)}$) but it does change the sign of the permutation (it introduces an extra swap). □

**Lemma 9.9.** *If $A'$ is obtained from $A$ by a row operation $R_i \mapsto R_i + \lambda R_j$ then* $\det(A') = \det(A)$.

*Proof.* We have

$$\begin{aligned}\det(A') &= \sum_\sigma sgn(\sigma)(\cdots)(A_{i\sigma(i)} + \lambda A_{j\sigma(i)})(\cdots)\\ &= \sum_\sigma sgn(\sigma)(\cdots)A_{i\sigma(i)}(\cdots) + \lambda \sum_\sigma sgn(\sigma)(\cdots)A_{j\sigma(i)}(\cdots)\\ &= \det(A) + \lambda\det(B),\end{aligned}$$

where $B$ is the matrix obtained from $A$ by replacing the $i$th row with the $j$th row. Since $B$ has two rows equal, its determinant vanishes, so $\det(A') = \det(A)$. □

**Theorem 9.10.** *Suppose we put $A$ into echelon form using only row operations $R_i \mapsto R_i + \lambda R_j$. Then* $\det(A)$ *is the product of the diagonal entries in the echelon form.*

*Proof.* These row operations do not change the determinant, so if $C$ is the echelon form of $A$ thus obtained, we have $\det(A) = \det(C)$. By definition, matrices in echelon form are upper triangular, so $\det(C)$ is just the product of its diagonal entries, by Example 9.5. □

**Lemma 9.11.** *If $A'$ is obtained by swapping two of the rows of $A$ then* $\det(A') = -\det(A)$.

*Proof.* Each term (for a permutation $\sigma$) in $\det(A')$ also appears in $\det(A)$ for a permutation $\sigma$ followed by the swap, and hence with the opposite sign. □

*Remark* 9.12. This means you can also swap rows around to reach echelon form and compute the determinant, *provided you multiply by $-1$ each time you swap two rows.* This can be useful, for example:

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

immediately without the mess of adding row 4 to row 1 and subtracting row 1 from row 4.

# 10 Computing determinants

## 10.1 Examples of computing determinants

**Example 10.1.** Let $A = \begin{pmatrix} 1 & 4 & -4 \\ -2 & -2 & -4 \\ 3 & -3 & 3 \end{pmatrix}$.

Clear row 2 using $R_2 \mapsto R_2 + (2)R_1$

$$\begin{pmatrix} 1 & 4 & -4 \\ 0 & 6 & -12 \\ 3 & -3 & 3 \end{pmatrix}$$

Clear row 3 using $R_3 \mapsto R_3 + (-3)R_1$

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

Clear row 3 using $R_3 \mapsto R_3 + (5/2)R_2$

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

This is now in echelon form and has the same determinant as the matrix we began with, so the determinant is the product of the diagonal entries, which is $-90$.

**Example 10.2.** Let $B = \begin{pmatrix} 2 & -3 & -1 & 4 \\ 2 & -3 & 2 & 4 \\ 2 & -1 & -4 & -3 \\ 2 & -3 & 4 & 2 \end{pmatrix}$.

Clear row 2 using $R_2 \mapsto R_2 + (-1)R_1$

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

Clear row 3 using $R_3 \mapsto R_3 + (-1)R_1$

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

Clear row 4 using $R_4 \mapsto R_4 + (-1)R_1$

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

Add row 3 to row 2

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

Clear row 3 using $R_3 \mapsto R_3 + (-1)R_2$

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

Clear row 4 using $R_4 \mapsto R_4 + (5/3)R_3$

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

This is now in echelon form and has the same determinant as the matrix we began with, so the determinant is the product of the diagonal entries, which is 24.

**Example 10.3.** Let $C = \begin{pmatrix} 3 & -3 & -5 & -4 \\ 2 & -5 & 2 & 0 \\ 2 & 3 & -5 & -2 \\ 0 & 3 & -1 & 0 \end{pmatrix}$.

Clear row 2 using $R_2 \mapsto R_2 + (-2/3)R_1$

$$\begin{pmatrix} 3 & -3 & -5 & -4 \\ 0 & -3 & 16/3 & 8/3 \\ 2 & 3 & -5 & -2 \\ 0 & 3 & -1 & 0 \end{pmatrix}$$

Clear row 3 using $R_3 \mapsto R_3 + (-2/3)R_1$

$$\begin{pmatrix} 3 & -3 & -5 & -4 \\ 0 & -3 & 16/3 & 8/3 \\ 0 & 5 & -5/3 & 2/3 \\ 0 & 3 & -1 & 0 \end{pmatrix}$$

Clear row 3 using $R_3 \mapsto R_3 + (5/3)R_2$

$$\begin{pmatrix} 3 & -3 & -5 & -4 \\ 0 & -3 & 16/3 & 8/3 \\ 0 & 0 & 65/9 & 46/9 \\ 0 & 3 & -1 & 0 \end{pmatrix}$$

Clear row 4 using $R_4 \mapsto R_4 + (1)R_2$

$$\begin{pmatrix} 3 & -3 & -5 & -4 \\ 0 & -3 & 16/3 & 8/3 \\ 0 & 0 & 65/9 & 46/9 \\ 0 & 0 & 13/3 & 8/3 \end{pmatrix}$$

Clear row 4 using $R_4 \mapsto R_4 + (-3/5)R_3$

$$\begin{pmatrix} 3 & -3 & -5 & -4 \\ 0 & -3 & 16/3 & 8/3 \\ 0 & 0 & 65/9 & 46/9 \\ 0 & 0 & 0 & -2/5 \end{pmatrix}$$

This is now in echelon form and has the same determinant as the matrix we began with, so the determinant is the product of the diagonal entries, which is 26.

# 11 Formulas for determinants and for inverses

## 11.1 Inductive formula for determinants

We can expand the determinant as follows. First make your choice of entry from the first row, say in the $j$th column. Now remove the first row and the $j$th column. You're left with a smaller square matrix, which we'll call $C_{1j}$, from which you have to select the remaining entries. The picture below shows how to extract $C_{12}$ from a 4-by-4 matrix.

$$\begin{pmatrix} A_{11} & A_{12} & A_{13} & A_{14} \\ A_{21} & A_{22} & A_{23} & A_{24} \\ A_{31} & A_{32} & A_{33} & A_{34} \\ A_{41} & A_{42} & A_{43} & A_{44} \end{pmatrix} \rightsquigarrow C_{12} = \begin{pmatrix} A_{21} & A_{23} & A_{24} \\ A_{31} & A_{33} & A_{34} \\ A_{41} & A_{43} & A_{44} \end{pmatrix}$$

As you run over these choices, you obtain the determinant of this submatrix $C_{1j}$. Now allow the choice of $j$ to vary, and you obtain the following useful inductive formula for the determinant:

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

In fact, we could have started from any row (say the $i$th) and obtained a similar expression

$$\det(A) = (-1)^{i+1}\left(A_{i1}\det(C_{i1}) - A_{i2}\det(C_{i2}) + A_{i3}\det(C_{i3}) + \cdots + (-1)^n A_{in}\det(C_{in})\right),$$

where $C_{ij}$ is the submatrix obtained by deleting the $i$th row and the $j$th column.

In fact, we could have expanded by going down the $j$th column instead:

$$\det(A) = (-1)^{j+1}\left(A_{1j}\det(C_{1j}) - A_{2j}\det(C_{2j}) + \cdots + (-1)^n A_{nj}\det(C_{nj})\right)$$

The only non-obvious thing about these formulas is how to get the signs. The contribution to $A_{ij}\det(C_{ij})$ to one of these formulas is the sign $(-1)^{i+j}$ in the $ij$ position of the grid below:

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

You can prove this using index notation if you start from our formula for the determinant, but rather than go through this, we will simply use the formula to compute some determinants.

*Remark* 11.1. The determinants of submatrices are called *minors*. Historically, the mathematician Sylvester introduced the word "matrix" (the Latin word for *womb*) because...

> I have in previous papers defined a "Matrix" as a rectangular array of terms, out of which different systems of determinants may be engendered as from the womb of a common parent.

Let it never be said that mathematicians don't have vivid imaginations.

**Example 11.2.** Let's calculate the determinant of

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

using this inductive formula. We have

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

**Example 11.3.** Let's calculate the determinant of

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

using the inductive formula. Note that every entry on the first row is nonzero, so expanding along the first row would involve calculating four 3-by-3 minors. If, instead, we expand along the second *column* then we only have two nonzero entries, so only need to compute two 3-by-3 minors (first column, second row or fourth row would also have this advantage; I picked the second column because it makes the signs more interesting). This gives

$$\begin{aligned} \det(B) &= -\det\begin{pmatrix} 0 & 4 & 5 \\ -1 & 1 & 1 \\ 0 & 2 & 3 \end{pmatrix} - 2\det\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 2 & 3 \end{pmatrix} \\ &= -\left(-\left(-\det\begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix}\right)\right) - 2\det\begin{pmatrix} 4 & 5 \\ 2 & 3 \end{pmatrix} \\ &= -3(4 \times 3 - 5 \times 2) \\ &= -6. \end{aligned}$$

Let's check we did it right using row operations. Add row 1 to row 3:

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

Switch rows 2 and 3 (picking up a minus sign in the determinant)

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

Subtract twice row 4 from row 3, then switch them (another sign, which cancels the previous one).

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

The determinant is the product of the diagonal entries, which is indeed $-6$.

## 11.2 Inverses in terms of determinants

**Definition 11.4.** Define the *adjugate matrix* of $A$ to be the matrix

$$\operatorname{adj}(A) := \begin{pmatrix} +\det(C_{11}) & -\det(C_{12}) & +\det(C_{13}) & \cdots \\ \det(C_{21}) & +\det(C_{22}) & -\det(C_{23}) & \cdots \\ +\det(C_{31}) & -\det(C_{32}) & +\det(C_{33}) & \cdots \\ \vdots & \vdots & \vdots & \end{pmatrix}^T.$$

**Theorem 11.5.** *If* $\det(A) \neq 0$ *then* $A^{-1} = \frac{1}{\det(A)}\operatorname{adj}(A)$.

*Proof.* We can compute $A\operatorname{adj}(A)$. The $ij$th entry is precisely the expression

$$\pm(A_{i1}\det(C_{j1}) - A_{i2}\det(C_{j2}) + \cdots \pm A_{in}\det(C_{jn}))$$

which equals $\det(A)$ if $i = j$. If $i \neq j$ then this expression is the determinant of the matrix obtained from $A$ by replacing the $j$th row with the $i$th row, so two rows coincide and the determinant vanishes.

Therefore $A\,\mathrm{adj}(A) = \begin{pmatrix} \det(A) & 0 & \cdots & 0 \\ 0 & \det(A) & & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \det(A) \end{pmatrix}$, and so $\frac{1}{\det(A)}\,\mathrm{adj}\,A$ is an inverse for $A$. $\square$

# 12 More on determinants

## 12.1 Further properties of determinants

**Lemma 12.1.** *If $A'$ is obtained from $A$ by a row operation of the form $R_i \mapsto \lambda R_i$ then* $\det(A') = \lambda \det(A)$.

*Proof.*

$$\begin{aligned}\det(A') &= \sum_\sigma sgn(\sigma) A_{1\sigma(1)} \cdots (\lambda A_{i\sigma(i)}) \cdots A_{n\sigma(n)} \\ &= \lambda \sum_\sigma sgn(\sigma) A_{1\sigma(1)} \cdots A_{i\sigma(i)} \cdots A_{n\sigma(n)} \\ &= \lambda \det(A).\end{aligned}$$

□

**Theorem 12.2.** *An $n$-by-$n$ matrix $A$ is invertible if and only if its determinant is nonzero.*

*Proof.* Put $A$ into echelon form using only row operations of type $R_i \mapsto R_i + \lambda R_j$. You don't change the determinant. Now use row operations of type $R_i \mapsto \lambda R_i$ ($\lambda \neq 0$) to put $A$ into reduced echelon form. You change the determinant by a nonzero factor (the product of all the $\lambda$s that you used in the row operations). By Theorem 7.6, a matrix is invertible if and only if its reduced echelon form is the identity matrix, which has determinant 1, so:

- if $A$ is invertible then its determinant differs from 1 by a nonzero factor, and
- if $A$ is not invertible then its reduced echelon form has a zero row somewhere, so the reduced echelon form has determinant zero and $\det(A)$ is a multiple of zero, hence zero. □

**Theorem 12.3.** *If $A$ and $B$ are $n$-by-$n$ matrices then*

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

*Proof.* First, we show this under the assumption that $A$ is an elementary matrix.

- If $A = E_{ij}(\lambda)$ then $AB$ is the result of the row operation $R_i \mapsto R_i + \lambda R_j$ on $B$, so $\det(AB) = \det(B)$ by Lemma 9.9. Moreover, $\det(A) = 1$ by Example 9.6. Therefore $\det(A)\det(B) = \det(B)$ too, so the theorem is proved in this case.
- If $A = E_i(\lambda)$ then $AB$ is the result of the row operation $R_i \mapsto \lambda R_i$ on $B$, so $\det(AB) = \lambda \det(B)$ by Lemma 9.9. Moreover, $\det(A) = \lambda$ by Example 9.7. Therefore $\det(A)\det(B) = \lambda \det(B)$ too, so the theorem is proved in this case.

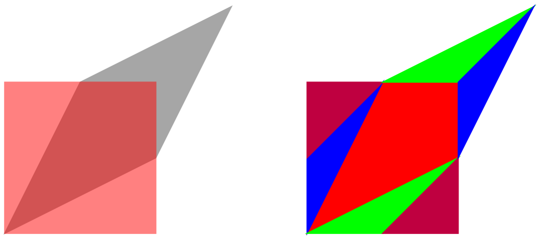
Now, if we assume that $A$ is a product of elementary matrices then the theorem follows from these two special cases by induction.

If $A$ is not a product of elementary matrices then $A$ is not invertible, so its determinant is zero by Theorem 12.2. Moreover, $AB$ is also noninvertible because $A$ is not invertible, so $\det(AB) = 0$ by Theorem 12.2, so $\det(AB) = 0 = \det(A)\det(B)$, and the theorem is proved in this case too. □

## 12.2 Geometric interpretation of determinants

**Theorem 12.4.** *Suppose that $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Let $S$ be the unit square sitting in the plane and let $A(S)$ denote the image of $S$ under the linear map defined by $A$. Then $|\det(A)|$ is the area of $A(S)$.*

*Proof.* The shape $A(S)$ is a parallelogram with sides parallel to the vectors $\begin{pmatrix} a \\ c \end{pmatrix}$ and $\begin{pmatrix} b \\ d \end{pmatrix}$. This parallelogram has area $ad - bc$ as we can see by dissection, using the following picture:



Take the grey parallelogram $A(S)$, draw the rectangle with sidelengths $a$ (along) and $d$ (up) over it. Move the green and blue pieces of the parallelogram inside the rectangle as shown. Now the red, green and blue areas inside the square have the same area as $A(S)$. The remaining (purple) part comprises two triangles which have height $c$ and base $b$, so the area of $A(S)$ is $ad - bc$. $\square$

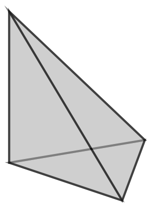
It is much harder to see the following theorems, but they are true:

**Theorem 12.5.** *If $A$ is an $n$-by-$n$ matrix then $|\det(A)|$ is the volume of $A(S)$, where $S$ is the unit cube in $n$-dimensions.*

*Remark* 12.6. The shape $A(S)$ is called a *parallelopiped*, the higher-dimensional analogue of a parallelogram.

**Theorem 12.7.** *Let $a_1, \ldots, a_n$ be $n$ vectors in $\mathbb{R}^n$. Consider the* simplex *with vertices at the origin and at $a_1, \ldots, a_n$. The volume of this simplex is $\frac{1}{n!}|\det(A)|$, where $A$ is the matrix with columns $a_1, \ldots, a_n$.*

**Example 12.8.** If $A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ then the simplex we get is the tetrahedron shown below. Its volume is $1/6$ (because of the above formula, but also because you can dissect and rearrange a cube into six such tetrahedra).



**Example 12.9.** The regular tetrahedron (or d4, for any Dungeons & Dragons fans out there) has vertices

$$a_0 = \begin{pmatrix} 1/2 \\ 0 \\ 1/\sqrt{2} \end{pmatrix}, \quad a_1 = \begin{pmatrix} -1/2 \\ 0 \\ -1/\sqrt{2} \end{pmatrix}, \quad a_2 \begin{pmatrix} 0 \\ 1/2 \\ -1/\sqrt{2} \end{pmatrix}, \quad a_3 \begin{pmatrix} 0 \\ -1/2 \\ 1/\sqrt{2} \end{pmatrix}.$$

By translating this so that the vertex $a_0$ is at the origin, we get the vertices

$$a_1 - a_0, \quad a_2 - a_0, \quad a_3 - a_0,$$

so the volume is

$$\frac{1}{6}|\det(a_1 - a_0, a_2 - a_0, a_3 - a_0)|,$$

or

$$\frac{1}{6}\left|\det \begin{pmatrix} -1 & -1/2 & -1/2 \\ 0 & 1/2 & -1/2 \\ 0 & \sqrt{2} & \sqrt{2} \end{pmatrix}\right| = \frac{1}{6\sqrt{2}}.$$

*Remark* 12.10. From this geometric point of view, the fact that $\det(AB) = \det(A)\det(B)$ is obvious: $\det(M)$ is the scaling factor for volumes under the linear map $M$, so under the composite $AB$ we first scale by $\det(B)$ and then by $\det(A)$, so $\det(AB) = \det(A)\det(B)$. Unfortunately, we haven't proved the theorems above which establish the connection between determinants and scaling of volumes.

# 13 Eigenvectors and eigenvalues

## 13.1 Definition and basic ideas

If someone gives you a complicated matrix $A$, it can be very difficult to determine salient information about the underlying linear map associated to $A$. For example, $A$ could be a very simple transformation like a rotation, but happening around an axis that points in some random direction, which makes the matrix very complicated. For this reason, we want to look for geometric features like fixed vectors of $A$, just as we did for 2-by-2 matrices. It turns out that the most fruitful thing to study is the following notion.

**Definition 13.1** (Eigenvectors, eigenvalues)**.** Let $A$ be a matrix. A vector $v$ is called an *eigenvector* for $A$ with *eigenvalue* $\lambda$ if $v \neq 0$ and

$$Av = \lambda v.$$

For example, a fixed vector is an eigenvector with eigenvalue 1. "Eigen" is a German prefix meaning "self". An eigenvector is mapped by $A$ back to itself rescaled by its eigenvalue.

*Remark* 13.2. It is hard to overemphasise the importance of eigenvectors and eigenvalues. We will see some fun applications in this course, but you will encounter them again and again in courses throughout your time as an undergraduate. They are one of the most important notions in mathematics and science. For example, in quantum mechanics the most important equation is the Schrödinger equation, which is the eigenvector equation[5] $H\psi = E\psi$. Here $H$ is a linear map called the Hamiltonian, $\psi$ is a vector describing the state of the quantum system, and $E$ is the energy of the state $\psi$. For example, if $H$ is the Hamiltonian for the hydrogen atom then the eigenvalues of $H$ are the possible energies of light that can be absorbed/emitted by hydrogen (the *spectrum* of the hydrogen atom). The fact that Schrödinger's equation predicts the hydrogen spectrum so well was an early confirmation that quantum mechanics was on the right track.

## 13.2 Finding the eigenvectors

Suppose someone tells you that some matrix $A$ (say $\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$) has some eigenvectors with eigenvalue $\lambda$ (say 1). It's now very easy to find all the eigenvectors with this eigenvalue: you just need to solve the simultaneous equations $Av = \lambda v$, in our case

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

or

$$\begin{aligned} 2x - y &= x \\ x &= y. \end{aligned}$$

These equations both reduce to $y = x$, so the eigenvectors must be $\begin{pmatrix} x \\ x \end{pmatrix}$.

*Remark* 13.3. Note that if $v$ is an eigenvector with eigenvalue $\lambda$ then any rescaling $\mu v$ is also an eigenvector with eigenvalue $\lambda$ because

$$A(\mu v) = \mu A v = \mu \lambda v \Rightarrow A(\mu v) = \lambda(\mu v).$$

Therefore you shouldn't be surprised that we have found a one-parameter family of eigenvectors instead of just one!

"By George!" you might say, "the fellow is right, there is indeed an eigenvector with eigenvalue $\lambda$... but how did he know which $\lambda$ to tell me?" For example, if we tried $\lambda = 2$, we would fail:

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

[5]Caveat: $H$ is more like an infinite-by-infinite matrix, i.e. an operator on a Hilbert space, which makes the Schrödinger equation into a differential equation.

have no solution other than $x = y = 0$, because the first implies $y = 0$ and the second implies $x = y = 0$. Remember than eigenvectors are required to be nonzero. Therefore there is no eigenvector of $\begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ with eigenvalue 2.

## 13.3 Finding the eigenvalues

**Theorem 13.4** (Characteristic polynomial)**.** *The eigenvalues of a matrix $A$ are the roots of the* characteristic polynomial $\chi_A(t)$ *of $A$. This is the polynomial defined by $\chi_A(t) = \det(A - tI)$.*

*Proof of Theorem 13.4.* If $v \neq 0$ and $Av = \lambda v$ then $(A - \lambda I)v = 0$, so $A - \lambda I$ has nontrivial kernel and fails to be invertible. In particular, $\det(A - \lambda I) = 0$. Conversely, if $\det(A - \lambda I) = 0$ then $A - \lambda I$ has nontrivial kernel, so there exists a vector $v$ such that $Av - \lambda v = 0$. □

**Example 13.5.** For the matrix $A = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}$ above, we have

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

This polynomial has 1 as a repeated root, so the only eigenvalue is 1 and, as we saw above, the only eigenvector (up to scaling) is $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

**Example 13.6.** The matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ has characteristic polynomial

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

which has roots $\lambda_1 = \frac{3+\sqrt{5}}{2}$ and $\lambda_2 = \frac{3-\sqrt{5}}{2}$. As eigenvectors, we can take

$$v_1 = \begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix}, \qquad v_2 = \begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix}.$$

**Example 13.7.** The matrix $A = \begin{pmatrix} \frac{3}{2} & \frac{5}{2} & 3 \\ -\frac{1}{2} & -\frac{3}{2} & -3 \\ 1 & 1 & 2 \end{pmatrix}$ has characteristic polynomial

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

What are the roots of this polynomial? With cubics, the easiest method is to guess one of the roots (say $\alpha$), divide the polynomial by $t-\alpha$ (using polynomial long division) and then solve the quadratic equation you get. Here, we can see that $t=1$ is a solution[6] and dividing $-t^3+2t^2+t-2$ by $t-1$ gives $-t^2+t+2$, which has solutions $\frac{-1\pm\sqrt{3}}{-2}=-1,2$. Therefore the eigenvalues are $-1,1,2$. The corresponding eigenvectors are $\begin{pmatrix}-1\\1\\0\end{pmatrix}$, $\begin{pmatrix}1\\-5\\4\end{pmatrix}$ and $\begin{pmatrix}1\\-1\\1\end{pmatrix}$. For example, to get the 1-eigenvector, we solve $v=Av$

$$\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}\frac{3}{2}&\frac{5}{2}&3\\-\frac{1}{2}&-\frac{3}{2}&-3\\1&1&2\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix},$$

that is

$$\begin{aligned}\frac{3x}{2}+\frac{5y}{2}+3z&=x\\-\frac{1}{2}x-\frac{3}{2}y-3z&=y\\x+y+2z&=z.\end{aligned}$$

These equations imply $x+5y+6z=0$ and $x+y+z=0$, so $4y+5z=0$, therefore if we pick $y=-5$ we get $z=4$ and $x=-y-z=1$.

[6]You'd be surprised how often that happens in carefully-constructed examples.

# 14 Applications of eigenvectors

## 14.1 Application I: Differential equations

Let $v(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{pmatrix}$ be a vector-valued function, let $A$ be an $n$-by-$n$ matrix, and consider the system of differential equations

$$\begin{aligned} \dot{x}_1 &= A_{11}x_1 + \cdots + A_{1n}x_n \\ &\vdots \\ \dot{x}_n &= A_{n1}x_1 + \cdots + A_{nn}x_n, \end{aligned}$$

or, more succinctly,

$$\dot{v} = Av.$$

**Example 14.1.** Consider the system of differential equations

$$\begin{aligned} \dot{x} &= 2x + y \\ \dot{y} &= x + y. \end{aligned}$$

We can rewrite this as

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

Suppose that $A$ has $n$ eigenvalues $\lambda_1, \ldots, \lambda_n$ with eigenvectors $v_1, \ldots, v_n$. We can write $v$ in terms of the basis of eigenvectors:

$$v = \sum_{i=1}^{n} f_i v_i$$

for some collection of numbers $f_1, \ldots, f_n$. We have

$$\dot{v} = \sum_{i=1}^{n} \dot{f}_i v_i$$

and

$$Av = A\sum_{i=1}^{n} f_i v_i = \sum_{i=1}^{n} f_i A v_i = \sum_{i=1}^{n} f_i \lambda_i v_i.$$

Since $\dot{v} = Av$, we can equate the coefficients of the vectors $v_1, \ldots, v_n$ in these two expressions. We get the much simpler equation

$$\dot{f}_i = \lambda_i f_i,$$

with solution $f_i(t) = C_i e^{\lambda_i t}$ for some constant $C_i$. The general solution to the differential equation is therefore

$$v = \sum_{i=1}^{n} C_i e^{\lambda_i t} v_i.$$

Let's apply this to solve the differential equations from Example 14.1

**Example 14.2.** The matrix $A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$ has eigenvalues $\lambda_1 = \frac{3+\sqrt{5}}{2}$, $\lambda_2 = \frac{3-\sqrt{5}}{2}$ and eigenvectors $v_1 = \begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix}$ and $v_2 = \begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix}$. Therefore, the general solution is

$$C_1 e^{(3+\sqrt{5})t/2}\begin{pmatrix} 1 \\ \frac{1+\sqrt{5}}{2} \end{pmatrix} + C_2 e^{(3-\sqrt{5})t/2}\begin{pmatrix} 1 \\ \frac{1-\sqrt{5}}{2} \end{pmatrix},$$

or

$$\begin{aligned} x(t) &= C_1 e^{(3+\sqrt{5})t/2} + C_2 e^{(3-\sqrt{5})t/2}, \\ y(t) &= \frac{1+\sqrt{5}}{2} C_1 e^{(3+\sqrt{5})t/2} + \frac{1-\sqrt{5}}{2} C_2 e^{(3-\sqrt{5})t/2}. \end{aligned}$$

**Example 14.3.** Consider the system of differential equations

$$\dot{x} = 2x + y$$
$$\dot{y} = 2y - x.$$

We can rewrite this as

$$\frac{d}{dt}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}.$$

The matrix $A = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix}$ has characteristic polynomial

$$\det(A - tI) = \det\begin{pmatrix} 2-t & 1 \\ -1 & 2-t \end{pmatrix} = (2-t)(2-t) + 1 = t^2 - 4t + 5,$$

which has roots $2 \pm i$. The eigenvectors for these eigenvalues are the solutions to

$$\begin{pmatrix} (2+i)x \\ (2+i)y \end{pmatrix} = \begin{pmatrix} 2x + y \\ 2y - x \end{pmatrix},$$

(i.e. $\begin{pmatrix} 1 \\ i \end{pmatrix}$) and

$$\begin{pmatrix} (2-i)x \\ (2-i)y \end{pmatrix} = \begin{pmatrix} 2x + y \\ 2y - x \end{pmatrix}$$

(i.e. $\begin{pmatrix} 1 \\ -i \end{pmatrix}$). Therefore the general solution to the system of differential equations in this example is

$$C_1 e^{(2+i)t}\begin{pmatrix} 1 \\ i \end{pmatrix} + C_2 e^{(2-i)t}\begin{pmatrix} 1 \\ -i \end{pmatrix},$$

or

$$x(t) = C_1 e^{(2+i)t} + C_2 e^{(2-i)t}, \quad y(t) = iC_1 e^{(2+i)t} - iC_2 e^{(2-i)t}.$$

You should not worry about the appearance of imaginary numbers here: if the initial condition you pick is real then all the imaginary terms will group together to give trigonometric functions, using the facts that

$$\cos(t) = \frac{e^{it} + e^{-it}}{2}, \qquad \sin(t) = \frac{e^{it} - e^{-it}}{2i}.$$

For example, let's try and find the solution for the initial condition $x(0) = 0$, $y(0) = 1$. This means

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

that is, $C_1 = -C_2 = -i/2$. Substituting these values for $C_1, C_2$ we get

$$x(t) = -\frac{i}{2}(e^{(2+i)t} - e^{(2-i)t}) = e^{2t}\frac{e^{it} - e^{-it}}{2i} = e^{2t}\sin(t)$$

and

$$y(t) = \frac{1}{2}(e^{(2+i)t} + e^{(2+it)}) = e^{2t}\cos(t).$$

Finally, we should investigate what happens when $A$ has fewer than $n$ eigenvectors.

**Example 14.4.** Suppose that $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The differential equations we get out of $A$ are

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

We can solve the second equation immediately and get $y = C_1 e^t$. Substituting back into the first, we get

$$\dot{x} = x + C_1 e^t.$$

Rearranging gives

$$\dot{x}e^{-t} - xe^{-t} = C_1,$$

and we note (using the Leibniz rule for differentiation) that

$$\frac{d}{dt}(xe^{-t}) = \dot{x}e^{-t} - xe^{-t},$$

so

$$\frac{d}{dt}(xe^{-t}) = C_1,$$

which gives

$$x = (C_1 t + C_2)e^t.$$

In a later course on linear algebra, you will see the *Jordan normal form* theorem for matrices, which tells you that, as long as you work over $\mathbb{C}$, viewed in suitable coordinates, your matrix always looks like a bunch of blocks which look like this:

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

For such matrices, you can do something similar to the previous example.

## 14.2 Application II: Ellipsoids

**Definition 14.5.** We say that $A$ is a *positive definite matrix* if $v^T A v > 0$ for any vector $v \neq 0$.

**Example 14.6.** The identity matrix is positive definite because $v^T I v = v \cdot v \geq 0$ with equality if and only if $v = 0$.

**Example 14.7.** The matrix $A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ is not positive definite because $\begin{pmatrix} 0 & 1 \end{pmatrix} A \begin{pmatrix} 0 \\ 1 \end{pmatrix} = -1$.

**Definition 14.8.** An *ellipsoid* is a subset in $n$-dimensional space having the form

$$\{v \in \mathbb{R}^n \ : \ v^T A v = c\},$$

where $A$ is a positive definite symmetric matrix with real entries and $c > 0$ is a positive real constant.

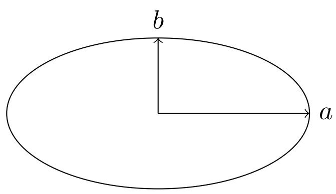
**Example 14.9.** Given two numbers $a, b \in \mathbb{R}$, the matrix

$$A = \begin{pmatrix} \frac{1}{a^2} & 0 \\ 0 & \frac{1}{b^2} \end{pmatrix}$$

is positive definite. If $c = 1$ then the corresponding ellipsoid is the ellipse

$$\left\{(x, y) \in \mathbb{R}^2 \ : \ \frac{x^2}{a^2} + \frac{y^2}{b^2} = 1\right\}$$

having semimajor axis $a$ and semiminor axis $b$.



**Theorem 14.10.** *An ellipsoid defined by a positive definite symmetric matrix $A$ can be rotated to the ellipsoid*

$$\{(u_1, \ldots, u_n) \in \mathbb{R}^n \ : \ \sum_{i=1}^{n} \lambda_i u_i^2 = c\}$$

*where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of $A$ and $c$ is some positive number.*

We won't prove this theorem in full, because it relies on the fact that a positive definite symmetric matrix has a basis of eigenvectors (which is beyond what we have time for). But we'll at least check that *if $A$ has a basis of eigenvectors* then the result holds. First, an important lemma.

**Lemma 14.11.** *Suppose that $A$ is a symmetric matrix with real entries. Then the eigenvalues of $A$ are real and if $\lambda, \mu$ are distinct eigenvalues with eigenvectors $v, w$ respectively then $v \cdot w = 0$.*

*Proof.* Suppose that $Av = \lambda v$. Consider the expression $\bar{v}^T Av$, where $\bar{v}$ denotes complex conjugation. Then, because $A = A^T = \bar{A}^T$, we have

$$\bar{\lambda}\bar{v}^T v = (\overline{Av})^T v = \bar{v}^T Av = \lambda \bar{v}^T v.$$

Note that if $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ then $\bar{v}^T v = \sum |x_1|^2 + \cdots + |x_n|^2 > 0$ if $v \neq 0$, so dividing through by $\bar{v}^T v$ we get $\bar{\lambda} = \lambda$ and deduce that $\lambda$ is real.

If $v$ and $w$ are two eigenvectors for distinct eigenvalues $\lambda, \mu$ then

$$\begin{aligned} \lambda w^T v &= w^T (Av) \\ &= (Aw)^T v \\ &= \mu w^T v \end{aligned}$$

so, since $\lambda \neq \mu$, we must have $w^T v = 0$, i.e. $v \cdot w = 0$. □

Now suppose that $A$ is a real symmetric matrix which has a basis of $n$ eigenvectors $v_1, \ldots, v_n$ with eigenvalues $\lambda_1, \ldots, \lambda_n$. By the lemma above, these eigenvalues are all real and the eigenvectors are orthogonal. Let's rescale the eigenvectors so that they each have unit length. If we write a vector $v$ as $\sum_{i=1}^n u_i v_i$ then we have

$$v^T Av = \sum_{i=1}^{n} \sum_{j=1}^{n} u_i x_j v_i^T Av_j = \sum_{i,j=1}^{n} \lambda_i u_i^2,$$

since $v_i^T v_j = \delta_{ij}$. These $u_1, \ldots, u_n$ are the coordinates referred to in Theorem 14.10. In particular, the *principal axes* (the higher-dimensional analogues of the semi-major and semi-minor axes) are the eigenvectors of $A$ and the principal radii are $\frac{1}{\sqrt{\lambda_i}}$, $i = 1, \ldots, n$.

**Example 14.12.** Let $A = \begin{pmatrix} 3/2 & -1/2 \\ -1/2 & 3/2 \end{pmatrix}$. This defines an ellipse $v^T Av = 1$, in other words

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

The characteristic polynomial of $A$ is

$$\det \begin{pmatrix} 3/2 - t & -1/2 \\ -1/2 & 3/2 - t \end{pmatrix} = t^3 - 3t + 2,$$

so the eigenvalues are 1 and 2. The (unit length) eigenvectors are $v_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$ and $v_2 \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$. If we work with coordinates $u_1, u_2$ related to $x, y$ via

$$\begin{pmatrix} x \\ y \end{pmatrix} = u_1 \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} + u_2 \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

(that is, $x = \frac{u_1 + u_2}{\sqrt{2}}$, $y = \frac{u_1 - u_2}{\sqrt{2}}$) then the equation of the ellipse $v^T Av = 1$ becomes $u_1^2 + 2u_2^2 = 1$. We see that the change of coordinates between $x, y$ and $u_1, u_2$ is actually a 45 degree rotation.

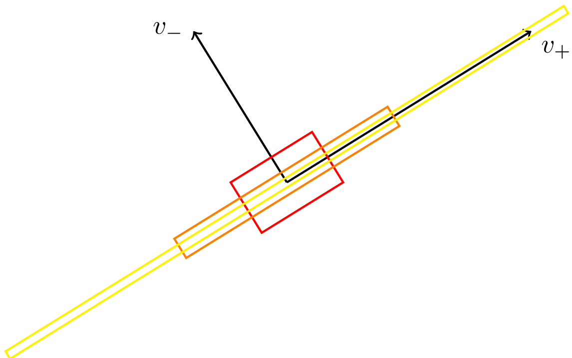
## 14.3 Application III: Dynamics

Consider the matrix $\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}$. We have seen (Example 13.6) that its eigenvalues are $\lambda_\pm := \frac{3\pm\sqrt{5}}{2}$, with eigenvectors $v_\pm = \begin{pmatrix} 1 \\ \frac{1\pm\sqrt{5}}{2} \end{pmatrix}$.

Suppose we pick a point $v \in \mathbb{R}^2$ and write it as $v = av_+ + bv_-$. Then $Av = \lambda_+ av_+ + \lambda_- bv_-$. Suppose that $a \neq 0$ and $b \neq 0$. Since $\lambda_+ > 1$ and $\lambda_- < 1$, this means that the point moves inwards along $v_-$ and outwards along $v_+$. If we apply $A$ again and again, we get

$$A^n v = \lambda_+^n av + \lambda_-^n bv_-.$$

As $n \to \infty$, $\lambda_+^n \to \infty$ and $\lambda_-^n \to 0$, so the point gets closer and closer to the $v_+$-eigenline, but gets pushed outwards along the eigenline. If we draw a rectangle in $\mathbb{R}^2$ and apply $A$ many times, this square will get stretched outwards in the $v_+$ direction and squished inwards in the $v_-$-direction.




This is typical behaviour of a "hyperbolic" dynamical system. Here are two fun facts which are not unrelated to this.

**Example 14.13** (Fibonacci numbers)**.** The Fibonacci sequence

$$F_1, F_2, F_3, F_4, F_5, F_6, F_7, \ldots = 1, 1, 2, 3, 5, 8, 13, \ldots$$

satisfies the recursion $F_{n+2} = F_{n+1} + F_n$, which we can write as a matrix equation:

$$\begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad \begin{pmatrix} F_{n+1} \\ F_{n+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix}.$$

The eigenvalues of $\begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$ are $\lambda_\pm = \frac{1\pm\sqrt{5}}{2}$ with eigenvectors $v_\pm = \begin{pmatrix} 1 \\ \frac{1\pm\sqrt{5}}{2} \end{pmatrix}$. Although $\lambda_-$ is negative, its magnitude is nonetheless $< 1$, so $\lambda_-^n \to 0$. Also, $\lambda_+^n \to \infty$. Therefore $\begin{pmatrix} F_n \\ F_{n+1} \end{pmatrix} = A^n \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ tends in the limit $n \to \infty$ to a vector pointing along the $v_+$-eigenline, which has slope $\frac{1+\sqrt{5}}{2}$. This means

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

This number is known as the golden ratio.

**Example 14.14** (Arnol'd's cat map)**.** Let $A$ be the example above. If you take a square picture of a cat and use it to tile the plane, then you apply $A^n$ to the plane and let $n$ increase, the picture will get distorted very quickly. However, at some point, the picture will reappear more-or-less exactly as you had it to begin with. In fact, if you have a digital image, it will reappear exactly how it started (because there's only a finite number of pixels involved). This is due to a phenomenon called *ergodicity* of the flow, whereby every point, at some time, comes back close to where it started (except possibly in a different tile). Eventually, many points come back close to where they started (except possibly in a different tile) and you see something resembling the image you started with.

You can see dramatic realisations of this in videos and applets online.

# 15 Subspaces I

## 15.1 Subspaces

**Definition 15.1** (Subspaces)**.** A subset $V \subset \mathbb{R}^n$ is called a *linear subspace* (or just subspace) if it satisfies the following conditions:

- $v, w \in W$ implies $v + w \in V$.
- $v \in V$, $\lambda \in \mathbb{R}$ implies $\lambda v \in V$.

In other words, $V$ is closed under addition and rescaling. Subspaces are the natural higher-dimensional generalisation of lines and planes through the origin in 3-d.

Sometimes you want to consider lines or planes which don't pass through the origin, in which case the following definition comes in handy:

**Definition 15.2** (Affine subspaces)**.** A subset $V \subset \mathbb{R}^n$ is called an *affine subspace* if there exists a vector $w \in \mathbb{R}^n$ and a linear subspace $V' \subset \mathbb{R}^n$ such that $V = \{w + v \ : \ v \in V'\}$. In other words, $V$ is obtained by translating $V'$ by the vector $w$.

*Remark* 15.3. A *line* is a 1-dimensional subspace. A *plane* is a 2-dimensional subspace.

**Definition 15.4** (Codimension)**.** The *codimension* of a subspace $V \subset \mathbb{R}^n$ is $p$ if $\dim V = n - p$.

**Example 15.5.** A line in $\mathbb{R}^3$ has codimension 2. A plane in $\mathbb{R}^4$ has codimension 2, while a line in $\mathbb{R}^4$ has codimension 3.

**Definition 15.6.** A *hyperplane* is a subspace of codimension 1. For example, a line in $\mathbb{R}^2$, or a plane in $\mathbb{R}^3$.

Suppose someone asks you to give them a subspace of $\mathbb{R}^n$. You can answer them in one of two ways:

- You can write down equations for the subspace, for example you can say something like:
  - "it's the line $x + y = 0$ in $\mathbb{R}^2$",
  - "it's the plane in $\mathbb{R}^3$ cut out by the equation $z = 0$".
- You can give them a collection of vectors which "span" the subspace, for example you can say something like:
  - "it's the line through the origin pointing in the $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$-direction".
  - "it's the plane in $\mathbb{R}^3$ spanned by the vectors $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$".

We'll focus on these two methods in order, then talk about how to relate them.

## 15.2 Equations for subspaces; kernel

**Example 15.7.** A linear hyperplane is cut out by a single linear equation. More precisely, a row vector $r = \begin{pmatrix} r_1 & \cdots & r_n \end{pmatrix}$ defines a linear hyperplane in $\mathbb{R}^n$, namely:

$$\left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \ : \ rx = 0 \right\}.$$

Equivalently, this is the hyperplane orthogonal to the column vector $r^T$.

**Definition 15.8.** Given a linear subspace $V \subset \mathbb{R}^n$ and a vector $w$, we define the *translate* $w + V = \{v + w \in \mathbb{R}^n \ : \ v \in V\}$ of $V$ by $w$ to be the affine subspace obtained by translating the elements of $V$ along the vector $w$.

**Example 15.9.** A row vector $r = \begin{pmatrix} r_1 & \cdots & r_n \end{pmatrix}$ together with a number $b$ defines an *affine hyperplane* in $\mathbb{R}^n$, namely:

$$\left\{ x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n \ : \ rx = b \right\}.$$

Equivalently, this is the hyperplane orthogonal to $r^T$ translated by $\frac{br^T}{|r|^2}$, i.e. translated a certain amount in the $r^T$ direction. Note that this is a linear subspace if and only if $b = 0$.

**Example 15.10.** An $m$-by-$n$ matrix $A$ define $m$ linear hyperplanes, cut out by the equations

$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= 0 \\ \vdots \qquad\quad \vdots \qquad\quad & \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= 0. \end{aligned}$$

A *solution* $v = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ to this system of equations is then a vector $v \in \mathbb{R}^n$ satisfying $Av = 0$; in other words a vector $v$ which belongs to all $m$ of the hyperplanes; in other words a point where the hyperplanes intersect.

**Definition 15.11.** There is a fancy name for the linear subspace given by $\{v \in \mathbb{R}^n \ : \ Av = 0\}$. It is called the kernel of $A$, written $\ker(A)$.

**Example 15.12.** An $m$-by-$n$ matrix $A$ and a vector $b \in \mathbb{R}^m$ define $m$ affine hyperplanes, cut out by the equations

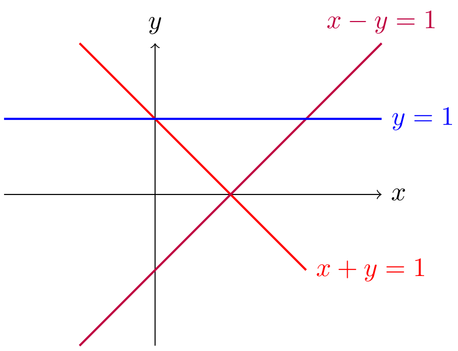
$$\begin{aligned} A_{11}x_1 + \cdots + A_{1n}x_n &= b_1 \\ \vdots \qquad\quad \vdots \qquad\quad & \vdots \\ A_{m1}x_1 + \cdots + A_{mn}x_n &= b_m. \end{aligned}$$

The set of solutions to $Av = b$ is the intersection of these affine hyperplanes.

**Example 15.13.** Consider the matrix $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 0 & 1 \end{pmatrix}$ and the vector $b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$. The equations $Av = b$ define three *lines* (hyperplanes in $\mathbb{R}^2$):

$$x + y = 1, \quad x - y = 1, \quad y = 1$$

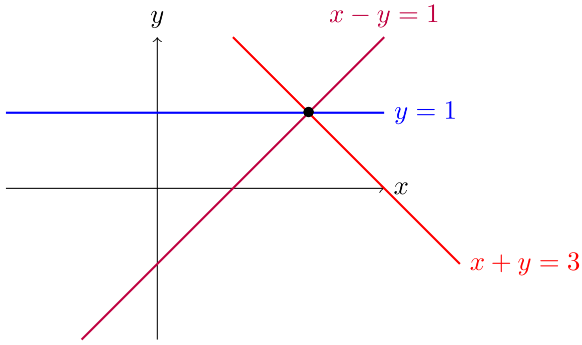
drawn red, purple and blue respectively in the diagram below.




Since the lines don't have a common intersection, we know the system of equations has no solutions (the lines intersect in pairs, so any two of the equations admit a solution, but there is no one point contained in all three lines).

We see that the intuition that an overdetermined system (more hyperplanes than dimensions) has no solutions is justified, because you need your $n + 1$ hyperplanes in $\mathbb{R}^n$ to be in very special position to make them have a common intersection. Nonetheless, it can happen.

**Example 15.14.** Let $A$ be as before but $b = \begin{pmatrix} 3 \\ 1 \\ 1 \end{pmatrix}$. This has the effect of translating one of the lines from Example 15.13 so that it becomes $x + y = 3$. As we see below, these lines below have a common intersection at $\begin{pmatrix} 2 \\ 1 \end{pmatrix}$, so the overdetermined system has a solution $x = 2$, $y = 1$ (marked with a dot below).




*Remark* 15.15. Given a subspace $V \subset \mathbb{R}^n$ of dimension $n - p$ (codimension $p$) and a subspace $W \subset \mathbb{R}^n$ of dimension $n - q$ (codimension $q$), we "expect" the intersection $V \cap W$ to have dimension $n - p - q$ (codimension $p + q$). In other words, codimension is *usually* additive under intersection. For example, in $\mathbb{R}^3$, a plane (codimension 1) and a line (codimension 2) will usually intersect at a point (codimension 3), unless you're in the exceptional situation that the line is contained inside the plane. As a corollary of this, we *expect* the space of solutions to a system of $m$ equations in $n$ unknowns to be $n - m$ (each equation cuts down the set of solutions by one dimension)...except when it isn't!

Having made this remark, let us give a more precise characterisation of the dimension of the space of solutions.

**Theorem 15.16.** *Let $A$ be an $m$-by-$n$ matrix and $b \in \mathbb{R}^m$ be a vector. Suppose that* $\ker(A)$ *has dimension $k$ (this number is called the* nullity *of $A$). Then, the dimension of the space of solutions to $Av = b$, assuming it is nonempty, is equal to $k$. Indeed, the space of solutions is a translate of* $\ker(A)$.

*Proof.* If $v_1, v_2$ are solutions to $Av = b$ then $A(v_1 - v_2) = b - b = 0$, so the difference $v_1 - v_2$ is in the kernel of $A$. Similarly, if $Av_1 = b$ and $Av = 0$ then $A(v_1 + v) = b + 0 = b$, so adding elements of the kernel to a solution gives another solution. Therefore, if we fix one solution $v_1$, the space of solutions is $v_1 + \ker(A) = \{v_1 + v \ : \ v \in \ker(A)\}$, i.e. a translate of $\ker(A)$. ☐

**Theorem 15.17.** *Given a matrix $A$, its nullity is equal to the number of free indices once $A$ has been put into reduced echelon form.*

*Proof.* We saw that the general solution to $Av = b$ has one parameter for each free index. Therefore it is a space with dimension equal to the number of free indices. ☐

# 16 Subspaces II

## 16.1 Spanning sets for subspaces

**Definition 16.1.** Given a collection of vectors $v_1, \ldots, v_k$, a *linear combination* of these vectors is an expression of the form

$$v = \lambda_1 v_1 + \cdots + \lambda_k v_k$$

for some choice of coefficients $\lambda_1, \ldots, \lambda_k$. We define the *linear subspace spanned by* $v_1, \ldots, v_k$ (or the span of $v_1, \ldots, v_k$, written $\mathrm{span}(v_1, \ldots, v_k)$) to be the set of all linear combinations of $v_1, \ldots, v_k$.

**Lemma 16.2.** *For any collection of vectors $v_1, \ldots, v_k \in \mathbb{R}^n$, the set $\mathrm{span}(v_1, \ldots, v_k)$ is a linear subspace of $\mathbb{R}^n$.*

*Proof.* If we rescale a linear combination $\sum_i \lambda_i v_i$ by $\mu$ then we get the linear combination $\sum_i (\mu\lambda_i) v_i$. If we add two linear combinations $\sum_i \lambda_i v_i$ and $\sum_i \mu_i v_i$ then we get the linear combination $\sum_i (\lambda_i + \mu_i) v_i$. Therefore linear combinations form a linear subspace. □

**Example 16.3.** The set of all linear combinations of $v_1$ is just the set of all vectors $\lambda_1 v_1$, $\lambda_1 \in \mathbb{R}$. In other words, it's the set of all rescalings of $v_1$, otherwise known as the line that points in the $v_1$-direction.

**Example 16.4.** Let $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$ and $v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$. The subspace spanned by $v_1, v_2$ is the set of all vectors $\lambda_1 v_1 + \lambda_2 v_2 = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ 0 \end{pmatrix}$, in other words, it is the $xy$-plane.

**Example 16.5.** The plane spanned by $v_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, $v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$ and $v_3 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ is *also* the $xy$-plane, because adding multiples of $v_3$ doesn't take you out of this plane. The issue here is that $v_3$ is itself a linear combination of $v_1$ and $v_2$ ($v_3 = v_1 + v_2$) so it doesn't change the spanning set.

**Definition 16.6.** A spanning set is called a *basis* if it has minimal size.

**Theorem 16.7.** *All bases for the same subspace have the same size. This size is called the dimension of the subspace. (I haven't actually given you a formal definition of dimension until now).*

*Proof.* This will be proved in your next course on linear algebra, next year. □

## 16.2 Image of a matrix

**Definition 16.8.** The *image* of an $m$-by-$n$ matrix $A$ is the set of all $b \in \mathbb{R}^m$ such that $Av = b$ has a solution $v \in \mathbb{R}^n$.

**Lemma 16.9.** *The image of $A$ is spanned by the columns of $A$.*

*Proof.* If the columns of $A$ are $a_1, \ldots, a_n \in \mathbb{R}^m$ then

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = x_1 a_1 + \cdots + x_n a_n,$$

so the image of $A$ is the set of linear combinations of the columns, as required. □

**Example 16.10.** If $A = \begin{pmatrix} 1 & 1 \\ 2 & 0 \\ 0 & 1 \end{pmatrix}$ then the image of $A$ is the plane spanned by $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ (see Example 1.17).

**Definition 16.11.** The *rank* of $A$ is defined to be the dimension of the image.

**Theorem 16.12.** *The rank of $A$ is equal to the number of leading indices when $A$ is put into reduced echelon form.*

*Proof.* First note that row operations do not change the rank: if $A$ and $A'$ are related by a row operation then $A' = EA$ for some elementary matrix $E$, and now the map $b \mapsto Eb$ gives an isomorphism between the image of $A$ and the image of $A'$ (isomorphism in the sense that $E$ is an invertible linear map). Therefore we may assume that $A$ is in reduced echelon form by Theorem 6.5.

So suppose that $A$ is in reduced echelon form with the first $k$ rows nonzero (so that $k$ equals the number of leading indices). The equation $Av = b$ has a solution if and only if $b_{k+1} = \cdots = b_m = 0$, so the image of $A$ is equal to the subspace spanned by the first $k$ basis vectors, which has dimension $k$. □

Here is a useful theorem relating the rank and the nullity of an $m$-by-$n$ matrix:

**Theorem 16.13** (Rank-nullity theorem)**.** *If $A$ is an $m$-by-$n$ matrix, the rank and the nullity of $A$ sum to $n$.*

*Proof.* In reduced echelon form, the number of leading indices and free indices sum to $n$ (number of columns), so this follows from Theorem 15.17 and Theorem 16.12. □

## 16.3 Kernel, image and simultaneous equations

To relate this to what we said about simultaneous equations, we can summarise everything we've said as follows:

**Theorem 16.14.** *Let $A$ be an $m$-by-$n$ matrix and $b \in \mathbb{R}^m$ be a vector. Then $Av = b$ has a solution if and only if $b \in \mathrm{im}(A)$. If $Av = b$ has a solution then the space of solutions is a translate of $\ker(A)$.*

The following diagram may help you to remember whereabouts the kernel and image of an $m$-by-$n$ matrix $A$ live:

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \\ \cup\!\!| & & \cup\!\!| \\ \ker(A) & & \mathrm{im}(A) \end{array}$$

# 17 Linear maps

This lecture is intended as a foretaste of things to come. We introduce an extra layer of abstraction, which suddenly elevates us above the clouds and we see how to apply linear algebra in contexts we had not formerly imagined.

## 17.1 Linearity

We defined a linear map $\mathbb{R}^n \to \mathbb{R}^m$ to be a map of the form $v \mapsto Av$ where $A$ is an $m$-by-$n$ matrix. There is a different way to characterise linear maps, which we now discuss.

**Definition 17.1.** An ($\mathbb{R}$-)vector space[7] is a set $V$ together with:

- a map $V \times V \to V$, written $(v, w) \mapsto v + w$,
- a map $\mathbb{R} \times V \to V$, written $(\lambda, v) \mapsto \lambda v$,
- an element $0 \in V$,

such that:

$$\begin{aligned} u + (v + w) &= (u + v) + w & v + w &= w + v \\ v &= 0 + v = v + 0, & v + (-v) &= 0 \\ 1v &= v & \lambda(\mu v) &= (\lambda\mu)v \\ (\lambda + \mu)v &= \lambda v + \mu v & \lambda(v + w) &= \lambda v + \lambda w \end{aligned}$$

for all $u, v, w \in V$ and $\lambda, \mu \in \mathbb{R}$.

For example, $\mathbb{R}^n$ equipped with the usual addition and rescaling action of $\mathbb{R}$ is a vector space.

**Definition 17.2.** Let $V, W$ be vector spaces. A map $T \colon V \to W$ is called *linear*[8] if the following conditions are satisfied:

- for all $v, w \in \mathbb{R}^n$ and we have $T(v + w) = T(v) + T(w)$.
- for all $\lambda \in \mathbb{R}$ and $v \in \mathbb{R}^n$, we have $T(\lambda v) = \lambda T(v)$.

**Theorem 17.3.** *If $T \colon \mathbb{R}^n \to \mathbb{R}^m$ is linear then there exists an $n$-by-$m$ matrix $A$ such that $T(v) = Av$ for all $v \in \mathbb{R}^n$. Conversely, if $A$ is an $m$-by-$n$ matrix then a map $\mathbb{R}^n \to \mathbb{R}^m$ of the form $v \mapsto Av$ is linear.*

*Proof.* If $T$ is linear then it is determined by its values on the basis vectors $e_1, \ldots, e_n$. To see this, observe that if $v = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \sum_{i=1}^n v_i e_i$ then $T(v) = T(\sum_{i=1}^n v_i e_i) = \sum_{i=1}^n v_i T(e_i)$ by linearity, so the vectors $T(e_1), \ldots, T(e_n)$ determine $T$ completely. If we pick $A$ to be the matrix whose columns are $T(e_1), \ldots, T(e_n)$ then $Av = \sum_{i=1}^n v_i T(e_i) = T(v)$, so the matrix we were looking for exists (and is uniquely specified by $T$).

Conversely, if $A$ is a matrix then the identity $A(v + w) = Av + Aw$ is just the distributivity of matrix multiplication and $A(\lambda v) = \lambda Av$ is easy to check. □

In fact, one can prove that any finite-dimensional vector space $V$ is isomorphic to $\mathbb{R}^n$ for some $n$. (Isomorphic here means that there is an invertible linear map $V \to \mathbb{R}^n$; finite-dimensional means that there is a finite spanning set). However, there's nothing to stop you using *infinite-dimensional* vector spaces, and then things get interesting.

[7] You can replace $\mathbb{R}$ by any field $k$ (like $\mathbb{Q}$ or $\mathbb{C}$) in this definition and get a $k$-vector space. Usually we just omit the field from the notation and call it a *vector space*.

[8] Again, if we're working with $k$-vector spaces (e.g $k = \mathbb{Q}, \mathbb{C}$) then you need to talk about $k$-linear maps and replace $\mathbb{R}$ with $k$ everywhere in this definition.

**Example 17.4.** The space of continuous functions $f\colon \mathbb{R} \to \mathbb{R}$ is a vector space, usually called $\mathcal{C}^0(\mathbb{R})$. You can add two functions $(f+g)(x) = f(x)+g(x)$ and you can rescale a function $(\lambda f)(x) = \lambda f(x)$ and these operations satisfy the conditions required of a vector space (the zero function is $f(x) = 0$).

**Example 17.5.** The space of once-continuously-differentiable functions is a subspace of $\mathcal{C}^0(\mathbb{R})$, usually written $\mathcal{C}^1(\mathbb{R}) \subset \mathcal{C}^0(\mathbb{R})$.

**Example 17.6.** Differentiation defines a linear map $\frac{d}{dx}\colon \mathcal{C}^1(\mathbb{R}) \to \mathcal{C}^0(\mathbb{R})$. It is linear because

$$\frac{d}{dx}(f+g)(x) = \frac{df}{dx}(x) + \frac{dg}{dx}(x), \qquad \frac{d}{dx}(\lambda f)(x) = \lambda \frac{df}{dx}(x).$$

Can we write a matrix for differentiation? We need to pick a basis for $\mathcal{C}^1(\mathbb{R})$, which is a highly nontrivial task. Let's be a little less ambitious and restrict to the subspace of *analytic functions*, i.e. functions $f$ whose Taylor series converges to $f$. This is usually written $\mathcal{C}^\omega(\mathbb{R})$. The functions $f_n(x) = x^n$, $n = 0, 1, 2, \ldots$, form a *Schauder basis* for this space, which means that any function $f \in \mathcal{C}^\omega(\mathbb{R})$ can be written as an infinite sum of these functions (namely its Taylor series!). In other words, we are thinking of the coefficients of the Taylor expansion as coordinates on the space $\mathcal{C}^\omega(\mathbb{R})$. That is, a function $f$ can be thought of as an infinite vector $\begin{pmatrix} f(0) \\ \frac{df}{dx}(0) \\ \frac{1}{2}\frac{d^2f}{dx^2}(0) \\ \frac{1}{3!}\frac{d^3f}{dx^3}(0) \\ \vdots \end{pmatrix}$.

If $f(x) = \sum_{n\geq 0} a_n x^n$ then $\frac{df}{dx} = \sum_{n\geq 1} n a_n x^{n-1} = \sum_{n\geq 0}(n+1)a_{n+1}x^n$, so our "matrix" for differentiation is

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 2 & 0 & \cdots \\ 0 & 0 & 0 & 3 & \\ \vdots & \vdots & \vdots & & \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \end{pmatrix} = \begin{pmatrix} a_1 \\ 2a_2 \\ 3a_3 \\ \vdots \end{pmatrix}.$$

If one restricts instead to *periodic functions* $f(x+2\pi) = f(x)$ then there is an alternative basis, coming from the functions $\sin(nx), \cos(nx)$. The expansion of a function in terms of this basis is called its Fourier expansion, and again differentiation of a function can be thought of as a linear transformation of its Fourier series. This leads to the powerful method of *Fourier transform*, which allows you to convert differential equations into much simpler linear equations.

**Example 17.7.** What is the kernel of differentiation? It is the set of functions whose derivative is identically zero, in other words, the constant functions. What is the inverse of differentiation? Well, because there is a kernel it has no inverse, strictly speaking, but clearly integration should define an inverse in some sense. This is why it doesn't make sense to say "the integral of $f$" unless you also say "plus an unknown constant".

**Example 17.8.** Consider the linear map $\frac{d}{dx}\colon \mathcal{C}^\omega(\mathbb{R}) \to \mathcal{C}^\omega(\mathbb{R})$. What are the eigenvalues and eigenvectors of this map? A $\lambda$-eigenvector will be a function $f$ which solves the equation

$$\frac{df}{dx} = \lambda f.$$

We can solve this by dividing through by $f$ and integrating:

$$\ln f = \int \frac{df}{f} = \int \lambda dx = \lambda x + c,$$

i.e. $f = Ce^{\lambda x}$. So the $\lambda$-eigenline is spanned by $f(x) = e^{\lambda x}$ and every $\lambda \in \mathbb{R}$ arises as an eigenvalue.

**Example 17.9.** Similarly, $\cos(x\sqrt{-\lambda})$ and $\sin(x\sqrt{-\lambda})$ are $\lambda$-eigenvectors for $\frac{d^2}{dx^2}$, that is they solve the differential equation

$$\frac{d^2f}{dx^2} = \lambda f.$$

We often say "eigenfunction" rather than eigenvector in this context. Finding eigenfunctions and eigenvalues of differential operators is an incredibly important problem; essentially all of quantum mechanics boils down to solving this problem for particular operators.