MT182 MATRIX ALGEBRA

MARK WILDON

These notes are intended to give the logical structure of the course; proofs and further examples and remarks will be given in lectures. Further installments will be issued as they are ready. All handouts and problem sheets will be put on Moodle.

These notes are based in part on notes for similar courses run by Dr Rainer Dietmann, Dr Stefanie Gerke and Prof. James McKee. I would very much appreciate being told of any corrections or possible improvements.

You are warmly encouraged to ask questions in lectures, and to talk to me after lectures and in my office hours. I am also happy to answer questions about the lectures or problem sheets by email. My email address is `mark.wildon@rhul.ac.uk`.

Lectures: Monday noon (QBLT), Friday 10am (QBLT), Friday 3pm (BLT1).

Office hours in McCrea 240: Monday 4pm, Wednesday 10am and Friday 4pm.

Workshops for MT172/MT182: Mondays or Wednesdays, from Week 2.

Date: Second Term 2015/16.

MATRIX ALGEBRA

This course is on vectors and matrices, the fundamental objects of algebra. You will learn the material mainly by solving problems, in lectures, in workshops and in your own time.

Outline.

(A) *Vectors in* $\mathbb{R}^3$*:* vector notation, displacement vectors, dot product and vector product. Lines and planes. Use in solving geometric problems.
(B) *Introduction to matrices:* matrix addition and multiplication. Rotations and reflections in $\mathbb{R}^2$ and $\mathbb{R}^3$. Eigenvectors and eigenvalues, characteristic polynomial, trace and determinant. Diagonalization with applications.
(C) *Determinants:* permutations and disjoint cycle decomposition. Definition of the determinant of an $n \times n$ matrix. Application: inverting a matrix.
(D) *Solving equations:* elementary row operations, echelon form and row-reduced echelon form, rank. Application to solving equation and computing inverse of a matrix.
(E) *Vector spaces:* abstract vector spaces, linear independence, spanning, dimension, subspaces. Bases of vector spaces.

Recommended Reading. All these books are available on short-term loan from the library. If you find there are not enough copies, email me.

[1] *Linear algebra (Schaum's outlines)*, Seymour Lipschutz, McGraw-Hill (1968), 510.76 LIP. Clear and concise with lots of examples.

[2] *Undergraduate algebra: A first course*, Christopher Norman, Oxford University Press (1986), 512.11 NOR. Very clear, good for Part (E) of the course.

[3] *Linear algebra: a modern introduction*, David Poole, Brooks/Cole (2011), 3rd edition, 512.3 POO. Simple and straightforward, good for first four parts of course.

[4] *Linear algebra*, Stephen H. Freidberg, Arnold J. Insel, Laurence E. Spence, Pearson Education (2002), 515.5 FRI. More advanced, good for further reading.

Also you will find a link on Moodle to Prof. James McKee's notes. These will give you a different view of the course material with more detail than these notes.

Problem sheets. There will be 8 marked problem sheets; the first is due in on Friday 22nd January. To encourage you to work hard during the term, **each problem sheet is worth 1.25% of your overall grade**. Note that this mark is awarded for *any reasonable attempt* at the sheet. (There is a link on Moodle to the document explaining this policy in more detail.)

Moodle. All handouts, problem sheets and answers will be posted on Moodle. You should find a link under 'My courses', but if not, go to `moodle.rhul.ac.uk/course/view.php?id=406`.

Exercises in these notes. Exercises set in these notes are mostly simple tests that you are following the material. Some will be used for quizzes in lectures. **Doing the others will help you to review your notes.**

Optional questions and extras. The 'Bonus question' at the end of each problem sheet, any 'optional' questions, and any 'extras' in these notes are included for interest only, and to show you some mathematical ideas beyond the scope of this course. You should not worry if you find them difficult.

If you can do the compulsory questions on problem sheets, know the definitions and main results from lectures, and can prove the results whose proofs are marked as examinable in these notes, then you should do very well in the examination.

(A) Vectors in $\mathbb{R}^3$

1. Introduction to vectors

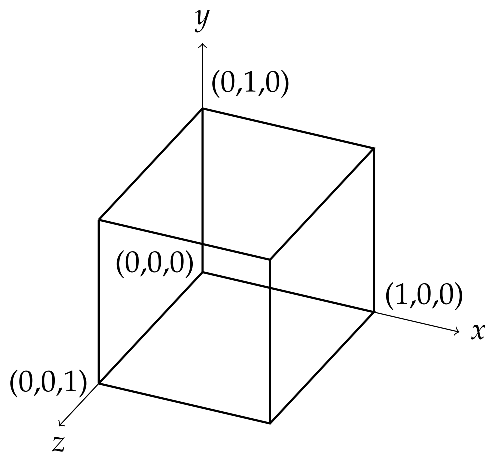
Lecture 1

VECTORS AND DISPLACEMENT VECTORS.

Definition 1.1. $\mathbb{R}^3$ is the set of all *ordered triples* (x, y, z) of real numbers.

The notation (x, y, z) means that we care about the order of the entries. For example $(1, 2, 3) \neq (2, 1, 3)$. Ordered triples are not sets.

We think of $(x, y, z) \in \mathbb{R}^3$ as the point in three-dimensional space with *coordinates* x, y and z in the x, y and z directions. For example, the diagram below shows a cube with one vertex at the *origin* $O = (0, 0, 0)$.



We can also write elements of $\mathbb{R}^3$ as *vectors* in column notation. As vectors, $(0, 0, 0)$, $(1, 0, 0)$, $(0, 0, 1)$ and $(0, 0, 1)$ are

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{i} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{j} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{k} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

Vectors are written in bold; when handwritten, they are underlined. We will see a more general definition of vectors later in the course.

Exercise 1.2. Label each vertex of the cube by the corresponding vector.

Suppose that A and B are distinct points in $\mathbb{R}^3$. Starting from A we can walk to B. The *displacement vector* $\overrightarrow{AB}$, gives the distance we move in each coordinate direction. For example if $A = (1, 2, 3)$ and $B = (3, 2, 1)$ then

$$A = (1,2,3) \xrightarrow{\overrightarrow{AB} = \begin{pmatrix} 2 \\ 0 \\ -2 \end{pmatrix}} (3,2,1) = B$$

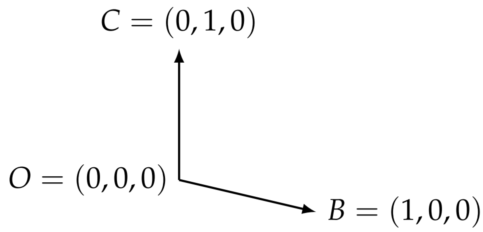
since we move $3 - 1 = 2$ in the x-direction, $2 - 2 = 0$ in the y-direction and $1 - 3 = -2$ in the z-direction.

Problem 1.3.

(a) What is the displacement vector from $(1,0,0)$ to $(0,0,1)$?
(b) If we apply this displacement starting at $(1,1,0)$, where do we finish?
(c) If we finish at $(12,-3,2)$ after applying this displacement, where must we have started?
(d) What is the displacement vector from the origin O to $(1,1,0)$?

VECTOR SUM AND PARALLELOGRAM RULE.

Problem 1.4. Let $B = (1,0,0)$ and let $C = (0,1,0)$. Start at the origin O and apply the displacement vector $\overrightarrow{OB}$. Where do we finish? Now apply $\overrightarrow{OC}$. Let D be the finishing point. Find D.



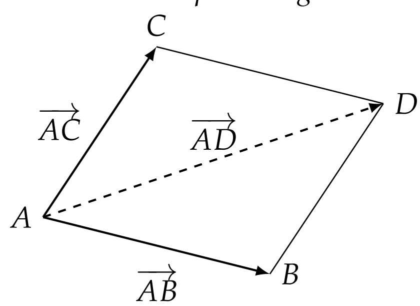
$$\overrightarrow{OB} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \overrightarrow{OC} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

This motivates the following definition.

Definition 1.5. Let $\mathbf{u} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ be vectors. We define the *sum* of $\mathbf{u}$ and $\mathbf{v}$ by

$$\mathbf{u} + \mathbf{v} = \begin{pmatrix} a+x \\ b+y \\ c+z \end{pmatrix}.$$

Generalizing Problem 1.4, let A, B, C, D be points. If we start at A, and apply the displacement vectors $\overrightarrow{AB}$ then $\overrightarrow{AC}$, we end up at D where $\overrightarrow{AD} = \overrightarrow{AB} + \overrightarrow{AC}$. This is called the *parallelogram rule*.



Exercise 1.6. Which displacement vectors label the sides BD and CD of the parallelogram? What is $\overrightarrow{AB} + \overrightarrow{BA}$? What is $\overrightarrow{AB} + \overrightarrow{BC} + \overrightarrow{CD} + \overrightarrow{DA}$?

SCALAR MULTIPLICATION AND LINEAR COMBINATIONS. Elements of $\mathbb{R}$ are called *scalars*.

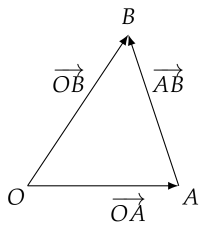
Definition 1.7. Let $\mathbf{v}$ be as in Definition 1.6, and let $\alpha \in \mathbb{R}$. We define the *scalar multiplication* of α and $\mathbf{v}$ by $\alpha\mathbf{v} = \begin{pmatrix} \alpha x \\ \alpha y \\ \alpha z \end{pmatrix}$.

Using this we can perform more complicated vector computations.

Lecture 2

Example 1.8. Let $A = (1, 0, 1)$ and $B = (1, 2, 3)$ and let $\mathbf{u}$ and $\mathbf{v}$ be the corresponding vectors. Then

$$2\mathbf{u} - 3\mathbf{v} = 2\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - 3\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -1 \\ -6 \\ -7 \end{pmatrix} = -\begin{pmatrix} 1 \\ 6 \\ 7 \end{pmatrix}$$



Exercise: Find $\mathbf{v} - \mathbf{u}$. Note that $\mathbf{v} - \mathbf{u} = \overrightarrow{OB} - \overrightarrow{OA} = \overrightarrow{AB}$. Find α such that the z-coordinate of $\alpha\mathbf{u} + \mathbf{v}$ is zero.

A sum of the form $\alpha\mathbf{u} + \beta\mathbf{v}$ is called a *linear combination* of the vectors $\mathbf{u}$ and $\mathbf{v}$. More generally, $\alpha_1\mathbf{u}_1 + \cdots + \alpha_r\mathbf{v}_r$ is a linear combination of the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$.

Problem 1.9. Let $\mathbf{u}$ and $\mathbf{v}$ be as in Example 1.8. Express $\mathbf{u}$ as a linear combination of the vectors $\mathbf{i}$, $\mathbf{j}$, $\mathbf{k}$. Express $\mathbf{v}$ as a linear combination of $\mathbf{i}, \mathbf{i} + \mathbf{j}, \mathbf{k}$.

Problem 1.10. There is a unique plane Π containing $(0, 0, 0)$, $(1, -1, 0)$ and $(0, 1, -1)$. Is $(1, 0, -1)$ in this plane? Is $(1, 1, -3)$ in this plane?

DOT PRODUCT. You might have seen another definition of vector, as something having both magnitude and direction. **Exercise:** criticize this definition.

Definition 1.11. The *length* of a vector $\mathbf{v} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is $||v|| = \sqrt{x^2 + y^2 + z^2}$.

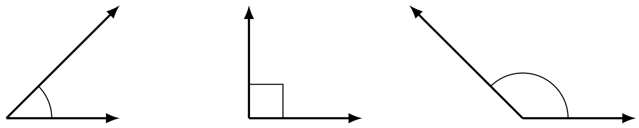
The notation $|v|$ is also used. This can be confused with the absolute value of a real number, or the modulus of a complex number, so $||v||$ is probably better.

Example 1.12. Consider the cube shown on page 4. The length of a diagonal across a face, and the length of a space diagonal are

$$\left|\left|\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}\right|\right| = \sqrt{2} \quad \text{and} \quad \left|\left|\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}\right|\right| = \sqrt{3}.$$

How can we compute the angle between two non-zero vectors?

By convention, the angle is between 0 and π. It is $\pi/2$ if and only if the vectors are *orthogonal*.



Example 1.13. We will find the angle between $\sqrt{3}\mathbf{i}+\mathbf{j}$ and $\sqrt{3}\mathbf{i}-\mathbf{j}$ by considering the triangle with vertices at $(0,0,0)$, $(\sqrt{3},1,0)$ and $(\sqrt{3},-1,0)$.

The key property that we used was $||\sqrt{3}\mathbf{i}+\mathbf{j}|| = ||\sqrt{3}\mathbf{i}-\mathbf{j}|| = 2$, so the triangle is isosceles. We can reduce to this case by scaling each vector.

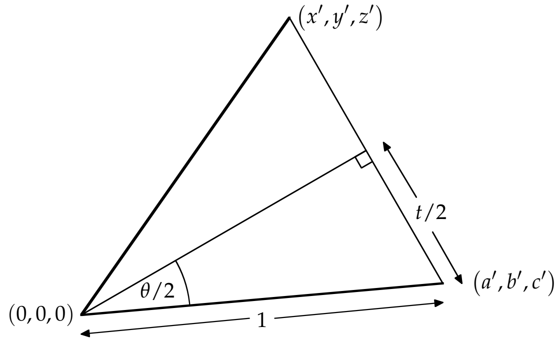
Exercise 1.14. Show that if $\mathbf{v} \in \mathbb{R}^3$ and $\alpha \in \mathbb{R}$ then $||\alpha\mathbf{v}|| = |\alpha|\,||\mathbf{v}||$. Deduce that $\mathbf{v}/||\mathbf{v}||$ has length 1. (Such vectors are said to be *unit vectors*.)

Lecture 3

Theorem 1.15. *Let* $\mathbf{u} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ *and let* $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}$ *be non-zero vectors. Let* θ *be the angle between* $\mathbf{u}$ *and* $\mathbf{v}$. *Then*

$$\cos\theta = \frac{ax+by+cz}{||\mathbf{u}||||\mathbf{v}||}.$$

The diagram used in the proof is below: $\widehat{\mathbf{u}} = \mathbf{u}/||\mathbf{u}|| = a'\mathbf{i} + b'\mathbf{j} + c'\mathbf{k}$ and $\widehat{\mathbf{v}} = \mathbf{v}/||\mathbf{v}|| = x'\mathbf{i} + y'\mathbf{j} + z'\mathbf{k}$.



This motivates the following definition.

Definition 1.16. Let $\mathbf{u} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ and $\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} \in \mathbb{R}^3$. The *dot product* of $\mathbf{u}$ and $\mathbf{v}$ is

$$\mathbf{u}\cdot\mathbf{v} = ax + by + cz.$$

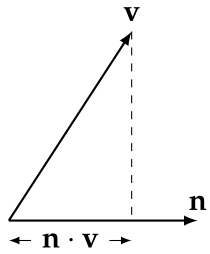
PROPERTIES OF THE DOT PRODUCT.

Lemma 1.17. *Let* $\mathbf{n}, \mathbf{v} \in \mathbb{R}^3$ **be non-zero vectors [corrected in lecture]**. *Let* θ *be the angle between* $\mathbf{n}$ *and* $\mathbf{v}$.

(a) $\mathbf{v}\cdot\mathbf{v} = ||\mathbf{v}||^2$.
(b) $\mathbf{n}\cdot\mathbf{v} = ||\mathbf{n}||\,||\mathbf{v}||\cos\theta$
(c) $\mathbf{n}\cdot\mathbf{v} = 0$ *if and only if* $\mathbf{n}$ *and* $\mathbf{v}$ *are orthogonal.*

(d) *Suppose that* $||\mathbf{n}|| = 1$. *Let* $\mathbf{w} = (\mathbf{n} \cdot \mathbf{v})\mathbf{n}$. *Then* $\mathbf{v} = \mathbf{w} + (\mathbf{v} - \mathbf{w})$ *where* $\mathbf{w}$ *is parallel to* $\mathbf{n}$ *and* $\mathbf{v} - \mathbf{w}$ *is orthogonal to* $\mathbf{n}$.

(e) *Let* $\mathbf{u} \in \mathbb{R}^3$ *and let* $\alpha, \beta \in \mathbb{R}$. *Then* $\mathbf{n} \cdot (\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha\mathbf{n} \cdot \mathbf{u} + \beta\mathbf{n} \cdot \mathbf{v}$.



In words, (d) says that the component of $\mathbf{v}$ in the direction of $\mathbf{n}$ has length $\mathbf{n} \cdot \mathbf{v}$. Note $||\mathbf{n}|| = 1$ is required. (You can always scale $\mathbf{n}$.)

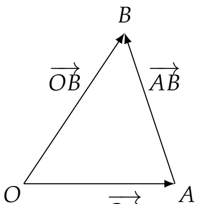
Example 1.18. Let ℓ be the line through the points $A = (1, 0, 0)$ and $B = (0, 0, 1)$. Let ℓ' be a line with direction $\mathbf{i} + \sqrt{2}\mathbf{j} - \mathbf{k}$ intersecting ℓ. The angle between ℓ and ℓ' does not depend on where the intersection is. Since the direction of ℓ is $\overrightarrow{AB} = -\mathbf{i} + \mathbf{k}$, the angle θ is determined by

$$\cos\theta = \frac{(-\mathbf{i} + \mathbf{k}) \cdot (\mathbf{i} + \sqrt{2}\mathbf{j} - \mathbf{k})}{||-\mathbf{i} + \mathbf{k}||\,||\mathbf{i} + \sqrt{2}\mathbf{j} - \mathbf{k}||} = \frac{-2}{\sqrt{2} \times 2} = -\frac{\sqrt{2}}{2}.$$

So the obtuse angle is $3\pi/4$ and the angle we want is $\pi/4$.

θ	0	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$
$\sin\theta$	0	$\frac{1}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{\sqrt{3}}{2}$	1
$\cos\theta$	1	$\frac{\sqrt{3}}{2}$	$\frac{\sqrt{2}}{2}$	$\frac{1}{2}$	0

You need to know the standard values of sine and cosine. The table in the margin may be a useful reminder.



Part (a) of Lemma 1.17 is surprisingly useful. As an application of (a) and (b) we prove the cosine rule. We need Example 1.8: $\overrightarrow{OB} - \overrightarrow{OA} = \overrightarrow{AB}$.

Lecture 4

Exercise 1.19. Let $A, B \in \mathbb{R}^3$ be points such that OAB is a triangle. Let $\mathbf{u} = \overrightarrow{OA}$ and $\mathbf{v} = \overrightarrow{OB}$ and let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Find the length of the side AB by using (a) to compute

$$||\overrightarrow{AB}||^2 = ||\overrightarrow{OB} - \overrightarrow{OA}||^2 = ||\mathbf{v} - \mathbf{u}||^2.$$

A more challenging exercise is to use (d) and (e) in Lemma 1.17 to show that the altitudes of a triangle meet at a point. This is a bonus question on Sheet 2.

LINES AND PLANE ANGLE AND INTERSECTION. At A-level you saw that the plane through $\mathbf{a} \in \mathbb{R}^3$ with normal direction $\mathbf{n} \in \mathbb{R}^3$ is $\{\mathbf{v} \in \mathbb{R}^3 : \mathbf{n} \cdot \mathbf{v} = \mathbf{n} \cdot \mathbf{a}\}$.

Problem 1.20. What is the angle θ between the planes

$$\{\mathbf{v} \in \mathbb{R}^3 : (-\mathbf{i} + \mathbf{j} + 2\mathbf{k}) \cdot \mathbf{v} = 1\} \text{ and } \{\mathbf{v} \in \mathbb{R}^3 : (-\mathbf{i} + \mathbf{k}) \cdot \mathbf{v} = -1\}?$$

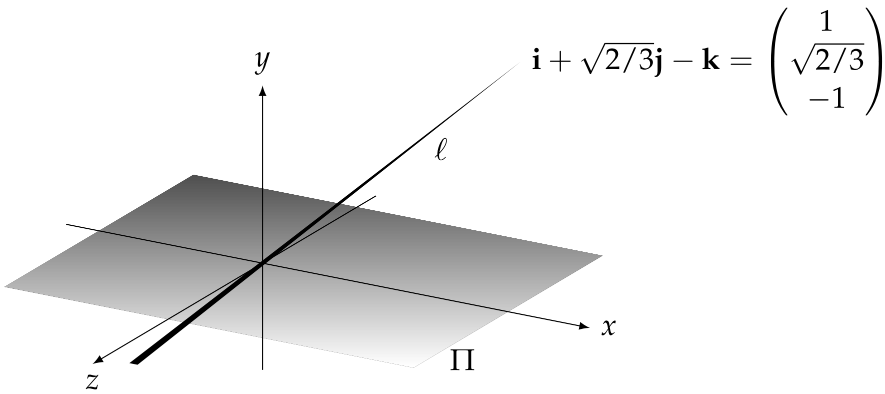
A line always makes an angle between 0 and $\pi/2$ (a right angle) with a plane.

Lemma 1.21. *Let Π be a plane with normal* $\mathbf{n}$. *Let ℓ be a line with direction* $\mathbf{c}$ *meeting Π at a unique point. The angle θ between Π and ℓ satisfies*

$$\sin\theta = |\widehat{\mathbf{n}} \cdot \widehat{\mathbf{c}}|$$

where $\widehat{\mathbf{n}} = \mathbf{n}/||\mathbf{n}||$ *and* $\widehat{\mathbf{c}} = \mathbf{c}/||\mathbf{c}||$.

Problem 1.22. Let Π be the xz-plane, so $\Pi = \{x\mathbf{i} + z\mathbf{k} : x, z \in \mathbb{R}\}$. Let ℓ be the line passing through $\mathbf{0}$ and $\mathbf{i} + \sqrt{2/3}\mathbf{j} - \mathbf{k}$. What are the minimum and maximum angles between ℓ and a line in Π passing through $\mathbf{0}$?



Where does Π meet the line through $(-1, -1, 0)$ and $(0, 1, 1)$? (You can use the same method for Question 4(b) on Sheet 1.)

2. The vector product

MOTIVATION. To solve Problem 1.22 we needed a vector orthogonal to the xz-plane. In this case the normal vector $\mathbf{j}$ was obvious. **Lecture 5**

Problem 2.1. Let Π be the plane containing $\mathbf{i}$, $2\mathbf{i} + \mathbf{j}$ and $4\mathbf{i} + 2\mathbf{j} + \mathbf{k}$. Find a normal vector to Π.

By Lemma 1.17(c), $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$ is orthogonal to both $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and $\begin{pmatrix} r \\ s \\ t \end{pmatrix}$ if and only if

$$\begin{aligned} ax + by + cz &= 0 \\ rx + sy + tz &= 0. \end{aligned}$$

Multiply the first equation by t, the second by c and subtract to get

$$(at - cr)x + (bt - cs)y = 0.$$

This suggests we might take $x = bt - cs$ and $y = cr - at$. Substituting in we find that both equations hold when $z = as - br$.

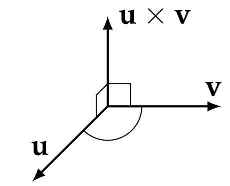
DEFINITION AND PROPERTIES OF THE VECTOR PRODUCT.

Definition 2.2. The *vector product* of vectors $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ and $\begin{pmatrix} r \\ s \\ t \end{pmatrix}$ is defined by

$$\begin{pmatrix} a \\ b \\ c \end{pmatrix} \times \begin{pmatrix} r \\ s \\ t \end{pmatrix} = \begin{pmatrix} bt - cs \\ cr - at \\ as - br \end{pmatrix}.$$

If you remember the top entry is $bt - cs$ you can obtain the others using the cyclic permutations $a \mapsto b \mapsto c \mapsto a$ and $r \mapsto s \mapsto t \mapsto r$. Sometimes $\wedge$ is used rather than $\times$.

Exercise 2.3.
(a) Show that $\mathbf{v} \times \mathbf{v} = \mathbf{0}$ for all $\mathbf{v} \in \mathbb{R}^3$.
(b) Show that $\mathbf{i} \times \mathbf{j} = \mathbf{k}$, $\mathbf{j} \times \mathbf{k} = \mathbf{i}$ and $\mathbf{k} \times \mathbf{i} = \mathbf{j}$. Hence find $(\mathbf{i} + \mathbf{j}) \times (\mathbf{i} - \mathbf{k})$.



Note that in each case, if $\mathbf{u} \times \mathbf{v} = \mathbf{w}$ then the vectors $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ form a right-handed system with (as expected) $\mathbf{w}$ orthogonal to $\mathbf{u}$ and $\mathbf{v}$. Changing $\mathbf{u}$ and $\mathbf{v}$ by a small displacement does not change the orientation of the system, so the system is always right-handed.

Lecture 6

It remains to find the length of $\mathbf{u} \times \mathbf{v}$.

Theorem 2.4. *Let $\mathbf{u}$ and $\mathbf{v}$ be non-zero vectors. Let θ be the angle between $\mathbf{u}$ and $\mathbf{v}$. Then*

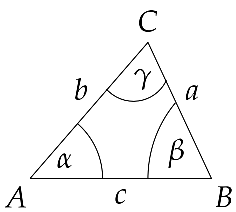
$$||\mathbf{u} \times \mathbf{v}|| = ||\mathbf{u}||\,||\mathbf{v}|| \sin\theta.$$

AREA OF TRIANGLES. The identity

$$||\mathbf{u} \times \mathbf{v}||^2 = ||\mathbf{u}||^2||\mathbf{v}||^2 - (\mathbf{u} \cdot \mathbf{v})^2$$

seen in the proof of Theorem 2.4 is useful and of independent interest.

Problem 2.5. Let $B = (1, 8, 2)$ and $C = (8, 2, 1)$. Let $\overrightarrow{OD} = \overrightarrow{OB} + \overrightarrow{OC}$.
(a) What is the area of the parallelogram $OBDC$?
(b) What is the area of the triangle OBC?



Suppose that ABC is a triangle with sides and angles as shown in the margin. Using Theorem 2.4 and the argument for (b), its area is

$$\tfrac{1}{2}||\overrightarrow{AB}||\,||\overrightarrow{AC}|| \sin\alpha = \tfrac{1}{2}bc\sin\alpha.$$

Repeating this argument with the other sides gives $\frac{1}{2}bc\sin\alpha = \frac{1}{2}ca\sin\beta = \frac{1}{2}ab\sin\gamma$. Now divide through by $\frac{1}{2}abc$ to get the sine rule.

Lecture 7

EQUATION OF A PLANE. If Π is a plane and $\mathbf{n}$ is a normal vector to Π then we have already used that

$$\Pi = \{\mathbf{v} \in \mathbb{R}^3 : \mathbf{n} \cdot \mathbf{v} = \alpha\}$$

for some $\alpha \in \mathbb{R}$. (This will be proved later in the course.) We can find $\mathbf{v}$ and α using the vector product.

Example 2.6. Let Π be the plane through $A = (1,2,3)$, $B = (3,1,2)$ and $C = (2,3,1)$. Then

$$\Pi = \{\overrightarrow{OA} + \lambda\overrightarrow{AB} + \mu\overrightarrow{AC} : \lambda, \mu \in \mathbb{R}\}.$$

(Compare Problem 1.10, where we could take $A = O$, so $\overrightarrow{OA} = \mathbf{0}$.) A normal vector is

$$\overrightarrow{AB} \times \overrightarrow{AC} = \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \\ -2 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix}.$$

We can scale the normal vector as we like, so dividing by 3, we take $\mathbf{n} = \mathbf{i} + \mathbf{j} + \mathbf{k}$. Since

$$\mathbf{n} \cdot \overrightarrow{OA} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = 6$$

we have $\Pi = \{\mathbf{v} \in \mathbb{R}^3 : (\mathbf{i} + \mathbf{j} + \mathbf{k}) \cdot \mathbf{v} = 6\}$.

INTERSECTION OF PLANES.

Problem 2.7. Let Π be the plane through $\mathbf{0}$ with normal $\mathbf{i} + \mathbf{j}$. Let Π' be the plane through $\mathbf{j}$ with normal $\mathbf{i} + 2\mathbf{j} + \mathbf{k}$. Let $\ell = \Pi \cap \Pi'$. Find $\mathbf{a}$ and $\mathbf{c} \in \mathbb{R}^3$ such that

$$\ell = \{\mathbf{a} + \lambda\mathbf{c} : \lambda \in \mathbb{R}\}.$$

SHORTEST DISTANCES.

Problem 2.8. Let $A = (2,3,1)$. What is the shortest distance between A and a point P on the plane Π through $\mathbf{0}$ with normal direction $\mathbf{i} + \mathbf{j} + \mathbf{k}$?

The solution using Lemma 1.17(d) needed that $\mathbf{0} \in \Pi$. One can always translate the plane and point to reduce to this case: see Question 4(b) on Sheet 2.

Recall that the line through $\mathbf{a}$ with direction $\mathbf{c}$ is $\{\mathbf{a} + \lambda\mathbf{c} : \lambda \in \mathbb{R}\}$.

Lecture 8

Problem 2.9. Let ℓ be the line through $-\mathbf{j}$ and $\mathbf{i} + \mathbf{j} + \mathbf{k}$.

(a) Does ℓ meet the line through $\mathbf{0}$ with direction $\mathbf{i} + \mathbf{k}$? **[Corrected from $\mathbf{i} - \mathbf{j}$.]**

(b) Does ℓ meet the line ℓ' through $\mathbf{0}$ with direction $\mathbf{i} + \mathbf{j}$?

(c) What is the shortest distance between ℓ and ℓ'?

The argument used for (c) generalizes to prove the following lemma.

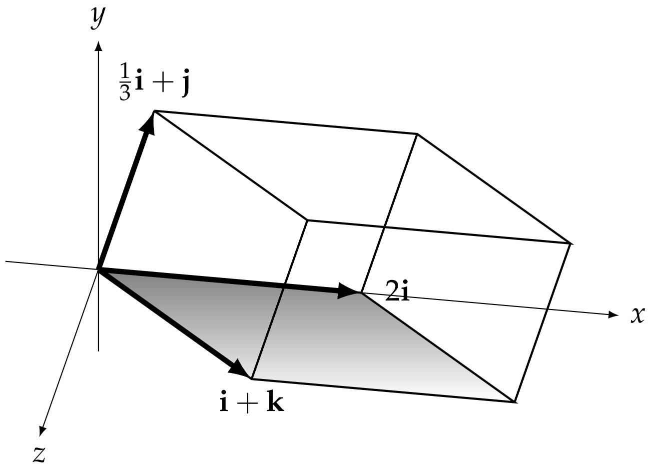
Lemma 2.10. *Let ℓ be the line through* $\mathbf{a}$ *with direction* $\mathbf{c}$*. Let ℓ' be the line through* $\mathbf{a}'$ *with direction* $\mathbf{c}'$*. The shortest distance between ℓ and ℓ' is*

$$\Big|\frac{\mathbf{c}\times\mathbf{c}'}{||\mathbf{c}\times\mathbf{c}'||}\cdot(\mathbf{a}'-\mathbf{a})\Big|.$$

In particular, ℓ and ℓ' intersect if and only if $(\mathbf{c}\times\mathbf{c}')\cdot(\mathbf{a}'-\mathbf{a}) = 0$. This motivates the following section.

THE SCALAR TRIPLE PRODUCT.

Problem 2.11. What is the volume of the parallelepiped formed by the vectors $\mathbf{i}+\mathbf{k}, 2\mathbf{i}, \frac{1}{3}\mathbf{i}+\mathbf{j}$, shown in the figure below?



More generally there is the following result.

Lecture 9

Theorem 2.12. *Let* $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$*. The volume of the parallelepiped formed by* $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^3$ *is* $|(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}|$.

Exercise 2.13. Deduce from Theorem 2.12 that

$$(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c} = (\mathbf{b}\times\mathbf{c})\cdot\mathbf{a} = (\mathbf{c}\times\mathbf{a})\cdot\mathbf{b}.$$

What is the geometric interpretation of the sign of $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$?

The scalar $(\mathbf{a}\times\mathbf{b})\cdot\mathbf{c}$ is called the *scalar triple product* of $\mathbf{a}, \mathbf{b}, \mathbf{c}$.

(B) Introduction to matrices

3. Matrices and vectors

VECTORS IN ANY DIMENSION. We now generalize the definition of 'vector' to mean an element of $\mathbb{R}^n$, for any $n \in \mathbb{N}$. If $\mathbf{u} \in \mathbb{R}^n$ we write u_i for the ith coordinate of $\mathbf{u}$. As usual vectors are written in column form.

Definition 3.1. The *length* of $\mathbf{u} \in \mathbb{R}^n$ is defined by

$$||\mathbf{u}|| = \sqrt{u_1^2 + \cdots + u_n^2}.$$

The *dot product* of vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ is defined by

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n.$$

The *sum* of vectors $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^n$ and the *scalar multiplication* of $\mathbf{v} \in \mathbb{R}^n$ by $\alpha \in \mathbb{R}$ are defined by the obvious generalization of Definition 1.7.

For example if $P = (0,1,2,3)$ and $Q = (3,2,1,0) \in \mathbb{R}^4$ and $\mathbf{u} = \overrightarrow{OP}$, $\mathbf{v} = \overrightarrow{OQ}$, then $v_1 = 3$, $v_2 = 2$, $v_3 = 1$, $v_4 = 0$ and

$$\mathbf{u} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix}, \ \mathbf{v} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix}, \mathbf{v} - \mathbf{u} = \begin{pmatrix} 3 \\ 2 \\ 1 \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ -1 \\ -3 \end{pmatrix}, \ 2\mathbf{v} = \begin{pmatrix} 0 \\ 2 \\ 4 \\ 6 \end{pmatrix}.$$

All the properties of the dot product proved in Lemma 1.17 hold in any dimension. In particular $||\mathbf{u}||^2 = \mathbf{u} \cdot \mathbf{u}$ for any vector $\mathbf{u}$, and if θ is the angle between $\mathbf{u}$ and $\mathbf{v}$ then $||\mathbf{u}|| \, ||\mathbf{v}|| \cos\theta = \mathbf{u} \cdot \mathbf{v}$.

Problem 3.2. Let $S = \big\{(x,y,z,w) : x,y,z,w \in \{0,1\}\big\} \subseteq \mathbb{R}^4$ be the set of vertices of the hypercube of side 1.

(a) What is the angle between $(1,1,0,0)$ and $(0,1,1,0)$?
(b) What is the angle between $(1,1,0,0)$ and $(0,0,1,1)$?
(c) What is the set of distances between points in S?

Shapes in high dimensions can be unintuitive. For instance, the maximum distance between two vertices in an n-dimensional cube with side length 1 is $\sqrt{n}$: it is maybe surprising that this tends to infinity with n.

Part of the power of algebra is that it lets us reason about these shapes without having to visualize them.

MATRICES. **Lecture 10**

Definition 3.3. Let m, $n \in \mathbb{N}$. An $m \times n$ *matrix* is an array with m rows and n columns of real numbers.

Matrices are useful simply as containers.

Problem 3.4. The matrix below records the stock prices of British Land, Glencore and Whitbread in the first weeks of 2012 and 2016, all in pence.[1]

$$\begin{array}{c} 2012 \\ 2016 \end{array} \begin{pmatrix} 360 & 395 & 1589 \\ 741 & 77 & 4130 \end{pmatrix}.$$

Suppose an investor has a portfolio consisting of $(5, 2, 1)$ units of each stock. What is her portfolio worth in 2012? In 2016?

LINEAR MAPS FROM $\mathbb{R}^2$ TO $\mathbb{R}^2$ AND 2×2 MATRICES.

Definition 3.5. A *linear map* from $\mathbb{R}^2$ to $\mathbb{R}^2$ is a function $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ of the form

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

where a, b, c, d are some fixed real numbers.

Problem 3.6.

(a) Take $a = 0$, $b = 1$, $c = 1$, $d = 0$. We obtain the linear map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} y \\ x \end{pmatrix}.$$

What, geometrically, does this linear map do to a vector $\mathbf{v} \in \mathbb{R}^2$?

(b) Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined so that $f\begin{pmatrix} x \\ y \end{pmatrix}$ records the values, in 2012 and 2016, of a portfolio consisting of x Glencore shares and y Whitbread shares. Show that f is a linear map. What are the coefficients a, b, c, d?

The coefficients a, b, c, d in a linear map f can be recorded in a 2×2 matrix. We define the product of a 2×2 matrix and vector in $\mathbb{R}^2$ by

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

By definition of the product, if $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ then $\begin{pmatrix} x \\ y \end{pmatrix} \overset{f}{\mapsto} A \begin{pmatrix} x \\ y \end{pmatrix}$.

[1] British Land is a real estate investment trust: they own the 'Cheesegrater'; Glencore is a mining conglomerate; Whitbread owns Costa Coffee, Premier Inn, and other brands.

Example 3.7. The matrix $\begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix}$ represents the linear map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} 2x + y \\ 3x + y \end{pmatrix}$$

Note that the entries a, b in the first row of the matrix appear in the first row of the result, and similarly for c and d.

Problem 3.8. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be defined so that $f(\mathbf{v})$ is $\mathbf{v}$ rotated by $\pi/4$. (Remember that a positive angle means an anticlockwise rotation.) For example,

$$f \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\sqrt{2}}{2} \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Find a formula for $f \begin{pmatrix} x \\ y \end{pmatrix}$ and show that f is a linear map.

The solution using complex numbers to this problem generalizes to show that rotation by θ is the linear map

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

You are asked to prove this on Question 2 on Problem Sheet 4.

Lecture 11

MATRIX MULTIPLICATION FOR 2×2 MATRICES. In Problem 3.6(a), suppose that two investors have portfolios $\begin{pmatrix} r \\ t \end{pmatrix}$ and $\begin{pmatrix} s \\ u \end{pmatrix}$. Using approximate prices in pounds, the values of their portfolios in 2012 and 2016 are

$$\begin{pmatrix} 4 & 16 \\ 1 & 40 \end{pmatrix} \begin{pmatrix} r \\ t \end{pmatrix} = \begin{pmatrix} 4r + 16t \\ r + 40t \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 4 & 16 \\ 1 & 40 \end{pmatrix} \begin{pmatrix} s \\ u \end{pmatrix} = \begin{pmatrix} 4s + 16u \\ s + 40u \end{pmatrix}.$$

It would be more convenient if we could compute all four values in one operation.

Definition 3.9. We define the *product* of 2×2 matrices $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $\begin{pmatrix} r & s \\ t & u \end{pmatrix}$ by

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} r & s \\ t & u \end{pmatrix} = \begin{pmatrix} ar + bt & as + bu \\ cr + dt & cs + du \end{pmatrix}.$$

For example,

$$\begin{pmatrix} 4 & 16 \\ 1 & 40 \end{pmatrix} \begin{pmatrix} r & s \\ t & u \end{pmatrix} = \begin{pmatrix} 4r + 16t & 4s + 16u \\ r + 40t & s + 40u \end{pmatrix}$$

agreeing with the matrix times vector calculations above.

Example 3.10. Let $\theta, \phi \in \mathbb{R}$. The matrix representing rotation by $\theta + \phi$ is

$$M_{\theta+\phi} = \begin{pmatrix} \cos(\theta+\phi) & -\sin(\theta+\phi) \\ \sin(\theta+\phi) & \cos(\theta+\phi) \end{pmatrix}.$$

We can rotate by $\theta + \phi$ by first rotating by ϕ, then by θ. The corresponding matrices are

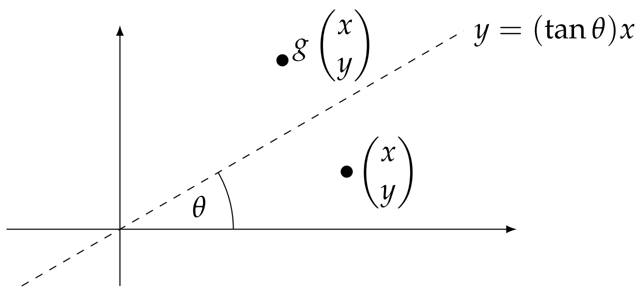
$$M_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{and} \quad M_\phi = \begin{pmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{pmatrix}$$

Exercise: Check that $M_{\theta+\phi} = M_\theta M_\phi$.

More generally, we have the following lemma.

Lemma 3.11. *Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ and $g : \mathbb{R}^2 \to \mathbb{R}^2$ be linear maps represented by the matrices A and B, respectively. Then $gf : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map, represented by the matrix BA.*

Problem 3.12. Let $\theta \in \mathbb{R}$. Let $g : \mathbb{R}^2 \to \mathbb{R}^2$ be defined so that $g(\mathbf{v})$ is $\mathbf{v}$ reflected in the line $y = (\tan\theta)x$. Show that g is a linear map and find the matrix representing g.



In the special case $\theta = 0$, we reflect in the x-axis, and so

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

We can reduce the general case to this by using matrix multiplication to 'conjugate' by a rotation by $-\theta$.

Lecture 12

INVERSES OF LINEAR MAPS AND 2×2-MATRICES. Recall that if S and T are sets and $f : S \to T$ is a bijective function then the *inverse* of f is the function $f^{-1} : T \to S$ defined by $f^{-1}(t) = s \iff f(s) = t$.

For example, the inverse of rotation by θ is rotation by $-\theta$. A reflection is its own inverse. But the linear map

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

is not bijective, so has no inverse.

Lemma 3.13. *The linear map $f : \mathbb{R}^2 \to \mathbb{R}^2$ represented by the matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. In this case the inverse $f^{-1} : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear map represented by the matrix $\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.*

We say that $\frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ is the *inverse* of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$. This will be related to inverses in the sense of rings in the following section.

DETERMINANTS OF 2×2 MATRICES. The quantity $ad - bc$ appearing in Lemma 3.12 has a geometric interpretation.

Problem 3.14. Let $\beta > 0$ be a real number. Let f, g, and $h : \mathbb{R}^2 \to \mathbb{R}^2$ be rotation by $\pi/4$, a stretch of β in the x-direction, and reflection in the x axis.

(a) What is the matrix representing hgf?
(b) What is the image of the square with vertices at $(0,0)$, $(1,0)$, $(0,1)$ and $(1,1)$ under hgf?
(c) How does the area change?

More generally we have the following lemma. Perhaps surprisingly, the shortest proof for us will use vectors in $\mathbb{R}^3$.

Lemma 3.15. *The image of the unit square under the linear transformation represented by $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ has area $|ad - bc|$. Moreover, $ad - bc > 0$ if and only if the vectors*

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

form a right-handed system.

We define the *determinant* of $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ by $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = ad - bc$.

By Lemma 3.13, a 2×2 matrix M is invertible if and only if $\det M \neq 0$.

Let A and B be 2×2-matrices. By Lemma 3.14, when we apply AB to the unit square, the image has signed area $\det AB$. On the other hand, when we apply B the signed area is multiplied by $\det B$, and when we apply A the signed area is multiplied by $\det A$. Hence

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

You are asked to give an algebraic proof of this on Sheet 4.

The determinant gives a convenient criterion for a matrix to send a non-zero vector to $\mathbf{0}$.

Lecture 13

Lemma 3.16. *Let M be a 2×2-matrix. There is a non-zero vector $\mathbf{v}$ such that $M\mathbf{v} = \mathbf{0}$ if and only if $\det M = \mathbf{0}$.*

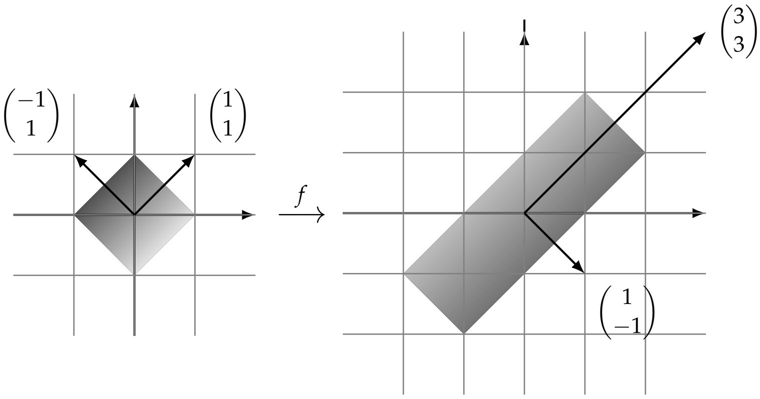
EIGENVECTORS AND EIGENVALUES OF 2×2-MATRICES. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the linear map defined by

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

for $x, y \in \mathbb{R}$. Observe that

$$f\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} = 3\begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad f\begin{pmatrix} 1 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} = -\begin{pmatrix} 1 \\ -1 \end{pmatrix}.$$

The diagram below shows the images of these special vectors.



Thus f is a stretch by 3 in the direction $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ followed by a reflection in the line with this direction.

Definition 3.17. Let A be a 2×2-matrix. We say that a non-zero vector $\mathbf{v} \in \mathbb{R}^2$ is an *eigenvector* of A with *eigenvalue* λ if $A\mathbf{v} = \lambda\mathbf{v}$.

For example, $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$ has eigenvectors $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ with eigenvalues 1 and -1. It is a special feature of this example that the eigenvectors are orthogonal. (This happens if and only if A is symmetric.)

Exercise: why do we require $\mathbf{v}$ to be non-zero in Definition 3.17?

The sum of 2×2-matrices and the product of a 2×2-matrix by a scalar were defined on Question 1 of Sheet 4. Let I denote the 2×2 *identity matrix* $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. We have

$$\begin{aligned} \mathbf{v} \text{ is an eigenvector of } A \text{ with eigenvalue } \lambda &\iff A\mathbf{v} = \lambda\mathbf{v} \\ &\iff (\lambda I - A)\mathbf{v} = 0 \end{aligned}$$

Hence, by Lemma 3.16,

$$A \text{ has an eigenvector with eigenvalue } \lambda \iff \det(\lambda I - A) = 0.$$

This gives a practical way to find eigenvectors and eigenvalues.

Lecture 14

Example 3.18. Let $A = \begin{pmatrix} 8 & 5 \\ -10 & -7 \end{pmatrix}$. Then

$$\begin{aligned} \det(\lambda I - A) &= \det \begin{pmatrix} \lambda - 8 & -5 \\ 10 & \lambda + 7 \end{pmatrix} \\ &= (\lambda - 8)(\lambda + 7) + 50 \\ &= \lambda^2 - \lambda - 6 \\ &= (\lambda + 2)(\lambda - 3). \end{aligned}$$

Hence A has eigenvalues -2 and 3. To find corresponding eigenvectors, we put $\lambda = -2$ and $\lambda = 3$ in the equation

$$(\lambda I - A) \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

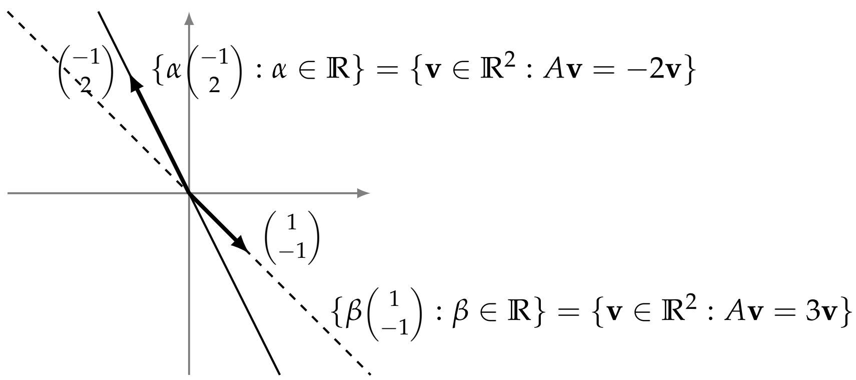
and solve for x and y. For $\lambda = -2$ the equations are

$$\begin{pmatrix} 10 & 5 \\ -10 & -5 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

with solutions $\begin{pmatrix} x \\ y \end{pmatrix} = \alpha \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ for $\alpha \in \mathbb{R}$. So $\mathbf{u} = \begin{pmatrix} -1 \\ 2 \end{pmatrix}$ is an eigenvector for the eigenvalue -2.

Exercise 3.19. Show that $\mathbf{v} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ is an eigenvector of A with eigenvalue 3.

It is no surprise that we found infinitely many solutions to both equations: if $A\mathbf{v} = \lambda \mathbf{v}$ then $A(\alpha \mathbf{v}) = \lambda(\alpha \mathbf{v})$ for any $\alpha \in \mathbb{R}$. The eigenvectors of A are the non-zero vectors on the two lines below.



Since eigenvectors are not unique, it is good style to write 'an eigenvector with eigenvalue 3', rather than 'the eigenvector with eigenvalue 3'.

DIAGONALIZATION. Let A be as in Example 3.18 and let

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

be the matrix whose columns are the eigenvectors $\mathbf{u}$ and $\mathbf{v}$ of A found above. Writing $(\mathbf{u}\ \mathbf{v})$ for the matrix with columns $\mathbf{u}$ and $\mathbf{v}$, and so on, we have

$$AP = A(\mathbf{u}\ \mathbf{v}) = (A\mathbf{u}\ \ A\mathbf{v}) = (-2\mathbf{u}\ 3\mathbf{v}) = (\mathbf{u}\ \ \mathbf{v}) \begin{pmatrix} -2 & 0 \\ 0 & 3 \end{pmatrix} = P \begin{pmatrix} -2 & 0 \\ 0 & 3 \end{pmatrix}.$$

Since $\det P = 1$, Lemma 3.13 implies that P has an inverse. Multiplying on the right by P^{-1}, we get

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

We say that A is *diagonalized* by P. (Compare the conjugation trick used in Problem 3.12.)

Problem 3.20. Compute A^{2016}.

Not all matrices can be diagonalized. A necessary and sufficient condition is that the matrix has two non-proportional eigenvectors, as defined below. (If time permits, this will be proved in Part E.)

Definition 3.21. Let $n \in \mathbb{N}$. Vectors $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^n$ are *proportional* if there exist α and $\beta \in \mathbb{R}$, not both zero, such that $\alpha\mathbf{u} = \beta\mathbf{v}$.

Thus $\mathbf{0}$ and $\mathbf{v}$ are proportional for any $\mathbf{v} \in \mathbb{R}^n$, and non-zero vectors (such as eigenvectors in $\mathbb{R}^2$) are proportional if and only if they point in the same direction.

We will prove that if a matrix has two distinct eigenvalues then it is diagonalizable. For this we need the following two lemmas.

Lemma 3.22. *Let A be a 2×2 matrix. Suppose that A has distinct real eigenvalues λ and μ with eigenvectors $\mathbf{u}$ and $\mathbf{v}$, respectively. Then $\mathbf{u}$ and $\mathbf{v}$ are not proportional.*

So eigenvectors with different eigenvalues point in different directions.

Lecture 15

Lemma 3.23. *If $\mathbf{u}$ and $\mathbf{v} \in \mathbb{R}^2$ are not proportional then the matrix $(\mathbf{u}\ \ \mathbf{v})$ with columns $\mathbf{u}$ and $\mathbf{v}$ is invertible.*

The converse of Lemma 3.23 is also true: you are asked to prove this on Sheet 5.

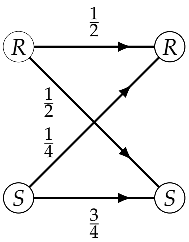
Proposition 3.24. *Let A be a 2×2 matrix. Suppose that A has distinct eigenvalues λ and μ with eigenvectors $\mathbf{u}$ and $\mathbf{v}$, respectively. Let $P = (\mathbf{u}\ \ \mathbf{v})$ be the matrix formed by these eigenvectors. Then P is invertible and*

$$A = P \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix} P^{-1}.$$

A VERY SIMPLE WEATHER MODEL. If today is sunny then tomorrow is sunny with probability $3/4$. If today is rainy then tomorrow is equally likely to be sunny and rainy.

Example 3.25. Suppose Monday is sunny. Then Tuesday is sunny with probability $3/4$. The probability that Wednesday is sunny is

$$\mathbf{P}[\text{Tuesday is sunny}]\tfrac{3}{4} + \mathbf{P}[\text{Tuesday is rainy}]\tfrac{1}{2}.$$



Since $\mathbf{P}[\text{Tuesday is rainy}] = 1 - \frac{3}{4} = \frac{1}{4}$, the probability that Wednesday is sunny is $\frac{3}{4} \times \frac{3}{4} + \frac{1}{4} \times \frac{1}{2} = \frac{9}{16} + \frac{1}{8} = \frac{11}{16}$.

More generally,

$$\mathbf{P}[\text{day } n+1 \text{ is sunny}] = \mathbf{P}[\text{day } n \text{ is sunny}]\tfrac{3}{4} + \mathbf{P}[\text{day } n \text{ is rainy}]\tfrac{1}{2}.$$

So, setting $p_n = \mathbf{P}[\text{day } n \text{ is sunny}]$, we get $p_{n+1} = \frac{3}{4}p_n + \frac{1}{2}(1 - p_n)$.

Exercise 3.26. Show that $1 - p_{n+1} = \frac{1}{4}p_n + \frac{1}{2}(1 - p_n)$.

The equations for p_{n+1} and $1 - p_{n+1}$ can be written in matrix form:

$$\begin{pmatrix} p_{n+1} \\ 1 - p_{n+1} \end{pmatrix} = A \begin{pmatrix} p_n \\ 1 - p_n \end{pmatrix} \quad \text{where } A = \begin{pmatrix} \frac{3}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix}.$$

We have $p_n = p_{n+1}$ if and only if $\begin{pmatrix} p_n \\ 1 - p_n \end{pmatrix}$ is an eigenvector of A with eigenvalue 1. Equivalently,

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = (I - A) \begin{pmatrix} p_n \\ 1 - p_n \end{pmatrix} = \begin{pmatrix} -\frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} \end{pmatrix} \begin{pmatrix} p_n \\ 1 - p_n \end{pmatrix} = \begin{pmatrix} -\frac{1}{4}p_n + \frac{1}{2}(1 - p_n) \\ \frac{1}{4}p_n - \frac{1}{2}(1 - p_n) \end{pmatrix}.$$

Hence $p_n = 2(1 - p_n)$ and $p_n = \frac{2}{3}$. So the long-term proportion of sunny days is $\frac{2}{3}$.

4. MATRICES AND LINEAR MAPS IN HIGHER DIMENSIONS

DEFINITIONS. Matrices were defined in Definition 3.3. If A is an $m \times n$ matrix we write A_{ij} for the entry of A in row i and column j. For example if A is the 2×3 matrix **Lecture 16**

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

then $A_{23} = 5$ and, more generally, $A_{ij} = i + j$ for all $i \in \{1, 2\}$ and $j \in \{1, 2, 3\}$.

Definition 4.1. Let $m, n \in \mathbb{N}$ and let A be an $m \times n$ matrix. Let $\mathbf{v} \in \mathbb{R}^n$. We define the *product* of A and $\mathbf{v}$ by

$$A \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} A_{11}v_1 + A_{12}v_2 + \cdots + A_{1n}v_n \\ A_{21}v_1 + A_{22}v_2 + \cdots + A_{2n}v_n \\ \vdots \\ A_{m1}v_1 + A_{m2}v_2 + \cdots + A_{mn}v_n \end{pmatrix}$$

Equivalently, $(A\mathbf{v})_i = \sum_{j=1}^{n} A_{ij}v_j$ for each $i \in \{1, 2, \ldots, m\}$.

Exercise 4.2. Check that Definition 4.1 generalizes the definition on page 14 of the product of a 2×2 matrix and a vector in $\mathbb{R}^2$.

We also generalize Definition 3.5.

Definition 4.3. Let $m, n \in \mathbb{N}$. A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear map* if there is an $m \times n$ matrix A such that $f(\mathbf{v}) = A\mathbf{v}$ for all $\mathbf{v} \in \mathbb{R}^n$.

Note that $\mathbb{R}^n$ is the domain of f and $\mathbb{R}^m$ is the codomain of f.

We saw earlier that if $f : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map then the matrix representing f is

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{where} \begin{pmatrix} a \\ c \end{pmatrix} = f\begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{and} \begin{pmatrix} b \\ d \end{pmatrix} = f\begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

This is true more generally: if $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear map represented by the matrix A then column j of A is the image of the vector in $\mathbb{R}^n$ with 1 in position j and zero in all other positions.

EXAMPLES.

Example 4.4. Let M be the stock price matrix in Problem 3.4. Suppose an investor has x, y and z shares in each company. The function $f : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$f\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 360 & 395 & 1589 \\ 741 & 77 & 4130 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 360x + 395y + 1589z \\ 741x + 77y + 4130z \end{pmatrix}$$

is linear; the top component is the value of the portfolio in 2012, and the bottom component is the value of the portfolio in 2016. Note that the columns of the matrix are $f(\mathbf{i})$, $f(\mathbf{j})$ and $f(\mathbf{k})$.

Example 4.5. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $f(\mathbf{v}) = (\mathbf{i} + 2\mathbf{j} + 3\mathbf{k}) \times \mathbf{v}$. We have

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

Hence f is a linear map, represented by the 3×3 matrix

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

Note that the columns of A are $f(\mathbf{i})$, $f(\mathbf{j})$ and $f(\mathbf{k})$. Now let $g : \mathbb{R}^3 \to \mathbb{R}$ be defined by $g(\mathbf{v}) = (\mathbf{i} - \mathbf{j}) \cdot \mathbf{v}$. We will show that the 1×3 matrix $B = (1 \quad -1 \quad 0)$ represents g, and answer these questions:

Lecture 17

(a) What is the matrix representing the composition gf?
(b) What is the matrix representing the composition ff?

PRODUCTS OF MATRICES.

Definition 4.6. If B is an $m \times n$ matrix and A is an $n \times p$ matrix then the *product* BA is the $m \times p$ matrix defined by

$$(BA)_{ij} = \sum_{k=1}^{n} B_{ik} A_{kj} \quad \text{for } 1 \le i \le m \text{ and } 1 \le j \le p.$$

Written out without the Sigma notation,

$$(BA)_{ij} = B_{i1}A_{1j} + B_{i2}A_{2j} + \cdots + B_{in}A_{nj}.$$

So to calculate $(BA)_{ij}$ go left-to-right along row i of B and top-to-bottom down column j of A, multiplying each pair of entries. Then take the sum.

Exercise 4.7. Continuing Example 4.5, check that the matrix $\begin{pmatrix} -3 & -3 & 3 \end{pmatrix}$ found in (a) representing gf is BA.

In each case matrix multiplication corresponds to composition of linear maps. The general proof is non-examinable, and will be skipped if time is pressing. (See Lemma 3.11 for the 2×2 case.)

Lemma 4.8. *Let $f : \mathbb{R}^p \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ be linear maps, represented by the $n \times p$ matrix A and the $m \times n$ matrix B, respectively. Then $gf : \mathbb{R}^p \to \mathbb{R}^m$ is a linear map, represented by the matrix BA.*

Proof. By definition of A, if $\mathbf{v} \in \mathbb{R}^p$ then $f(\mathbf{v}) = A\mathbf{v}$. Similarly by definition of B, if $\mathbf{w} \in \mathbb{R}^n$ then $f(\mathbf{w}) = B\mathbf{w}$. Hence

$$(gf)(\mathbf{v}) = g(f(\mathbf{v})) = g(A\mathbf{v}) = B(A\mathbf{v}) \quad \text{for all } \mathbf{v} \in \mathbb{R}^p.$$

So to show that gf is represented by BA, it is sufficient to prove that $(BA)\mathbf{v} = B(A\mathbf{v})$ for all $\mathbf{v} \in \mathbb{R}^p$. Let $1 \le i \le m$. Then

$$\begin{aligned}\big(B(A\mathbf{v})\big)_i &= \sum_{j=1}^{n} B_{ij}(A\mathbf{v})_j = \sum_{j=1}^{n} B_{ij}\sum_{k=1}^{p} A_{jk}v_k \\ &= \sum_{k=1}^{p}\big(\sum_{j=1}^{n} B_{ij}A_{jk}\big)v_k = \sum_{k=1}^{p}(BA)_{ik}v_k = \big((BA)\mathbf{v}\big)_i\end{aligned}$$

Therefore $B(A\mathbf{v}) = (BA)\mathbf{v}$, as required. □

RINGS OF MATRICES. The sum of $m \times n$ matrices B and C is defined, as you would expect, by $(B + C)_{ij} = B_{ij} + C_{ij}$. The set of all $n \times n$ matrices (for a fixed n) forms a ring, as defined in 181 Number Systems. The zero element is the all-zeros matrix, and the one element is the $n \times n$ identity matrix

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

The proposition below shows that two of the ring axioms hold. We will assume the rest.

Proposition 4.9.

(a) *If A, B, C are matrices then $C(BA) = (CB)A$, whenever either side is defined. (Associativity.)*

(b) *If A is an $n \times p$ matrix and B and C are $m \times n$ matrices then $(B + C)A = BA + CA$. (Distributivity)*

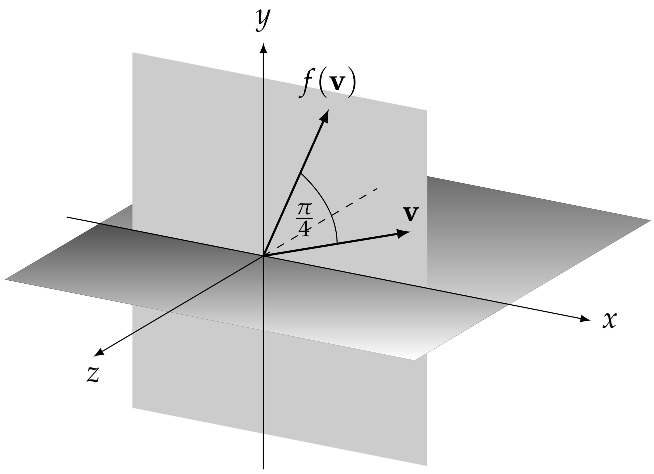
For the left-handed version of (b) see Problem Sheet 6. Since matrix multiplication is not commutative, an independent proof is needed.

Lecture 18

EIGENVECTORS AND EIGENVALUES. Eigenvectors and eigenvalues of a square matrix are defined by the expected generalization of Definition 3.17.

Definition 4.10. Let A be a $n \times n$ matrix. A non-zero vector $\mathbf{v} \in \mathbb{R}^n$ is an *eigenvector* of A with *eigenvalue* λ if $A\mathbf{v} = \lambda\mathbf{v}$.

Example 4.11. Let $f : \mathbb{R}^3 \to \mathbb{R}^3$ be rotation by $\pi/4$ with axis $\mathbf{k}$ and $g : \mathbb{R}^3 \to \mathbb{R}^3$ be rotation by $\pi/2$ with axis $\mathbf{j}$. The planes orthogonal to these vectors are shown in the diagram overleaf.

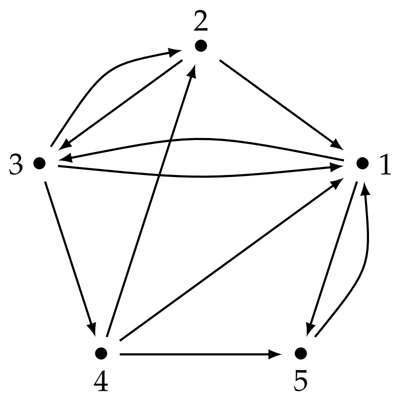


It is a theorem (no need to remember it beyond this example) that the composition of two rotations of $\mathbb{R}^3$ is another rotation. Moreover, if C is a 3×3 matrix representing a rotation by θ then the sum of the diagonal entries of C is $1 + 2\cos\theta$. (You can check this holds for f and g.)

Assuming these results, we will find the axis of the rotation gf and its angle.

The final example is included for interest only, and will be skipped it time is pressing.

Example 4.12. The diagram below shows a tiny part of the World Wide Web. Each dot represents a website; an arrow from site A to site B means that site A links to site B. (Internal links are not considered.)



Suppose we start at site 1 and surf the web by following links at random. So we might visit 3, then 4, then 2, then 1, and so on.

Exercise 4.13. Simulate a surfer who clicks 10 links. Where do you finish?

Let $p(n)_i$ be the probability we are at site i after n steps. For example, since we are equally likely to visit 3 and 5 from 1, and from 3 we can go to 1, 2 or 4, whereas from 5 we must go back to 1, we have

$$\mathbf{p}(0) = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{p}(1) = \begin{pmatrix} 0 \\ 0 \\ \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix}, \quad \mathbf{p}(2) = \begin{pmatrix} \frac{4}{6} \\ \frac{1}{6} \\ 0 \\ \frac{1}{6} \\ 0 \end{pmatrix}.$$

More generally, a similar argument to the Purple Fever problem on Sheet 5 shows that

$$\mathbf{p}(n+1) = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 1 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{3} & 0 \end{pmatrix} \mathbf{p}(n)$$

for all $n \in \mathbb{N}_0$. For n large, $\mathbf{p}(n)$ is close to an eigenvector of this 5×5 matrix with eigenvalue 1. (This can be checked using the MATHEMATICA notebook available from Moodle.) We find this eigenvector by solving the equations

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{3} & 1 \\ 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{3} & 0 \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix}, \quad p_1 + p_2 + p_3 + p_4 + p_5 = 1.$$

The unique solution is

$$\begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \end{pmatrix} = \frac{1}{38} \begin{pmatrix} 14 \\ 4 \\ 9 \\ 3 \\ 8 \end{pmatrix}.$$

A measure of how important a website is the proportion of time we spend at it. As you might expect, this analysis shows site 1 is the most important (every other site links to it) and site 4 (reachable only via site 3) is the least important.

This is essentially how Google's Pagerank Algorithm works.

(C) Solving equations

5. Row operations

MOTIVATION. Many mathematical problems come down to solving linear equations. In this section we will see how matrices gives a systematic way to do this. Here are two illustrative examples. **Lecture 19**

Example 5.1. Suppose that

$$\begin{aligned} 2x + y + 3z &= 11 \\ 2x + 4y + 6z &= 20. \end{aligned}$$

There are infinitely many solutions, since for each $z \in \mathbb{R}$, the equations $2x + y = 11 - 3z$ and $2x + 4y = 20 - 6z$ have the unique solution

$$x = 4 - z, \quad y = 3 - z$$

(*Proof:* subtract the two equations, to get $3y = 9 - 3z$, hence $y = 3 - z$, then substitute to find x.) Is there a solution with $x - y = 2$?

It is routine to convert simultaneous equations into matrix form. For example, the equations above can be written as

$$\begin{pmatrix} 2 & 1 & 3 \\ 2 & 4 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 11 \\ 20 \end{pmatrix}.$$

An important special case is the equation $A\mathbf{x} = \mathbf{0}$, where A is an $m \times n$ matrix, $\mathbf{x} \in \mathbb{R}^m$ and $\mathbf{0} \in \mathbb{R}^n$ denotes, as usual, the all-zero vector.

Example 5.2. Consider the linear equations $A\mathbf{x} = \mathbf{0}$ and $A'\mathbf{x} = \mathbf{0}$ below

$$\begin{pmatrix} 4 & 8 & 0 & -4 \\ 2 & 4 & -1 & -1 \\ -1 & -2 & 3 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \mathbf{0} \tag{1}$$

$$\begin{pmatrix} \mathbf{1} & 2 & 0 & -1 \\ 0 & 0 & \mathbf{1} & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \mathbf{0}. \tag{2}$$

The solutions to (2) can be found very easily: using the **1**s marked in bold, we see that x_2 and x_4 uniquely determine x_1 and x_3. The solution set is

$$\left\{ \begin{pmatrix} -2x_2 + x_4 \\ x_2 \\ -x_4 \\ x_4 \end{pmatrix} : x_2, x_4 \in \mathbb{R} \right\}.$$

Exercise: show that (1) has the same set of solutions.

Remarkably, *any* system $A\mathbf{x} = \mathbf{0}$ can be transformed, as in Example 5.2, so that it has the form of the second system, *without changing its solution set*. We prove this in Corollary 5.5 below. This gives a systematic way to solve linear equations.

ROW OPERATIONS. Fix $m, n \in \mathbb{N}$ and let A be an $m \times n$ matrix.

Definition 5.3. An elementary row operation (ERO) on an $m \times n$ matrix A is one of the following:

(a) Multiply a row of A by a non-zero scalar.
(b) Swap two rows of A.
(c) Add a multiple of one row of A to another row of A.

For example, the matrix A in Example 5.2(1) can be converted to the matrix A' in Example 5.2(2) by a sequence of EROs beginning as follows:

$$2\begin{pmatrix} 4 & 8 & 0 & -4 \\ 2 & 4 & -1 & -1 \\ -1 & -2 & 3 & -2 \end{pmatrix} \xmapsto{\textcircled{1} \mapsto \textcircled{1}/4} \begin{pmatrix} 1 & 2 & 0 & -1 \\ 2 & 4 & -1 & -1 \\ -1 & -2 & 3 & -2 \end{pmatrix}$$

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

$$\xmapsto{\textcircled{3} \mapsto \textcircled{3} + \textcircled{1}} \begin{pmatrix} 1 & 2 & 0 & -1 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 3 & -3 \end{pmatrix}.$$

Exercise: apply two more EROs to reach A'.

Elementary row operations on an $m \times n$ matrix are performed by *left* multiplication by suitable $m \times m$ matrices. Importantly, these matrices are all invertible.

Lemma 5.4. *Let A be an $m \times n$ matrix. Let $1 \leq k, \ell \leq m$ with $k \neq \ell$. Let I be the identity $m \times m$ matrix. Let $\alpha \in \mathbb{R}$, with $\alpha \neq 0$.*

(a) *Let S be I with the 1 in row k replaced with α. Then SA is A with row k scaled by α.*
(b) *Let P be I with rows k and ℓ swapped. Then PA is A with rows k and ℓ swapped.*

Lecture 20

(c) *Let Q be I with the entry in row k and column ℓ changed from 0 to α. Then QA is A with row k replaced with the sum of row k and α times row ℓ.*

Moreover, S, P and Q are invertible.

For example, applying the row operations corresponding to

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

of type S, Q, Q, S, Q reduce A in Example 5.2(1) to A'.

Corollary 5.5. *Let A be an $m \times n$ matrix and let A' be obtained from A by a sequence of elementary row operations. Then*

$$\{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} = 0\} = \{\mathbf{x} \in \mathbb{R}^n : A'\mathbf{x} = 0\}.$$

The analogous result holds for $A\mathbf{x} = \mathbf{b}$, provided $\mathbf{b}$ is transformed along with A. See Questions 2 and 3 on Sheet 7 and Example 5.9 below.

ECHELON FORM AND ROW-REDUCED ECHELON FORM. Example 5.2 suggests what to aim for when doing row operations.

Definition 5.6. Let A be an $m \times n$ matrix. We say that A is in *echelon form* if both of the following conditions hold.

(i) All zero rows are at the bottom.
(ii) Suppose $i < m$. If row i of A is non-zero, and the leftmost non-zero entry of row i is in column j then all the non-zero entries of row $i+1$ of A are in columns $j+1, \ldots, n$.

We say that A is in *row-reduced echelon form* if A is in echelon form and

(iii) If row i is non-zero and its leftmost non-zero entry is in column j, then $A_{ij} = 1$, and this 1 is the unique non-zero entry in column j.

If j is a column as in (iii) then we say that j is a *pivot column*.

For example,

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

is in echelon form, but not row-reduced echelon form, because of the entry 2. It can be put in row-reduced echelon form by the row operation ① $\mapsto$ ① $- 2$②. The matrix in Example 5.2(2) is in row-reduced echelon form with pivot columns 1 and 3.

Lecture 21

Proposition 5.7. *Let A be an $m \times n$ matrix in row-reduced echelon form. Suppose that the first r rows of A are non-zero and the rest are zero. Thus A has r pivot columns and $n - r$ non-pivot columns, say columns $\ell_1, \ldots, \ell_{n-r}$. Given any $\alpha_1, \ldots, \alpha_{n-r} \in \mathbb{R}$ there is a unique solution $\mathbf{x} \in \mathbb{R}^n$ to the equation*

$$A\mathbf{x} = \mathbf{0}$$

such that $x_{\ell_s} = \alpha_s$ for each s with $1 \leq s \leq n - r$.

Stated informally, the solution space of a system of n equations in m variables is $n - r$ dimensional, for some $r \leq m, n$.

Example 5.8. If A' is an in Example 5.2 then A' has pivot columns 1 and 3, and non-pivot columns $\ell_1 = 2$ and $\ell_2 = 4$. So $r = 2$. The unique solution $\mathbf{x} \in \mathbb{R}^4$ to the equation $A'\mathbf{x} = \mathbf{0}$ with given x_2 and x_4 was found earlier to be $\mathbf{x} = \begin{pmatrix} -2x_2 + x_4 \\ x_2 \\ -x_4 \\ x_4 \end{pmatrix}$. Note x_1 and x_3 are determined by x_2 and x_4.

The more general version of Proposition 5.7 for the equation $A\mathbf{x} = \mathbf{b}$ is illustrated in the problem below.

Problem 5.9. Find all solutions to the system

$$\begin{aligned} x_1 \qquad - \;\; x_3 - x_4 &= -4 \\ x_1 + \;\; x_2 \qquad\quad - x_4 &= -1 \\ x_1 - \;\; x_2 - 2x_3 \qquad &= -1 \\ 2x_1 - 3x_2 - 5x_3 \qquad &= \beta \end{aligned}$$

where $\beta \in \mathbb{R}$. Row reducing the augmented matrix

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

we get

$$\left(\begin{array}{cccc|c} \mathbf{1} & 0 & -1 & 0 & 2 \\ 0 & \mathbf{1} & 1 & 0 & 3 \\ 0 & 0 & 0 & \mathbf{1} & 6 \\ 0 & 0 & 0 & 0 & \beta+5 \end{array}\right)$$

The equation for the final row is $0x_1 + 0x_2 + 0x_3 + 0x_4 = \beta + 5$. So if $\beta \neq -5$ then the equations are inconsistent and there is no solution.

When $\beta = -5$ the pivot columns are 1, 2 and 4, so given x_3, there exist unique x_1, x_2 and x_4 such that $\mathbf{x} \in \mathbb{R}^4$ is a solution. From row 3 of the reduced matrix, we get $x_4 = 6$. From row 2 we get $x_2 + x_3 = 3$, hence $x_2 = 3 - x_4$. From row 1 we get $x_1 - x_3 = 2$, hence $x_1 = 2 + x_3$. The solution set is

$$\left\{ \begin{pmatrix} 2 + x_3 \\ 3 - x_3 \\ x_3 \\ 6 \end{pmatrix} : x_3 \in \mathbb{R} \right\}$$

Intuitively, we can say that the solution set is 2-dimensional, corresponding to the 2 non-pivot columns.

Lecture 22 APPLICATION: INVERTING A MATRIX.

Problem 5.10. Let $A = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \\ 1 & 2 & 4 \end{pmatrix}$. What is A^{-1}? We form the augmented matrix $(A \mid I)$ and perform row operations putting it into row-reduced echelon form, getting $(I \mid B)$. As seen in the proof of Corollary 5.5, this corresponds to multiplying $(A \mid I)$ by a sequence of invertible 3×3 matrices $E_1, \dots, E_t$. Setting $R = E_t \dots E_2 E_1$, we have

$$R(A \mid I) = (I \mid B).$$

Hence $RA = I$ and so $R = A^{-1}$.

What happens if we apply the method of Problem 5.10 to a non-invertible matrix?

Example 5.11. By row operations we can transform

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix} \quad \text{to} \quad A' = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}$$

in row-reduced echelon form. So there is an invertible matrix R such that $RA = A'$. If A is invertible then so is RA, since

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

But A' is not invertible, because it is not surjective. For example

$$A' \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} x+z \\ y+z \\ 0 \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

for any $x, y, z \in \mathbb{R}$.

EXTRAS: LEFT VERSUS RIGHT INVERSES (NON-EXAMINABLE). By definition, the inverse of an $n \times n$ matrix A is an $n \times n$ matrix B such that $AB = BA = I$, the $n \times n$ identity matrix.

Exercise: show that B is unique if it exists.

Suppose that, by applying the method in Problem 5.10 to a matrix A, we get $R = E_t \dots E_2 E_1$ such that $RA = I$. Each E_i is invertible, by Lemma 5.4, so multiplying on the left by $E_1^{-1} E_2^{-1} \dots E_t^{-1}$, we get

$$A = E_1^{-1} E_2^{-1} \dots E_t^{-1}.$$

Hence $AR = (E_1^{-1} E_2^{-1} \dots E_t^{-1})(E_t \dots E_2 E_1) = I$ and R is indeed the inverse of A.

Generalizing this argument and Example 5.11 gives most of the proof of the following proposition. Say that a matrix is *elementary* if it is one of the matrices in Lemma 5.4.

Proposition 5.12. *Let $n \in \mathbb{N}$. Let A be an $n \times n$ matrix. The following are equivalent:*

(i) *A is invertible,*
(ii) *A is surjective,*
(iii) *A is equal to a product of elementary matrices.*

Proof. It is sufficient to prove that (i) $\Longrightarrow$ (ii) $\Longrightarrow$ (iii) $\Longrightarrow$ (i). Suppose (i) holds. Then $A(A^{-1}\mathbf{v}) = \mathbf{v}$ for each $\mathbf{v} \in \mathbb{R}^3$, so A is surjective.

Suppose (ii) holds. Let $A' = E_t \ldots E_2 E_1 A = RA$ be in row-reduced echelon form. If A' has a zero row then A' is not surjective. Choose $\mathbf{v}$ such that $A'\mathbf{u} \neq \mathbf{v}$ for any $\mathbf{u} \in \mathbb{R}^n$. Since

$$A\mathbf{u} = R^{-1}\mathbf{v} \Longrightarrow A'\mathbf{u} = RA\mathbf{u} = R(R^{-1}\mathbf{v}) = \mathbf{v}$$

we see that A is not surjective, a contradiction. Hence A' is a square matrix in row-reduced echelon form with no zero rows, and so $A' = I$. Therefore $A = E_1^{-1}E_2^{-1} \ldots E_t^{-1}$ is a product of elementary matrices.

Suppose (iii) holds. Then A is a product of invertible matrices, so invertible. $\square$

Exercise 5.13. Show that (ii) can be replaced with (ii)$'$ A is injective.

Therefore a linear map $f : \mathbb{R}^n \to \mathbb{R}^n$ is injective if and only if it is surjective if and only if it is bijective. Nothing like this is true for general functions.

(D) Determinants

Lecture 23

6. Permutations and determinants

MOTIVATION. We saw in Lemma 3.15 that if A is a 2×2 matrix then the image of the unit square with vertices at

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

under A has area $|\det A|$.

Now suppose that A is a 3×3 matrix, with columns $\mathbf{a}, \mathbf{b}, \mathbf{c}$. As seen in Example 4.5, the columns of A are the images of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ under A. That is,

$$A\mathbf{i} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} A_{11} \\ A_{21} \\ A_{31} \end{pmatrix} = \mathbf{a}$$

and similarly $A\mathbf{j} = \mathbf{b}$ and $A\mathbf{k} = \mathbf{c}$. Thus the unit cube is sent to the parallelepiped formed by $\mathbf{a}, \mathbf{b}, \mathbf{c}$. By Theorem 2.12 and Exercise 2.13 the volume of this parallelepiped is $|\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})|$.

Definition 6.1. Let A be a 3×3 matrix with columns $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$. We define

$$\det A = \mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}).$$

Two alternative forms for $\det A$ are given in the next lemma.

Lemma 6.2. *Let A be a 3×3 matrix. Then*

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

The diagram below shows the patterns of signs: products coming from lines on the left get $+$, products from lines on the right get $-$.

$$\begin{array}{ccccccc} & A_{11} & & A_{12} & & A_{13} & \\ \diagdown & & \diagdown & & \diagdown & & \diagdown \\ & A_{21} & & A_{22} & & A_{23} & \\ \diagdown & & \diagdown & & \diagdown & & \diagdown \\ & A_{31} & & A_{32} & & A_{33} & \end{array} \qquad \qquad \begin{array}{ccccccc} & A_{11} & & A_{12} & & A_{13} & \\ \diagup & & \diagup & & \diagup & & \diagup \\ & A_{21} & & A_{22} & & A_{23} & \\ \diagup & & \diagup & & \diagup & & \diagup \\ & A_{31} & & A_{32} & & A_{33} & \end{array}$$

Note that each of the six summands takes one entry from each row and each column of A. For example $-A_{11}A_{23}A_{32}$ takes the entries in columns 1, 3 and 2 from rows 1, 2 and 3 respectively. So we have

$$\det A = \sum_{\sigma} \pm A_{1\sigma(1)}A_{2\sigma(2)}A_{3\sigma(3)}$$

where the sums is over all bijections $\sigma : \{1, 2, 3\} \to \{1, 2, 3\}$ and $\pm$ is the appropriate sign for each of these bijections.

PERMUTATIONS. To define the determinant for $n \times n$ matrices we need to know the correct sign to attach to each bijection $\sigma : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$. We call such bijections *permutations*.

Example 6.3. It is convenient to write permutations in *two-line notation*. For example, the permutation σ of $\{1, 2, 3, 4, 5, 6\}$ defined by $\sigma(1) = 6$, $\sigma(2) = 1$, $\sigma(3) = 5$, $\sigma(4) = 4$, $\sigma(5) = 3$, $\sigma(6) = 2$ is written as

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

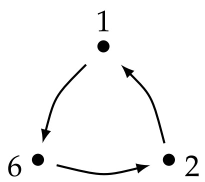
Another way to write permutations uses the following definition.

Definition 6.4. A permutation σ of $\{1, 2, \ldots, n\}$ is an *r-cycle* if there exist distinct $x_1, x_2, \ldots, x_r \in \{1, 2, \ldots, n\}$ such that

$$\sigma(x_1) = x_2,\ \sigma(x_2) = x_3,\ \ldots\ \sigma(x_r) = x_1$$

and $\sigma(y) = y$ for all $y \notin \{x_1, x_2, \ldots, x_r\}$. We write

$$\sigma = (x_1, x_2, \ldots, x_r).$$



For example, $(1, 6, 2)$ is a 3-cycle and $(3, 5)$ is a 2-cycle, or *transposition*. Note that the same cycle can be written in different ways. For example,

$$(1, 6, 2) = (6, 2, 1) = (2, 1, 6).$$

More informally, we can represent this 3-cycle by a diagram as in the margin. The notation for cycles clashes with the notation for tuples: one has to rely on the context to make it clear what is intended.

As a small example of composing cycles, consider

$$\begin{aligned}
\big((1,2)(2,3)\big)1 &= (1,2)\big((2,3)(1)\big) = (1,2)(1) = 2 \\
\big((1,2)(2,3)\big)2 &= (1,2)\big((2,3)(2)\big) = (1,2)(3) = 3 \\
\big((1,2)(2,3)\big)3 &= (1,2)\big((2,3)(3)\big) = (1,2)(2) = 1
\end{aligned}$$

hence $(1, 2)(2, 3) = (1, 2, 3)$.

There are $n!$ permutations of $\{1, 2, \ldots, n\}$. To see this, let $X = \{1, 2, \ldots, n\}$ and think about constructing a bijection $\sigma : X \to X$ step-by-step.

- We may choose any element of X for $\sigma(1)$.
- For $\sigma(2)$ we may choose any element of X except $\sigma(1)$.

  $\vdots$

- Finally for $\sigma(n)$ we have a unique choice, the unique element of X not equal to any of $\sigma(1), \ldots, \sigma(n-1)$.

Hence the number of permutations is $n(n-1)\dots 1 = n!$.

Any shuffle of a deck of cards can be achieved by repeatedly swapping two chosen cards. Correspondingly, any permutation is a composition of transpositions. For example, $(1,2,3,4) = (1,2)(2,3)(3,4)$. **Lecture 24**

Definition 6.5. Let σ be a permutation. We define the *sign* of σ by

- $\operatorname{sgn}(\sigma) = 1$ if σ is equal to a composition of an even number of transpositions;
- $\operatorname{sgn}(\sigma) = -1$ if σ is equal to a composition of an odd number of transpositions.

By Question 7 (optional) on Sheet 8 the sign of a permutation is well-defined. That is, no permutation is expressible both as a composition of evenly many and oddly many transpositions. The examples above show that $\operatorname{sgn}(1,2) = -1$,

$$\operatorname{sgn}(1,2,3) = \operatorname{sgn}\big((1,2)(2,3)\big) = (-1)^2 = 1$$

and $\operatorname{sgn}(1,2,3,4) = (-1)^3 = -1$. See Lemma 6.10 for the general result.

SLIDING BLOCK PUZZLE. A (somewhat unfair) sliding block puzzle has these initial and target positions.

1	2	3	4
5	6	7	8
9	10	11	12
13	15	14	$\square$

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	$\square$

Here $\square$ shows the empty space.

Claim 6.6. *There is no sequence of slides going from the initial position to the target position.*

Proof. A sequence of slides corresponds to a composition of transpositions each involving the empty space $\square$. For example, the permutation corresponding to sliding 14 right, 11 down, 12 left and 14 up is

$$(14,\square)(11,\square)(12,\square)(14,\square) = (14,12,11).$$

Any slide sequence leaving $\square$ in its original position has even length, since $\square$ must move right the same number of times it moves left, and up the same number of times it moves down. The corresponding permutation therefore has sign $+1$. But the initial and target positions differ by the transposition $(14,15)$ which has sign -1. $\square$

In fact any permutation of $\{1, 2, \ldots, 15, \square\}$ of sign $+1$ that fixes $\square$ can be obtained by a suitable sequence of slides. A special case is proved in (optional) Question 8 on Sheet 8.

Lemma 6.7. *Let σ, τ be permutations of $\{1, 2, \ldots, n\}$. Then*

$$\operatorname{sgn}(\sigma\tau) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\tau).$$

In particular, $1 = \operatorname{sgn}(\mathrm{id}) = \operatorname{sgn}(\sigma\sigma^{-1}) = \operatorname{sgn}(\sigma)\operatorname{sgn}(\sigma^{-1})$, and so $\operatorname{sgn}(\sigma) = \operatorname{sgn}(\sigma^{-1})$ for any permutation σ. Moreover, if σ is any permutation and τ is a transposition then $\operatorname{sgn}(\sigma) = -\operatorname{sgn}(\sigma\tau)$.

DISJOINT CYCLE DECOMPOSITION. We say that two cycles $(x_1, \ldots, x_k)$ and $(y_1, \ldots, y_\ell)$ are *disjoint* if $\{x_1, \ldots, x_k\} \cap \{y_1, \ldots, y_\ell\} = \varnothing$.

Exercise 6.8. Let $\tau = (4, 5)$ and let $\sigma = (1, 4, 7)(2, 5)(6, 8)$. To write $\tau\sigma$ as a composition of disjoint cycles we start with 1 and find that $(\tau\sigma)(1) = \tau\bigl(\sigma(1)\bigr) = \tau(4) = 5$, $(\tau\sigma)(5) = \tau\bigl(\sigma(5)\bigr) = \tau(2) = 2$, and so on, forming the cycle $(1, 5, 2, 4, 7)$. We then repeat with the smallest number not yet seen, namely 3, which is fixed. Finally 6 is in the 2-cycle $(6, 8)$. Hence

$$\tau\sigma = (1, 5, 2, 4, 7)(3)(6, 8) = (1, 5, 2, 4, 7)(6, 8).$$

(It is fine to omit the 1-cycle (3), since it is the identity permutation.)

Lecture 25

The algorithm in this exercise can be used to write any permutation as a composition of disjoint cycles.

Lemma 6.9. *A permutation σ of $\{1, 2, \ldots, n\}$ can be written as a composition of disjoint cycles. The cycles in this composition are uniquely determined by σ.*

COMPUTING SIGNS.

Lemma 6.10. *Let $\tau = (x_1, x_2, \ldots, x_\ell)$ be an ℓ-cycle. Then* $\operatorname{sgn}(\tau) = (-1)^{\ell-1}$.

Thus cycles of odd length have sign 1 and cycles of even length have sign -1.

Corollary 6.11. *If σ is a permutation of $\{1, 2, \ldots, n\}$ then*

$$\operatorname{sgn}(\sigma) = \begin{cases} 1 & \textit{if } \sigma \textit{ has an even number of cycles of even length} \\ -1 & \textit{if } \sigma \textit{ has an odd number of cycles of even length} \end{cases}$$

in its disjoint cycle decomposition.

Exercise 6.12. By Exercise 6.8, if $\tau = (4, 5), \sigma = (1, 4, 7)(2, 5)(6, 8)$ then $\tau\sigma = (4, 7, 1, 5, 2)(6, 8) = (1, 5, 2, 4, 7)(6, 8)$. Find $\operatorname{sgn}(\tau), \operatorname{sgn}(\sigma), \operatorname{sgn}(\sigma^{-1})$ and $\operatorname{sgn}(\tau\sigma)$ using either Lemma 6.7 or Corollary 6.11.

Lecture 26

DEFINITION OF THE DETERMINANT. We are now ready to define the determinant of a general $n \times n$ matrix.

Definition 6.13. Let A be an $n \times n$ matrix. We define the *determinant* of A by

$$\det A = \sum_{\sigma} \operatorname{sgn}(\sigma) A_{1\sigma(1)} A_{2\sigma(2)} \ldots A_{n\sigma(n)}$$

where the sum is over all permutations σ of $\{1, 2, \ldots, n\}$.

Thus a permutation σ of $\{1, 2, \ldots, n\}$ corresponds to the summand where we choose the entry in row i and column $\sigma(i)$ of A for each $i \in \{1, 2, \ldots, n\}$.

Exercise 6.14.

(a) Check that if A is a 2×2 matrix then Definition 6.13 agrees with the definition given immediately after Lemma 3.15.

(b) List all permutations of $\{1, 2, 3\}$ and state their signs. Hence check that Definition 6.13 agrees with Lemma 6.2.

(c) Compute $\det \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ \beta & 0 & 0 & 1 \end{pmatrix}$ for each $\beta \in \mathbb{R}$.

[*Hint:* only a few of the $4! = 24$ permutations of $\{1, 2, 3, 4\}$ give a non-zero summand.]

7. PROPERTIES OF THE DETERMINANT

In this section we see some effective way to compute the determinant of a general $n \times n$ matrix. Some proofs will be omitted or given only in a special case that shows the key idea.

ROW OPERATIONS AND DETERMINANTS. Recall that A^T denotes the transpose of A, defined on Question 3 of Sheet 6 by $(A^T)_{ij} = A_{ji}$.

Proposition 7.1. *Let A be an $n \times n$ matrix.*

(i) *If A' is obtained from A by swapping two rows then* $\det A' = -\det A$.

(ii) *If A' is obtained from A by scaling a row by $\alpha \in \mathbb{R}$ then* $\det A' = \alpha \det A$.

(iii) $\det A^T = \det A$.

In particular, if E is an elementary matrix (see Lemma 5.4) that swaps two rows then, since E is obtained from the identity matrix by swapping these rows, $\det E = -1$. Similarly if E scales a row by α then $\det E = \alpha$.

By (iii) applied to (i) and (ii), we get the analogous results for columns. You could use them in the following exercise, or use the alternative definition of the 3×3 determinant using the vector product (see Definition 6.1).

Exercise 7.2. Let A be a 3×3 matrix with columns $\mathbf{a}, \mathbf{b}, \mathbf{c}$. Let $A' = (\mathbf{a} \;\; \mathbf{c} \;\; \mathbf{b})$ and let $A'' = (\mathbf{b} \;\; 3\mathbf{c} \;\; 2\mathbf{a})$ Express $\det A'$ and $\det A''$ in terms of $\det A$.

For the third elementary row operation, adding a multiple of one row to another, we need the following two lemmas.

Lemma 7.3. *If A is an $n \times n$ matrix with two equal rows then* $\det A = 0$.

Lecture 27

Lemma 7.4. *Let $A_{ij} \in \mathbb{R}$ for $2 \leq i \leq n$ and $1 \leq j \leq n$. If $\alpha_1, \alpha_2, \ldots, \alpha_n$ and $\beta_1, \beta_2, \ldots, \beta_n \in \mathbb{R}$ then*

$$\det \begin{pmatrix} \alpha_1 & \alpha_2 & \ldots & \alpha_n \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix} + \det \begin{pmatrix} \beta_1 & \beta_2 & \ldots & \beta_n \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix} =$$

$$\det \begin{pmatrix} \alpha_1 + \beta_1 & \alpha_2 + \beta_2 & \ldots & \alpha_n + \beta_n \\ A_{21} & A_{22} & \ldots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \ldots & A_{nn} \end{pmatrix}.$$

A more general version of Lemma 7.4 holds when each entry in an arbitrary row k is given as a sum. It can be proved by swapping row 1 with row k using Proposition 7.2(i), applying Lemma 7.4 as above, and then swapping back.

Proposition 7.5. *Let A be an $n \times n$ matrix. Let $1 \leq k, \ell \leq n$ with $k \neq \ell$. Let A' be obtained from A by replacing row k with the sum of row k and α times row ℓ. Then* $\det A' = \det A$.

In particular, the elementary matrix corresponding to the row operation in Proposition 7.5 has determinant 1.

Example 7.6. We will use row operations to show that

$$\det \begin{pmatrix} \alpha & \beta & \gamma \\ \beta & \gamma & \alpha \\ \gamma & \alpha & \beta \end{pmatrix} = (\alpha + \beta + \gamma)(\alpha\beta + \beta\gamma + \gamma\alpha - \alpha^2 - \beta^2 - \gamma^2).$$

Question 4 on Problem Sheet 9 can be done in a similar way.

The proof of the following corollary is non-examinable and will be skipped if time is pressing.

Corollary 7.7. *Let B be an $n \times n$ matrix.*

(i) *If E is one of the elementary matrices in Lemma 5.4 then* $\det EB = \det E \det B$. **Lecture 28**

(ii) *If Q is a product of elementary $n \times n$ matrices then* $\det QB = \det Q \det B$.

(iii) *If A is an $n \times n$ matrix then* $\det AB = \det A \det B$.

(iv) $\det B \neq 0$ *if and only if B is invertible.*

Proof. (i) If E scales a row by $\alpha \in \mathbb{R}$ then, by Proposition 7.1(ii) and the following remark, $\det EB = \alpha \det B = \det E \det B$.

The proof is similar if E swaps two rows, or adds one multiple of one row to another, using Proposition 7.1(i) and Proposition 7.5.

(ii) Suppose $Q = E_1 \ldots E_t$. By repeated applications of (i), we have

$$\begin{aligned}\det QB &= \det E_1(E_2 \ldots E_t A) = \det E_1 \det(E_2 \ldots E_t B) = \ldots \\ &= \det E_1 \det E_2 \ldots \det E_{t-1}(\det E_t B) = \det E_1 \det E_2 \ldots \det E_t \det B.\end{aligned}$$

In particular, taking $B = I$, we get $\det Q = \det E_1 \det E_2 \ldots \det E_t$. Hence, $\det QB = \det Q \det B$, as required.

(iii) Let $E_1, \ldots, E_t$ be elementary matrices such that $E_t \ldots E_1 A = A'$ is in row-reduced echelon form. Let $Q = E_1^{-1} \ldots E_t^{-1}$; by Lemma 5.4, Q is a product of elementary matrices. By (ii) we have

$$(\star) \qquad \det AB = \det QA'B = \det Q \det A'B.$$

We now consider two cases.

- If A' has no zero rows then A has a pivot in every column and so $A' = I$ and $Q = A$. Hence, by $(\star)$, $\det AB = \det QB = \det Q \det B = \det A \det B$.
- If A' has a zero row then so does $A'B$. Hence $\det A'B = 0$ for any B. So by $(\star)$, $\det AB = 0$. Taking $B = I$, we get $\det A = 0$. Hence $\det AB = \det A \det B = 0$.

(iv) If B is invertible then, by (iii), $\det I = \det(BB^{-1}) = \det B \det B^{-1}$. Hence $\det B \neq 0$. Conversely, if B is not invertible then we saw in Example 5.11 and Sheet 8, Question 1(c), that there is a product of elementary matrices Q such that QB has a zero row. Hence $\det Q \det B = \det QB = 0$. Since Q is invertible, $\det Q \neq 0$. Hence $\det B = 0$. $\square$

LAPLACE EXPANSION AND THE ADJUGATE.

Definition 7.8. Let A be an $n \times n$ matrix with $n \geq 2$ and let $1 \leq k, \ell \leq n$. The *minor of A in row k, column ℓ*, denoted $M(k, \ell)$, is the $(n-1) \times (n-1)$ matrix obtained from A by deleting row k and column ℓ.

For example, the minors of the 3×3 matrix $A = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$ are

$$M(1,1) = \begin{pmatrix} A_{22} & A_{23} \\ A_{32} & A_{33} \end{pmatrix}, M(1,2) = \begin{pmatrix} A_{21} & A_{23} \\ A_{31} & A_{33} \end{pmatrix}, \ldots, M(3,3) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}.$$

Proposition 7.9. *Let A be an $n \times n$ matrix. Then*

$$\det A = A_{11} \det M(1,1) - A_{12} \det M(1,2) + \cdots + (-1)^{n+1} A_{1n} \det M(1,n).$$

The proof is slightly fiddly so will be omitted: see the bonus questions on Sheet 10 for an outline.

Definition 7.10. Let A be an $n \times n$ matrix. The *adjugate of A* is the $n \times n$ matrix defined by

$$(\operatorname{adj} A)_{k\ell} = (-1)^{\ell+k} M(\ell, k).$$

$$\begin{pmatrix} + & - & + \\ - & + & - \\ + & - & + \end{pmatrix}$$

Note that the k and ℓ are swapped on the right-hand side. The pattern of signs is shown in the margin for the 3×3 case.

Example 7.11. Let $A = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 4 & 4 \\ 3 & 5 & 6 \end{pmatrix}$ be the matrix in Sheet 8, Question 1(a). We have

$$\begin{aligned} \operatorname{adj} A &= \begin{pmatrix} \det \begin{pmatrix} 4 & 4 \\ 5 & 6 \end{pmatrix} & -\det \begin{pmatrix} 1 & 1 \\ 5 & 6 \end{pmatrix} & \det \begin{pmatrix} 1 & 1 \\ 4 & 4 \end{pmatrix} \\ -\det \begin{pmatrix} 2 & 4 \\ 3 & 6 \end{pmatrix} & \det \begin{pmatrix} 1 & 1 \\ 3 & 6 \end{pmatrix} & -\det \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix} \\ \det \begin{pmatrix} 2 & 4 \\ 3 & 5 \end{pmatrix} & -\det \begin{pmatrix} 1 & 1 \\ 3 & 5 \end{pmatrix} & \det \begin{pmatrix} 1 & 1 \\ 2 & 4 \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} 4 & -1 & 0 \\ 0 & 3 & -2 \\ -2 & -2 & 2 \end{pmatrix}. \end{aligned}$$

Comparing with $A^{-1} = \begin{pmatrix} 2 & -\frac{1}{2} & 0 \\ 0 & \frac{3}{2} & -1 \\ -1 & -1 & 1 \end{pmatrix}$, we see that $\operatorname{adj} A = 2A^{-1} = (\det A)A^{-1}$.

This is a special case of the following result. While of some interest, for large matrices, computing the inverse using the adjugate is far slower than using row operations.

Proposition 7.12. *If A is a square matrix then*

$$A \operatorname{adj}(A) = (\det A) I.$$

Hence if $\det A \neq 0$ *then* $A^{-1} = \frac{1}{\det A} \operatorname{adj} A$.

Exercise 7.13. Find $\operatorname{adj} \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and hence check that Proposition 7.12 holds in the 2×2 case.

EIGENVALUES AND EIGENVECTORS. We saw earlier on page 19 that the eigenvalues of a 2×2 matrix are the roots of the *characteristic polynomial* $\det(\lambda I - A)$. This is true in general.

Proposition 7.14. *Let A be a square matrix. Then λ is an eigenvalue of A if and only if* $\det(\lambda I - A) = 0$.

In the 2×2 case we saw that if A has distinct eigenvalues then it is diagonalizable. For $n \geq 3$ if A has n distinct eigenvalues then it is diagonalizable; the proof is beyond the scope of this course. [**Correction:** this was worded rather vaguely in the issued version of these notes, and might have suggested the wrong result.]

The following example is illustrative.

Example 7.15. Let $\alpha \in \mathbb{R}$ and let $A = \begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix}$. Hence

$$\lambda I - A = \begin{pmatrix} \lambda - 2 & 0 & 0 \\ -1 & \lambda - 1 & -\alpha \\ 0 & 0 & \lambda - 1 \end{pmatrix}.$$

Since we have to take the $\lambda - 2$ in the top-left corner to get a non-zero summand in $\det(\lambda I - A)$ we have

$$\det(\lambda I - A) = (\lambda - 2) \det \begin{pmatrix} \lambda - 1 & -\alpha \\ 0 & \lambda - 1 \end{pmatrix} = (\lambda - 2)(\lambda - 1)^2$$

Hence the eigenvalues of A are 1 and 2. When $\lambda = 2$ we have

$$(2I - A) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & -\alpha \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ -x + y - \alpha z \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

if and only if $z = 0$ and $x = y$. So $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ is an eigenvector with eigenvalue 2. When $\lambda = 1$ we have

$$(I-A)\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}-1&0&0\\-1&0&-\alpha\\0&0&0\end{pmatrix}\begin{pmatrix}x\\y\\z\end{pmatrix}=\begin{pmatrix}-x\\-x-\alpha z\\0\end{pmatrix}.$$

(i) If $\alpha = 0$ then all we need is $x = 0$, so there are two non-parallel eigenvectors with eigenvalue 1, for example, $\mathbf{j}$ and $\mathbf{k}$ in the usual notation. Let P be the matrix whose columns are these chosen eigenvectors, so

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

Then, for the same reasons seen in the 2×2 case on page 20,

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

and $P^{-1}AP$ is diagonal. Thus A is diagonalizable.

(ii) If $\alpha \neq 0$ then we need $x = z = 0$, so the only eigenvectors with eigenvalue 1 are the non-zero scalar multiples of $\mathbf{j}$. In this case A is not diagonalizable.

(E) Vector Spaces

8. Vector spaces and subspaces

The following definition abstracts the key properties of sets such as $\mathbb{R}^n$ or planes in $\mathbb{R}^3$. You do not need to memorize it.

Definition 8.1. A vector space is a set V of *vectors* with an addition rule and zero element $\mathbf{0}$ such that

(A1) (*Addition is associative*) $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$,
(A2) (*Addition is commutative*) $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$,
(A3) (*Zero is the identity for addition*) $\mathbf{u} + \mathbf{0} = \mathbf{0} + \mathbf{u} = \mathbf{u}$,

for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$. Moreover there is a rule for multiplying a vector by a scalar such that

(M1) (*Scalar multiplication is associative*) $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$,
(M2) (*One is the identity for scalar multiplication*) $1\mathbf{v} = \mathbf{v}$,
(M3) (*Action of zero*) $0\mathbf{v} = \mathbf{0}$,
(D) (*Distributivity*) $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$,

for all $\mathbf{v} \in V$ and $\alpha, \beta \in \mathbb{R}$.

Note that (A1), (A2), (A3) refer only to the additive structure on V, while (M1), (M2), (M3) refer only to scalar multiplication. The only axiom combining both is (D). Compare the axioms for rings you saw in MT181.

By (M2), (D) and (M3), applied in that order, if V is a vector space and $\mathbf{v} \in V$ then

$$(-1)\mathbf{v} + \mathbf{v} = (-1)\mathbf{v} + (1\mathbf{v}) = \big((-1) + 1\big)\mathbf{v} = 0\mathbf{v} = \mathbf{0}.$$

Thus each $\mathbf{v} \in V$ has an additive inverse, namely $(-1)\mathbf{v}$. This vector is usually written $-\mathbf{v}$.

In this part we will follow the convention that Greek letters are used for scalars.

Example 8.2.

(1) Let $n \in \mathbb{N}$. Then $\mathbb{R}^n$ is a vector space.
(2) Let $n \in \mathbb{N}$. The set of all $n \times n$ matrices is a vector space.
(3) The set of all real polynomials,

$$\mathbb{R}[x] = \{a_0 + a_1x + \cdots + a_dx^d : a_0, a_1, \ldots, a_d \in \mathbb{R}, d \in \mathbb{N}_0\}$$

is a vector space.
(4) The line $\ell = \{\lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} : \lambda \in \mathbb{R}\}$ is a vector space.
(5) The translated line $\ell' = \{\begin{pmatrix} 1 \\ 0 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix} : \lambda \in \mathbb{R}\}$ is not a vector space.

LINEAR INDEPENDENCE, SPANNING AND BASES. These definitions require some thinking about, and will be explored in much greater depth in MT280 linear algebra.

Definition 8.3. Let V be a vector space. Vectors $\mathbf{u}_1, \ldots, \mathbf{u}_r \in V$

(i) are *linearly dependent* if there exist $\alpha_1, \alpha_2, \ldots, \alpha_r \in \mathbb{R}$, not all equal to zero, such that

$$\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \cdots + \alpha_r \mathbf{u}_r = \mathbf{0};$$

(ii) are *linearly independent* if they are not linearly dependent;

(iii) *span* V if for all $\mathbf{v} \in V$ there exist $\beta_1, \ldots, \beta_r \in \mathbb{R}$ such that

$$\mathbf{v} = \beta_1 \mathbf{u}_1 + \beta_2 \mathbf{u}_2 + \cdots + \beta_r \mathbf{u}_r;$$

(iv) are a *basis* for V if they are linearly independent and span V.

Finally if V has a basis of size d we say that V has *dimension* d.

Any two bases of a finite-dimensional vector space have the same size. (A proof is given in the optional extras below.) So dimension is well-defined.

Some vector spaces, for example $\mathbb{R}[x]$, are not finite-dimensional.

Example 8.4. Let

$$V = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 : x + z = 0 \right\}$$

be the plane in $\mathbb{R}^3$ through the origin with normal $\mathbf{i} + \mathbf{k}$. Let

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}, \; \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \; \mathbf{u}_3 = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}.$$

Then

(i) $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ are linearly dependent;

(ii) $\mathbf{u}_1, \mathbf{u}_2$ are linearly independent;

(iii) $\mathbf{u}_1, \mathbf{u}_2$ span V;

(iv) $\mathbf{u}_1, \mathbf{u}_2$ are a basis of W, so V has dimension 2.

Note that we found a linear dependence between $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ by solving the equation $\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2 + \alpha_3 \mathbf{u}_3 = 0$. As a matrix equation it is

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & 0 & -1 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

and it can solved systematically by putting the matrix above into row-reduced echelon form. (Here only one row operation is needed.) More generally, the row-reduced echelon form of $(\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ has a zero row if and only if the vectors $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are linearly dependent.

We also saw a good way to find a basis for a vector space: write down vectors that span it, then delete vectors until those that remain are linearly independent.

SUBSPACES.

Definition 8.5. Let W be a vector space. A subset V of W is a *subspace of* W if

(i) $\mathbf{0} \in V$ *and*
(ii) for any $\mathbf{u}, \mathbf{v} \in V$ and $\alpha, \beta \in \mathbb{R}$, we have $\alpha\mathbf{u} + \beta\mathbf{v} \in V$.

Note the universal quantifier in (ii): it is *not* enough to check that $\alpha\mathbf{u} + \beta\mathbf{v} \in V$ for some particular choice of α, β and $\mathbf{u}, \mathbf{v}$.

Example 8.6. Let $\mathbf{n} \in \mathbb{R}^3$ be non-zero and let

$$V = \{\mathbf{v} \in \mathbb{R}^3 : \mathbf{n} \cdot \mathbf{v} = 0\}$$

be the plane in $\mathbb{R}^3$ through the origin with normal $\mathbf{n}$. Then V is a subspace of $\mathbb{R}^3$ of dimension 2.

Note the requirement that a subspace contains $\mathbf{0}$. For example, the translated plane $\{\mathbf{v} \in \mathbb{R}^3 : (\mathbf{i} + \mathbf{k}) \cdot \mathbf{v} = 1\}$ is not a subspace of $\mathbb{R}^3$, because it does not contain $\mathbf{0}$.

Proposition 8.7. *If V is a subspace of a vector space W then V is a vector space.*

This gives an effective way to prove that a subset of $\mathbb{R}^n$ is a vector space. For example, since $\mathbb{R}^n$ is itself a vector space, the plane in Example 8.4 is now proved to be a vector space.

EXTRAS: DIMENSION IS WELL-DEFINED. There is a very beautiful proof that the dimension of a finite-dimensional vector space does not depend on the choice of basis. Like many important proofs, it is the seed of something more general: the very keen might like to search for 'matroid' on the web. We need the following lemma.

Lemma 8.8 (Steinitz Exchange Lemma). *Let V be a vector space. Suppose that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ are linearly independent vectors in V and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s$ span V. Either $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_s$ is a basis of V, or there exists $j \in \{1, \ldots, s\}$ such that $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r, \mathbf{v}_j$ are linearly independent.*

Proof. Suppose $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_s$ is not a basis of V. Then the set

$$U = \{\beta_1\mathbf{u}_1 + \beta_2\mathbf{u}_2 + \cdots + \beta_s\mathbf{u}_s\}$$

is a subspace of V not equal to V. (*Exercise:* check U is a subspace.) If $\mathbf{v}_j \in U$ for all $j \in \{1, 2, \ldots, s\}$ then, since U is a subspace of V, we have

$\gamma_1\mathbf{v}_1 + \gamma_2\mathbf{v}_2 + \cdots + \gamma_s\mathbf{v}_s \in U$ for all $\gamma_1, \gamma_2, \ldots, \gamma_s \in \mathbb{R}$. But $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s$ span V, so this implies that U contains V, a contradiction.

Hence there exists j such that $\mathbf{v}_j \notin U$. Since $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r$ are linearly independent, if

$$\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_r\mathbf{u}_r + \beta\mathbf{v}_j = \mathbf{0}$$

where not all the scalars are zero, then $\beta \neq 0$. Hence

$$\mathbf{v}_j = -\frac{1}{\beta}(\alpha_1\mathbf{u}_1 + \alpha_2\mathbf{u}_2 + \cdots + \alpha_r\mathbf{u}_r)$$

and so $\mathbf{v}_j \in U$, contrary to the choice of j. Therefore $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_r, \mathbf{v}_j$ are linearly independent. $\square$

By repeatedly applying Lemma 8.8, any linearly independent set can be extended to a basis. Note this goes in the opposite direction to the way we found a basis in Example 8.4.

Corollary 8.9. *Any two bases of a finite-dimensional vector space V have the same size.*

Proof. Let $\mathbf{u}_1, \ldots, \mathbf{u}_r$ and $\mathbf{v}_1, \ldots, \mathbf{v}_s$ be bases of V. Thinking of $\mathbf{u}_1, \ldots, \mathbf{u}_r$ as linearly independent and the $\mathbf{v}_1, \ldots, \mathbf{v}_s$ as spanning, the remark after Lemma 8.8 shows that $r \leq s$. By symmetry $s \leq r$. Hence $r = s$. $\square$

It would perhaps be more in the spirit of this course to prove the following related result which also implies Corollary 8.9.

Lemma 8.10. *Let U be the subspace of $\mathbb{R}^d$ spanned by $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$. Let $A = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n)$ be the matrix having these vectors as columns. Let A' be the row-reduced echelon form of A. If the pivot columns in A' are columns $j_1, \ldots, j_r$ then U has $\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_r}$ as a basis.*

Proof of Lemma 8.10 using row-reduced echelon form. Let $\ell_1, \ldots, \ell_{n-r}$ be the non-pivot columns of A'. By Proposition 5.7, given any $\alpha_1, \ldots, \alpha_{n-r}$ there exist unique $x_1, x_2, \ldots, x_n$ such that

$$(\star) \qquad x_1\mathbf{u}_1 + x_2\mathbf{u}_2 + \cdots + x_n\mathbf{u}_n = \mathbf{0}$$

and $x_{\ell_s} = \alpha_s$ for each s. Taking $\alpha_1 = 1$, $\alpha_2 = \ldots = \alpha_{n-r} = 0$ gives a linear relation expressing the non-pivot column $\mathbf{u}_{\ell_1}$ as a linear combination of the pivot columns $\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_r}$. We argue similarly for $\mathbf{u}_{\ell_2}, \ldots, \mathbf{u}_{\ell_{n-r}}$. Hence $\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_r}$ span U. By the exercise below, they are linearly independent, and so form a basis. $\square$

Exercise 8.11. Taking $\alpha_1 = \ldots = \alpha_{n-r} = 0$, Proposition 5.9 implies that the unique solution to $(\star)$ is $x_1 = \ldots = x_n = 0$. Deduce that $\mathbf{u}_{j_1}, \ldots, \mathbf{u}_{j_r}$ are linearly independent.