# Vectors & Matrices, 2020–2021

Lecturers: Prof. Vito Latora and Dr. Abhishek Saha
These lecture notes are based on previous notes by Prof. Oliver Jenkinson

December 1, 2020

# Contents

# Chapter 1

# Vectors

## 1.1 Bound Vectors and Free Vectors

**Definition 1.1.** A **bound vector** is a directed line segment in 3-space. If $A$ and $B$ are points in 3-space, we denote the bound vector with starting point $A$ and endpoint $B$ by $\overrightarrow{AB}$.

As the notation $\overrightarrow{AB}$ suggests, an ordered pair of points $A, B$ in 3-space determines a bound vector. Alternatively, a bound vector is determined by its:

- starting point.
- length,
- direction (provided that the length is not 0),

We denote the length of the bound vector $\overrightarrow{AB}$ by $|\overrightarrow{AB}|$. If a bound vector has length 0 then it is of the form $\overrightarrow{AA}$ (where $A$ is some point in $3$-space) and has undefined direction.

If we ignore the starting point we get the notion of a **free vector** (or simply a **vector**). So a free vector is determined by its:

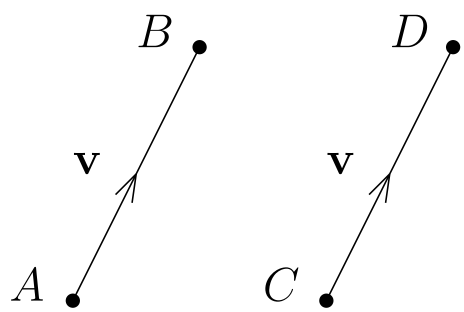
- length,
- direction (provided that the length is not 0).

We will use letters in **bold type** for free vectors ($\mathbf{u}, \mathbf{v}, \mathbf{w}$ etc.)[1] The length of the free vector $\mathbf{v}$ will be denoted by $|\mathbf{v}|$.

**Definition 1.2.** We say that the bound vector $\overrightarrow{AB}$ **represents** the free vector $\mathbf{v}$ if it has the same length and direction as $\mathbf{v}$.

[1] In handwritten notes underlining would be used: $\underline{u}$,$\underline{v}$,$\underline{w}$ etc.

In the figure below, the two bound vectors $\overrightarrow{AB}$ and $\overrightarrow{CD}$ represent the same free vector $\mathbf{v}$.




**Definition 1.3.** The **zero vector** is the free vector with length 0 and undefined direction. It is denoted by $\mathbf{0}$.

For any point $A$, the bound vector $\overrightarrow{AA}$ represents $\mathbf{0}$.

It is important to be aware of the difference between bound vectors and free vectors. In particular you should never write something like $\overrightarrow{AB} = \mathbf{v}$. The problem with this is that the two things we are asserting are equal are different types of mathematical object. It would be correct to say that $\overrightarrow{AB}$ represents $\mathbf{v}$.

One informal analogy that might be helpful is that a free vector is a bit like an instruction (go 20 miles Northeast say) while a bound vector is like the path you trace out when you follow that instruction. Notice that there is no way to draw the instruction on a map (similarly we cannot really draw a free vector) and that if you follow the same instruction from different starting points you get different paths (just as we have many different bound vectors representing the same free vector).

In what follows we will mainly be working with free vectors and when we write vector we will always mean free vector.

## 1.2 Vector Negation

If $\mathbf{v}$ is a non-zero vector we define its negation $-\mathbf{v}$ to be the vector with the same length as $\mathbf{v}$ and opposite direction. We define $-\mathbf{0} = \mathbf{0}$. Negation is a function from the set of vectors to itself. If $\overrightarrow{AB}$ represents $\mathbf{v}$ then $\overrightarrow{BA}$ represents $-\mathbf{v}$.

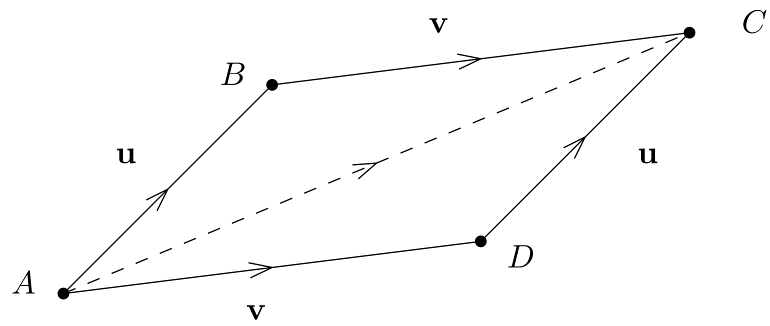
## 1.3 Vector Addition

To define the sum of two vectors we need the notion of a parallelogram.

**Definition 1.4.** The figure $ABCD$ is a **parallelogram** if $\overrightarrow{AB}$ and $\overrightarrow{DC}$ represent the same vector.

**Fact 1.5** (The Parallelogram Axiom)**.** *If $\overrightarrow{AB}$ and $\overrightarrow{DC}$ represent the same vector ($\mathbf{u}$, say), then $\overrightarrow{BC}$ and $\overrightarrow{AD}$ represent the same vector ($\mathbf{v}$, say). Note that we need not have $\mathbf{u} = \mathbf{v}$.*

**Definition 1.6.** Given vectors $\mathbf{u}$ and $\mathbf{v}$ we define the sum $\mathbf{u} + \mathbf{v}$ as follows. Pick any point $A$ and let $B, C, D$ be points such that $\overrightarrow{AB}$ represents $\mathbf{u}$, $\overrightarrow{AD}$ represents $\mathbf{v}$ and $ABCD$ is a parallelogram. Then $\mathbf{u} + \mathbf{v}$ is the vector represented by $\overrightarrow{AC}$.




In the figure, $\overrightarrow{DC}$ represents $\mathbf{u}$ because we chose $C$ in order to make $ABCD$ a parallelogram. Now by the parallelogram axiom $\overrightarrow{AD}$ and $\overrightarrow{BC}$ represent the same vector $\mathbf{v}$. By definition $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$.

**Remark 1.7.** In order for this to be a sensible definition it needs to specify $\mathbf{u}+\mathbf{v}$ in a completely unambiguous way. In other words, if two people follow the recipe above to find $\mathbf{u}+\mathbf{v}$ they should come up with the same vector. For this to be true we need to check that $B, C, D$ are uniquely specified by the rule we gave (this is obvious) and that the answer we get does not depend on the choice of $A$. This last point requires bit of checking which I give as an exercise (with hints) at the end of the chapter.

From this definition we get the following useful interpretation of vector addition[2]:

**Proposition 1.8** (The Triangle Rule for Vector Addition)**.** *If $\overrightarrow{AB}$ represents $\mathbf{u}$ and $\overrightarrow{BC}$ represents $\mathbf{v}$ then $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$.*

*Proof.* Suppose that $\overrightarrow{AB}$ represents $\mathbf{u}$ and $\overrightarrow{BC}$ represents $\mathbf{v}$. Let $D$ be the point such that $\overrightarrow{DC}$ represents $\mathbf{u}$. Since $\overrightarrow{AB}$ and $\overrightarrow{DC}$ both represent $\mathbf{u}$, the figure $ABCD$ is a parallelogram. By the parallelogram axiom $\overrightarrow{BC}$ and $\overrightarrow{AD}$ represent the same free vector and so $\overrightarrow{AD}$ represents $\mathbf{v}$. It follows that the figure $ABCD$ is precisely the parallelogram constructed in the definition of $\mathbf{u}+\mathbf{v}$ and so by definition $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$ as required. $\square$

Vector addition shares several properties with ordinary addition of numbers.

**Proposition 1.9.** *[Properties of Vector Addition[3]] If $\mathbf{u}, \mathbf{v}, \mathbf{w}$ are vectors then:*

1. $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$ *(that is vector addition is commutative),*
2. $\mathbf{u}+\mathbf{0} = \mathbf{u}$ *(that is $\mathbf{0}$ is an identity for vector addition),*
3. $\mathbf{u}+(\mathbf{v}+\mathbf{w}) = (\mathbf{u}+\mathbf{v})+\mathbf{w}$ *(that is vector addition is associative),*
4. $\mathbf{u}+(-\mathbf{u}) = \mathbf{0}$ *(that is $-\mathbf{u}$ is an additive inverse for $\mathbf{u}$).*

We usually write $\mathbf{u}-\mathbf{v}$ for $\mathbf{u}+(-\mathbf{v})$ (this defines vector subtraction) and so part (iv) could be written as $\mathbf{u}-\mathbf{u} = \mathbf{0}$.

*Proof.* 1. If $\overrightarrow{AB}$ represents $\mathbf{u}$, $\overrightarrow{AD}$ represents $\mathbf{v}$ and $ABCD$ is a parallelogram then $\overrightarrow{DC}$ represents $\mathbf{u}$. It follows from the triangle rule applied to the triangle $ADC$ that $\overrightarrow{AC}$ represents $\mathbf{v}+\mathbf{u}$. We know (by the definition of vector addition) that $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$. Hence $\mathbf{u}+\mathbf{v} = \mathbf{v}+\mathbf{u}$.

[2]Some people regard this as the definition of $\mathbf{u}+\mathbf{v}$ in which case the parallelogram interpretation becomes a result that needs a proof

[3]If you take the module Introduction To Algebra you will recognise that these properties mean that the set of free vectors forms an Abelian group under vector addition.

2. Let $\overrightarrow{AB}$ represent $\mathbf{u}$. Since $\overrightarrow{BB}$ represents $\mathbf{0}$, the triangle rule applied to the (degenerate) triangle $ABB$ gives that $\overrightarrow{AB}$ represents $\mathbf{u}+\mathbf{0}$. Hence $\mathbf{u}+\mathbf{0}=\mathbf{u}$.

3. See problem sheet 1.

4. Let $\overrightarrow{AB}$ represent $\mathbf{u}$. Since $\overrightarrow{BA}$ represents $-\mathbf{u}$, the triangle rule applied to the (degenerate) triangle $ABA$ gives that $\overrightarrow{AA}$ represents $\mathbf{u}+(-\mathbf{u})$. But $\overrightarrow{AA}$ represents $\mathbf{0}$ and so $\mathbf{u}-\mathbf{u}=\mathbf{0}$.

□

## 1.4 Scalar Multiplication

If $\alpha \in \mathbb{R}$ and $\mathbf{v}$ is a vector, we define $\alpha\mathbf{v}$ to be the vector with length[4] $|\alpha||\mathbf{v}|$ and direction the same as $\mathbf{v}$ if $\alpha > 0$, opposite to $\mathbf{v}$ if $\alpha < 0$, and undefined if $\alpha = 0$.

Multiplication of a vector by a scalar also has some nice properties:

**Proposition 1.10.** *[Properties of Scalar Multiplication] For any $\alpha, \beta \in \mathbb{R}$ and vectors $\mathbf{u}, \mathbf{v}$ we have:*

*(i)* $0\mathbf{u}=\mathbf{0}$, $\alpha\mathbf{0}=\mathbf{0}$, $1\mathbf{u}=\mathbf{u}$, $-1\mathbf{u}=-\mathbf{u}$,

*(ii)* $\alpha(\beta\mathbf{u}) = (\alpha\beta)\mathbf{u}$,

*(iii)* $(\alpha+\beta)\mathbf{u} = \alpha\mathbf{u}+\beta\mathbf{u}$,

*(iv)* $\alpha(\mathbf{u}+\mathbf{v}) = \alpha\mathbf{u}+\alpha\mathbf{v}$.

The last two properties in this proposition are called distributive laws.

We will show parts of the proofs of these but will not go through all cases in detail. **The proofs of this Proposition are not examinable.**

*Proof.* (i) Trivial.

(ii) By definition of scalar multiplication:

$$|\alpha(\beta\mathbf{u})| = |\alpha||\beta\mathbf{u}| = |\alpha||\beta||\mathbf{u}| = |\alpha\beta||\mathbf{u}| = |(\alpha\beta)\mathbf{u}|.$$

So $\alpha(\beta\mathbf{u})$ and $(\alpha\beta)\mathbf{u}$ have the same length.

If $\alpha = 0$ or $\beta = 0$ (or both) then both sides are equal to $\mathbf{0}$.

We consider cases according to whether $\alpha$ and $\beta$ are positive or negative.

If $\alpha > 0, \beta > 0$. Then both $\alpha(\beta\mathbf{u})$ and $(\alpha\beta)\mathbf{u}$ have the same direction as $\mathbf{u}$ and so are equal.

If $\alpha < 0, \beta > 0$. Then $\beta\mathbf{u}$ has the same direction as $\mathbf{u}$ and $\alpha(\beta\mathbf{u})$ has direction opposite to $\mathbf{u}$. Also $\alpha\beta < 0$ so $(\alpha\beta)\mathbf{u}$ has direction opposite to $\mathbf{u}$. It follows that both $\alpha(\beta\mathbf{u})$ and $(\alpha\beta)\mathbf{u}$ have the same direction as $-\mathbf{u}$ and so are equal.

The remaining cases of $\alpha > 0, \beta < 0$ and $\alpha < 0, \beta < 0$ are similar.

[4]Be careful with the notation here. In this expression $|\alpha|$ is the absolute value of the scalar $\alpha$ while $\mathbf{v}$ is the length of the vector $\mathbf{v}$.

(ii) Let $\overrightarrow{AB}$ represent $\alpha\mathbf{u}$ and $\overrightarrow{BC}$ represent $\beta\mathbf{u}$. Then by the triangle rule $\overrightarrow{AC}$ represents $\alpha\mathbf{u}+\beta\mathbf{u}$.

If $\alpha > 0, \beta > 0$ then $\overrightarrow{AC}$ is a bound vector of length $\alpha|\mathbf{u}|+\beta|\mathbf{u}| = (\alpha+\beta)|\mathbf{u}|$ in the same direction as $\mathbf{u}$. That is $\overrightarrow{AC}$ represents $(\alpha+\beta)\mathbf{u}$. It follows that $\alpha\mathbf{u}+\beta\mathbf{u} = (\alpha+\beta)\mathbf{u}$.

The remaining cases are similar.

(iv) If $\alpha = 0$ then both sides are equal to $\mathbf{0}$.

Suppose that $\alpha > 0$. Let $\overrightarrow{AB}$ represent $\mathbf{u}$, $\overrightarrow{BC}$ represent $\mathbf{v}$, $\overrightarrow{AD}$ represent $\alpha\mathbf{u}$, and $\overrightarrow{DE}$ represent $\alpha\mathbf{v}$ (draw a picture).

The triangles $ABC$ and $ADE$ are similar triangles and the edge $AB$ is in the same direction as the edge $AD$. It follows that the bound vector $\overrightarrow{AE}$ is in the same direction as $\overrightarrow{AC}$ and its length differs by a factor of $\alpha$. But $\overrightarrow{AC}$ represents $\mathbf{u}+\mathbf{v}$ and $\overrightarrow{AE}$ represents $\alpha\mathbf{u}+\alpha\mathbf{v}$. It follows that $\alpha\mathbf{u}+\alpha\mathbf{v} = \alpha(\mathbf{u}+\mathbf{v})$.

The $\alpha < 0$ case is similar.

□

## 1.5 Position Vectors

Suppose now that we fix a special point in space called the **origin** and denoted by $O$.

**Definition 1.11.** If $P$ is a point, the **position vector** of $P$ is the free vector represented by the bound vector $\overrightarrow{OP}$.

We will usually write $\mathbf{p}$ for the position vector of $P$, $\mathbf{q}$ for the position vector of $Q$ and so on.

Each point in space has a unique position vector and each vector is the position vector of a unique point in space.

If $A$ and $B$ are points with position vectors $\mathbf{a}$ and $\mathbf{b}$ respectively then by the triangle rule applied to the triangle $AOB$ we get that $\overrightarrow{AB}$ represents the vector $\mathbf{b}-\mathbf{a}$.

**Theorem 1.12.** *Let $A, B$ be points with position vectors $\mathbf{a}$ and $\mathbf{b}$ respectively. Let $P$ be the point on the line segment $AB$ with $|\overrightarrow{AP}| = \lambda|\overrightarrow{AB}|$. The position vector $\mathbf{p}$ of $P$ is $(1-\lambda)\mathbf{a}+\lambda\mathbf{b}$.*

*Proof.* Define $\mathbf{u}$ to be the free vector such that $\overrightarrow{AB}$ represents $\mathbf{u}$. It follows that the bound vector $\overrightarrow{AP}$ represents $\lambda\mathbf{u}$. The triangle rule applied to $OAP$ gives that $\mathbf{p} = \mathbf{a}+\lambda\mathbf{u}$. Also $\mathbf{u} = \mathbf{b}-\mathbf{a}$. Putting these together gives that $\mathbf{p} = \mathbf{a}+\lambda(\mathbf{b}-\mathbf{a})$ which after some manipulation (using the distributive laws of Proposition 1.3(iii,iv)) gives the result. □

In lectures we used this theorem to prove the following geometric fact about parallelograms.

**Example 1.13.** The diagonals of a parallelogram intersect at their midpoints.

## 1.6 The Definition of $\mathbf{u}+\mathbf{v}$ does not depend on $A$

**This section is non-examinable** but I encourage you to work through the argument as an exercise.

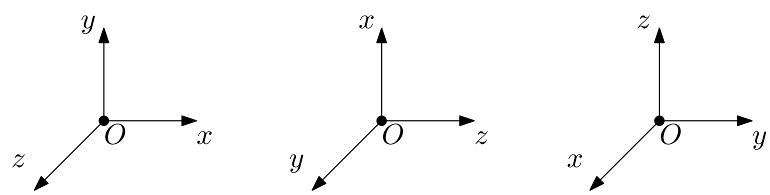
If you look back to the definition of vector addition you will see that we started by picking an arbitrary point $A$. This question leads you through the proof that the definition of vector addition does not depend on which point is chosen. The steps are roughly indicated with some gaps. Your task is to expand each step and fill in the gaps to get a complete proof:

- Consider the parallelogram needed to define $\mathbf{u} + \mathbf{v}$ with respect to point $A$ (name the points).
- Consider the parallelogram needed to define $\mathbf{u} + \mathbf{v}$ with respect to a different point $E$ (name the points).
- Fill in the gaps: "In order for it not to matter whether we used the parallelogram based at $A$ or the one based at $E$ in the definition of $\mathbf{u}+\mathbf{v}$, our task is to show that the bound vectors . . . and . . . represent the same free vector".
- Fill in the gaps: "The figure . . . is a parallelogram because . . . and . . . both represent the vector $\mathbf{u}$ and so by the parallelogram axiom . . . and . . . represent the same vector [give it a name]."
- Fill in the gaps: "The figure . . . is a parallelogram because . . . and . . . both represent the vector $\mathbf{v}$ and so by the parallelogram axiom . . . and . . . represent the same vector [what is that vector]."
- Make one more application of the parallelogram axiom to show that the bound vectors from step 3 really do represent the same free vector.

# Chapter 2

# Coordinates

Suppose now that we choose an origin $O$ and 3 mutually perpendicular axes (the $x-$, $y-$ and $z-$ axes) arranged in a right-handed system as in the figures below:




Let $\mathbf{i}, \mathbf{j}, \mathbf{k}$ denote vectors of unit length (i.e. length 1) in the directions of the $x-$, $y-$ and $z-$axes respectively.

We say that $R$ is **the point with coordinates** $(a, b, c)$ if the position vector of $R$ is

$$\mathbf{r} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

If $Q$ is the point with position vector $a\mathbf{i}+b\mathbf{j}$ and $P$ is the point with position vector $a\mathbf{i}$ then $OPQ$ is a right-angled triangle and $\overrightarrow{PQ}$ represents $b\mathbf{j}$. It follows from Pythyagoras's Theorem that

$$|\overrightarrow{OQ}|^2 = |a\mathbf{i}|^2 + |b\mathbf{j}|^2 = a^2 + b^2.$$

Further, $OQR$ is a right-angled triangle and $\overrightarrow{QR}$ represents $c\mathbf{k}$. So

$$|\mathbf{r}|^2 = |\overrightarrow{OR}|^2 = |\overrightarrow{OQ}|^2 + |\overrightarrow{QR}|^2 = a^2 + b^2 + c^2.$$

It follows that

$$|\mathbf{r}| = \sqrt{a^2 + b^2 + c^2}.$$

## 2.1 Unit Vectors

**Definition 2.1.** A **unit vector** is a vector of length 1.

For instance $\mathbf{i}, \mathbf{j}$ and $\mathbf{k}$ are unit vectors.

If $\mathbf{u}$ is any non-zero vector then $\hat{\mathbf{u}} = \left(\frac{1}{|\mathbf{u}|}\right)\mathbf{u}$ is a unit vector in the same direction as $\mathbf{u}$. We often write $\frac{\mathbf{u}}{|\mathbf{u}|}$ for $\frac{1}{|\mathbf{u}|}\mathbf{u}$.

## 2.2 Sums and Scalar Multiples in Coordinates

We will write $\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ for the vector $a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ where $a, b, c \in \mathbb{R}$.

Let $\mathbf{u} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} d \\ e \\ f \end{pmatrix}$ and $\alpha \in \mathbb{R}$.

$$\begin{aligned} \mathbf{u} + \mathbf{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} d \\ e \\ f \end{pmatrix} &= (a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) + (d\mathbf{i} + e\mathbf{j} + f\mathbf{k}) \\ &= (a\mathbf{i} + d\mathbf{i}) + (b\mathbf{j} + e\mathbf{j}) + (c\mathbf{k} + f\mathbf{k}) \\ &= (a+d)\mathbf{i} + (b+e)\mathbf{j} + (c+f)\mathbf{k} \\ &= \begin{pmatrix} a+d \\ b+e \\ c+f \end{pmatrix}. \end{aligned}$$

Also,

$$\begin{aligned} \alpha\mathbf{u} = \alpha \begin{pmatrix} a \\ b \\ c \end{pmatrix} &= \alpha(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}) \\ &= \alpha(a\mathbf{i}) + \alpha(b\mathbf{j}) + \alpha(c\mathbf{k}) \\ &= (\alpha a)\mathbf{i} + (\alpha b)\mathbf{j} + (\alpha c)\mathbf{k} = \begin{pmatrix} \alpha a \\ \alpha b \\ \alpha c \end{pmatrix}. \end{aligned}$$

So vector addition and scalar multiplication can be nicely expressed in coordinates.

Note that in deriving these expressions, we used our properties of vector addition and scalar multiplication (Propositions 1.9 and 1.10) repeatedly.

## 2.3 Equations of Lines

Let $l$ be the line through the point $P$ in the direction of the non-zero vector $\mathbf{u}$.

The point $R$ with position vector $\mathbf{r}$ is on the line $l$ if and only if the vector represented by $\overrightarrow{PR}$ is a multiple of $\mathbf{u}$. That is, $R$ is on $l$ if and only if $\mathbf{r} - \mathbf{p} = \lambda\mathbf{u}$ for some $\lambda \in \mathbb{R}$, or equivalently if and only if $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$. This is called the **vector equation** for $l$.

Note that in this equation, $\mathbf{p}$ and $\mathbf{u}$ are constant vectors (depending on the line), while $\mathbf{r}$ is a (vector) variable depending on the (real number) variable $\lambda$. The equation gives a condition which $\mathbf{r}$ satisfies if and only if $R$ lies on the line $l$. Specifically, suppose that $R$ is a point with position vector $\mathbf{r}$. If there is some $\lambda$ for which $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$ then $R$ lies on $l$; if there is no such $\lambda$ then $R$ does not lie on $l$.

Working in coordinates, let $\mathbf{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$ and $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$. We get that $R$ is on $l$ if and only if

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix} + \lambda \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = \begin{pmatrix} p_1 + \lambda u_1 \\ p_2 + \lambda u_2 \\ p_3 + \lambda u_3 \end{pmatrix}.$$

This is equivalent to the system of equations:

$$\left.\begin{array}{l} x = p_1 + \lambda u_1 \\ y = p_2 + \lambda u_2 \\ z = p_3 + \lambda u_3 \end{array}\right\} .$$

These are called the **parametric equations** for the line $l$.

The variable $\lambda$ is referred to as a **parameter**. Note that it appears in the above 3 equations (which we have called the parametric equations), but it also appears in the equation $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$ (which we called the vector equation, but could equally well be called a **parametric vector equation**).

If $u_1 \neq 0, u_2 \neq 0, u_3 \neq 0$ we can eliminate the parameter $\lambda$ from the parametric equations to get

$$\frac{x - p_1}{u_1} = \frac{y - p_2}{u_2} = \frac{z - p_3}{u_3},$$

called the **Cartesian equations** for the line $l$.

If $u_1 = 0, u_2 \neq 0, u_3 \neq 0$ then the Cartesian equations are

$$x = p_1, \quad \frac{y - p_2}{u_2} = \frac{z - p_3}{u_3}.$$

If $u_1 = u_2 = 0, u_3 \neq 0$ then the Cartesian equations are

$$x = p_1, \quad y = p_2$$

(with no constraint on $z$).

Note that we cannot have $u_1 = u_2 = u_3 = 0$ because we insisted that $\mathbf{u}$ was a non-zero vector.

Another natural way of describing a line is by giving two points that lie on it. If $P$ and $Q$ are distinct points[1] with position vectors $\mathbf{p}$ and $\mathbf{q}$ respectively then the line containing $P$ and $Q$ is in direction $\mathbf{q} - \mathbf{p}$. We can now use the method above with $\mathbf{u} = \mathbf{q} - \mathbf{p}$. For instance the line through $P$ and $Q$ has vector equation

$$\mathbf{r} = \mathbf{p} + \lambda(\mathbf{q} - \mathbf{p}).$$

We found this equation by noting that the line in question is the line through $P$ in direction $(\mathbf{q} - \mathbf{p})$. However we could equally have identified the same line as the line through $Q$ in direction $(\mathbf{q} - \mathbf{p})$. This yields the vector equation

$$\mathbf{r} = \mathbf{q} + \lambda(\mathbf{q} - \mathbf{p}).$$

In general there are many possible different vector equations all determining the same line.

Whether we choose to use the vector, parametric or Cartesian equations for the line, in each case we have described the geometric object (in this case a line) by giving a condition that position vectors of points on the line must satisfy. If we think of the line as being the set of points on it then this set is determined by the following set of position vectors:

$$\{\mathbf{r} : \mathbf{r} = \mathbf{p} + \lambda\mathbf{u} \text{ for some } \lambda \in \mathbb{R}\}.$$

More generally, any set of position vectors defined by giving a condition on $\mathbf{r}$ (or on $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$) determines a geometric object in $3$-space.

[1] Can you see what goes wrong if $P = Q$?

# Chapter 3

# Scalar Product and Vector Product

## 3.1 The scalar product

If $\mathbf{u}$ and $\mathbf{v}$ are non-zero vectors with $\overrightarrow{AB}$ representing $\mathbf{u}$ and $\overrightarrow{AC}$ representing $\mathbf{v}$, we define the **angle between $\mathbf{u}$ and $\mathbf{v}$** to be the angle $\theta$ (in radians) between the line segments $\overrightarrow{AB}$ and $\overrightarrow{AC}$ with $0 \leq \theta \leq \pi$.

**Definition 3.1.** The **scalar product** of $\mathbf{u}$ and $\mathbf{v}$ is denoted by $\mathbf{u} \cdot \mathbf{v}$ and defined by

$$\mathbf{u} \cdot \mathbf{v} = \begin{cases} |\mathbf{u}||\mathbf{v}| \cos\theta & \text{if } \mathbf{u} \neq \mathbf{0}, \mathbf{v} \neq \mathbf{0} \\ 0 & \text{if } \mathbf{u} = \mathbf{0} \text{ or } \mathbf{v} = \mathbf{0} \end{cases}$$

where $\theta$ is the angle between $\mathbf{u}$ and $\mathbf{v}$.

**Definition 3.2.** We say that $\mathbf{u}$ and $\mathbf{v}$ are **orthogonal** if $\mathbf{u} \cdot \mathbf{v} = 0$.

Note that $\mathbf{u}$ and $\mathbf{v}$ are orthogonal if either one or both of them is the zero vector, or they are perpendicular (the angle between them is $\pi/2$).

Working in coordinates we have the following very useful formula for the scalar product:

**Theorem 3.3.** *If* $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ *and* $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$. *Then*

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + u_3 v_3.$$

*Proof.* If $\mathbf{u} = \mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ or $\mathbf{v} = \mathbf{0}$ then $\mathbf{u} \cdot \mathbf{v} = 0$ by definition and $u_1 v_1 + u_2 v_2 + u_3 v_3 = 0$ and so the result is true.

Suppose that $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$ and let $\theta$ be the angle between $\mathbf{u}$ and $\mathbf{v}$.

We will use the fact that if we calculate $|\mathbf{u} + \mathbf{v}|^2$ in two different ways we must get the same answer.

First, let $\overrightarrow{AB}$ represent $\mathbf{u}$, $\overrightarrow{AD}$ represent $\mathbf{v}$, and $ABCD$ be a parallelogram. Let $E$ be the point on the line through $AB$ with $\overrightarrow{EC}$ perpendicular to $\overrightarrow{AE}$ (draw a picture!).

By the definition of vector addition we have that $\overrightarrow{AC}$ represents $\mathbf{u} + \mathbf{v}$ and $AEC$ is a right-angled triangle so

$$\begin{aligned} |\mathbf{u} + \mathbf{v}|^2 &= |\overrightarrow{AE}|^2 + |\overrightarrow{EC}|^2 \\ &= (|\mathbf{u}| + |\mathbf{v}| \cos\theta)^2 + (|\mathbf{v}| \sin\theta)^2 \\ &= |\mathbf{u}|^2 + |\mathbf{v}|^2 \left((\sin\theta)^2 + (\cos\theta)^2\right) + 2|\mathbf{u}||\mathbf{v}| \cos\theta \\ &= |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2\mathbf{u} \cdot \mathbf{v}. \end{aligned}$$

Secondly, in coordinates

$$\begin{aligned}
|\mathbf{u}+\mathbf{v}|^2 &= \left|\begin{pmatrix} u_1+v_1 \\ u_2+v_2 \\ u_3+v_3 \end{pmatrix}\right|^2 \\
&= (u_1+v_1)^2 + (u_2+v_2)^2 + (u_3+v_3)^2 \\
&= (u_1^2+u_2^2+u_3^2) + (v_1^2+v_2^2+v_3^2) + 2(u_1v_1+u_2v_2+u_3v_3).
\end{aligned}$$

Equating these two expressions for $|\mathbf{u}+\mathbf{v}|^2$ and rearranging gives the result. □

Theorem 3.3 can be used to find the angle $\theta$ between two non-zero vectors given in coordinates. Rearranging the definition of $\mathbf{u}\cdot\mathbf{v}$ and substituting the formula of Theorem 3.3 we get

$$\cos\theta = \frac{\mathbf{u}\cdot\mathbf{v}}{|\mathbf{u}||\mathbf{v}|} = \frac{u_1v_1+u_2v_2+u_3v_3}{\sqrt{(u_1^2+u_2^2+u_3^2)(v_1^2+v_2^2+v_3^2)}}.$$

This also shows that for non-zero $\mathbf{u}$ and $\mathbf{v}$

$$\mathbf{u}\cdot\mathbf{v}\begin{cases} \text{is positive if and only if} & 0\le\theta<\pi/2 \\ \text{is zero if and only if} & \theta=\pi/2 \\ \text{is negative if and only if} & \pi/2<\theta\le\pi \end{cases}$$

**Proposition 3.4** (Properties of Scalar Product)**.** *For any vectors* $\mathbf{u},\mathbf{v},\mathbf{w}$ *and* $\alpha\in\mathbb{R}$ *we have*

1. $\mathbf{u}\cdot\mathbf{v}=\mathbf{v}\cdot\mathbf{u}$,
2. $\mathbf{u}\cdot(\mathbf{v}+\mathbf{w})=\mathbf{u}\cdot\mathbf{v}+\mathbf{u}\cdot\mathbf{w}$,
3. $(\mathbf{u}+\mathbf{v})\cdot\mathbf{w}=\mathbf{u}\cdot\mathbf{w}+\mathbf{v}\cdot\mathbf{w}$,
4. $(\alpha\mathbf{u})\cdot\mathbf{v}=\mathbf{u}\cdot(\alpha\mathbf{v})=\alpha(\mathbf{u}\cdot\mathbf{v})$.

*Proof.* These are all easy consequences of Theorem 3.3. □

## 3.2 The Equation of a Plane

A plane $\Pi$ in 3-space can be specified by giving

- a point $P$ on $\Pi$
- a non-zero vector $\mathbf{n}$ orthogonal to $\Pi$.

By $\mathbf{n}$ being orthogonal to $\Pi$ we mean that for any two points $A$ and $B$ on $\Pi$, the vector represented by $\overrightarrow{AB}$ is orthogonal to $\mathbf{n}$.

Let $R$ be a point with position vector $\mathbf{r}$. The point $R$ is on $\Pi$ if and only if the vector represented by $\overrightarrow{PR}$ is orthogonal to $\mathbf{n}$. That is, if and only if $(\mathbf{r}-\mathbf{p})\cdot\mathbf{n}=0$. Rearranging this we get the **vector equation** for the plane through $P$ and orthogonal to $\mathbf{n}$ to be

$$\mathbf{r}\cdot\mathbf{n}=\mathbf{p}\cdot\mathbf{n}\,.$$

In coordinates, letting $\mathbf{r}=\begin{pmatrix} x \\ y \\ z \end{pmatrix}$, $\mathbf{n}=\begin{pmatrix} a \\ b \\ c \end{pmatrix}$ (with $a,b,c$ not all $0$) and $\mathbf{p}=\begin{pmatrix} p_1 \\ p_2 \\ p_3 \end{pmatrix}$, we get the **Cartesian equation** for the plane to be

$$ax+by+cz=d.$$

where $d=n_1p_1+n_2p_2+n_3p_3$. That is to say, the point with coordinates $(x,y,z)$ is on $\Pi$ if and only if it satisfies $ax+by+cz=d$.

## 3.3 Distance from a Point to a Plane

Let $\Pi$ be a plane and $Q$ be a point with position vector $\mathbf{q}$. We would like to determine the distance from $Q$ to $\Pi$; that is the distance from $Q$ to $M$ where $M$ is the point on $\Pi$ which is closest to $Q$.

Suppose that $\Pi$ has equation $\mathbf{n} \cdot \begin{pmatrix} x \\ y \\ z \end{pmatrix} = d$ (or equivalently $ax + by + cz = d$). For $M$ to be the point of $\Pi$ closest to $Q$ we need that the vector represented by $\overrightarrow{MQ}$ is orthogonal to the plane $\Pi$ and so is a scalar multiple of $\mathbf{n}$. That is, writing $\mathbf{m}$ for the position vector of $M$, we need $\mathbf{q} - \mathbf{m} = \alpha\mathbf{n}$ for some $\alpha \in \mathbb{R}$. This means that

$$\begin{aligned} (\mathbf{q} - \mathbf{m}) \cdot \mathbf{n} &= (\alpha\mathbf{n}) \cdot \mathbf{n} \\ \mathbf{q} \cdot \mathbf{n} - \mathbf{m} \cdot \mathbf{n} &= \alpha|\mathbf{n}|^2. \end{aligned}$$

but $M$ is on $\Pi$ and so $\mathbf{m} \cdot \mathbf{n} = d$. We get

$$\alpha = \frac{\mathbf{q} \cdot \mathbf{n} - d}{|\mathbf{n}|^2}.$$

Now, the distance from $M$ to $Q$ is

$$|\overrightarrow{MQ}| = |\mathbf{q} - \mathbf{m}| = |\alpha||\mathbf{n}| = \frac{|\mathbf{q} \cdot \mathbf{n} - d|}{|\mathbf{n}|}.$$

Note that, as you would expect, this is 0 if $\mathbf{q} \cdot \mathbf{n} = d$ since in this case $Q$ lies on the plane $\Pi$.

We could also use this method to find the position vector of $M$, the point on $\Pi$ closest to $Q$ by

$$\mathbf{m} = \mathbf{q} - \alpha\mathbf{n} = \mathbf{q} - \left(\frac{\mathbf{q} \cdot \mathbf{n} - d}{|\mathbf{n}|^2}\right)\mathbf{n}.$$

Summarising the above, we have proved the following result:

**Proposition 3.5.** *If the plane $\Pi$ has equation $\mathbf{r} \cdot \mathbf{n} = d$, and the point $Q$ has position vector $\mathbf{q}$, then the distance between $Q$ and $\Pi$ is*

$$\frac{|\mathbf{q} \cdot \mathbf{n} - d|}{|\mathbf{n}|},$$

*and the point on $\Pi$ that is closest to $Q$ has position vector*

$$\mathbf{q} - \left(\frac{\mathbf{q} \cdot \mathbf{n} - d}{|\mathbf{n}|^2}\right)\mathbf{n}\,.$$

## 3.4 The vector product

**Definition 3.6.** Given vectors $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, the **vector product** $\mathbf{u} \times \mathbf{v}$ is defined to be

$$\mathbf{u} \times \mathbf{v} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}.$$

In other words,

$$\mathbf{u} \times \mathbf{v} = (u_2v_3 - u_3v_2)\mathbf{i} - (u_1v_3 - u_3v_1)\mathbf{j} + (u_1v_2 - u_2v_1)\mathbf{k}\,.$$

Note in particular that the vector product of two vectors is itself a vector, and that in general $\mathbf{u} \times \mathbf{v} = -\mathbf{v} \times \mathbf{u}$.

**Example 3.7.** If $\mathbf{u} = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} -1 \\ 3 \\ 4 \end{pmatrix}$ then $\mathbf{u} \times \mathbf{v} = \begin{pmatrix} 11 \\ -3 \\ 5 \end{pmatrix}$ and $\mathbf{v} \times \mathbf{u} = \begin{pmatrix} -11 \\ 3 \\ -5 \end{pmatrix}$.

The slightly strange-looking definition of the vector product can be explained geometrically, by the following result:

**Proposition 3.8.** *Given vectors $\mathbf{u}$ and $\mathbf{v}$, the vector product $\mathbf{u} \times \mathbf{v}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$, and its length $|\mathbf{u} \times \mathbf{v}|$ satisfies*

$$|\mathbf{u} \times \mathbf{v}| = \begin{cases} |\mathbf{u}||\mathbf{v}| \sin\theta & \text{if } \mathbf{u} \neq \mathbf{0}, \mathbf{v} \neq \mathbf{0} \\ 0 & \text{if } \mathbf{u} = \mathbf{0} \text{ or } \mathbf{v} = \mathbf{0} \end{cases}$$

*where $\theta$ denotes the angle between $\mathbf{u}$ and $\mathbf{v}$ (in the case that they are both non-zero).*

*Proof.* To prove orthogonality we need to show that $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{u} = 0$ and $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0$. We calculate

$$(\mathbf{u}\times\mathbf{v})\cdot\mathbf{u} = \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix} \cdot \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix} = (u_2v_3 - u_3v_2)u_1 + (u_3v_1 - u_1v_3)u_2 + (u_1v_2 - u_2v_1)u_3 = 0\,,$$

as required, and the calculation showing $(\mathbf{u} \times \mathbf{v}) \cdot \mathbf{v} = 0$ is similar and left as an exercise.

If $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ then it is easily seen that $\mathbf{u} \times \mathbf{v} = \mathbf{0}$, so that $|\mathbf{u} \times \mathbf{v}| = 0$. If both $\mathbf{u}$ and $\mathbf{v}$ are non-zero then we note that

$$|\mathbf{u} \times \mathbf{v}|^2 = |\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u} \cdot \mathbf{v})^2\,, \tag{3.1}$$

because

$$\begin{aligned} |\mathbf{u} \times \mathbf{v}|^2 &= (u_2v_3 - u_3v_2)^2 + (u_3v_1 - u_1v_3)^2 + (u_1v_2 - u_2v_1)^2 \\ &= u_2^2v_3^2 + u_3^2v_2^2 + u_3^2v_1^2 + u_1^2v_3^2 + u_1^2v_2^2 + u_2^2v_1^2 - 2(u_2v_3u_3v_2 + u_3v_1u_1v_3 + u_1v_2u_2v_1) \\ &= (u_1^2 + u_2^2 + u_3^2)(v_1^2 + v_2^2 + v_3^2) - (u_1v_1 + u_2v_2 + u_3v_3)^2 \\ &= |\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u} \cdot \mathbf{v})^2\,. \end{aligned}$$

Now

$$|\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u} \cdot \mathbf{v})^2 = |\mathbf{u}|^2|\mathbf{v}|^2 - (|\mathbf{u}||\mathbf{v}|\cos\theta)^2 = |\mathbf{u}|^2|\mathbf{v}|^2(1 - \cos^2\theta) = |\mathbf{u}|^2|\mathbf{v}|^2\sin^2\theta\,,$$

so substituting into (3.1) gives $|\mathbf{u}\times\mathbf{v}|^2 = |\mathbf{u}|^2|\mathbf{v}|^2\sin^2\theta$, and hence $|\mathbf{u}\times\mathbf{v}| = |\mathbf{u}||\mathbf{v}|\sin\theta$. □

## 3.5 Vector equation of a plane given 3 points on it

Let $A, B, C$ be points in 3-space which do not all lie on a common line. Let $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ be their respective position vectors. Let $\Pi$ be the plane containing the three points $A, B, C$. Then $\mathbf{n} = (\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})$ is orthogonal to $\Pi$.

A vector equation for $\Pi$ is $\mathbf{r} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$ (since $A$ is on $\Pi$ and $\mathbf{n}$ is orthogonal to $\Pi$). This equation can be written as

$$\mathbf{r} \cdot ((\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})) = \mathbf{a} \cdot ((\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a}))\,,$$

or in other words

$$(\mathbf{r} - \mathbf{a}) \cdot ((\mathbf{b} - \mathbf{a}) \times (\mathbf{c} - \mathbf{a})) = 0\,.$$

## 3.6 Distance from a point to a line

Let $l$ be the line with vector equation $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$. (Recall that this means the point $R$ with position vector $\mathbf{r}$ lies on $l$ if and only if this equation is satisfied.) The line $l$ has the same direction as $\mathbf{u}$ and goes though the point $P$ with position vector $\mathbf{p}$. Let $X$ be a point with position vector $\mathbf{x}$. We wish to find the distance from $l$ to $X$. This means that if $M$ is the point on $l$ which is closest to $X$ we need to find $|\overrightarrow{MX}|$. (Draw a picture.)

Let $\mathbf{v}$ be the vector represented by $\overrightarrow{PX}$ so $\mathbf{v} = \mathbf{x} - \mathbf{p}$. We may assume that $\mathbf{v} \neq \mathbf{0}$ since if $\mathbf{v} = \mathbf{0}$ then $X$ lies on $l$ and we conclude that the distance in question is $0$. We now let $\theta$ be the angle between $\mathbf{u}$ and $\mathbf{v}$. We have that

$$|\overrightarrow{MX}| = |\mathbf{v}| \sin\theta = \frac{|\mathbf{u} \times \mathbf{v}|}{|\mathbf{u}|} = \frac{|\mathbf{u} \times (\mathbf{x} - \mathbf{p})|}{|\mathbf{u}|}.$$

Note that when $\mathbf{u}$ and $\mathbf{v}$ are parallel $X$ lies on $l$ and so the formula above is still valid (it correctly gives the distance as $0$).

## 3.7 Distance between two lines

Let $l_1$ be the line with vector equation $\mathbf{r} = \mathbf{p} + \lambda\mathbf{u}$ and $l_2$ be the line with vector equation $\mathbf{r} = \mathbf{q} + \mu\mathbf{v}$. We wish to find the distance between $l_1$ and $l_2$. Note that in contrast to lines in 2-space, two lines in 3-space will typically neither intersect nor be parallel.

If $\mathbf{u} = \alpha\mathbf{v}$ for some $\alpha \in \mathbb{R}$ then $l_1$ and $l_2$ lie in the same direction and we can find the distance between them by choosing any point $A$ on $l_1$, and then finding the distance from the point $A$ to the line $l_2$ (e.g. by using the method of §3.6).

Suppose that $\mathbf{u}$ and $\mathbf{v}$ are such that we cannot write $\mathbf{u} = \alpha\mathbf{v}$ for $\alpha \in \mathbb{R}$. We wish to choose the point $A$ on $l_1$, and the point $B$ on $l_2$, so that $|\overrightarrow{AB}|$ is minimised. Let $\mathbf{w}$ be the vector represented by $\overrightarrow{AB}$. Ensuring that $|\overrightarrow{AB}|$ is as small as possible means that $\mathbf{w}$ is orthogonal to both $\mathbf{u}$ and $\mathbf{v}$, and so $\mathbf{w} = \alpha(\mathbf{u} \times \mathbf{v})$ for some $\alpha \in \mathbb{R}$. Also

$$\mathbf{w} = \mathbf{b} - \mathbf{a} = \mathbf{q} + \mu\mathbf{v} - \mathbf{p} - \lambda\mathbf{u}$$

for some $\lambda, \mu \in \mathbb{R}$ (since $A$ is on $l_1$ and $B$ is on $l_2$).

Putting this together and taking the scalar product with $\mathbf{u} \times \mathbf{v}$ we get

$$\alpha(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) = (\mathbf{q} + \mu\mathbf{v} - \mathbf{p} - \lambda\mathbf{u}) \cdot (\mathbf{u} \times \mathbf{v}) = (\mathbf{q} - \mathbf{p}) \cdot (\mathbf{u} \times \mathbf{v})$$

(note that the $\mu\mathbf{v} \cdot (\mathbf{u} \times \mathbf{v})$ and $\lambda\mathbf{u} \cdot (\mathbf{u} \times \mathbf{v})$ terms are $0$ because $\mathbf{u}$ and $\mathbf{v}$ are orthogonal to $\mathbf{u} \times \mathbf{v}$.)

Dividing by $(\mathbf{u} \times \mathbf{v}) \cdot (\mathbf{u} \times \mathbf{v}) = |\mathbf{u} \times \mathbf{v}|^2$ in the above equality gives us

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

and therefore

$$|\mathbf{w}| = |\alpha||\mathbf{u} \times \mathbf{v}| = \frac{|(\mathbf{q} - \mathbf{p}) \cdot (\mathbf{u} \times \mathbf{v})|}{|\mathbf{u} \times \mathbf{v}|}$$

is the distance between the two lines $l_1$ and $l_2$.

## 3.8 Intersections of Planes and Systems of Linear Equations

We will write $(x, y, z)$ to mean the point with coordinates $(x, y, z)$ or equivalently the point with position vector $\begin{pmatrix} x \\ y \\ z \end{pmatrix}$.

Let $\Pi$ be the plane with Cartesian equation $ax + by + cz = d$. The set of points on $\Pi$ is

$$\{(x, y, z) : ax + by + cz = d\}.$$

We will be interested in the intersection of a collection of planes, that is the set of points which lie in all of them.

As a warm-up think about how a collection of lines in 2-space may intersect: two lines will typically intersect in one point but may not intersect (if they are parallel) or intersect in a whole line (if they are both the same line), three or more lines will typically not intersect but other configurations are possible.

Similarly, suppose we have $k$ planes in 3-space and want to find all points which lie on all $k$ of them. This intersection will typically be a line if $k = 2$, a point if $k = 3$, and empty if $k \geq 4$, although (as in the lines in 2-space example) there are other possibilities.

Algebraically, we have $k$ planes $\Pi_1, \ldots, \Pi_k$ with Cartesian equations

$$\begin{aligned} a_1x + b_1y + c_1z &= d_1 \\ a_2x + b_2y + c_2z &= d_2 \\ \vdots \quad\ \vdots \quad\ \vdots \\ a_kx + b_ky + c_kz &= d_k. \end{aligned}$$

A point $(p, q, r)$ is in the intersection of the $k$ planes precisely if it a common solution to these $k$ equations.

## 3.9 Intersections of other geometric objects

Finding intersections of geometric objects often reduces to finding solutions to collections of equations of various kinds. Here are two more instances.

- To find the intersection of the plane $\Pi$ with equation $ax + by + cz = d$ and the line $l$ with parametric equations

  $$\left.\begin{aligned} x &= p_1 + \lambda u_1 \\ y &= p_2 + \lambda u_2 \\ z &= p_3 + \lambda u_3 \end{aligned}\right\}$$

  we solve

  $$a(p_1 + \lambda u_1) + b(p_2 + \lambda u_2) + c(p_3 + \lambda u_3) = d$$

  for $\lambda$. Usually there will be a unique solution (reflecting the fact that a plane and a line in 3-space typically intersect in a *single point*). However there are some conditions on $a, b, c, d, \mathbf{p}, \mathbf{u}$ (can you work out these conditions?) which mean that either there are no solutions (corresponding to the case when the line is parallel to the plane), or that every point on the line gives a solution (corresponding to the case when the line is a subset of the plane). Substituting the obtained value of $\lambda$ back into the parametric equations for the line then gives the coordinates of the point of intersection.

- To find the intersection of the line $l_1$ with parametric equations

$$\left.\begin{array}{l} x = p_1 + \lambda u_1 \\ y = p_2 + \lambda u_2 \\ z = p_3 + \lambda u_3 \end{array}\right\}$$

and the line $l_2$ with parametric equations

$$\left.\begin{array}{l} x = q_1 + \mu v_1 \\ y = q_2 + \mu v_2 \\ z = q_3 + \mu v_3 \end{array}\right\}$$

we solve

$$\left.\begin{array}{l} p_1 + \lambda u_1 = q_1 + \mu v_1 \\ p_2 + \lambda u_2 = q_2 + \mu v_2 \\ p_3 + \lambda u_3 = q_3 + \mu v_3 \end{array}\right\}$$

or equivalently solve

$$\left.\begin{array}{l} \lambda u_1 - \mu v_1 = q_1 - p_1 \\ \lambda u_2 - \mu v_2 = q_2 - p_2 \\ \lambda u_3 - \mu v_3 = q_3 - p_3 \end{array}\right\}$$

for $\lambda$ and $\mu$. As there are three equations in two unknowns, there will typically be no solutions (reflecting the fact that two lines in 3-space typically do not intersect).

# Chapter 4

# Systems of Linear Equations

Systems of linear equations arise frequently in many areas of the sciences, including physics, engineering, business, economics, and sociology. Their systematic study also provided part of the motivation for the development of modern linear algebra at the end of the 19th century. Linear equations are extremely important and in particular in higher dimensions, one aims to have a systematic and efficient way to solve them.

## 4.1 Basic terminology and examples

A **linear equation in $n$ unknowns** is an equation of the form

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

where $a_1, \ldots, a_n$ and $b$ are given real numbers and $x_1, \ldots, x_n$ are variables.

A **system of $m$ linear equations in $n$ unknowns** is a collection of equations of the form

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

where the $a_{ij}$'s and $b_i$'s are all real numbers. We also call such systems $m \times n$ **systems**.

**Example 4.1.**

$$(a)\ \begin{array}{rcrcl} 2x_1 & + & x_2 & = & 4 \\ 3x_1 & + & 2x_2 & = & 7 \end{array} \qquad (b)\ \begin{array}{rcrcrcl} x_1 & + & x_2 & - & x_3 & = & 3 \\ 2x_1 & - & x_2 & + & x_3 & = & 6 \end{array} \qquad (c)\ \begin{array}{rcrcl} x_1 & - & x_2 & = & 0 \\ x_1 & + & x_2 & = & 3\,. \\ & & x_2 & = & 1 \end{array}$$

(a) is a $2 \times 2$ system, (b) is a $2 \times 3$ system, and (c) is a $3 \times 2$ system.

A **solution** of an $m \times n$ system is an ordered $n$-tuple $(x_1, x_2, \ldots, x_n)$ of specific real values of $x_1, \ldots, x_n$ that satisfies all equations of the system.

**Example 4.2.** $(1, 2)$ is a solution of Example 4.1 (a).
For each $\alpha \in \mathbb{R}$, the 3-tuple $(3, \alpha, \alpha)$ is a solution of Example 4.1 (b) (Exercise: check this!).
Example 4.1 (c) has no solution, since, on the one hand $x_2 = 1$ by the last equation, but the first equation implies $x_1 = 1$, while the second equation implies $x_1 = 2$, which is impossible.

A system with no solution is called **inconsistent**, while a system with at least one solution is called **consistent**.

The set of all solutions of a system is called its **solution set**, which may be empty if the system is inconsistent.

The basic problem we want to address in this section is the following: given an arbitrary $m \times n$ system, determine its solution set. Later on, we will discuss a procedure that provides a complete and practical solution to this problem (the so-called 'Gaussian algorithm'). Before we encounter this procedure, we require a bit more terminology.

**Definition 4.3.** Two $m \times n$ systems are said to be **equivalent**, if they have the same solution set.

**Example 4.4.** Consider the two systems

$$(a) \quad \begin{array}{rcrcrcr} 5x_1 & - & x_2 & + & 2x_3 & = & -3 \\ & & x_2 & & & = & 2 \\ & & & & 3x_3 & = & 6 \end{array} \qquad (b) \quad \begin{array}{rcrcrcr} 5x_1 & - & x_2 & + & 2x_3 & = & -3 \\ -5x_1 & + & 2x_2 & - & 2x_3 & = & 5 \\ 5x_1 & - & x_2 & + & 5x_3 & = & 3 \end{array} .$$

System (a) is easy to solve: looking at the last equation we find first that $x_3 = 2$; the second from bottom equation implies $x_2 = 2$; and finally the first equation yields $x_1 = (-3 + x_2 - 2x_3)/5 = -1$. So the solution set of this system is $\{(-1, 2, 2)\}$.

To find the solution of system (b), add the first and the second equation. Then $x_2 = 2$, while subtracting the first from the third equation gives $3x_3 = 6$, that is $x_3 = 2$. Finally, the first equation now gives $x_1 = (-3 + x_2 - 2x_3)/5 = -1$, so the solution set is again $\{(-1, 2, 2)\}$.

Thus the systems (a) and (b) are equivalent.

In solving system (b) above we have implicitly used the following important observation:

**Lemma 4.5.** *The following operations do not change the solution set of a linear system:*

*(i) interchanging two equations;*

*(ii) multiplying an equation by a non-zero scalar;*

*(iii) adding a multiple of one equation to another.*

*Proof.* (i) and (ii) are obvious. (iii) is a simple consequence of the fact that these equations are linear equations. □

We shall see shortly how to use the above operations systematically to obtain the solution set of any given linear system. Before doing so, however, we introduce a useful short-hand.

An $m \times n$ **matrix** is a rectangular array of real numbers:

$$\begin{pmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{m1} & \cdots & c_{mn} \end{pmatrix} .$$

Given an $m \times n$ linear system

$$\begin{array}{rcl} a_{11}x_1 + \cdots + a_{1n}x_n & = & b_1 \\ \vdots & & \\ a_{m1}x_1 + \cdots + a_{mn}x_n & = & b_m \end{array}$$

we call the array

$$\left(\begin{array}{ccc|c} a_{11} & \cdots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{mn} & b_m \end{array}\right)$$

the **augmented matrix** of the linear system, and the $m \times n$ matrix

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

the **coefficient matrix** of the linear system.

**Example 4.6.**

$$\text{system: } \begin{array}{rcrcrl} 3x_1 & + & 2x_2 & - & x_3 & = 5 \\ 2x_1 & & & + & x_3 & = -1 \end{array} \qquad \text{augmented matrix: } \left(\begin{array}{ccc|c} 3 & 2 & -1 & 5 \\ 2 & 0 & 1 & -1 \end{array}\right).$$

A system can be solved by performing operations on the augmented matrix. Corresponding to the three operations given in Lemma 4.5 we have the following three operations that can be applied to the augmented matrix, called **elementary row operations**.

**Definition 4.7** (Elementary row operations)**.**

- **Type I** interchanging two rows;
- **Type II** multiplying a row by a non-zero scalar;
- **Type III** adding a multiple of one row to another row.

## 4.2 Gaussian elimination

Gaussian elimination is a systematic procedure to determine the solution set of a given linear system. The basic idea is to perform elementary row operations on the corresponding augmented matrix bringing it to a simpler form from which the solution set is readily obtained.

The simple form alluded to above is given in the following definition.

**Definition 4.8.** A matrix is said to be in **row echelon form** if it satisfies the following three conditions:

(i) All zero rows (consisting entirely of zeros) are at the bottom.

(ii) The first non-zero entry from the left in each nonzero row is a 1, called the **leading** $1$ for that row.

(iii) Each leading 1 is to the right of all leading 1's in the rows above it.

A row echelon matrix is said to be in **reduced row echelon form** if, in addition it satisfies the following condition:

(iv) Each leading 1 is the only nonzero entry in its column

Roughly speaking, a matrix is in row echelon form if the leading 1's form an echelon (that is, a 'steplike') pattern.

**Example 4.9.** Matrices in row echelon form:

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

Matrices in reduced row echelon form:

$$\begin{pmatrix}1&2&0&1&0\\0&0&1&2&0\\0&0&0&0&1\end{pmatrix}, \qquad \begin{pmatrix}1&5&0&2\\0&0&1&-1\\0&0&0&0\end{pmatrix}, \qquad \begin{pmatrix}1&0&3\\0&1&2\\0&0&0\end{pmatrix}.$$

The variables corresponding to the leading $1$'s of the augmented matrix in row echelon form will be referred to as the **leading variables**, the remaining ones as the **free variables**.

**Example 4.10.**

(a) $\left(\begin{array}{cccc|c}1&2&3&-4&6\\0&0&1&2&3\end{array}\right)$.
Leading variables: $x_1$ and $x_3$; free variables: $x_2$ and $x_4$.

(b) $\left(\begin{array}{cc|c}1&0&5\\0&1&3\end{array}\right)$.
Leading variables: $x_1$ and $x_2$; no free variables.

Note that if the augmented matrix of a system is in row echelon form, the solution set is easily obtained.

**Example 4.11.** Determine the solution set of the systems given by the following augmented matrices in row echelon form:

$$\text{(a)} \left(\begin{array}{ccc|c}1&3&0&2\\0&0&0&1\end{array}\right), \qquad \text{(b)} \left(\begin{array}{cccc|c}1&-2&0&1&2\\0&0&1&-2&1\\0&0&0&0&0\end{array}\right).$$

*Solution.* (a) The corresponding system is

$$\begin{array}{ccccc}x_1&+&3x_2&=&2\\&&0&=&1\end{array}$$

so the system is inconsistent and the solution set is empty.

(b) The corresponding system is

$$\begin{array}{ccccccccc}x_1&-&2x_2&&&+&x_4&=&2\\&&&&x_3&-&2x_4&=&1\\&&&&&&0&=&0\end{array}$$

We can express the leading variables in terms of the free variables $x_2$ and $x_4$. So set $x_2 = \alpha$ and $x_4 = \beta$, where $\alpha$ and $\beta$ are arbitrary real numbers. The second line now tells us that $x_3 = 1 + 2x_4 = 1 + 2\beta$, and then the first line that $x_1 = 2 + 2x_2 - x_4 = 2 + 2\alpha - \beta$. Thus the solution set is $\{\, (2 + 2\alpha - \beta, \alpha, 1 + 2\beta, \beta) \mid \alpha, \beta \in \mathbb{R} \,\}$. $\square$

It turns out that every matrix can be brought into row echelon form using only elementary row operations. The procedure is known as the

**Gaussian algorithm**:

Step 1 If the matrix consists entirely of zeros, stop — it is already in row echelon form.

Step 2 Otherwise, find the first column from the left containing a non-zero entry (call it $a$), and move the row containing that entry to the top position.

Step 3 Now multiply that row by $1/a$ to create a leading 1.

Step 4 By subtracting multiples of that row from rows below it, make each entry below the leading 1 zero.

This completes the first row. All further operations are carried out on the other rows.

Step 5 Repeat steps 1-4 on the matrix consisting of the remaining rows

The process stops when either no rows remain at Step 5 or the remaining rows consist of zeros.

**Example 4.12.** Solve the following system using the Gaussian algorithm:

$$\begin{array}{rcrcrl} & & x_2 & + & 6x_3 & = 4 \\ 3x_1 & - & 3x_2 & + & 9x_3 & = -3 \\ 2x_1 & + & 2x_2 & + & 18x_3 & = 8 \end{array}$$

*Solution.* Performing the Gaussian algorithm on the augmented matrix gives:

$$\left(\begin{array}{ccc|c} 0 & 1 & 6 & 4 \\ 3 & -3 & 9 & -3 \\ 2 & 2 & 18 & 8 \end{array}\right) \sim^{R_1 \leftrightarrow R_2} \left(\begin{array}{ccc|c} 3 & -3 & 9 & -3 \\ 0 & 1 & 6 & 4 \\ 2 & 2 & 18 & 8 \end{array}\right) \sim^{\frac{1}{3}R_1} \left(\begin{array}{ccc|c} 1 & -1 & 3 & -1 \\ 0 & 1 & 6 & 4 \\ 2 & 2 & 18 & 8 \end{array}\right)$$

$$\sim_{R_3 - 2R_1} \left(\begin{array}{ccc|c} 1 & -1 & 3 & -1 \\ 0 & 1 & 6 & 4 \\ 0 & 4 & 12 & 10 \end{array}\right) \sim_{R_3 - 4R_2} \left(\begin{array}{ccc|c} 1 & -1 & 3 & -1 \\ 0 & 1 & 6 & 4 \\ 0 & 0 & -12 & -6 \end{array}\right) \sim_{-\frac{1}{12}R_3} \left(\begin{array}{ccc|c} 1 & -1 & 3 & -1 \\ 0 & 1 & 6 & 4 \\ 0 & 0 & 1 & \frac{1}{2} \end{array}\right),$$

where the last matrix is now in row echelon form. The corresponding system reads:

$$\begin{array}{rcrcrcr} x_1 & - & x_2 & + & 3x_3 & = & -1 \\ & & x_2 & + & 6x_3 & = & 4 \\ & & & & x_3 & = & \frac{1}{2} \end{array}$$

Leading variables are $x_1$, $x_2$ and $x_3$; there are no free variables. The last equation now implies $x_3 = \frac{1}{2}$; the second equation from bottom yields $x_2 = 4 - 6x_3 = 1$ and finally the first equation yields $x_1 = -1 + x_2 - 3x_3 = -\frac{3}{2}$. Thus the solution is $\left\{(-\frac{3}{2}, 1, \frac{1}{2})\right\}$. □

A variant of the Gauss algorithm is the Gauss-Jordan algorithm, which brings a matrix to reduced row echelon form:

**Gauss-Jordan algorithm**

Step 1 Bring matrix to row echelon form using the Gaussian algorithm.

Step 2 Find the row containing the first leading 1 from the right, and add suitable multiples of this row to the rows above it to make each entry above the leading 1 zero.

This completes the first non-zero row from the bottom. All further operations are carried out on the rows above it.

Step 3 Repeat steps 1-2 on the matrix consisting of the remaining rows.

**Example 4.13.** Solve the following system using the Gauss-Jordan algorithm:

$$\begin{array}{ccccccccccc} x_1 & + & x_2 & + & x_3 & + & x_4 & + & x_5 & = & 4 \\ x_1 & + & x_2 & + & x_3 & + & 2x_4 & + & 2x_5 & = & 5 \\ x_1 & + & x_2 & + & x_3 & + & 2x_4 & + & 3x_5 & = & 7 \end{array}$$

*Solution.* Performing the Gauss-Jordan algorithm on the augmented matrix gives:

$$\left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 4 \\ 1 & 1 & 1 & 2 & 2 & 5 \\ 1 & 1 & 1 & 2 & 3 & 7 \end{array}\right) \sim \begin{array}{c} \\ R_2 - R_1 \\ R_3 - R_1 \end{array} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 2 & 3 \end{array}\right) \sim \begin{array}{c} \\ \\ R_3 - R_2 \end{array} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 1 & 4 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right)$$

$$\begin{array}{c} R_1 - R_3 \\ \sim R_2 - R_3 \\ \\ \end{array} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right) \sim \begin{array}{c} R_1 - R_2 \\ \\ \\ \end{array} \left(\begin{array}{ccccc|c} 1 & 1 & 1 & 0 & 0 & 3 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 & 1 & 2 \end{array}\right),$$

where the last matrix is now in reduced row echelon form. The corresponding system reads:

$$\begin{array}{ccccccccc} x_1 & + & x_2 & + & x_3 & & & = & 3 \\ & & & & & x_4 & & = & -1 \\ & & & & & & x_5 & = & 2 \end{array}$$

Leading variables are $x_1$, $x_4$, and $x_5$; free variables $x_2$ and $x_3$. Now set $x_2 = \alpha$ and $x_3 = \beta$, and solve for the leading variables starting from the last equation. This yields $x_5 = 2$, $x_4 = -1$, and finally $x_1 = 3 - x_2 - x_3 = 3 - \alpha - \beta$. Thus the solution set is $\{\, (3 - \alpha - \beta, \alpha, \beta, -1, 2) \mid \alpha, \beta \in \mathbb{R} \,\}$. □

We have just seen that any matrix can be brought to (reduced) row echelon form using only elementary row operations, and moreover that there is an explicit procedure to achieve this (namely the Gaussian and Gauss-Jordan algorithm). We record this important insight for later use:

**Theorem 4.14.**

(a) *Every matrix can be brought to row echelon form by a series of elementary row operations.*

(b) *Every matrix can be brought to reduced row echelon form by a series of elementary row operations.*

*Proof.* For (a): apply the Gaussian algorithm; for (b): apply the Gauss-Jordan algorithm. □

**Remark 4.15.** It can be shown (but not in this module) that the reduced row echelon form of a matrix is unique. On the contrary, this is not the case for just the row echelon form.

The remark above implies that if a matrix is brought to reduced row echelon form by any sequence of elementary row operations (that is, not necessarily by those prescribed by the Gauss-Jordan algorithm) the leading ones will nevertheless always appear in the same positions.

## 4.3 Special classes of linear systems

In this last section of the chapter we'll have a look at a number of special types of linear systems and derive the first important consequences of the fact that every matrix can be brought to row echelon form by a series of elementary row operations.

We start with the following classification of linear systems:

**Definition 4.16.** An $m \times n$ linear system is said to be

- **overdetermined** if it has more equations than unknowns (i.e. $m > n$);
- **underdetermined** if it has fewer equations than unknowns (i.e. $m < n$).

Note that overdetermined systems are usually (but not necessarily) inconsistent. Underdetermined systems may or may not be consistent. However, if they are consistent, then they necessarily have infinitely many solutions:

**Theorem 4.17.** *If an underdetermined system is consistent, it must have infinitely many solutions.*

*Proof.* Note that the row echelon form of the augmented matrix of the system has $r \leq m$ non-zero rows. Thus there are $r$ leading variables, and consequently $n - r \geq n - m > 0$ free variables. □

Another useful classification of linear systems is the following:

**Definition 4.18.** A linear system

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} \tag{4.1}$$

is said to be **homogeneous** if $b_i = 0$ for all $i$. Otherwise it is said to be **inhomogeneous**.

Given an inhomogeneous system (4.1), call the system obtained by setting all $b_i$'s to zero, the **associated homogeneous system** .

**Example 4.19.**

$$\underbrace{\begin{array}{rcrcrcr} 3x_1 & + & 2x_2 & + & 5x_3 & = & 2 \\ 2x_1 & - & x_2 & + & x_3 & = & 5 \end{array}}_{\text{inhomogeneous system}} \qquad \underbrace{\begin{array}{rcrcrcr} 3x_1 & + & 2x_2 & + & 5x_3 & = & 0 \\ 2x_1 & - & x_2 & + & x_3 & = & 0 \end{array}}_{\text{associated homogeneous system}}$$

The first observation about homogeneous systems is that they always have a solution, the so-called **trivial** or **zero** solution: $(0, 0, \ldots, 0)$.

For later use we record the following useful consequence of the previous theorem on consistent homogeneous systems:

**Theorem 4.20.** *An underdetermined homogeneous system always has non-trivial solutions.*

*Proof.* We just observed that a homogeneous system is consistent. Thus, if the system is underdetermined and homogeneous, it must have infinitely many solutions by Theorem 4.17, hence, in particular, it must have a non-zero solution. □

Our final result in this section is devoted to the special case of $n \times n$ systems. For such systems there is a delightful characterisation of the existence and uniqueness of solutions of a given system in terms of the associated homogeneous systems. At the same time, the proof of this result serves as another illustration of the usefulness of the row echelon form for theoretical purposes.

**Theorem 4.21.** *An $n \times n$ system is consistent and has a unique solution, if and only if the only solution of the associated homogeneous system is the zero solution.*

*Proof.* Follows from the following two observations:

- The same sequence of elementary row operations that brings the augmented matrix of a system to row echelon form, also brings the augmented matrix of the associated homogeneous system to row echelon form, and vice versa.

- An $n \times n$ system in row echelon form has a unique solution precisely if there are $n$ leading variables.

Thus, if an $n \times n$ system is consistent and has a unique solution, the corresponding homogeneous system must have a unique solution, which is necessarily the zero solution.

Conversely, if the associated homogeneous system of a given system has the zero solution as its unique solution, then the original inhomogeneous system must have a solution, and this solution must be unique. $\square$

# Chapter 5

# Matrices

In this chapter we give basic rules and definitions that are necessary for doing calculations with matrices in an efficient way. We will then consider the inverse of a matrix, the transpose of a matrix, and what is meant by the concept of a symmetric matrix. A highlight in the later sections is the Invertible Matrix Theorem.

## 5.1 Matrices and basic properties

An $m \times n$ **matrix** $A$ is a rectangular array of scalars (real numbers)

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

We write $A = (a_{ij})_{m\times n}$ or simply $A = (a_{ij})$ to denote an $m \times n$ matrix whose $(i, j)$-entry is $a_{ij}$, i.e. $a_{ij}$ is the $i$-th row and in the $j$-th column.

If $A = (a_{ij})_{m\times n}$ we say that $A$ has **size** $m \times n$. An $n \times n$ matrix is said to be **square**.

**Example 5.1.** If

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

then $A$ is a matrix of size $2 \times 3$. The $(1, 2)$-entry of $A$ is 3 and the $(2, 3)$-entry of $A$ is 0.

**Definition 5.2** (Equality)**.** Two matrices $A$ and $B$ are **equal**, and we write $A = B$, if they have the same size and $a_{ij} = b_{ij}$ where $A = (a_{ij})$ and $B = (b_{ij})$.

**Definition 5.3** (Scalar multiplication)**.** If $A = (a_{ij})_{m\times n}$ and $\alpha$ is a scalar, then $\alpha A$ is the $m \times n$ matrix whose $(i, j)$-entry is $\alpha a_{ij}$.

**Definition 5.4** (Addition)**.** If $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{m\times n}$ then the **sum** $A + B$ of $A$ and $B$ is the $m \times n$ matrix whose $(i, j)$-entry is $a_{ij} + b_{ij}$.

**Example 5.5.** Let

$$A = \begin{pmatrix} 2 & 3 \\ -1 & 2 \\ 4 & 0 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ -2 & 1 \end{pmatrix}.$$

Then

$$3A + 2B = \begin{pmatrix} 6 & 9 \\ -3 & 6 \\ 12 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 2 \\ 4 & 6 \\ -4 & 2 \end{pmatrix} = \begin{pmatrix} 6 & 11 \\ 1 & 12 \\ 8 & 2 \end{pmatrix}.$$

**Definition 5.6** (Zero matrix)**.** We write $O_{m\times n}$ or simply $O$ (if the size is clear from the context) for the $m \times n$ matrix all of whose entries are zero, and call it a **zero matrix**.

Scalar multiplication and addition of matrices satisfy the following rules:

**Theorem 5.7.** *Let $A$, $B$ and $C$ be matrices of the same size, and let $\alpha$ and $\beta$ be scalars. Then:*

*(a)* $A + B = B + A$;

*(b)* $A + (B + C) = (A + B) + C$;

*(c)* $A + O = A$;

*(d)* $A + (-A) = O$, *where* $-A = (-1)A$;

*(e)* $\alpha(A + B) = \alpha A + \alpha B$;

*(f)* $(\alpha + \beta)A = \alpha A + \beta A$;

*(g)* $(\alpha\beta)A = \alpha(\beta A)$;

*(h)* $1A = A$.

*Proof.* We prove part (b) only, leaving the other parts as exercises.
For part (b), $B + C$ is an $m \times n$ matrix and so $A + (B + C)$ is an $m \times n$ matrix.
The $ij$-entry of $B + C$ is $b_{ij} + c_{ij}$ and so the $ij$-entry of $A + (B + C)$ is $a_{ij} + (b_{ij} + c_{ij})$.
Similarly, $A + B$ is an $m \times n$ matrix and so $(A + B) + C$ is an $m \times n$ matrix.
The $ij$-entry of $A + B$ is $a_{ij} + b_{ij}$ and so the $ij$-entry of $(A + B) + C$ is $(a_{ij} + b_{ij}) + c_{ij}$.
Since $a_{ij} + (b_{ij} + c_{ij}) = (a_{ij} + b_{ij}) + c_{ij}$ we have that $A + (B + C) = (A + B) + C$. □

**Example 5.8.** Simplify $2(A + 3B) - 3(C + 2B)$, where $A$, $B$, and $C$ are matrices with the same size.

*Solution.*

$$2(A + 3B) - 3(C + 2B) = 2A + 2 \cdot 3B - 3C - 3 \cdot 2B = 2A + 6B - 3C - 6B = 2A - 3C\,.$$

□

**Definition 5.9** (Matrix multiplication)**.** If $A = (a_{ij})$ is an $m \times n$ matrix and $B = (b_{ij})$ is an $n \times p$ matrix then the **product** $AB$ of $A$ and $B$ is the $m \times p$ matrix $C = (c_{ij})$ with

$$c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}\,.$$

**Example 5.10.** Compute the $(1, 3)$-entry and the $(2, 4)$-entry of $AB$, where

$$A = \begin{pmatrix} 3 & -1 & 2 \\ 0 & 1 & 4 \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 2 & 1 & 6 & 0 \\ 0 & 2 & 3 & 4 \\ -1 & 0 & 5 & 8 \end{pmatrix}.$$

*Solution.*

$(1, 3)$-entry: $3 \cdot 6 + (-1) \cdot 3 + 2 \cdot 5 = 25$;

$(2, 4)$-entry: $0 \cdot 0 + 1 \cdot 4 + 4 \cdot 8 = 36$. □

**Definition 5.11** (Identity matrix)**.** An **identity matrix** $I$ is a square matrix with 1's on the diagonal and zeros elsewhere. If we want to emphasise its size we write $I_n$ for the $n \times n$ identity matrix.

Matrix multiplication satisfies the following rules:

**Theorem 5.12.** *Let $A = (a_{ij})_{m\times n}$, $B = (b_{ij})_{m\times n}$, $C = (c_{ij})_{n\times p}$ and $D = (d_{ij})_{n\times p}$ be matrices and $\alpha \in \mathbb{R}$. Then,*

*(a)* $(A+B)C = AC + BC$ *and* $A(C+D) = AC + AD$*;*

*(b)* $\alpha(AC) = (\alpha A)C = A(\alpha C)$*;*

*(c)* $I_m A = A I_n = A$*;*

*Let $X = (x_{ij})_{m\times n}$, $Y = (y_{ij})_{n\times p}$ and $Z = (z_{ij})_{p\times q}$. Then,*

*(d)* $(XY)Z = X(YZ)$*.*

*Proof.* Again we will just prove selected parts, the remainder are similar and left as exercises.

(a) Let $A + B = M = (m_{ij})_{m\times n}$ so $m_{ij} = a_{ij} + b_{ij}$. Now, $MC$ is an $m \times p$ matrix. Also $AC$ and $BC$ are $m \times p$ matrices and so $AC + BC$ is an $m \times p$ matrix.

The $ij$-entry of $MC$ is

$$\begin{aligned}\sum_{k=1}^{n} m_{ik}c_{kj} &= \sum (a_{ik} + b_{ik})c_{kj} \\ &= \sum_{k=1}^{n} a_{ik}c_{kj} + \sum_{k=1}^{n} b_{ik}c_{kj} \\ &= (ij\text{-entry of } AC) + (ij\text{-entry of } BC) \\ &= (ij\text{-entry of } AC + BC)\end{aligned}$$

It follows that $(A + B)C = AC + BC$.

The second identity in part (a) is proved in a similar way

(c) $I_m A$ is an $m \times n$ matrix with $ij$-entry

$$0 \times a_{1j} + 0 \times a_{2j} + \cdots + 1 \times a_{ij} + \cdots + 0 \times a_{mj} = a_{ij}.$$

(where we are multiplying the $a_{ij}$ by the entries in row $i$ of $I_n$). So $I_m A = A$.

Similarly $AI_n$ is an $m \times n$ matrix with $ij$-entry

$$a_{i1} \times 0 + a_{i2} \times 0 + \cdots + a_{ij} \times 1 + \cdots + a_{in} \times 0 = a_{ij}.$$

(where we are multiplying the $a_{ij}$ by the entries in column $j$ of $I_n$). So $AI_n = A$.

(d) Both $(XY)Z$ and $X(YZ)$ are $m \times q$ matrices.

Let $XY = T = (t_{ij})_{m\times p}$ so

$$t_{ij} = x_{i1}y_{1j} + x_{i2}y_{2j} + \cdots + x_{in}y_{nj}.$$

Now $(XY)Z = TZ$ has $ij$-entry

$$\begin{aligned}t_{i1}z_{1j} + t_{i2}z_{2j} + \cdots + t_{ip}z_{pj} = &(x_{i1}y_{11} + x_{i2}y_{21} + \cdots + x_{in}y_{n1})z_{1j} \\ &+ (x_{i1}y_{12} + x_{i2}y_{22} + \cdots + x_{in}y_{n2})z_{2j} \\ &+ \ldots \\ &(x_{i1}y_{1p} + x_{i2}y_{2p} + \cdots + x_{in}y_{np})z_{pj}.\end{aligned}$$

Expanding out the brackets we get that this sum consists of all terms $x_{ir}y_{rs}z_{sj}$ where $r$ ranges over $1, \ldots, n$ and $s$ ranges over $1, \ldots, p$. Equivalently,

$$\text{The } ij\text{-entry of } TZ = \sum_{r=1}^{n} \sum_{s=1}^{p} x_{ir}y_{rs}z_{sj}.$$

A similar calculation of $X(YZ) = ZS$ where $YZ = S$ gives that the $ij$-entry of $X(YZ)$ is the same sum.

This completes the proof.

□

**Notation 5.13.**

- Since $X(YZ) = (XY)Z$, we can omit the brackets and simply write $XYZ$ and similarly for products of more than three factors.
- If $A$ is a square matrix we write $A^k = \underbrace{AA \cdots A}_{k \text{ factors}}$ for the $k$-**th power** of $A$.

$\boxed{\textbf{Warning: } \text{In general } AB \neq BA, \text{ even if } AB \text{ and } BA \text{ have the same size!}}$

**Example 5.14.**

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

but

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

**Definition 5.15.** If $A$ and $B$ are two matrices with $AB = BA$, then $A$ and $B$ are said to **commute**.

We now come to the important notion of an inverse of a matrix.

**Definition 5.16.** If $A$ is a square matrix, a matrix $B$ is called an **inverse** of $A$ if

$$AB = I \quad \text{and} \quad BA = I\,.$$

A matrix that has an inverse is called **invertible**.

Note that not every matrix is invertible. For example the matrix

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

cannot have an inverse since for any $2 \times 2$ matrix $B = (b_{ij})$ we have

$$AB = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ 0 & 0 \end{pmatrix} \neq I_2\,.$$

Later on in this chapter we shall discuss an algorithm that lets us decide whether a matrix is invertible. If the matrix is invertible then this algorithm also tells us exactly what the inverse is. If a matrix is invertible then its inverse is unique, by the following result:

**Theorem 5.17.** *If $B$ and $C$ are both inverses of $A$, then $B = C$.*

*Proof.* Since $B$ and $C$ are inverses of $A$ we have $AB = I$ and $CA = I$. Thus

$$B = IB = (CA)B = C(AB) = CI = C\,.$$

□

If $A$ is an invertible matrix, the unique inverse of $A$ is denoted by $A^{-1}$. Hence $A^{-1}$ (if it exists!) is a square matrix of the same size as $A$ with the property that

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

Note that the above equality implies that if $A$ is invertible, then its inverse $A^{-1}$ is also invertible with inverse $A$, that is,

$$(A^{-1})^{-1} = A\,.$$

Slightly deeper is the following result:

**Theorem 5.18.** *If $A$ and $B$ are invertible matrices of the same size, then $AB$ is invertible and*

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

*Proof.* Observe that

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

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

Thus, by definition of invertibility, $AB$ is invertible with inverse $B^{-1}A^{-1}$. □

## 5.2 Transpose of a matrix

The first new concept we encounter is the following:

**Definition 5.19.** The **transpose** of an $m \times n$ matrix $A = (a_{ij})$ is the $n \times m$ matrix $B = (b_{ij})$ given by

$$b_{ij} = a_{ji}$$

The transpose of $A$ is denoted by $A^T$.

**Example 5.20.**

(a) $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \Rightarrow A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$

(b) $B = \begin{pmatrix} 1 & 2 \\ 3 & -1 \end{pmatrix} \Rightarrow B^T = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}$

Matrix transposition satisfies the following rules:

**Theorem 5.21.** *Assume that $\alpha$ is a scalar and that $A$, $B$, and $C$ are matrices so that the indicated operations can be performed. Then:*

*(a)* $(A^T)^T = A$;

*(b)* $(\alpha A)^T = \alpha(A^T)$;

*(c)* $(A + B)^T = A^T + B^T$;

*(d)* $(AB)^T = B^T A^T$.

*Proof.* (a) is obvious while (b) and (c) are proved as a Coursework exercise. For the proof of (d) assume $A = (a_{ij})_{m\times n}$ and $B = (b_{ij})_{n\times p}$ and write $A^T = (\tilde{a}_{ij})_{n\times m}$ and $B^T = (\tilde{b}_{ij})_{p\times n}$ where

$$\tilde{a}_{ij} = a_{ji} \quad \text{and} \quad \tilde{b}_{ij} = b_{ji}\,.$$

Notice that $(AB)^T$ and $B^T A^T$ have the same size, so it suffices to show that they have the same entries. Now, the $(i, j)$-entry of $B^T A^T$ is

$$\sum_{k=1}^{n} \tilde{b}_{ik}\tilde{a}_{kj} = \sum_{k=1}^{n} b_{ki}a_{jk} = \sum_{k=1}^{n} a_{jk}b_{ki}\,,$$

which is the $(j, i)$-entry of $AB$, that is, the $(i, j)$-entry of $(AB)^T$. Thus $B^T A^T = (AB)^T$. □

Transposition ties in nicely with invertibility:

**Theorem 5.22.** *Let $A$ be invertible. Then $A^T$ is invertible and*

$$(A^T)^{-1} = (A^{-1})^T\,.$$

*Proof.* See a Coursework exercise. □

## 5.3 Special types of square matrices

In this section we briefly introduce a number of special classes of matrices which will be studied in more detail later in this course.

**Definition 5.23.** A matrix is said to be **symmetric** if $A^T = A$.

Note that a symmetric matrix is necessarily square.

**Example 5.24.**

$$\text{symmetric:} \quad \begin{pmatrix} 1 & 2 & 4 \\ 2 & -1 & 3 \\ 4 & 3 & 0 \end{pmatrix}, \quad \begin{pmatrix} 5 & 2 \\ 2 & -1 \end{pmatrix}.$$

$$\text{not symmetric:} \quad \begin{pmatrix} 2 & 2 & 4 \\ 2 & 2 & 3 \\ 1 & 3 & 5 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

Symmetric matrices play an important role in many parts of pure and applied Mathematics as well as in some other areas of science, for example in quantum physics. Some of the reasons for this will become clearer towards the end of this course, when we shall study symmetric matrices in much more detail.

Some other useful classes of square matrices are the triangular ones, which will also play a role later on in the course.

**Definition 5.25.** A square matrix $A = (a_{ij})$ is said to be

| | |
|---|---|
| **upper triangular** | if $a_{ij} = 0$ for $i > j$; |
| **strictly upper triangular** | if $a_{ij} = 0$ for $i \geq j$; |
| **lower triangular** | if $a_{ij} = 0$ for $i < j$; |
| **strictly lower triangular** | if $a_{ij} = 0$ for $i \leq j$; |
| **diagonal** | if $a_{ij} = 0$ for $i \neq j$. |

If $A = (a_{ij})$ is a square matrix of size $n \times n$, we call $a_{11}, a_{22}, \ldots, a_{nn}$ the **diagonal entries** of $A$. So, informally speaking, a matrix is upper triangular if all the entries below the diagonal entries are zero, and it is strictly upper triangular if all entries below the diagonal entries and the diagonal entries itself are zero. Similarly for (strictly) lower triangular matrices.

**Example 5.26.**

$$\text{upper triangular:} \quad \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}, \quad \text{diagonal:} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}$$

$$\text{strictly lower triangular:} \quad \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 2 & 3 & 0 \end{pmatrix}.$$

We close this section with the following two observations:

**Theorem 5.27.** *The sum and product of two upper triangular matrices of the same size is upper triangular.*

*Proof.* See a Coursework exercise. □

## 5.4 Column vectors of dimension $n$

Although vectors in 3-space were originally defined geometrically, recall that the introduction of coordinates allowed us to think of vectors as lists of numbers.

Let us write $\mathbb{R}^3 = \left\{ \begin{pmatrix} a \\ b \\ c \end{pmatrix} : a, b, c \in \mathbb{R} \right\}$ for the set of all vectors in 3-space thought of in coordinate form.

More generally, let us write

$$\mathbb{R}^n = \left\{ \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} : a_1, a_2, \ldots, a_n \in \mathbb{R} \right\}.$$

We call this the set of all **(column) vectors of dimension** $n$ (or the set of $n$-**dimensional (column) vectors**). In particular, a column vector of dimension $n$ is just an $n \times 1$ matrix.

We can extend our definitions of how to add two vectors and multiply a vector by a scalar to $\mathbb{R}^n$ by letting:

$$\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix} \quad \text{and} \quad \alpha \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} \alpha a_1 \\ \alpha a_2 \\ \vdots \\ \alpha a_n \end{pmatrix}$$

(that is define addition and scalar multiplication coordinate-wise).

We denote the zero vector in $\mathbb{R}^n$ by $\mathbf{0}_n$ (or simply $\mathbf{0}$ if $n$ is clear from the context).

Note that our definition of the scalar product $\mathbf{u}\cdot\mathbf{v}$ was only made for vectors in $\mathbb{R}^3$ although we could extend it to work in $\mathbb{R}^n$ (using the formula in coordinates). Our definition of the vector product $\mathbf{u} \times \mathbf{v}$ was only made for vectors in $\mathbb{R}^3$ and, in contrast, cannot be extended to $\mathbb{R}^n$.

Working in $\mathbb{R}^n$ for $n > 3$ we lose some geometric intuition. However, the mathematics still makes sense and can be useful. For instance under some circumstances we may want to use the vector $\begin{pmatrix} 1/2 \\ 1/4 \\ 1/8 \\ 1/8 \end{pmatrix}$ to represent the probability distribution

| $k$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $P(X = k)$ | 1/2 | 1/4 | 1/8 | 1/8 |

## 5.5 Linear systems in matrix notation

We shall now have another look at systems of linear equations, this time using the language of matrices to study them.

Suppose that we are given an $m \times n$ linear system

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} . \tag{5.1}$$

The reformulation is based on the observation that we can write this system as a single matrix equation

$$A\mathbf{x} = \mathbf{b}\,, \tag{5.2}$$

where

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n\,, \text{ and } \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m\,,$$

and where $A\mathbf{x}$ is interpreted as the matrix product of $A$ and $\mathbf{x}$.

**Example 5.28.** Using matrix notation the system

$$\begin{array}{rcrcrcr} 2x_1 & - & 3x_2 & + & x_3 & = & 2 \\ 3x_1 & & & - & x_3 & = & -1 \end{array}$$

can be written

$$\underbrace{\begin{pmatrix} 2 & -3 & 1 \\ 3 & 0 & -1 \end{pmatrix}}_{=A} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{=\mathbf{x}} = \underbrace{\begin{pmatrix} 2 \\ -1 \end{pmatrix}}_{=\mathbf{b}} .$$

Apart from obvious notational economy, writing (5.1) in the form (5.2) has a number of other advantages which will become clearer shortly.

## 5.6 Elementary matrices and the Invertible Matrix Theorem

Using the reformulation of linear systems discussed in the previous section we shall now have another look at the process of solving them. Instead of performing elementary row operations we shall now view this process in terms of matrix multiplication. This will shed some light on both matrices and linear systems and will be useful for formulating and proving the main result of this chapter, the Invertible Matrix Theorem, which will be presented towards the end of this section. Before doing so, however, we shall consider the effect of multiplying both sides of a linear system in matrix form by an invertible matrix.

**Lemma 5.29.** *Let $A$ be an $m \times n$ matrix and let $\mathbf{b} \in \mathbb{R}^m$. Suppose that $M$ is an invertible $m \times m$ matrix. The following two systems are equivalent (i.e. they have the same set of solutions):*

$$A\mathbf{x} = \mathbf{b} \tag{5.3}$$

$$MA\mathbf{x} = M\mathbf{b} \tag{5.4}$$

*Proof.* Note that if $\mathbf{x}$ satisfies (5.3), then it clearly satisfies (5.4). Conversely, suppose that $\mathbf{x}$ satisfies (5.4), that is,

$$MA\mathbf{x} = M\mathbf{b} .$$

Since $M$ is invertible, we may multiply both sides of the above equation by $M^{-1}$ from the left to obtain

$$M^{-1}MA\mathbf{x} = M^{-1}M\mathbf{b} ,$$

so $IA\mathbf{x} = I\mathbf{b}$, and hence $A\mathbf{x} = \mathbf{b}$, that is, $\mathbf{x}$ satisfies (5.3). □

We now come back to the idea outlined at the beginning of this section. It turns out that we can 'algebraize' the process of applying an elementary row operation to a matrix $A$ by left-multiplying $A$ by a certain type of matrix, defined as follows:

**Definition 5.30.** An **elementary matrix** of **type I** (respectively, **type II**, **type III**) is a matrix obtained by applying an elementary row operation of type I (respectively, type II, type III) to an identity matrix.

**Example 5.31.**

$$\text{type I:}\quad E_1 = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{(take } I_3 \text{ and swap rows 1 and 2)}$$

$$\text{type II:}\quad E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \qquad \text{(take } I_3 \text{ and multiply row 3 by 4)}$$

$$\text{type III:}\quad E_3 = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \text{(take } I_3 \text{ and add 2 times row 3 to row 1)}$$

Let us now consider the effect of left-multiplying an arbitrary $3 \times 3$ matrix $A$ in turn by each of the three elementary matrices given in the previous example.

**Example 5.32.** Let $A = (a_{ij})_{3\times 3}$ and let $E_l$ $(l = 1, 2, 3)$ be defined as in the previous example. Then

$$E_1 A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

$$E_2 A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 4a_{31} & 4a_{32} & 4a_{33} \end{pmatrix},$$

$$E_3 A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} + 2a_{31} & a_{12} + 2a_{32} & a_{13} + 2a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

You should now pause and marvel at the following observation: interchanging rows 1 and 2 of $A$ produces $E_1 A$, multiplying row 3 of $A$ by 4 produces $E_2 A$, and adding 2 times row 3 to row 1 of $A$ produces $E_3 A$.

This example should convince you of the truth of the following theorem, the proof of which will be omitted as it is straightforward, slightly lengthy and not particularly instructive.

**Theorem 5.33.** *If $E$ is an $m \times m$ elementary matrix obtained from $I$ by an elementary row operation, then left-multiplying an $m \times n$ matrix $A$ by $E$ has the effect of performing that same row operation on $A$.*

Slightly deeper is the following:

**Theorem 5.34.** *If $E$ is an elementary matrix, then $E$ is invertible and $E^{-1}$ is an elementary matrix of the same type.*

*Proof.* The assertion follows from the previous theorem and the observation that an elementary row operation can be reversed by an elementary row operation of the same type. More precisely,

- if two rows of a matrix are interchanged, then interchanging them again restores the original matrix;
- if a row is multiplied by $\alpha \neq 0$, then multiplying the same row by $1/\alpha$ restores the original matrix;

- if $\alpha$ times row $q$ has been added to row $r$, then adding $-\alpha$ times row $q$ to row $r$ restores the original matrix.

Now, suppose that $E$ was obtained from $I$ by a certain row operation. Then, as we just observed, there is another row operation of the same type that changes $E$ back to $I$. Thus there is an elementary matrix $F$ of the same type as $E$ such that $FE = I$. A moment's thought shows that $EF = I$ as well, since $E$ and $F$ correspond to reverse operations. All in all, we have now shown that $E$ is invertible and its inverse $E^{-1} = F$ is an elementary matrix of the same type. $\square$

**Example 5.35.** Determine the inverses of the elementary matrices $E_1$, $E_2$, and $E_3$ in Example 5.31.

*Solution.* In order to transform $E_1$ into $I$ we need to swap rows 1 and 2 of $E_1$. The elementary matrix that performs this feat is

$$E_1^{-1} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Similarly, in order to transform $E_2$ into $I$ we need to multiply row 3 of $E_2$ by $\frac{1}{4}$. Thus

$$E_2^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{4} \end{pmatrix}.$$

Finally, in order to transform $E_3$ into $I$ we need to add $-2$ times row 3 to row 1, and so

$$E_3^{-1} = \begin{pmatrix} 1 & 0 & -2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

$\square$

Before we come to the main result of this chapter we need some more terminology:

**Definition 5.36.** A matrix $B$ is **row equivalent** to a matrix $A$ if there exists a finite sequence $E_1, E_2, \ldots, E_k$ of elementary matrices such that

$$B = E_k E_{k-1} \cdots E_1 A\,.$$

In other words, $B$ is row equivalent to $A$ if and only if $B$ can be obtained from $A$ by a finite number of row operations. In particular, two augmented matrices $(A|\mathbf{b})$ and $(B|\mathbf{c})$ are row equivalent if and only if $A\mathbf{x} = \mathbf{b}$ and $B\mathbf{x} = \mathbf{c}$ are equivalent systems.

The following properties of row equivalent matrices are easily established:

**Fact 5.37.**

(a) $A$ is row equivalent to itself;

(b) if $A$ is row equivalent to $B$, then $B$ is row equivalent to $A$;

(c) if $A$ is row equivalent to $B$, and $B$ is row equivalent to $C$, then $A$ is row equivalent to $C$.

Property (b) follows from Theorem 5.34. Details of the proof of (a), (b), and (c) are left as an exercise.

We are now able to formulate and prove a delightful characterisation of invertibility of matrices. More precisely, the following theorem provides three equivalent conditions for a matrix to be invertible (and later on in this module we will encounter one further equivalent condition).

Before stating the theorem we recall that the **zero vector**, denoted by $\mathbf{0}$, is the column vector all of whose entries are zero.

**Theorem 5.38** (Invertible Matrix Theorem)**.** *Let $A$ be a square $n \times n$ matrix. The following are equivalent:*

*(a) $A$ is invertible;*

*(b) $A\mathbf{x} = \mathbf{0}$ has only the trivial solution;*

*(c) $A$ is row equivalent to $I$;*

*(d) $A$ is a product of elementary matrices.*

*Proof.* We shall prove this theorem using a cyclic argument: we shall first show that (a) implies (b), then (b) implies (c), then (c) implies (d), and finally that (d) implies (a). This is a frequently used trick to show the logical equivalence of a list of assertions.

(a) $\Rightarrow$ (b): Suppose that $A$ is invertible. If $\mathbf{x}$ satisfies $A\mathbf{x} = \mathbf{0}$, then

$$\mathbf{x} = I\mathbf{x} = (A^{-1}A)\mathbf{x} = A^{-1}\mathbf{0} = \mathbf{0}\,,$$

so the only solution of $A\mathbf{x} = \mathbf{0}$ is the trivial solution.

(b) $\Rightarrow$ (c): Use elementary row operations to bring the system $A\mathbf{x} = \mathbf{0}$ to the form $U\mathbf{x} = \mathbf{0}$, where $U$ is in row echelon form. Since, by hypothesis, the solution of $A\mathbf{x} = \mathbf{0}$ and hence the solution of $U\mathbf{x} = \mathbf{0}$ is unique, there must be exactly $n$ leading variables. Thus $U$ is upper triangular with $1$'s on the diagonal, and hence, the reduced row echelon form of $U$ is $I$. Thus $A$ is row equivalent to $I$.

(c) $\Rightarrow$ (d): If $A$ is row equivalent to $I$, then there is a sequence $E_1, \ldots, E_k$ of elementary matrices such that

$$A = E_k E_{k-1} \cdots E_1 I = E_k E_{k-1} \cdots E_1\,,$$

that is, $A$ is a product of elementary matrices.

(d) $\Rightarrow$ (a). If $A$ is a product of elementary matrices, then $A$ must be invertible, since elementary matrices are invertible by Theorem 5.34 and since the product of invertible matrices is invertible by Theorem 5.18. □

An immediate consequence of the previous theorem is the following perhaps surprising result:

**Corollary 5.39.** *Suppose that $A$ and $C$ are square matrices such that $CA = I$. Then also $AC = I$; in particular, both $A$ and $C$ are invertible with $C = A^{-1}$ and $A = C^{-1}$.*

*Proof.* To show that $A$ is invertible, by the Invertible Matrix Theorem it suffices to show that the only solution of $A\mathbf{x} = \mathbf{0}$ is the trivial one. To show this, note that if $A\mathbf{x} = \mathbf{0}$ then $\mathbf{x} = I\mathbf{x} = CA\mathbf{x} = C\mathbf{0} = \mathbf{0}$, as required, so $A$ is indeed invertible. Then note that $C = CI = CAA^{-1} = IA^{-1} = A^{-1}$, so both $A$ and $C$ are invertible, and are the inverses of each other. □

What is surprising about this result is the following: suppose we are given a square matrix $A$. If we want to check that $A$ is invertible, then, by the definition of invertibility, we need to produce a matrix $B$ such that $AB = I$ *and* $BA = I$. The above corollary tells us that if we have a candidate $C$ for an inverse of $A$ it is enough to check that *either* $AC = I$ *or* $CA = I$ in order to guarantee that $A$ is invertible with inverse $C$. This is a non-trivial fact about matrices, which is often useful.

## 5.7 Gauss-Jordan inversion

The Invertible Matrix Theorem provides a simple method for inverting matrices. Recall that the theorem states (amongst other things) that if $A$ is invertible, then $A$ is row equivalent to $I$. Thus there is a sequence $E_1, \ldots\ E_k$ of elementary matrices such that

$$E_k E_{k-1} \cdots E_1 A = I\,.$$

Multiplying both sides of the above equation by $A^{-1}$ from the right yields

$$E_k E_{k-1} \cdots E_1 = A^{-1}\,,$$

that is,

$$E_k E_{k-1} \cdots E_1 I = A^{-1}\,.$$

Thus, the **same sequence** of elementary row operations that brings an invertible matrix to $I$, will bring $I$ to $A^{-1}$. This gives a practical algorithm for inverting matrices, known as Gauss-Jordan inversion.

Note that in the following we use a slight generalisation of the augmented matrix notation. Given an $m \times n$ matrix $A$ and an $m$-dimensional vector $\mathbf{b}$ we currently use $(A|\mathbf{b})$ to denote the $m \times (n+1)$ matrix consisting of $A$ with $\mathbf{b}$ attached as an extra column to the right of $A$, and a vertical line in between them. Suppose now that $B$ is an $m \times r$ matrix then we write $(A|B)$ for the $m \times (n+r)$ matrix consisting of $A$ with $B$ attached to the right of $A$, and a vertical line separating them.

**Gauss-Jordan inversion**

Bring the augmented matrix $(A|I)$ to reduced row echelon form. If $A$ is row equivalent to $I$, then $(A|I)$ is row equivalent to $(I|A^{-1})$. Otherwise, $A$ does not have an inverse.

**Example 5.40.** Show that

$$A = \begin{pmatrix} 1 & 2 & 0 \\ 2 & 5 & 3 \\ 0 & 3 & 8 \end{pmatrix}$$

is invertible and compute $A^{-1}$.

*Solution.* Using Gauss-Jordan inversion we find

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 2 & 5 & 3 & 0 & 1 & 0 \\ 0 & 3 & 8 & 0 & 0 & 1 \end{array}\right) \sim R_2 - 2R_1 \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & -2 & 1 & 0 \\ 0 & 3 & 8 & 0 & 0 & 1 \end{array}\right)$$

$$\underset{R_3 - 3R_2}{\sim} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & -2 & 1 & 0 \\ 0 & 0 & -1 & 6 & -3 & 1 \end{array}\right) \underset{(-1)R_3}{\sim} \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 3 & -2 & 1 & 0 \\ 0 & 0 & 1 & -6 & 3 & -1 \end{array}\right)$$

$$\sim R_2 - 3R_3 \left(\begin{array}{ccc|ccc} 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 16 & -8 & 3 \\ 0 & 0 & 1 & -6 & 3 & -1 \end{array}\right) \overset{R_1 - 2R_2}{\sim} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -31 & 16 & -6 \\ 0 & 1 & 0 & 16 & -8 & 3 \\ 0 & 0 & 1 & -6 & 3 & -1 \end{array}\right).$$

Thus $A$ is invertible (because it is row equivalent to $I_3$) and

$$A^{-1} = \begin{pmatrix} -31 & 16 & -6 \\ 16 & -8 & 3 \\ -6 & 3 & -1 \end{pmatrix} .$$

$\square$

# Chapter 6

# Determinants

We will define the important concept of a determinant, which is a useful invariant for general $n \times n$ matrices. We will discuss the most important properties of determinants, and illustrate what they are good for and how calculations involving determinants can be simplified.

## 6.1 Determinants of $2 \times 2$ and $3 \times 3$ matrices

To every $2 \times 2$ matrix $A$ we associate a scalar, called the determinant of $A$, which is given by a certain sum of products of the entries of $A$:

**Definition 6.1.** Let $A = (a_{ij})$ be a $2 \times 2$ matrix. The **determinant** of $A$, denoted $\det(A)$, is defined by

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12} \,. \tag{6.1}$$

Although this definition of the determinant may look strange and non-intuitive, one of the main motivations for introducing it is that it allows us to decide whether a matrix is invertible or not.

**Theorem 6.2.** *If $A$ and $B$ are $2 \times 2$ matrices then*

*(a)* $\det(AB) = \det(A)\det(B)$*,*

*(b)* $\det(A) \neq 0$ *if and only if $A$ is invertible,*

*(c) If* $A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ *is invertible then*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix} . \tag{6.2}$$

*Proof.* (a) can be proved by a direct calculation.

To prove (b) first note that if $\det(A) = 0$ but $A$ is invertible then $1 = \det(I) = \det(AA^{-1}) = \det(A)\det(A^{-1}) = 0$, which is a contradiction, so $A$ is not invertible. If on the other hand $\det(A) \neq 0$ then the matrix $C$ given by the righthand side of (6.2) is well-defined, and we can calculate directly that $CA = I$. From Corollary 5.39 it follows that $AC = I$ (or alternatively we could show that $AC = I$ by direct calculation as well). Thus $A$ is invertible, so (b) is proved. In fact we have proved (c) as well: the fact that $AC = I = CA$ means that $C = A^{-1}$. $\square$

Our goal in this chapter is to introduce determinants for square matrices of any size, study some of their properties, and then prove the generalisation of the above theorem. However, before considering this very general definition, let us move to the case of $3 \times 3$ determinants:

**Definition 6.3.** If $A = (a_{ij})$ is a $3 \times 3$ matrix, its determinant $\det(A)$ is defined by

$$\begin{aligned}\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}\,.\end{aligned} \tag{6.3}$$

Notice that the determinant of a $3 \times 3$ matrix $A$ is given in terms of the determinants of certain $2 \times 2$ submatrices of $A$.

## 6.2 General definition of determinants

In general, we shall see that the determinant of a $4 \times 4$ matrix is given in terms of the determinants of $3 \times 3$ submatrices, and so forth. Before stating the general definition we introduce a convenient short-hand:

**Notation 6.4.** For any square matrix $A$, let $A_{ij}$ denote the submatrix formed by deleting the $i$-th row and the $j$-th column of $A$.

**Example 6.5.** If

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

then

$$A_{23} = \begin{pmatrix} 3 & 2 & -1 \\ 7 & -2 & 1 \\ 4 & -5 & -4 \end{pmatrix}.$$

If we now define the determinant of a $1 \times 1$ matrix $A = (a_{ij})$ by $\det(A) = a_{11}$, we can re-write (6.1) and (6.3) as follows:

- if $A = (a_{ij})_{2\times 2}$ then

$$\det(A) = a_{11}\det(A_{11}) - a_{21}\det(A_{21})\,;$$

- if $A = (a_{ij})_{3\times 3}$ then

$$\det(A) = a_{11}\det(A_{11}) - a_{21}\det(A_{21}) + a_{31}\det(A_{31})\,.$$

This observation motivates the following recursive definition:

**Definition 6.6.** Let $A = (a_{ij})$ be an $n \times n$ matrix. The **determinant** of $A$, written $\det(A)$, is defined as follows:

- If $n = 1$, then $\det(A) = a_{11}$.

- If $n > 1$ then $\det(A)$ is the sum of $n$ terms of the form $\pm a_{i1}\det(A_{i1})$, with plus and minus signs alternating, and where the entries $a_{11}, a_{21}, \ldots, a_{n1}$ are from the first column of $A$. In symbols:

$$\begin{aligned}\det(A) &= a_{11}\det(A_{11}) - a_{21}\det(A_{21}) + \cdots + (-1)^{n+1}a_{n1}\det(A_{n1}) \\ &= \sum_{i=1}^{n}(-1)^{i+1}a_{i1}\det(A_{i1}).\end{aligned}$$

**Example 6.7.** Compute the determinant of

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

*Solution.*

$$\begin{vmatrix} 0 & 0 & 7 & -5 \\ -2 & 9 & 6 & -8 \\ 0 & 0 & -3 & 2 \\ 0 & 3 & -1 & 4 \end{vmatrix} = -(-2)\begin{vmatrix} 0 & 7 & -5 \\ 0 & -3 & 2 \\ 3 & -1 & 4 \end{vmatrix} = 2\cdot 3\begin{vmatrix} 7 & -5 \\ -3 & 2 \end{vmatrix} = 2\cdot 3\cdot[7\cdot 2-(-3)\cdot(-5)] = -6.$$

□

To state the next theorem, it will be convenient to write the definition of $\det(A)$ in a slightly different form.

**Definition 6.8.** Given a square matrix $A = (a_{ij})$, the $(i,j)$-**cofactor** of $A$ is the number $C_{ij}$ defined by

$$C_{ij} = (-1)^{i+j}\det(A_{ij}).$$

Thus, the definition of $\det(A)$ reads

$$\det(A) = a_{11}C_{11} + a_{21}C_{21} + \cdots + a_{n1}C_{n1}.$$

This is called the **cofactor expansion down the first column of** $A$. There is nothing special about the first column, as the next theorem shows:

**Theorem 6.9** (Cofactor Expansion Theorem)**.** *The determinant of an $n \times n$ matrix $A$ can be computed by a cofactor expansion across any column or row. The expansion down the $j$-th column is*

$$\det(A) = a_{1j}C_{1j} + a_{2j}C_{2j} + \cdots + a_{nj}C_{nj}$$

*and the cofactor expansion across the $i$-th row is*

$$\det(A) = a_{i1}C_{i1} + a_{i2}C_{i2} + \cdots + a_{in}C_{in}.$$

Although this theorem is fundamental for the development of determinants, we shall not prove it here, as it would lead to a rather lengthy workout.

Before moving on, notice that the plus or minus sign in the $(i,j)$-cofactor depends on the position of $a_{ij}$ in the matrix, regardless of $a_{ij}$ itself. The factor $(-1)^{i+j}$ determines the following checkerboard pattern of signs

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

**Example 6.10.** Use a cofactor expansion across the second row to compute $\det(A)$, where

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

*Solution.*

$$\begin{aligned} \det(A) &= a_{21}C_{21} + a_{22}C_{22} + a_{23}C_{23} \\ &= (-1)^{2+1}a_{21}\det(A_{21}) + (-1)^{2+2}a_{22}\det(A_{22}) + (-1)^{2+3}a_{23}\det(A_{23}) \\ &= -0\begin{vmatrix} -1 & 3 \\ 0 & 7 \end{vmatrix} + 0\begin{vmatrix} 4 & 3 \\ 1 & 7 \end{vmatrix} - 2\begin{vmatrix} 4 & -1 \\ 1 & 0 \end{vmatrix} \\ &= -2[4 \cdot 0 - 1 \cdot (-1)] = -2\,. \end{aligned}$$

□

**Example 6.11.** Compute $\det(A)$, where

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

*Solution.* Notice that all entries but the first of row 1 are 0. Thus it will shorten our labours if we expand across the first row:

$$\det(A) = 3\begin{vmatrix} 5 & 0 & 0 & 0 \\ -6 & 4 & -1 & 3 \\ 4 & 0 & 0 & 2 \\ 3 & 1 & 0 & 7 \end{vmatrix}.$$

Again it is advantageous to expand this $4 \times 4$ determinant across the first row:

$$\det(A) = 3 \cdot 5 \cdot \begin{vmatrix} 4 & -1 & 3 \\ 0 & 0 & 2 \\ 1 & 0 & 7 \end{vmatrix}.$$

We have already computed the value of the above $3 \times 3$ determinant in the previous example and found it to be equal to $-2$. Thus $\det(A) = 3 \cdot 5 \cdot (-2) = -30$. □

Notice that the matrix in the previous example was almost lower triangular. The method of this example is easily generalised to prove the following theorem:

**Theorem 6.12.** *If $A$ is either an upper or a lower triangular matrix, then $\det(A)$ is the product of the diagonal entries of $A$.*

## 6.3 Properties of determinants

At several points in this module we have seen that elementary row operations play a fundamental role in matrix theory. It is only natural to enquire how $\det(A)$ behaves when an elementary row operation is applied to $A$.

**Theorem 6.13.** *Let $A$ be a square matrix.*

*(a) If two rows of $A$ are interchanged to produce $B$, then $\det(B) = -\det(A)$.*

*(b) If one row of $A$ is multiplied by $\alpha$ to produce $B$, then $\det(B) = \alpha \det(A)$.*

*(c) If a multiple of one row of $A$ is added to another row to produce a matrix $B$ then $\det(B) = \det(A)$.*

*Proof.* These assertions follow from a slightly stronger result to be proved later in this chapter (see Theorem 6.23). □

**Example 6.14.**

(a) $\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{vmatrix} = -\begin{vmatrix} 4 & 5 & 6 \\ 1 & 2 & 3 \\ 7 & 8 & 9 \end{vmatrix}$ by (a) of the previous theorem.

(b) $\begin{vmatrix} 0 & 1 & 2 \\ 3 & 12 & 9 \\ 1 & 2 & 1 \end{vmatrix} = 3\begin{vmatrix} 0 & 1 & 2 \\ 1 & 4 & 3 \\ 1 & 2 & 1 \end{vmatrix}$ by (b) of the previous theorem.

(c) $\begin{vmatrix} 3 & 1 & 0 \\ 4 & 2 & 9 \\ 0 & -2 & 1 \end{vmatrix} = \begin{vmatrix} 3 & 1 & 0 \\ 7 & 3 & 9 \\ 0 & -2 & 1 \end{vmatrix}$ by (c) of the previous theorem.

The following examples show how to use the previous theorem for the effective computation of determinants:

**Example 6.15.** Compute

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

*Solution.* Perhaps the easiest way to compute this determinant is to spot that when adding two times row 1 to row 3 we get two identical rows, which, by another application of the previous theorem, implies that the determinant is zero:

$$\begin{vmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ -6 & 7 & -7 & 4 \\ -5 & -8 & 0 & 9 \end{vmatrix} \underset{R_3 + 2R_1}{=} \begin{vmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 5 & -3 & -6 \\ -5 & -8 & 0 & 9 \end{vmatrix} \underset{R_3 - R_2}{=} \begin{vmatrix} 3 & -1 & 2 & -5 \\ 0 & 5 & -3 & -6 \\ 0 & 0 & 0 & 0 \\ -5 & -8 & 0 & 9 \end{vmatrix} = 0\,,$$

by a cofactor expansion across the third row. □

**Example 6.16.** Compute $\det(A)$, where

$$A = \begin{pmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ -2 & -5 & 4 & -2 \end{pmatrix}.$$

*Solution.* Here we see that the first column already has two zero entries. Using the previous theorem we can introduce another zero in this column by adding row 2 to row 4. Thus

$$\det(A) = \begin{vmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ -2 & -5 & 4 & -2 \end{vmatrix} = \begin{vmatrix} 0 & 1 & 2 & -1 \\ 2 & 5 & -7 & 3 \\ 0 & 3 & 6 & 2 \\ 0 & 0 & -3 & 1 \end{vmatrix}.$$

If we now expand down the first column we see that

$$\det(A) = -2 \begin{vmatrix} 1 & 2 & -1 \\ 3 & 6 & 2 \\ 0 & -3 & 1 \end{vmatrix}.$$

The $3 \times 3$ determinant above can be further simplified by subtracting 3 times row 1 from row 2. Thus

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

Finally we notice that the above determinant can be brought to triangular form by swapping row 2 and row 3, which changes the sign of the determinant by the previous theorem. Thus

$$\det(A) = (-2) \cdot (-1) \begin{vmatrix} 1 & 2 & -1 \\ 0 & -3 & 1 \\ 0 & 0 & 5 \end{vmatrix} = (-2) \cdot (-1) \cdot 1 \cdot (-3) \cdot 5 = -30\,,$$

by Theorem 6.12. □

We are now able to prove the first important general result about determinants, allowing us to decide whether a matrix is invertible or not by computing its determinant (as such it is a generalisation of the $2 \times 2$ case treated in Theorem 6.2(b)).

**Theorem 6.17.** *A matrix $A$ is invertible if and only if $\det(A) \neq 0$.*

*Proof.* Bring $A$ to row echelon form $U$ (which is then necessarily upper triangular) using elementary row operations. In the process we only ever multiply a row by a non-zero scalar, so Theorem 6.13 implies that $\det(A) = \gamma \det(U)$ for some $\gamma \neq 0$. If $A$ is invertible, then $\det(U) = 1$ by Theorem 6.12, since $U$ is upper triangular with $1$'s on the diagonal, and hence $\det(A) = \gamma \det(U) = \gamma \neq 0$. If $A$ is not invertible then at least one diagonal entry of $U$ is zero, so $\det(U) = 0$ by Theorem 6.12, and hence $\det(A) = \gamma \det(U) = 0$. □

**Definition 6.18.** A square matrix $A$ is called **singular** if $\det(A) = 0$. Otherwise it is said to be **nonsingular**.

**Corollary 6.19.** *A matrix is invertible if and only if it is nonsingular*

Our next result shows what effect transposing a matrix has on its determinant:

**Theorem 6.20.** *If $A$ is an $n \times n$ matrix, then* $\det(A) = \det(A^T)$.

*Proof.* The proof is by induction on $n$ (that is, the size of $A$). The theorem is obvious for $n = 1$. Suppose now that it has already been proved for $k \times k$ matrices for some integer $k$. Our aim now is to show that the assertion of the theorem is true for $(k+1) \times (k+1)$ matrices as well. Let $A$ be a $(k+1) \times (k+1)$ matrix. Note that the $(i, j)$-cofactor of $A$ equals the $(i, j)$-cofactor of $A^T$, because the cofactors involve $k \times k$ determinants only, for which we assumed that the assertion of the theorem holds. Hence

$$\begin{aligned} &\text{cofactor expansion of } \det(A) \text{ across first row} \\ =&\text{cofactor expansion of } \det(A^T) \text{ down first column} \end{aligned}$$

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

Let's summarise: the theorem is true for $1 \times 1$ matrices, and the truth of the theorem for $k \times k$ matrices for some $k$ implies the truth of the theorem for $(k+1) \times (k+1)$ matrices. Thus, the theorem must be true for $2 \times 2$ matrices (choose $k = 1$); but since we now know that it is true for $2 \times 2$ matrices, it must be true for $3 \times 3$ matrices as well (choose $k = 2$); continuing with this process, we see that the theorem must be true for matrices of arbitrary size. □

By the previous theorem, each statement of the theorem on the behaviour of determinants under row operations (Theorem 6.13) is also true if the word 'row' is replaced by 'column', since a row operation on $A^T$ amounts to a column operation on $A$.

**Theorem 6.21.** *Let $A$ be a square matrix.*

*(a) If two columns of $A$ are interchanged to produce $B$, then* $\det(B) = -\det(A)$.

*(b) If one column of $A$ is multiplied by $\alpha$ to produce $B$, then* $\det(B) = \alpha \det(A)$.

*(c) If a multiple of one column of $A$ is added to another column to produce a matrix $B$ then* $\det(B) = \det(A)$.

**Example 6.22.** Find $\det(A)$ where

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

*Solution.* Adding column 1 to column 2 gives

$$\det(A) = \begin{vmatrix} 1 & 3 & 4 & 8 \\ -1 & 2 & 1 & 9 \\ 2 & 5 & 7 & 0 \\ 3 & -4 & -1 & 5 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 4 & 8 \\ -1 & 1 & 1 & 9 \\ 2 & 7 & 7 & 0 \\ 3 & -1 & -1 & 5 \end{vmatrix}.$$

Now subtracting column 3 from column 2 the determinant is seen to vanish by a cofactor expansion down column 2.

$$\det(A) = \begin{vmatrix} 1 & 0 & 4 & 8 \\ -1 & 0 & 1 & 9 \\ 2 & 0 & 7 & 0 \\ 3 & 0 & -1 & 5 \end{vmatrix} = 0.$$

□

Our next aim is to prove that determinants are multiplicative, that is, $\det(AB) = \det(A)\det(B)$ for any two square matrices $A$ and $B$ of the same size. We start by establishing a baby-version of this result, which, at the same time, proves the theorem on the behaviour of determinants under row operations stated earlier (see Theorem 6.13).

**Theorem 6.23.** *If $A$ is an $n \times n$ matrix and $E$ an elementary $n \times n$ matrix, then*

$$\det(EA) = \det(E)\det(A)$$

*with*

$$\det(E) = \begin{cases} -1 & \text{if } E \text{ is of type I (interchanging two rows)} \\ \alpha & \text{if } E \text{ is of type II (multiplying a row by } \alpha\text{)} \\ 1 & \text{if } E \text{ is of type III (adding a multiple of one row to another)} \end{cases} .$$

*Proof.* By induction on the size of $A$. The case where $A$ is a $2 \times 2$ matrix follows from Theorem 6.2(a). Suppose now that the theorem has been verified for determinants of $k \times k$ matrices for some $k$ with $k \geq 2$. Let $A$ be $(k+1) \times (k+1)$ matrix and write $B = EA$. Expand $\det(EA)$ across a row that is unaffected by the action of $E$ on $A$, say, row $i$. Note that $B_{ij}$ is obtained from $A_{ij}$ by the same type of elementary row operation that $E$ performs on $A$. But since these matrices are only $k \times k$, our hypothesis implies that

$$\det(B_{ij}) = r\det(A_{ij}),$$

where $r = -1, \alpha, 1$ depending on the nature of $E$.

Now by a cofactor expansion across row $i$

$$\begin{aligned} \det(EA) = \det(B) &= \sum_{j=1}^{k+1} a_{ij}(-1)^{i+j}\det(B_{ij}) \\ &= \sum_{j=1}^{k+1} a_{ij}(-1)^{i+j} r\det(A_{ij}) \\ &= r\det(A). \end{aligned}$$

In particular, taking $A = I_{k+1}$ we see that $\det(E) = -1, \alpha, 1$ depending on the nature of $E$.

To summarise: the theorem is true for $2 \times 2$ matrices and the truth of the theorem for $k \times k$ matrices for some $k \geq 2$ implies the truth of the theorem for $(k+1) \times (k+1)$ matrices. By the principle of induction the theorem is true for matrices of any size. □

Using the previous theorem we are now able to prove the second important general result of this chapter (and a generalisation of the $2 \times 2$ case treated in Theorem 6.2(a)):

**Theorem 6.24.** *If $A$ and $B$ are square matrices of the same size, then*

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

*Proof.* Case I: If $A$ is not invertible, then neither is $AB$ (for otherwise $A(B(AB)^{-1}) = I$, which by the corollary to the Invertible Matrix Theorem would force $A$ to be invertible). Thus, by Theorem 6.17,

$$\det(AB) = 0 = 0 \cdot \det(B) = \det(A)\det(B).$$

Case II: If $A$ is invertible, then by the Invertible Matrix Theorem $A$ is a product of elementary matrices, that is, there exist elementary matrices $E_1, \ldots, E_k$, such that

$$A = E_k E_{k-1} \cdots E_1 \,.$$

For brevity, write $|A|$ for $\det(A)$. Then, by the previous theorem,

$$\begin{aligned} |AB| &= |E_k \cdots E_1 B| = |E_k||E_{k-1} \cdots E_1 B| = \ldots \\ &= |E_k| \cdots |E_1||B| = \ldots = |E_k \cdots E_1||B| \\ &= |A||B| \,. \end{aligned}$$

□

**Corollary 6.25.** *If $A$ is an invertible matrix then*

$$\textit{det}(A^{-1}) = \frac{1}{\textit{det}(A)}.$$

*Proof.* Since $A$ is invertible, we have $A^{-1}A = I$. Taking determinants of both sides gives $\det(A^{-1}A) = \det(I) = 1$. By Theorem 6.24 we know that $\det(A^{-1}A) = \det(A^{-1})\det(A)$, and so in fact we have $\det(A^{-1})\det(A) = 1$. Moreover, $\det(A) \neq 0$ because $A$ is invertible (by Theorem 6.17), and so we can divide both sides of the preceding equation by $\det(A)$ to obtain the required property

$$\det(A^{-1}) = \frac{1}{\det(A)}.$$

□

# Index