A Summary of

MATH1131 Mathematics 1A

ALGEBRA NOTES

For MATH1231 students, this summary is an extract from the Algebra Notes in the MATH1131/41 Course Pack. It contains all the definitions and theorems, etc., which you have learnt.

If you found any mistakes or typos, please send me an email to chi.mak@unsw.edu.au

Chapter 1

INTRODUCTION TO VECTORS

1.1 Vector quantities

A **scalar** quantity is anything that can be specified by a single number.
A **vector** quantity is one which is specified by both a magnitude and a direction.

Definition 1. *The **zero vector** is the vector* $\mathbf{0}$ *of magnitude zero, and undefined direction.*

1.1.1 Geometric vectors

In figure 1, a vector $\mathbf{a}$ is represented by an arrow with initial point P and terminal point Q. We denote this arrow by $\overrightarrow{PQ}$.

Two vectors are said to be equal if they have the same magnitude and direction. In the figure,

$$\overrightarrow{PQ} = \overrightarrow{AB} = \overrightarrow{EF}.$$

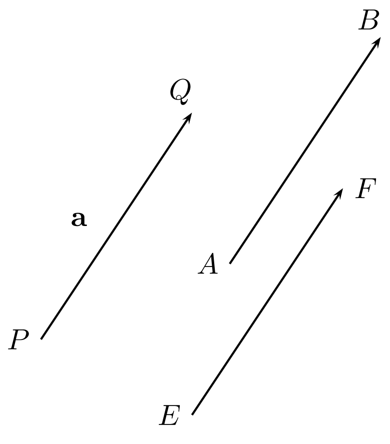

Figure 1.

Definition 2. (The addition of vectors). *On a diagram drawn to scale, draw an arrow representing the vector* $\mathbf{a}$*. Now draw an arrow representing* $\mathbf{b}$ *whose initial point lies at the terminal point of the arrow representing* $\mathbf{a}$*. The arrow which goes from the initial point of* $\mathbf{a}$ *to the terminal point of* $\mathbf{b}$ *represents the vector* $\mathbf{c} = \mathbf{a} + \mathbf{b}$.

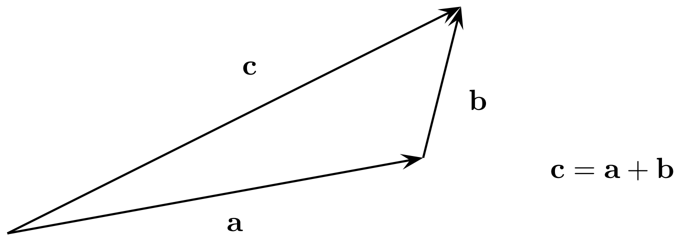

Figure 2: Addition of Vectors.

The second addition rule is known as the parallelogram law for vector addition.

Definition 3. (Alternate definition of vector addition). *On a diagram drawn to scale, draw an arrow representing the vector* $\mathbf{a}$*. Now draw an arrow representing* $\mathbf{b}$ *whose initial point lies at the initial point of the arrow representing* $\mathbf{a}$*. Draw the parallelogram with the two arrows as adjacent sides. The initial point and the two terminal points are vertices of the parallelogram. The arrow which goes from the initial point of* $\mathbf{a}$ *to the fourth vertex of the parallelogram represents the vector* $\mathbf{c} = \mathbf{a} + \mathbf{b}$.

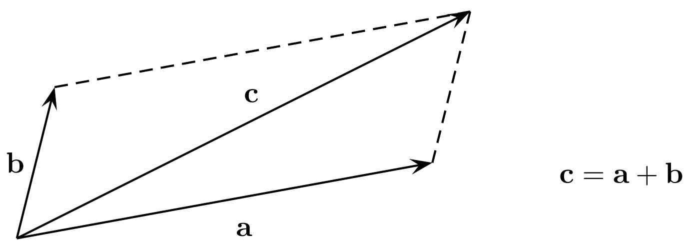

Figure 3: Addition of Vectors.

For any vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$,

$$\mathbf{a} + \mathbf{b} = \mathbf{b} + \mathbf{a}, \qquad (\textbf{Commutative law of vector addition})$$

and
$$(\mathbf{a} + \mathbf{b}) + \mathbf{c} = \mathbf{a} + (\mathbf{b} + \mathbf{c}). \qquad (\textbf{Associative law of vector addition})$$

The law of addition can easily be extended to cover the zero vector by the natural condition that $\mathbf{a} + \mathbf{0} = \mathbf{0} + \mathbf{a} = \mathbf{a}$ for all vectors $\mathbf{a}$. We then can introduce the negative of a vector and the subtraction of vectors.

Definition 4. (Negative of a vector and subtraction). *The* ***negative*** *of* $\mathbf{a}$*, written* $-\mathbf{a}$ *is a vector such that*

$$\mathbf{a} + (-\mathbf{a}) = -\mathbf{a} + \mathbf{a} = \mathbf{0}.$$

If $\mathbf{a}$ *and* $\mathbf{b}$ *are vectors, we define the subtraction by*

$$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b}).$$

Suppose we represent the vectors $\mathbf{a}$ and $\mathbf{b}$ by arrows with the same initial point O. Let P and Q be the terminal points of the two vectors, respectively. From definition, we have

$$\begin{aligned}\overrightarrow{OQ}+\overrightarrow{QP} &= \overrightarrow{OP}\\ \mathbf{b}+\overrightarrow{QP} &= \mathbf{a}\\ \overrightarrow{QP} &= \mathbf{a}-\mathbf{b}\end{aligned}$$

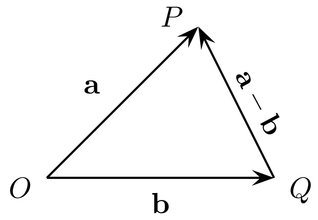

Figure 4: Subtraction of Vectors.

Definition 5. (Multiplication of a vector by a scalar). *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be vectors and let* $\lambda \in \mathbb{R}$.

1. *If* $\lambda > 0$, *then* $\lambda\mathbf{a}$ *is the vector whose magnitude is* $\lambda|\mathbf{a}|$ *and whose direction is the same as that of* $\mathbf{a}$.
2. *If* $\lambda = 0$, *then* $\lambda\mathbf{a} = \mathbf{0}$.
3. *If* $\lambda < 0$, *then* $\lambda\mathbf{a}$ *is the vector whose length is* $|\lambda||\mathbf{a}|$ *and whose direction is the opposite of the direction of* $\mathbf{a}$.

The negative of a vector is the same as the vector multiplied by -1.

$$-\mathbf{a} = (-1)\mathbf{a}.$$

Let $\mathbf{a}$ and $\mathbf{b}$ be vectors, λ and μ be real numbers, then:

$$\lambda(\mu\mathbf{a}) = (\lambda\mu)\mathbf{a}, \qquad \text{(\textbf{Associative law of multiplication by a scalar})}$$
$$(\lambda+\mu)\mathbf{a} = \lambda\mathbf{a}+\mu\mathbf{a}, \qquad \text{(\textbf{Scalar distributive law})}$$
$$\lambda(\mathbf{a}+\mathbf{b}) = \lambda\mathbf{a}+\lambda\mathbf{b}. \qquad \text{(\textbf{Vector distributive law})}$$

1.2 Vector quantities and $\mathbb{R}^n$

1.2.1 Vectors in $\mathbb{R}^2$

We first choose two vectors, conventionally denoted by $\mathbf{i}$ and $\mathbf{j}$, of unit length, and at right angles to each other so that $\mathbf{j}$ is pointing at an angle of $\frac{\pi}{2}$ anticlockwise from $\mathbf{i}$. The vectors $\mathbf{i}$ and $\mathbf{j}$ are known as the standard **basis vectors** for $\mathbb{R}^2$.

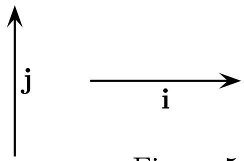

Figure 5.

As shown in Figure 6, every vector **a** can be 'resolved' (in a unique way) into the sum of a scalar multiple of **i** plus a scalar multiple of **j**. That is there are unique real numbers a_1 and a_2 such that $\mathbf{a} = a_1\mathbf{i} + a_2\mathbf{j}$. If the direction θ of **a** is measured from the direction of **i**, then these scalars can be easily found using the formulae

$$a_1 = |\mathbf{a}|\cos\theta, \quad \text{and} \quad a_2 = |\mathbf{a}|\sin\theta.$$

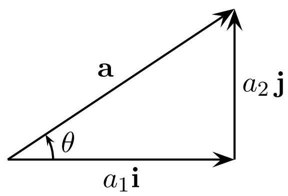

Figure 6.

We call $a_1\mathbf{i}$, $a_2\,\mathbf{j}$ the **components** of **a**.

Now, every vector **a** in the plane can be specified by these two unique real numbers. We can write **a** in form of a **column vector** or a **2-vector** $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$. We call the numbers a_1, a_2 the **components** of $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$. In this case, $\mathbf{i} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{j} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

The column vector $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ is also called the **coordinate vector** with respect to the basis vectors $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$ and a_1, a_2 are also called the **coordinates** of the vector.

Theorem 1. *Let* **a** *and* **b** *be (geometric) vectors, and let* $\lambda \in \mathbb{R}$*. Suppose that* $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ *and* $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$*. Then*

1. *the coordinate vector for* $\mathbf{a} + \mathbf{b}$ *is* $\begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix}$*;*

2. *the coordinate vector for* $\lambda\mathbf{a}$ *is* $\begin{pmatrix} \lambda a_1 \\ \lambda a_2 \end{pmatrix}$*.*

We then can define the mathematics structure $\mathbb{R}^2$, which is the set of 2-vectors, by

$$\mathbb{R}^2 = \left\{ \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} : a_1, a_2 \in \mathbb{R} \right\},$$

with addition and multiplication by a scalar defined by — for any $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$,

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \end{pmatrix} \qquad \text{and} \qquad \lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \end{pmatrix}.$$

The elements in $\mathbb{R}^2$ are called **vectors** and sometimes **column vectors**. It is obvious that the set $\mathbb{R}^2$ is closed under addition and scalar multiplication. Like geometric vectors, the vectors in $\mathbb{R}^2$ also obey the commutative, associative, and distributive laws. There is a zero vector $\mathbf{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and a negative $\begin{pmatrix} -a_1 \\ -a_2 \end{pmatrix}$ for any vector $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$.

1.2.2 Vectors in $\mathbb{R}^n$

Definition 1. *Let n be a positive integer. The set $\mathbb{R}^n$ is defined by*

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

An element $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ *in $\mathbb{R}^n$ is called an n-vector or simply a vector; and $a_1, a_2, \ldots, a_n$ are called the components of the vector.*

Note. We say that two vectors in $\mathbb{R}^n$ are equal if the corresponding components are equal. In other words $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$ if and only if $a_1 = b_1, \ldots, a_n = b_n$.

Definition 2. *Let* $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$ *be vectors in $\mathbb{R}^n$ and λ be a real number.*

We define the sum of $\mathbf{a}$ and $\mathbf{b}$ by $\mathbf{a} + \mathbf{b} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix}$.

We define the scalar multiplication of $\mathbf{a}$ by λ by $\lambda\,\mathbf{a} = \begin{pmatrix} \lambda a_1 \\ \lambda a_2 \\ \vdots \\ \lambda a_n \end{pmatrix}$.

Proposition 2. Let $\mathbf{u}$, $\mathbf{v}$ and $\mathbf{w}$ be vectors in $\mathbb{R}^n$.

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$. (Commutative Law of Addition)

2. $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$. (Associative Law of Addition)

Definition 3. *Let n be a positive integer.*

1. *The **zero vector** in $\mathbb{R}^n$, denoted by $\mathbf{0}$, is the vector with all n components 0.*
2. *Let $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n$. The **negative** of $\mathbf{a}$, denoted by $-\mathbf{a}$ is the vector $\begin{pmatrix} -a_1 \\ \vdots \\ -a_n \end{pmatrix}$.*
3. *Let $\mathbf{a}$, $\mathbf{b}$ be vectors in $\mathbb{R}^n$. We define the difference, $\mathbf{a} - \mathbf{b}$, by*
$$\mathbf{a} - \mathbf{b} = \mathbf{a} + (-\mathbf{b}).$$

NOTE. Let $\mathbf{a}$, $\mathbf{b}$ be vectors in $\mathbb{R}^n$ and $\mathbf{0}$ be the zero vector in $\mathbb{R}^n$.

1. $\mathbf{a} + \mathbf{0} = \mathbf{0} + \mathbf{a} = \mathbf{a}$.
2. $\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} - \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = \begin{pmatrix} a_1 - b_1 \\ \vdots \\ a_n - b_n \end{pmatrix}$.
3. $\mathbf{a} + (-\mathbf{a}) = (-\mathbf{a}) + \mathbf{a} = \mathbf{0}$.

Proposition 3. Let λ, μ be scalars and $\mathbf{u}$, $\mathbf{v}$ be vectors in $\mathbb{R}^n$.

1. $\lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}$ (Associative Law of Scalar Multiplication)
2. $(\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}$ (Scalar Distributive Law)
3. $\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v}$ (Vector distributive Law)

1.3 $\mathbb{R}^n$ and analytic geometry

Definition 1. *Two non-zero vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^n$ are said to be **parallel** if $\mathbf{b} = \lambda\,\mathbf{a}$ for some non-zero real number λ.*
*They are said to be in the same **direction** if $\lambda > 0$.*

Definition 2. *Let $\mathbf{e}_j$ be the vector in $\mathbb{R}^n$ with a 1 in the jth component and 0 for all the other components. Then the n vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ are called the **standard basis vectors** for $\mathbb{R}^n$.*

Definition 3. *The **length** of a vector $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n$ is defined by*
$$|\mathbf{a}| = \sqrt{a_1^2 + \cdots + a_n^2}.$$

Definition 4. *The **distance** between two points A and B with position vectors in $\mathbb{R}^n$ is the length of the vector $\overrightarrow{AB}$.*

Hence, the distance between A and B is

$$\left|\overrightarrow{AB}\right| = |\mathbf{b} - \mathbf{a}| = \sqrt{(b_1 - a_1)^2 + \ldots + (b_n - a_n)^2}.$$

1.4 Lines

Definition 1. *Let $\mathbf{v}$ be any non-zero vector in $\mathbb{R}^n$. The set*

$$S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \lambda \mathbf{v},\ \text{for some } \lambda \in \mathbb{R}\}.$$

*is the **line** in $\mathbb{R}^n$ **spanned** by $\mathbf{v}$, and we call the expression $\mathbf{x} = \lambda\mathbf{v}$, a **parametric vector form** for this line.*

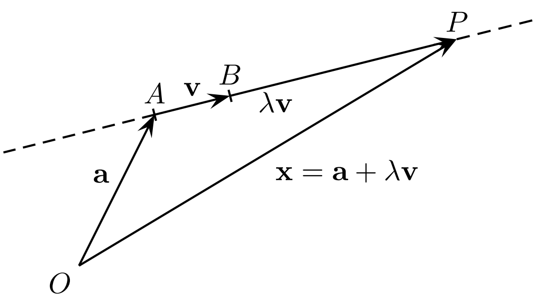

Figure 7.

Definition 2. *A **line** in $\mathbb{R}^n$ is any set of vectors of the form*

$$S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{a} + \lambda \mathbf{v},\ \text{for some } \lambda \in \mathbb{R}\},$$

where $\mathbf{a}$ and $\mathbf{v} \neq \mathbf{0}$ are fixed vectors in $\mathbb{R}^n$. The expression

$$\mathbf{x} = \mathbf{a} + \lambda \mathbf{v},\ \text{for some } \lambda \in \mathbb{R},$$

*is a **parametric vector form** of the line through $\mathbf{a}$ parallel to $\mathbf{v}$.*

1.4.1 Lines in $\mathbb{R}^2$

We can write the equation of a line in Cartesian form $y = mx + d$ and also in parametric vector form $\mathbf{x} = \mathbf{a} + \lambda\mathbf{v}$. You should be able to convert between these two forms.

1.4.2 Lines in $\mathbb{R}^3$

A line in $\mathbb{R}^3$ is still of the form $\mathbf{x} = \mathbf{a} + \lambda\mathbf{v}$, where $\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ are vectors in $\mathbb{R}^3$. Eliminating the parameter λ yields (if all $v_i \neq 0$), the **Cartesian form**

$$\frac{x - a_1}{v_1} = \frac{y - a_2}{v_2} = \frac{z - a_3}{v_3}\ (= \lambda).$$

If v_1, v_2 or v_3 is 0, then x, y or z will, respectively, be constant.

To convert a line from the Cartesian form to parametric vector form, we find a point on the line and a vector parallel to the line. Alternatively, we can introduce a parameter. Put

$$\frac{x-a_1}{v_1} = \frac{y-a_2}{v_2} = \frac{z-a_3}{v_3} = \lambda.$$

1.4.3 Lines in $\mathbb{R}^n$

Parametric vector form of a line through two points which have position vectors **a** and **b** is

$$\mathbf{x} = \mathbf{a} + \lambda\mathbf{v}, \quad \text{where } \mathbf{v} = \mathbf{b} - \mathbf{a}.$$

If none of the components of $\mathbf{v}$ is 0, the **symmetric form**, or the **Cartesian form** of a line through $(a_1, \ldots, a_n)$ parallel $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ is

$$\frac{x_1-a_1}{v_1} = \frac{x_2-a_2}{v_2} = \cdots = \frac{x_n-a_n}{v_n}, \qquad \text{where} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

1.5 Planes

1.5.1 Linear combination and span

Definition 1. *A* ***linear combination*** *of two vectors* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *is a sum of scalar multiples of* $\mathbf{v}_1$ *and* $\mathbf{v}_2$*. That is, it is a vector of the form*

$$\lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2,$$

where λ_1 *and* λ_2 *are scalars.*

Definition 2. *The* ***span*** *of two vectors* $\mathbf{v}_1$ *and* $\mathbf{v}_2$*, written span*$(\mathbf{v}_1, \mathbf{v}_2)$*, is the set of all linear combinations of* $\mathbf{v}_1$ *and* $\mathbf{v}_2$*. That is, it is the set*

$$\mathit{span}(\mathbf{v}_1, \mathbf{v}_2) = \{\mathbf{x} : \ \mathbf{x} = \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2, \text{ for some } \lambda_1, \lambda_2 \in \mathbb{R}\ \}.$$

Theorem 1. *In* $\mathbb{R}^3$*, the span of any two non-zero non-parallel vectors is a plane through the origin.*

We can extend this to $\mathbb{R}^n$ using:

Definition 3. *A* ***plane through the origin*** *is the span of any two (non-zero) non-parallel vectors.*

1.5.2 Parametric vector form of a plane

The span of two non-zero non-parallel vectors is a plane through the origin. For planes that do not pass through the origin we use a similar approach to that we use for lines.

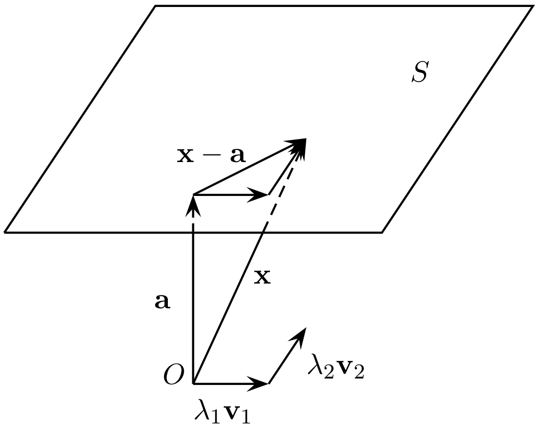

Figure 8: $\mathbf{x} = \mathbf{a} + \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2$.

Definition 4. *Let* $\mathbf{a}$, $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *be fixed vectors in* $\mathbb{R}^n$, *and suppose that* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *are not parallel. Then the set*

$$S = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{a} + \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2,\ \textit{for some } \lambda_1, \lambda_2 \in \mathbb{R}\},$$

is the ***plane*** *through the point with position vector* $\mathbf{a}$, *parallel to the vectors* $\mathbf{v}_1$ *and* $\mathbf{v}_2$. *The expression*

$$\mathbf{x} = \mathbf{a} + \lambda_1\mathbf{v}_1 + \lambda_2\mathbf{v}_2,\ \textit{for } \lambda_1, \lambda_2 \in \mathbb{R},$$

is called a ***parametric vector form*** *of the plane.*

1.5.3 Cartesian form of a plane in $\mathbb{R}^3$

In $\mathbb{R}^3$, any equation of the form

$$ax_1 + bx_2 + cx_3 = d$$

represents plane when not all a, b, c are zero.

In $\mathbb{R}^n$, $n > 3$, if not all a_1, a_2, $\ldots, a_n$ are zero, an equation of the form

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

is not a plane. We call it a hyperplane.

Chapter 2

VECTOR GEOMETRY

Many of the two and three dimensional results can be easily generalised to $\mathbb{R}^n$. The key idea is to use **theorems** in $\mathbb{R}^2$ and $\mathbb{R}^3$ to motivate **definitions** in $\mathbb{R}^n$ for $n > 3$.

2.1 Lengths

Proposition 1. For all $\mathbf{a} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$,

1. $|\mathbf{a}|$ is a real number,
2. $|\mathbf{a}| \geqslant 0$,
3. $|\mathbf{a}| = 0$ if and only if $\mathbf{a} = \mathbf{0}$,
4. $|\lambda\mathbf{a}| = |\lambda|\,|\mathbf{a}|$.

2.2 The dot product

Proposition 1. Cosine Rule for Triangles. If the sides of a triangle in $\mathbb{R}^2$ or $\mathbb{R}^3$ are given by the vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$, then

$$|\mathbf{c}|^2 = |\mathbf{a}|^2 + |\mathbf{b}|^2 - 2|\mathbf{a}|\ |\mathbf{b}|\cos\theta,$$

where θ is the interior angle between $\mathbf{a}$ and $\mathbf{b}$.

If we write $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ and use the formula for the length of $\mathbf{c} = \mathbf{a} - \mathbf{b}$, we have

$$a_1b_1 + a_2b_2 + a_3b_3 = |\mathbf{a}|\ |\mathbf{b}|\cos\theta.$$

For vectors $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}$ and $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n$, we do not have the notion of angle between two vectors yet. We shall define the angle between two vectors via dot product.

> **Definition 1.** *The **dot product** of two vectors* $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ *is*
>
> $$\mathbf{a} \cdot \mathbf{b} = a_1b_1 + \cdots + a_nb_n = \sum_{k=1}^{n} a_kb_k.$$

Notice that $\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^T \mathbf{b}$.

For the special case of $\mathbb{R}^2$ or $\mathbb{R}^3$, we then have

$$\mathbf{a} \cdot \mathbf{b} = |\mathbf{a}|\ |\mathbf{b}| \cos\theta.$$

2.2.1 Arithmetic properties of the dot product

Proposition 2. For all vectors $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$ and scalars $\lambda \in \mathbb{R}$,

1. $\mathbf{a} \cdot \mathbf{a} = |\mathbf{a}|^2$, and hence $|\mathbf{a}| = \sqrt{\mathbf{a} \cdot \mathbf{a}}$;
2. $\mathbf{a} \cdot \mathbf{b}$ is a scalar, i.e., is a real number;
3. **Commutative Law:** $\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$;
4. $\mathbf{a} \cdot (\lambda \mathbf{b}) = \lambda(\mathbf{a} \cdot \mathbf{b})$;
5. **Distributive Law:** $\mathbf{a} \cdot (\mathbf{b} + \mathbf{c}) = \mathbf{a} \cdot \mathbf{b} + \mathbf{a} \cdot \mathbf{c}$;

2.2.2 Geometric interpretation of the dot product in $\mathbb{R}^n$

Definition 2. *If $\mathbf{a}$, $\mathbf{b}$ are non-zero vectors in $\mathbb{R}^n$, then the **angle** θ between $\mathbf{a}$ and $\mathbf{b}$ is given by*

$$\cos\theta = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}|\ |\mathbf{b}|}, \qquad \textit{where } \theta \in [0, \pi].$$

Theorem 3 (The Cauchy-Schwarz Inequality)**.** *If $\mathbf{a}$, $\mathbf{b} \in \mathbb{R}^n$, then $-|\mathbf{a}|\ |\mathbf{b}| \leqslant \mathbf{a} \cdot \mathbf{b} \leqslant |\mathbf{a}|\ |\mathbf{b}|$.*

Theorem 4 (Minkowski's Inequality (or the Triangle Inequality))**.** *For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$,*

$$|\mathbf{a} + \mathbf{b}| \leqslant |\mathbf{a}| + |\mathbf{b}|.$$

2.3 Applications: orthogonality and projection

2.3.1 Orthogonality of vectors

Definition 1. *Two vectors $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^n$ are said to be **orthogonal** if $\mathbf{a} \cdot \mathbf{b} = 0$.*

Note that the zero vector is orthogonal to every vector, including itself.Two non-zero vectors at right angles to each other are also said to be **perpendicular** to each other or to be **normal** to each other.

Definition 2. *An **orthonormal set of vectors** in $\mathbb{R}^n$ is a set of vectors which are of unit length and mutually orthogonal.*

2.3.2 Projections

The geometric idea of a projection of a vector $\mathbf{a}$ on a non-zero vector $\mathbf{b}$ in $\mathbb{R}^2$ or $\mathbb{R}^3$ is shown in Figure 1, where $\overrightarrow{OP}$ is the projection.

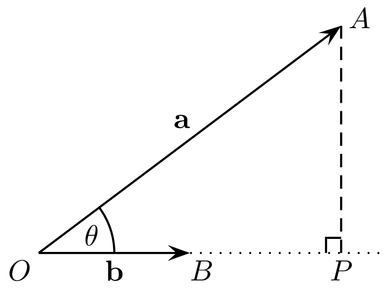

Figure 1: Projection of $\mathbf{a}$ on $\mathbf{b}$.

Definition 3. *For $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and $\mathbf{b} \neq \mathbf{0}$, the* ***projection of* a *on* b** *is*

$$proj_{\mathbf{b}}\mathbf{a} = \left(\frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{b}|^2}\right)\mathbf{b}.$$

Proposition 1. $\text{proj}_{\mathbf{b}}\mathbf{a}$ is the unique vector $\lambda\mathbf{b}$ parallel to the non-zero vector $\mathbf{b}$ such that

$$(\mathbf{a} - \lambda\mathbf{b}) \cdot \mathbf{b} = 0. \qquad (\#)$$

An alternative forms of writing the formula for a projection,

$$\text{proj}_{\mathbf{b}}\mathbf{a} = (\mathbf{a}\cdot\widehat{\mathbf{b}})\widehat{\mathbf{b}} = |\mathbf{a}|\cos\theta\,\widehat{\mathbf{b}},$$

where $\widehat{\mathbf{b}} = \dfrac{\mathbf{b}}{|\mathbf{b}|}$ is the unit vector in the direction of $\mathbf{b}$ and where θ is the angle between $\mathbf{a}$ and $\mathbf{b}$.

Note that a simple formula for the length of the projection of $\mathbf{a}$ on $\mathbf{b}$ is

$$|\text{proj}_{\mathbf{b}}\mathbf{a}| = |\mathbf{a}\cdot\widehat{\mathbf{b}}| = \frac{|\mathbf{a}\cdot\mathbf{b}|}{|\mathbf{b}|}.$$

2.3.3 Distance between a point and a line in $\mathbb{R}^3$

The distance between a point B and a line $\mathbf{x} = \mathbf{a} + \lambda\mathbf{d}$ is the shortest distance between the point and the line. In the diagram, the distance is $|\overrightarrow{PB}|$, where P is the point on the line such that $\angle APB$ is a right angle.

$$\overrightarrow{PB} = \overrightarrow{AB} - \overrightarrow{AP} = \mathbf{b} - \mathbf{a} - \text{proj}_{\mathbf{d}}(\mathbf{b} - \mathbf{a}),$$

and then the shortest distance is $|\overrightarrow{PB}|$.

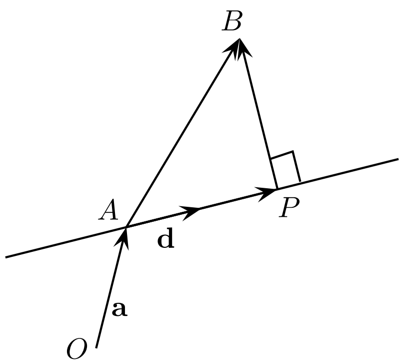

Figure 2: Shortest Distance between Point and Line.

2.4 The cross product

Definition 1. *The **cross product** of two vectors* $\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$ *and* $\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix}$ *in* $\mathbb{R}^3$ *is*

$$\mathbf{a} \times \mathbf{b} = \begin{pmatrix} a_2 b_3 - a_3 b_2 \\ a_3 b_1 - a_1 b_3 \\ a_1 b_2 - a_2 b_1 \end{pmatrix}.$$

Note that the cross product of two vectors is a **vector**. For this reason the cross product is often called the **vector product** of two vectors, in contrast to the dot product which is a scalar and is often called the **scalar product** of two vectors. Note also that the cross product $\mathbf{a} \times \mathbf{b}$ has the important property that it is perpendicular to the two vectors $\mathbf{a}$ and $\mathbf{b}$.

The most common trick for remembering the formula of cross product is to use determinants.

$$\mathbf{a} \times \mathbf{b} = \begin{vmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3 \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

The determinant is expanded along the first row and as usual $\mathbf{e}_1$, $\mathbf{e}_2$, $\mathbf{e}_3$ are the standard basis vectors of $\mathbb{R}^3$.

2.4.1 Arithmetic properties of the cross product

Proposition 1. For all $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$,

1. $\mathbf{a} \times \mathbf{a} = \mathbf{0}$, i.e., the cross product of a vector with itself is the zero vector.
2. $\mathbf{a} \times \mathbf{b} = -\mathbf{b} \times \mathbf{a}$. The cross product is **not commutative**. If the order of vectors in the cross product is reversed, then the sign of the product is also reversed.
3. $\mathbf{a} \times (\lambda \mathbf{b}) = \lambda(\mathbf{a} \times \mathbf{b})$ and $(\lambda \mathbf{a}) \times \mathbf{b} = \lambda(\mathbf{a} \times \mathbf{b})$.
4. $\mathbf{a} \times (\lambda \mathbf{a}) = \mathbf{0}$, i.e., the cross product of parallel vectors is zero.
5. **Distributive Laws**. $\mathbf{a} \times (\mathbf{b} + \mathbf{c}) = \mathbf{a} \times \mathbf{b} + \mathbf{a} \times \mathbf{c}$ and $(\mathbf{a} + \mathbf{b}) \times \mathbf{c} = \mathbf{a} \times \mathbf{c} + \mathbf{b} \times \mathbf{c}$.

Proposition 2. The three standard basis vectors in $\mathbb{R}^3$ satisfy the relations

1. $\mathbf{e}_1 \times \mathbf{e}_1 = \mathbf{e}_2 \times \mathbf{e}_2 = \mathbf{e}_3 \times \mathbf{e}_3 = \mathbf{0}$,
2. $\mathbf{e}_1 \times \mathbf{e}_2 = \mathbf{e}_3, \quad \mathbf{e}_2 \times \mathbf{e}_3 = \mathbf{e}_1, \quad \mathbf{e}_3 \times \mathbf{e}_1 = \mathbf{e}_2$.

NOTE. The cross product is **not** associative. In fact,

$$\begin{aligned} & (\mathbf{e}_1 \times \mathbf{e}_2) \times \mathbf{e}_2 = -\mathbf{e}_1, \\ \text{but} \quad & \mathbf{e}_1 \times (\mathbf{e}_2 \times \mathbf{e}_2) = \mathbf{0}. \end{aligned}$$

Proposition 3. Suppose A, B are points in $\mathbb{R}^3$ that have coordinate vectors $\mathbf{a}$ and $\mathbf{b}$, and $\angle AOB = \theta$ then $|\mathbf{a} \times \mathbf{b}| = |\mathbf{a}|\ |\mathbf{b}| \sin\theta$.

2.4.2 A geometric interpretation of the cross product

Let θ denote the angle between $\mathbf{a}$ and $\mathbf{b}$. $\mathbf{a} \times \mathbf{b}$ *is a vector of length* $|\mathbf{a}|\ |\mathbf{b}| \sin\theta$ *in the direction perpendicular to both* $\mathbf{a}$ and $\mathbf{b}$.

2.4.3 Areas

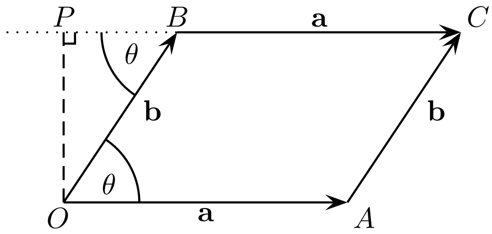

Figure 3: The Cross Product and the Area of a Parallelogram.

The area of the parallelogram $OACB$ in Figure 3 is given by

$$\text{Area} = |\overrightarrow{OA}||\overrightarrow{OP}| = |\mathbf{a}|\ |\mathbf{b}| \sin\theta = |\mathbf{a} \times \mathbf{b}|.$$

2.5 Scalar triple product and volume

Definition 1. *For* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \mathbb{R}^3$, *the* ***scalar triple product*** *of* $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$ *is* $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c})$.

Definition 2. *For* $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \mathbb{R}^3$, *the* ***vector triple product*** *of* $\mathbf{a}$, $\mathbf{b}$ *and* $\mathbf{c}$ *is* $\mathbf{a} \times (\mathbf{b} \times \mathbf{c})$.

Proposition 1. For $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c} \in \mathbb{R}^3$,

1. $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = (\mathbf{a} \times \mathbf{b}) \cdot \mathbf{c}$, that is, the dot and cross can be interchanged.

2. $\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = -\mathbf{a} \cdot (\mathbf{c} \times \mathbf{b})$, that is, the sign is reversed if the order of two vectors is reversed.

3. $\mathbf{a} \cdot (\mathbf{a} \times \mathbf{b}) = (\mathbf{a} \times \mathbf{a}) \cdot \mathbf{b} = 0$, that is, the scalar triple product is zero if any two vectors are the same.

4. The scalar triple product can be written as a determinant.

$$\mathbf{a} \cdot (\mathbf{b} \times \mathbf{c}) = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix}.$$

2.5.1 Volumes of parallelepipeds

A parallelepiped is a three-dimensional analogue of a parallelogram. Given three vectors $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{c}$ we can form a parallelepiped as shown in Figure 4.

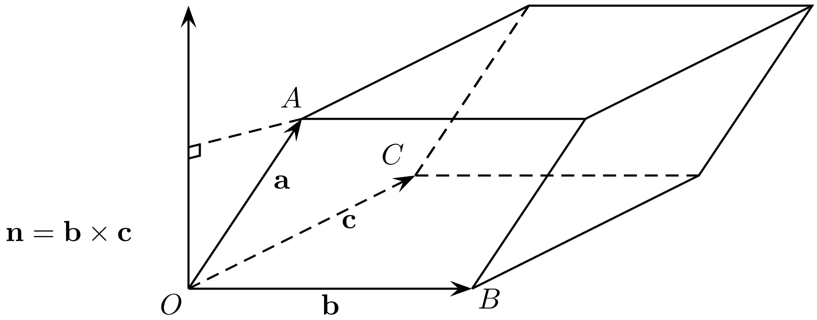

Figure 4: The Scalar Triple Product and the Volume of a Parallelepiped.

The volume of the parallelepiped is given by the formula

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

2.6 Planes in $\mathbb{R}^3$

2.6.1 Equations of planes in $\mathbb{R}^3$

Parametric Vector Form. The equation of a plane through a given point with position vector $\mathbf{c}$ and parallel to two given non-parallel vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ could be written in a parametric vector form

$$\mathbf{x} = \mathbf{c} + \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 \text{ for } \lambda_1, \lambda_2 \in \mathbb{R}.$$

Cartesian Form. A linear equation in three unknowns, $a_1x_1 + a_2x_2 + a_3x_3 = b$, where not all a_1, a_2, a_3 are zero.

Point-Normal Form. The point-normal form is the equation $\mathbf{n} \cdot (\mathbf{x} - \mathbf{c}) = 0$, where $\mathbf{n}$ and $\mathbf{c}$ are fixed vectors and $\mathbf{x}$ is the position vector of a point in space.

2.6.2 Distance between a point and a plane in $\mathbb{R}^3$

The distance from the point B to the plane is $\left|\overrightarrow{PB}\right|$, where $\overrightarrow{PB}$ is normal to the plane. If A is any point on the plane, $\overrightarrow{PB}$ is the projection of $\overrightarrow{AB}$ on any vector normal to the plane. The shortest distance is the "perpendicular distance" to the plane, and it is the length of the projection on a normal to the plane of a line segment from any point on the plane to the given point.

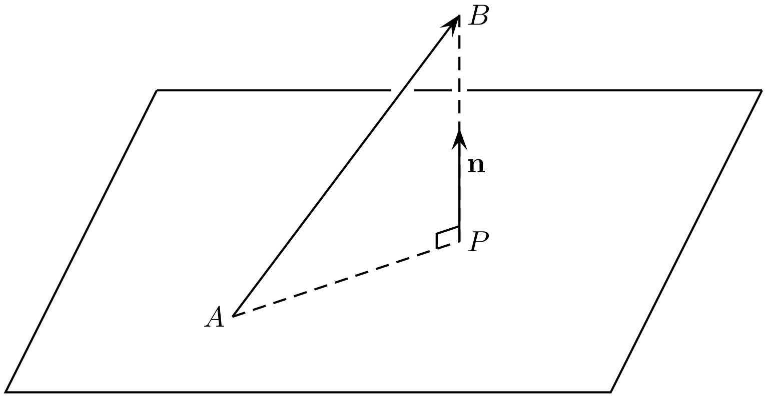

Figure 5: Shortest Distance between Point and Plane.

Chapter 3

3.1 A review of number systems

The set of **natural numbers** (or counting numbers): $\mathbb{N} = \{0, 1, 2, \ldots\}$.
The set of **integers**: $\mathbb{Z} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$.
The set of **rational numbers**: $\mathbb{Q} = \left\{ \frac{p}{q} : p, q \in \mathbb{Z} \text{ for } q \neq 0 \right\}$.
The set of real numbers, $\mathbb{R}$, contains all the rationals, along with numbers such as $\sqrt{2}, \sqrt{3}, \pi, e$, etc. which are **not** rational.

The definition of a field:

Definition 1. *Let $\mathbb{F}$ be a non-empty set of elements for which a rule of addition ($+$) and a rule of multiplication are defined. Then the system is a* ***field*** *if the following twelve axioms (or fundamental number laws) are satisfied.*

1. **Closure under Addition**. If $x, y \in \mathbb{F}$ then $x + y \in \mathbb{F}$.
2. **Associative Law of Addition**. $(x + y) + z = x + (y + z)$ for all $x, y, z \in \mathbb{F}$.
3. **Commutative Law of Addition**. $x + y = y + x$ for all $x, y \in \mathbb{F}$.
4. **Existence of a Zero**. There exists an element of $\mathbb{F}$ (usually written as 0) such that $0 + x = x + 0 = x$ for all $x \in \mathbb{F}$.
5. **Existence of a Negative**. For each $x \in \mathbb{F}$, there exists an element $w \in \mathbb{F}$ (usually written as $-x$) such that $x + w = w + x = 0$.
6. **Closure under Multiplication**. If $x, y \in \mathbb{F}$ then $xy \in \mathbb{F}$.
7. **Associative Law of Multiplication**. $x(yz) = (xy)z$ for all $x, y, z \in \mathbb{F}$.
8. **Commutative Law of Multiplication**. $xy = yx$ for all $x, y \in \mathbb{F}$.
9. **Existence of a One**. There exists a non-zero element of $\mathbb{F}$ (usually written as 1) such that $x1 = 1x = x$ for all $x \in \mathbb{F}$.
10. **Existence of an Inverse for Multiplication**. For each non-zero $x \in \mathbb{F}$, there exists an element w of $\mathbb{F}$ (usually written as $1/x$ or x^{-1}) such that $xw = wx = 1$.
11. **Distributive Law**. $x(y + z) = xy + xz$ for all $x, y, z \in \mathbb{F}$.
12. **Distributive Law**. $(x + y)z = xz + yz$, for all $x, y, z \in \mathbb{F}$.

3.2 Introduction to complex numbers

We now define the set $\mathbb{C}$ of **complex numbers** by:

$$\mathbb{C} = \{a + bi \,|\, a, b \in \mathbb{R},\ i^2 = -1\}.$$

A complex number written in the form $a + bi$, where $a,\ b \in \mathbb{R}$, is said to be in **Cartesian form**. The real number a is called the *real part* of $a + bi$, and b is called the *imaginary part*. The set $\mathbb{C}$ contains all the real numbers (when $b = 0$). Numbers of the form bi, with b real ($b \neq 0$), are called **purely imaginary numbers**. The set of complex numbers also satisfies the twelve number laws, and so it also forms a field.

3.3 The rules of arithmetic for complex numbers

Let $z = a + bi$ and $w = c + di$, where $a, b, c, d \in \mathbb{R}$.

Addition and subtraction. We define the sum, $z + w$, by

$$z + w = (a + c) + (b + d)i.$$

and the difference, $z - w$, by

$$z - w = (a - c) + (b - d)i.$$

That is, we add or subtract the real parts and the imaginary parts separately.

Multiplication. Expanding out, we have

$$(a + bi)(c + di) = ac + bci + adi + (bi)(di) = (ac - bd) + (bc + ad)i.$$

Hence we define $(a + bi)(c + di) = (ac - bd) + (bc + ad)i$.

Division. Since

$$\frac{z}{w} = \frac{a + bi}{c + di} = \frac{(a + bi)(c - di)}{(c + di)(c - di)} = \frac{ac + bd + (bc - ad)i}{c^2 + d^2} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}\, i,$$

we define the quotient $\frac{z}{w}$, $(w \neq 0)$ by

$$\frac{z}{w} = \frac{ac + bd}{c^2 + d^2} + \frac{bc - ad}{c^2 + d^2}\, i.$$

Proposition 1. [**X**] *The following properties hold for addition of complex numbers:*

1. **Uniqueness of Zero.** There is one and only one zero in $\mathbb{C}$.

2. **Cancellation Property.** If $z, v, w \in \mathbb{C}$ satisfy $z + v = z + w$, then $v = w$.

Proposition 2. [**X**] *The following properties hold for multiplication of complex numbers:*

1. $0z = 0$ for all complex numbers z.

2. $(-1)z = -z$ for all complex numbers z.

3. **Cancellation Property**. If $z, v, w \in \mathbb{C}$ satisfy $zv = zw$ and $z \neq 0$, then $v = w$.

4. If $z, w \in \mathbb{C}$ satisfy $zw = 0$, then either $z = 0$ or $w = 0$ or both.

3.4 Real parts, imaginary parts and complex conjugates

Definition 1. *The* ***real part*** *of $z = a + bi$ (written $\mathrm{Re}(z)$), where $a, b \in \mathbb{R}$, is given by*

$$\mathrm{Re}(z) = a.$$

Definition 2. *The* ***imaginary part*** *of $z = a + bi$ (written $\mathrm{Im}(z)$), where $a, b \in \mathbb{R}$, is given by*

$$\mathrm{Im}(z) = b.$$

NOTE.

1. The imaginary part of a complex number is a real number.

2. Two complex numbers are equal if and only if their real parts are equal and their imaginary parts are equal. That is

$$a + bi = c + di \quad \text{if and only if} \quad a = c \text{ and } b = d,$$

where $a, b, c, d \in \mathbb{R}$.

Definition 3. *If $z = a + bi$, where $a, b \in \mathbb{R}$, then the* ***complex conjugate*** *of z is $\overline{z} = a - bi$.*

Properties of the Complex Conjugate.

1. $\overline{\overline{z}} = z$.

2. $\overline{z + w} = \overline{z} + \overline{w}$ and $\overline{z - w} = \overline{z} - \overline{w}$.

3. $\overline{zw} = \overline{z}\,\overline{w}$ and $\overline{\left(\dfrac{z}{w}\right)} = \dfrac{\overline{z}}{\overline{w}}$.

4. $\mathrm{Re}(z) = \frac{1}{2}(z + \overline{z})$ and $\mathrm{Im}(z) = \frac{1}{2i}(z - \overline{z})$.

5. If $z = a + bi$, then $z\overline{z} = a^2 + b^2$, so $z\overline{z} \in \mathbb{R}$ and $z\overline{z} \geqslant 0$.

3.5 The Argand diagram

We identify a complex number $z = a + bi$ with the point in the xy-plane whose coordinates are (a, b).

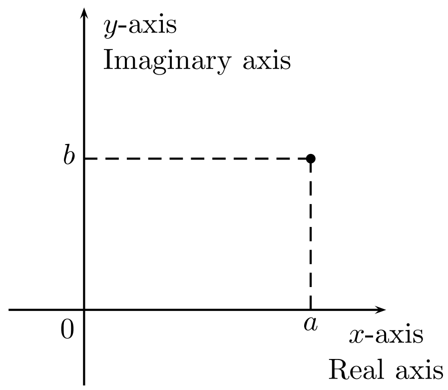

Figure 1: The Argand Diagram.

Addition and subtraction is done by adding or subtracting x-coordinates and y-coordinates separately, as shown in Figure 2.

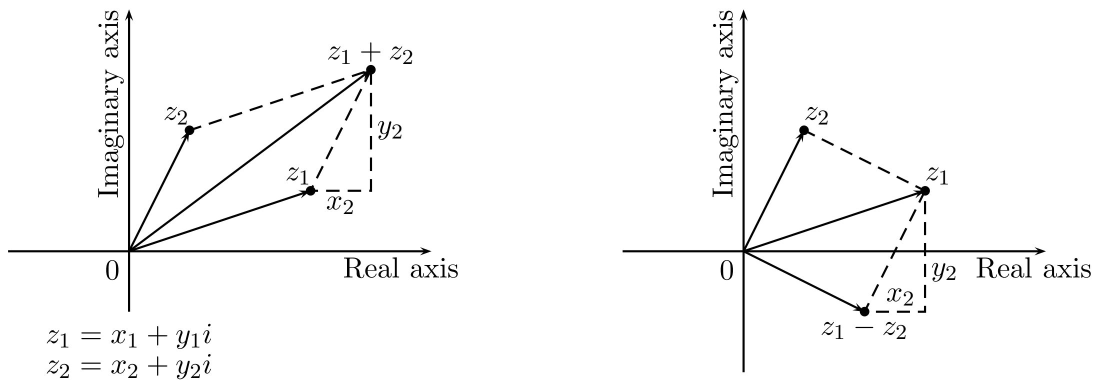

Figure 2: Addition and Subtraction of Complex Numbers.

3.6 Polar form, modulus and argument

An alternative representation for complex numbers is obtained by using plane polar coordinates r and θ.

Work out r and θ using a diagram.

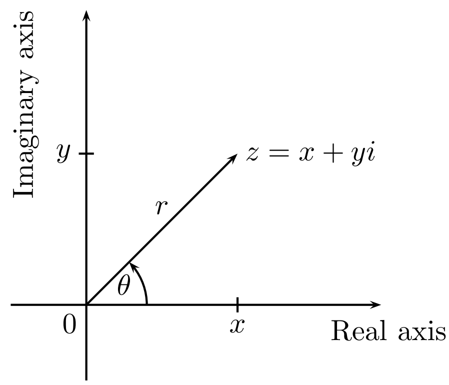

Figure 3: Polar Coordinates of a Complex Number.

A complex number $z \neq 0$ can be written using the polar coordinates r and θ as:

$$z = r(\cos\theta + i\sin\theta).$$

Proposition 1 (Equality of Complex Numbers)**.** *Two complex numbers*

$$z_1 = r_1(\cos\theta_1 + i\sin\theta_1) \qquad \text{and} \qquad z_2 = r_2(\cos\theta_2 + i\sin\theta_2), \quad z_1,\ z_2 \neq 0$$

are equal if and only if $r_1 = r_2$ and $\theta_1 = \theta_2 + 2k\pi$, where k is any integer.

The polar coordinate r that we have associated with a complex number is often called the **modulus** or the **magnitude** of the complex number. The formal definition is:

Definition 1. *For $z = x + yi$, where $x, y \in \mathbb{R}$, we define the **modulus** of z to be*

$$|z| = \sqrt{x^2 + y^2}.$$

The polar coordinate θ that we have associated with a complex number is called an **argument** of the complex number and is written as $\arg(z)$. The argument such that $-\pi < \theta \leqslant \pi$ is called the **principal argument** of z and is written as $\mathrm{Arg}(z)$.

3.7 Properties and applications of the polar form

Lemma 1. *For any real numbers θ_1 and θ_2*

$$(\cos\theta_1 + i\sin\theta_1)(\cos\theta_2 + i\sin\theta_2) = \cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2).$$

Theorem 2 (De Moivre's Theorem)**.** *For any real number θ and integer n*

$$(\cos\theta + i\sin\ \theta)^n = \cos n\theta + i\sin n\theta. \tag{\#}$$

De Moivre's Theorem provides a simple formula for integer powers of complex numbers. However, it can also be used to suggest a meaning for complex powers of complex numbers. To make this extension to complex powers it is actually sufficient just to give a meaning to the exponential function for imaginary exponents. We first make the following definition.

Definition 1. (Euler's Formula). *For real θ, we define*

$$e^{i\theta} = \cos\theta + i\sin\theta.$$

3.7.1 The arithmetic of polar forms

Definition 2. *The **polar form** for a non-zero complex number z is*

$$z = re^{i\theta}$$

where $r = |z|$ and $\theta = Arg(z)$.

Using Euler's formula, we can rewrite the equality proposition for polar forms (Proposition 1 of Section 3.6) as

$$z_1 = r_1 e^{i\theta_1} = r_2 e^{i\theta_2} = z_2$$

if and only if $r_1 = r_2$ and $\theta_1 = \theta_2 + 2k\pi$ for $k \in \mathbb{Z}$.

The formulae for multiplication and division of polar forms are:

$$z_1 z_2 = r_1 e^{i\theta_1} r_2 e^{i\theta_2} = r_1 r_2 e^{i(\theta_1+\theta_2)},$$

$$\frac{z_1}{z_2} = \frac{r_1 e^{i\theta_1}}{r_2 e^{i\theta_2}} = \frac{r_1}{r_2} e^{i(\theta_1-\theta_2)}.$$

It is sometimes useful to express these results in terms of the modulus and argument.

$$|z_1 z_2| = r_1 r_2 = |z_1|\,|z_2| \qquad \text{and} \qquad \operatorname{Arg}(z_1 z_2) = \operatorname{Arg}(z_1) + \operatorname{Arg}(z_2) + 2k\pi,$$

$$\left|\frac{z_1}{z_2}\right| = \frac{r_1}{r_2} = \frac{|z_1|}{|z_2|} \qquad \text{and} \qquad \operatorname{Arg}\left(\frac{z_1}{z_2}\right) = \operatorname{Arg}(z_1) - \operatorname{Arg}(z_2) + 2k\pi.$$

We need to choose a suitable integer value for k to obtain the principal argument.

3.7.2 Powers of complex numbers

If $z = re^{i\theta}$. then $z^n = r^n e^{in\theta}$.

3.7.3 Roots of complex numbers

Definition 3. *A complex number z is an* **nth root** *of a number z_0 if z_0 is the nth power of z, that is, z is the nth root of z_0 if $z^n = z_0$.*

3.8 Trigonometric applications of complex numbers

Sine and cosine of multiples of θ.

To express $\cos n\theta$ or $\sin n\theta$ in terms of powers of $\cos\theta$ or $\sin\theta$, we expand the right hand side of

$$\cos n\theta + i\sin n\theta = (\cos\theta + i\sin\ \theta)^n,$$

by the Binomial Theorem then separately equate the real and imaginary parts.

Powers of sine and cosine.

Using Euler's formula, we note that

$$e^{in\theta} = \cos n\theta + i\sin n\theta,$$

$$e^{-in\theta} = \cos(-n\theta) + i\sin\ (-n\theta) = \cos n\theta - i\sin n\theta.$$

On first adding and then subtracting these formulae, we obtain the important formulae

$$\cos n\theta = \frac{1}{2}\left(e^{in\theta} + e^{-in\theta}\right), \qquad \sin n\theta = \frac{1}{2i}\left(e^{in\theta} - e^{-in\theta}\right).$$

In particular, we have

$$\cos\theta = \frac{1}{2}\left(e^{i\theta} + e^{-i\theta}\right), \qquad \sin\theta = \frac{1}{2i}\left(e^{i\theta} - e^{-i\theta}\right).$$

We can apply the above formulae to derive trigonometric formulae which relate powers of $\sin\theta$ or $\cos\theta$ to sines or cosines of multiples of θ.

3.9 Geometric applications of complex numbers

Relations between complex numbers can be given a geometric interpretation in the complex plane, and, conversely, the geometry of a plane can be represented by algebraic relations between complex numbers.

The complex numbers 0, z_1, z_2 and $z_1 + z_2$ form a parallelogram. The complex number $z - w$ can be represented by the directed line segment (the arrow) from w to z. We plot z and w on the Argand diagram. The magnitude $|z - w|$ is the distance between the points z and w and that $\operatorname{Arg}(z - w)$ is the angle between the arrow from z to w and a line in the direction of the positive real axis.

3.10 Complex polynomials

Definition 1. *Suppose n is a natural number and $a_0, a_1, \ldots, a_n$ are complex numbers with $a_n \neq 0$. Then the function $p : \mathbb{C} \longrightarrow \mathbb{C}$ such that*

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

is called a polynomial of degree n.
The zero polynomial is defined to be the function $p(z) = 0$ and we do not define its degree.

3.10.1 Roots and factors of polynomials

Definition 2. *A number α is a* **root** *(or* **zero***) of a polynomial p if $p(\alpha) = 0$.*

Definition 3. *Let p be a polynomial. Then, if there exist polynomials p_1 and p_2 such that $p(z) = p_1(z)p_2(z)$ for all complex z, then p_1 and p_2 are called* **factors** *of p.*

Theorem 1 (Remainder Theorem). *The remainder r which results when $p(z)$ is divided by $z - \alpha$ is given by $r = p(\alpha)$.*

Theorem 2 (Factor Theorem). *A number α is a root of p if and only if $z - \alpha$ is a factor of $p(z)$.*

Theorem 3 (The Fundamental Theorem of Algebra). *A polynomial of degree $n \geqslant 1$ has at least one root in the complex numbers.*

Theorem 4 (Factorisation Theorem). *Every polynomial of degree $n \geqslant 1$ has a factorisation into n linear factors of the form*

$$p(z) = a(z - \alpha_1)(z - \alpha_2) \ldots (z - \alpha_n), \qquad (\#)$$

where the n complex numbers $\alpha_1, \alpha_2, \ldots, \alpha_n$ are roots of p and where a is the coefficient of z^n.

3.10.2 Factorisation of polynomials with real coefficients

Proposition 5. If α is a root of a polynomial p with **real** coefficients, then the complex conjugate $\overline{\alpha}$ is also a root of p.

Proposition 6. If p is a polynomial with real coefficients, then p can be factored into linear and quadratic factors all of which have real coefficients.

Chapter 4

LINEAR EQUATIONS AND MATRICES

4.1 Introduction to linear equations

4.2 Systems of linear equations and matrix notation

Definition 1. *A **system** of m **linear equations** in n variables is a set of m linear equations in n variables which must be simultaneously satisfied. Such a system is of the form*

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

A **solution** to a system of equations is the set of values of the variables which simultaneously satisfy all the equations. A system of equations is said to be **consistent** if it has at least one solution. Otherwise, the system is said to be **inconsistent**.

Definition 2. *The system (*) is said to be **homogeneous** if $b_1 = 0, \ \ldots, \ b_m = 0$.*

The **augmented matrix** for the system (*):

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

Let

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}. \qquad \text{and} \qquad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

The augmented matrix can be abbreviated by $(A|\mathbf{b})$.

The system of equations can also be written as $A\mathbf{x} = \mathbf{b}$, or as a **vector equation** or **vector form**.

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

This vector equation can be written more concisely as

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n = \mathbf{b},$$

where

$$\mathbf{a}_j = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix} \quad \text{for} \quad 1 \leqslant j \leqslant n \qquad \text{and} \qquad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}.$$

4.3 Elementary row operations

The three **elementary row operations** on (augmented) matrices:

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

Each of these operations gives us a new augmented matrix such that the system represented by the new matrix has the same solution set as the system represented by the old one.

4.4 Solving systems of equations

A process for solving systems of linear equations.

1. In the first stage, known as **Gaussian elimination**, we use two types of row operation (interchange of rows and adding a multiple of one row to another row) to produce an equivalent system in a simpler form which is known as **row-echelon form**. From the row-echelon form we can tell many things about solutions of the system. In particular, we can tell whether the system has no solution, a unique solution or infinitely many solutions.
2. If the system does have solutions, the second stage is to find them. It can be carried out by either of two methods.
   (a) We can use further row operations to obtain an even simpler form which is called **reduced row-echelon form**. From this form we can read off the solution(s).
   (b) From the row echelon form, we can read off the value (possibly in terms of parameters) for at least one of the variables. We substitute this value into one of the other equations and get the value for another variable, and so on. This process, which is called **back-substitution**, will be described fully later.

4.4.1 Row-echelon form and reduced row-echelon form

Definition 1. *In any matrix*

1. *a **leading row** is one which is not all zeros,*
2. *the **leading entry** in a leading row is the first (i.e. leftmost) non-zero entry,*
3. *a **leading column** is a column which contains the leading entry for some row.*

Definition 2. *A matrix is said to be in **row-echelon form** if*

1. *all leading rows are above all non-leading rows (so any all-zero rows are at the bottom of the matrix), and*
2. *in every leading row, the leading entry is further to the right than the leading entry in any row higher up in the matrix.*

Warning. The definition of row-echelon form varies from one book to another and the terminology in Definition 1 is not universally adopted.

Definition 3. *A matrix is said to be in **reduced row-echelon form** if it is in row-echelon form and in addition*

3. *every leading entry is 1, and*
4. *every leading entry is the only non-zero entry in its column.*

4.4.2 Gaussian elimination

In this subsection we shall see how the operations of interchanging two rows and of adding a multiple of a row to another row can be used to take a system of equations $A\mathbf{x} = \mathbf{b}$ and transform it into an equivalent system $U\mathbf{x} = \mathbf{y}$ such that the augmented matrix $(U|\mathbf{y})$ is in row-echelon form. The method we shall describe is called the Gaussian-elimination algorithm.

We now describe the steps in the algorithm and illustrate each step by applying it to the following augmented matrices.

$$\text{a)}\quad (A|\mathbf{b}) = \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 2 & 3 & 1 & 3 \\ 1 & 3 & 3 & 2 \end{array}\right), \qquad \text{b)}\quad (A|\mathbf{b}) = \left(\begin{array}{ccccc|c} 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 1 & 3 & -2 \\ 0 & 6 & -6 & 6 & 3 & -9 \end{array}\right).$$

Step 1. Select a pivot element.

We shall use the following rule to choose what is called a **pivot element**: start at the left of the matrix and move to the right until you reach a column which is not all zeros, then go down this column and choose the first non-zero entry as the pivot entry. The column containing the pivot entry is called the **pivot column** and the row containing the pivot entry is called the **pivot row**.

NOTE. When solving linear equations on a computer, a more complicated pivot selection rule is generally used in order to minimise "rounding errors".

In examples (a) and (b), our pivot selection rule selects the circled entries below as pivot entries for the first step of Gaussian elimination.

$$\text{a)}\quad \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 2 & 3 & 1 & 3 \\ 1 & 3 & 3 & 2 \end{array}\right), \qquad \text{b)}\quad \left(\begin{array}{ccccc|c} 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 1 & 3 & -2 \\ 0 & 6 & -6 & 6 & 3 & -9 \end{array}\right).$$

In example (a), row 1 is the pivot row and column 1 is the pivot column. In example (b), row 2 is the pivot row and column 2 is the pivot column.

Step 2. By a row interchange, swap the pivot row and the top row if necessary.

The rule is that the first pivot row of the augmented matrix must finish as the first row of $(U|\mathbf{y})$. You can achieve this by interchanging row 1 and the pivot row.

In example (a), the pivot row is already row 1, so no row interchange is needed. In example (b), the pivot row is row 2, so rows 1 and 2 of the augmented matrix must be interchanged. This is shown as

$$\text{b)}\quad \left(\begin{array}{ccccc|c} 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 1 & 3 & -2 \\ 0 & 6 & -6 & 6 & 3 & -9 \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_2} \left(\begin{array}{ccccc|c} 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 0 & 0 & 1 & 3 & -2 \\ 0 & 6 & -6 & 6 & 3 & -9 \end{array}\right).$$

Step 3. Eliminate (i.e., reduce to 0) all entries in the pivot column below the pivot element.

We can do this by adding suitable multiples of the pivot row to the *lower* rows. After this process the pivot column is in the correct form for the final row-echelon matrix $(U|\mathbf{y})$.

In example (a), we can use the row operations $R_2 = R_2 + (-2)\,R_1$ and $R_3 = R_3 + (-1)\,R_1$ to reduce the pivot column to the required form. In recording the row operations, we can simply write $R_2 = R_2 - 2R_1$ and $R_3 = R_3 - R_1$.

$$\text{a)}\quad \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 2 & 3 & 1 & 3 \\ 1 & 3 & 3 & 2 \end{array}\right) \xrightarrow[R_3 = R_3 - R_1]{R_2 = R_2 - 2R_1} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -1 & -1 & -3 \\ 0 & 1 & 2 & -1 \end{array}\right).$$

In example (b), the row operation $R_4 = R_4 - 2\,R_1$ reduces the pivot column to the required form. This gives

b) $$\left(\begin{array}{ccccc|c} 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 0 & 0 & 1 & 3 & -2 \\ 0 & 6 & -6 & 6 & 3 & -9 \end{array}\right) \xrightarrow{R_4 = R_4 - 2R_1} \left(\begin{array}{ccccc|c} 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 & -3 & 3 \end{array}\right).$$

Step 4. Repeat steps 1 to 3 on the submatrix of rows and columns strictly to the right of and below the pivot element and stop when the augmented matrix is in row-echelon form.

Note that the top row in step 2 here should mean the top row of the submatrix. In example (a), the required operation is

$$\left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -1 & -1 & -3 \\ 0 & 1 & 2 & -1 \end{array}\right) \xrightarrow{R_3 = R_3 + R_2} \left(\begin{array}{ccc|c} 1 & 2 & 1 & 3 \\ 0 & -1 & -1 & -3 \\ 0 & 0 & 1 & -4 \end{array}\right).$$

Then we have reduced the matrix to row-echelon from

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

In example (b), the required operations are

$$\left(\begin{array}{ccccc|c} 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 0 & 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 & -3 & -3 \end{array}\right) \xrightarrow{R_3 = R_3 - \frac{1}{2}R_2} \left(\begin{array}{ccccc|c} 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 0 & 0 & 0 & \frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & 0 & 0 & -3 & -3 \end{array}\right)$$

$$\xrightarrow{R_4 = R_4 + 2R_3} \left(\begin{array}{ccccc|c} 0 & 3 & -3 & 3 & 3 & -6 \\ 0 & 0 & 0 & 2 & 3 & -1 \\ 0 & 0 & 0 & 0 & \frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array}\right) = (U|\mathbf{y}).$$

In both examples we have now reached a matrix which is in row-echelon form, so we have completed the process of Gaussian elimination for these examples.

4.4.3 Transformation to reduced row-echelon form

Given a matrix in row-echelon form, we can transform it into *reduced* row-echelon —

Start with the lowest row which is not all zeros. Multiply it by a suitable constant to make its leading entry 1. Then add multiples of this row to higher rows to get all zeros in the column above the leading entry of this row. Repeat this procedure with the second lowest non-zero row, and so on.

4.4.4 Back-substitution

When solving small systems by hand, you may prefer to use this procedure as an alternative.

Assign an arbitrary parameter value to each non-leading variable. Then read off from the last non-trivial equation an expression for the last leading variable in terms of your arbitrary parameters. Substitute this expression back into the second last equation to get an expression for the second last leading variable, and so on.

Example 1. We will redo example 4 of the last subsection, using back-substitution instead of transformation to reduced echelon form. The row-echelon form is

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

SOLUTION. The non-leading variables are x_2, x_4 and x_5 so we let $x_2 = \lambda_1$, $x_4 = \lambda_2$ and $x_5 = \lambda_3$. Then the second row corresponds to the equation

$$2x_3 + 3\lambda_2 + 2\lambda_3 = 1$$

which gives

$$x_3 = \frac{1}{2} - \frac{3}{2}\lambda_2 - \lambda_3.$$

Substituting this back into the equation represented by the first row gives

$$-5x_1 + 2\left(\frac{1}{2} - \frac{3}{2}\lambda_2 - \lambda_3\right) + 4\lambda_2 + \lambda_3 = 0.$$

We solve this for x_1 and get

$$x_1 = \frac{1}{5} + \frac{1}{5}\lambda_2 - \frac{1}{5}\lambda_3.$$

◊

4.5 Deducing solubility from row-echelon form

Proposition 1. If the augmented matrix for a system of linear equations can be transformed into an equivalent row-echelon form $(U|\mathbf{y})$ then:

1. The system has **no solution** if and only if the right hand column $\mathbf{y}$ is a **leading column**.
2. If the right hand column $\mathbf{y}$ is a NOT a leading column then the system has:
   a) A **unique** solution if and only if **every variable is a leading variable**.
   b) **Infinitely many** solutions if and only if there is **at least one non-leading variable**. In this case, each non-leading variable becomes a parameter in the general expression for all solutions and the number of parameters in the solution equals the number of non-leading variables.

Proposition 2. If A is an $m \times n$ matrix which can be transformed by elementary row operations into a row-echelon form U then the system $A\mathbf{x} = \mathbf{b}$ has

a) **at least one** solution for each $\mathbf{b}$ in $\mathbb{R}^m$ if and only if U has **no non-leading rows**,

b) **at most one** solution for each $\mathbf{b}$ in $\mathbb{R}^m$ if and only if U has **no non-leading columns**,

c) **exactly one** solution for each $\mathbf{b}$ in $\mathbb{R}^m$ if and only if U has **no non-leading rows** and **no non-leading columns**.

4.6 Solving $A\mathbf{x} = \mathbf{b}$ for indeterminate $\mathbf{b}$

4.7 General properties of the solution of $A\mathbf{x} = \mathbf{b}$

Proposition 1. $A\mathbf{x} = \mathbf{0}$ has $\mathbf{x} = \mathbf{0}$ as a solution.

Proposition 2. If $\mathbf{v}$ and $\mathbf{w}$ are solutions of $A\mathbf{x} = \mathbf{0}$ then so are $\mathbf{v} + \mathbf{w}$ and $\lambda\mathbf{v}$ for any scalar λ.

Proposition 3. Let $\mathbf{v}_1, \ldots, \mathbf{v}_k$ be solutions of $A\mathbf{x} = \mathbf{0}$ for $1 \leqslant j \leqslant k$. Then $\lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k$ is also a solution of $A\mathbf{x} = \mathbf{0}$ for all $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$.

Proposition 4. If $\mathbf{v}$ and $\mathbf{w}$ are solutions of $A\mathbf{x} = \mathbf{b}$ then $\mathbf{v} - \mathbf{w}$ is a solution of $A\mathbf{x} = \mathbf{0}$.

Proposition 5. Let $\mathbf{x}_p$ be a solution of $A\mathbf{x} = \mathbf{b}$.

1. If $\mathbf{x} = \mathbf{0}$ is the only solution of the homogeneous equation $A\mathbf{x} = \mathbf{0}$ then $\mathbf{x}_p$ is the unique solution of $A\mathbf{x} = \mathbf{b}$.

2. If the homogeneous equation has non-zero solutions $\mathbf{v}_1, \ldots, \mathbf{v}_k$ then

$$\mathbf{x} = \mathbf{x}_p + \lambda_1\mathbf{v}_1 + \cdots + \lambda_k\mathbf{v}_k,$$

   is also a solution of $A\mathbf{x} = \mathbf{b}$ for all $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$.

NOTE. The form of solutions in part 2 of Proposition 5 raises the question of what is the minimum number of vectors $\mathbf{v}_1, \cdots, \mathbf{v}_k$ required to yield all the solutions of $A\mathbf{x} = \mathbf{b}$. We will return to this question in Chapter 6 when we discuss the ideas of "spanning sets" and "linear independence". For the present, it is sufficient to note that the solution method we have used in this chapter does find **all** solutions of $\mathbf{Ax} = \mathbf{b}$.

Chapter 5

MATRICES

5.1 Matrix arithmetic and algebra

Definition 1. *An $m \times n$ (read "m by n") **matrix** is an array of m rows and n columns of numbers of the form*

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

*The number a_{ij} in the matrix A lies in row i and column j. It is called the ijth **entry** or ijth **element** of the matrix.*

NOTE.

1. An $m \times 1$ matrix is called a column vector, while a $1 \times n$ matrix is called a row vector.

2. We use M_{mn} to stand for the set of all $m \times n$ matrices, i.e., the set of all matrices with m rows and n columns.

   We also use $M_{mn}(\mathbb{R})$ for the set of all real $m \times n$ matrices and $M_{mn}(\mathbb{C})$ for the set of all complex $m \times n$ matrices. Likewise, $M_{mn}(\mathbb{Q})$ is used for the set of all rational $m \times n$ matrices.

3. We say an $m \times n$ matrix is of **size** $m \times n$.

4. When we say "let $A = (a_{ij})$", we are specifying a matrix of fixed size, in which, for each given i, j, the ijth entry is a_{ij}. On the other hand, for a given matrix A, we denote the entry in the ith row and jth column by $[A]_{ij}$.

5.1.1 Equality, addition and multiplication by a scalar

Definition 2. *Two matrices A and B are defined to be* ***equal*** *if*

1. *the number of rows of A equals the number of rows of B,*
2. *the number of columns of A equals the number of columns of B,*
3. *$[A]_{ij} = [B]_{ij}$ for all i and j.*

Definition 3. *If A and B are $m \times n$ matrices, then the* ***sum*** *$C = A + B$ is the $m \times n$ matrix whose entries are*

$$[C]_{ij} = [A]_{ij} + [B]_{ij} \qquad \text{for all} \quad i, j.$$

Proposition 1. Let A, B and C be $m \times n$ matrices.

1. $A + B = B + A$. (Commutative Law of Addition)
2. $(A + B) + C = A + (B + C)$. (Associative Law of Addition)

Definition 4. *A* ***zero*** *matrix (written $\mathbf{0}$) is a matrix in which every entry is zero.*

Proposition 2. Let A be a matrix and $\mathbf{0}$ be the zero matrix, both in M_{mn}. Then

$$A + \mathbf{0} = \mathbf{0} + A = A.$$

Definition 5. *For any matrix $A \in M_{mn}$, the* ***negative*** *of A is the $m \times n$ matrix $-A$ with entries*

$$[-A]_{ij} = -[A]_{ij} \quad \textit{for all } 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n.$$

Using the above definition, we can define subtraction by

$$A - B = A + (-B) \quad \text{for all } A, B \in M_{mn}.$$

Proposition 3. If A is an $m \times n$ matrix, then

$$A + (-A) = (-A) + A = \mathbf{0}.$$

Definition 6. *If A is an $m \times n$ matrix and λ is a scalar, then the* ***scalar multiple*** *$B = \lambda A$, of A is the $m \times n$ matrix whose entries are*

$$[B]_{ij} = \lambda [A]_{ij} \ \textit{for all } 1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n.$$

Proposition 4. Let λ, μ be scalars and A, B be matrices in M_{mn}.

1. $\lambda(\mu A) = (\lambda\mu)A$ (Associative Law of Scalar Multiplication)
2. $(\lambda + \mu)A = \lambda A + \mu A$ (Scalar Multiplication is Distributive over Scalar Addition)
3. $\lambda(A + B) = \lambda A + \lambda B$ (Scalar Multiplication is distributive over Matrix Addition)

5.1.2 Matrix multiplication

Definition 7. *If $A = (a_{ij})$ is an $m \times n$ matrix and $\mathbf{x}$ is an $n \times 1$ matrix with entries x_i, then the* ***product*** $\mathbf{b} = A\mathbf{x}$ *is the $m \times 1$ matrix whose entries are given by*

$$b_i = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = \sum_{k=1}^{n} a_{ik}x_k \quad \text{for} \quad 1 \leqslant i \leqslant m.$$

Definition 8. *Let A be an $m \times n$ matrix and X be an $n \times p$ matrix and let $\mathbf{x}_j$ be the jth column of X. Then the* ***product*** $B = AX$ *is the $m \times p$ matrix whose jth column $\mathbf{b}_j$ is given by*

$$\mathbf{b}_j = A\mathbf{x}_j \qquad \text{for} \quad 1 \leqslant j \leqslant p.$$

Definition 9. *(Alternative) If A is an $m \times n$ matrix and X is an $n \times p$ matrix, then the* ***product*** AX *is the $m \times p$ matrix whose entries are given by the formula*

$$[AX]_{ij} = [A]_{i1}[X]_{1j} + \cdots + [A]_{in}[X]_{nj} = \sum_{k=1}^{n} [A]_{ik}[X]_{kj} \quad \text{for } 1 \leqslant i \leqslant m, \ 1 \leqslant j \leqslant p.$$

NOTE.

1. Let A be an $m \times n$ matrix and B be an $r \times s$ matrix. The product AB is defined **only** when the number of columns of A is the same as the number of rows of B, i.e. $n = r$. If $n = r$, the size of AB will be $m \times s$.

$$\begin{array}{ccc} A & & B \\ m \times n & & r \times s \end{array}$$

if equal

size of AB

2. Matrix multiplication is a row-times-column process: we get the (row i, column j) entry of AX by going across the ith row of A and down the jth column of X multiplying and adding as we go.

Warning. For some A and X, $AX \neq XA$, i.e., matrix multiplication does **not** satisfy the commutative law.

A matrix is said to be **square** if it has the same number of rows as columns.

The **diagonal** of a square matrix consists of the positions on the line from the top left to the bottom right. More precisely, the diagonal entries of an $n \times n$ square matrix (a_{ij}) are $a_{11}, \ a_{22}, \ldots, \ a_{nn}$.

Definition 10. *An* ***identity*** *matrix (written I) is a square matrix with 1's on the diagonal and 0's off the diagonal.*

For each integer n there is one and only one $n \times n$ identity matrix. It is denoted by I_n, or just by I if there is no risk of ambiguity.

Proposition 5 (Properties of Matrix Multiplication)**.** Let A, B, C be matrices and λ be a scalar.

1. If the product AB exists, then $A(\lambda B) = \lambda(AB) = (\lambda A)B$.
2. **Associative Law of Matrix Multiplication**. If the products AB and BC exist, then $A(BC) = (AB)C$.
3. $AI = A$ and $IA = A$, where I represents identity matrices of the appropriate (possibly different) sizes.
4. **Right Distributive Law**. If $A + B$ and AC exist, then $(A + B)C = AC + BC$.
5. **Left Distributive Law**. If $B + C$ and AB exist, then $A(B + C) = AB + AC$.

5.1.3 Matrix arithmetic and systems of linear equations

Proposition 6. Let $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$, $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, $\mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$, $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$.

The vector $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$ is a solution to the matrix equation $A\mathbf{x} = \mathbf{b}$, i.e. $A\mathbf{v} = \mathbf{b}$ if and only if $x_1 = v_1, \ldots, x_n = v_n$ is a solution to the system of equations

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

5.2 The transpose of a matrix

Definition 1. *The* ***transpose*** *of an $m \times n$ matrix A is the $n \times m$ matrix A^T (read 'A transpose') with entries given by*

$$[A^T]_{ij} = [A]_{ji}.$$

NOTE. The columns of A^T are the rows of A and the rows of A^T are the columns of A.

Proposition 1. The transpose of a transpose is the original matrix, i.e., $(A^T)^T = A$.

Proposition 2. If $A, B \in M_{mn}$ and $\lambda, \mu \in \mathbb{R}$, then $(\lambda A + \mu B)^T = \lambda A^T + \mu B^T$.

Proposition 3. If AB exists, then $(AB)^T = B^T A^T$.

Definition 2. *A matrix is said to be a* ***symmetric*** *matrix if $A = A^T$.*

5.3 The inverse of a matrix

Definition 1. *A matrix X is said to be an* **inverse** *of a matrix A if both*

$$AX = I \qquad \text{and} \qquad XA = I,$$

where I is an identity (or unit) matrix of the appropriate size.

If a matrix A has an inverse, then A is said to be an **invertible** matrix. An invertible matrix is also called a **non-singular** matrix. If a matrix A is not an invertible matrix, then it is called a **singular** matrix.

Proposition 1. All invertible matrices are square.

Definition 2. *A matrix X is said to be a* **right inverse** *of A if A is $r \times c$, X is $c \times r$ and $AX = I_r$. A matrix Y is said to be a* **left inverse** *of A if A is $r \times c$, Y is $c \times r$ and $YA = I_c$.*

5.3.1 Some useful properties of inverses

Theorem 2. *If the matrix A has both a left inverse Y and a right inverse X, then $Y = X$. In particular, if both Y and X are inverses of A, then $Y = X$.*

That is, if A has a left and a right inverse, then A is both invertible and square.

Notation: Because the inverse of a matrix A (if one exists) is unique we can denote it by a special symbol: A^{-1} (read 'A inverse').

Proposition 3. If A is an invertible matrix, then A^{-1} is also an invertible matrix and the inverse of A^{-1} is A. That is, $(A^{-1})^{-1} = A$.

Proposition 4. If A and B are invertible matrices and the product AB exists, then AB is also an invertible matrix and $(AB)^{-1} = B^{-1}A^{-1}$.

5.3.2 Calculating the inverse of a matrix

Proposition 5. A matrix A is invertible if and only if it can be reduced by elementary row operations to an identity matrix I and if $(A \mid I)$ can be reduced to $(I \mid B)$ then $B = A^{-1}$.

The above proposition suggests a method of finding the inverse of a (invertible) matrix A.

5.3.3 Inverse of a 2×2 matrix

A simple formula of the inverse of an invertible 2×2 matrix:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}, \qquad \text{provided } ad - bc \neq 0.$$

[X] 5.3.4 Elementary row operations and matrix multiplication

5.3.5 Inverses and solution of $A\mathbf{x} = \mathbf{b}$

Proposition 10. If A is a square matrix, then A is invertible if and only if $A\mathbf{x} = 0$ implies $\mathbf{x} = 0$.

Proposition 11. Let A be an $n \times n$ square matrix. Then A is invertible if and only if the matrix equation $A\mathbf{x} = \mathbf{b}$ has a unique solution for all vectors $\mathbf{b} \in \mathbb{R}^n$. In this case, the unique solution is $\mathbf{x} = A^{-1}\mathbf{b}$.

NOTE. In principle, this proposition can be used to obtain a solution of $A\mathbf{x} = \mathbf{b}$ by finding an inverse A^{-1} and then forming $\mathbf{x} = A^{-1}\mathbf{b}$. However, there are several serious practical problems which occur if the proposition is used in this way.

Corollary 12. *Let A be a square matrix.*

1. *If X is a left inverse of A, then X is the (two-sided) inverse of A. That is,*

$$\text{if } XA = I, \text{ then } X = A^{-1}.$$

2. *If X is a right inverse of A, then X is the (two-sided) inverse of A.*

5.4 Determinants

Determinants are defined only for square matrices.

5.4.1 The definition of a determinant

Definition 1. *The **determinant** of a* 2×2 *matrix*

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \qquad \textit{is} \qquad \det(A) = a_{11}a_{22} - a_{12}a_{21}.$$

Definition 2. *For a matrix A, the (row i, column j) **minor** is the determinant of the matrix obtained from A by deleting row i and column j from A.*

Notation. We shall use the symbol $|A_{ij}|$ to represent the (row i, column j) minor in a matrix A.

Definition 3. *The **determinant** of an $n \times n$ matrix A is*

$$\begin{aligned} |A| &= a_{11}|A_{11}| - a_{12}|A_{12}| + a_{13}|A_{13}| - a_{14}|A_{14}| + \cdots + (-1)^{1+n}a_{1n}|A_{1n}| \\ &= \sum_{k=1}^{n}(-1)^{1+k}a_{1k}|A_{1k}|. \end{aligned}$$

5.4.2 Properties of determinants

Proposition 1. $\det(A^T) = \det(A)$

An immediate consequence of Proposition 1 is that a determinant can also be evaluated by 'expansion down the first column', that is

Proposition 2. If any two rows (or any two columns) of A are interchanged, then the sign of the determinant is reversed. More precisely if the matrix B is obtained from the matrix A by interchanging two rows (or columns), then $\det B = -\det A$.

Proposition 3. If a matrix contains a zero row or column then its determinant is zero.

Proposition 4. If a row (or column) of A is multiplied by a scalar, then the value of $\det A$ is multiplied by the same scalar. That is, if the matrix B is obtained from the matrix A by multiplying a row (or column) of A by the scalar λ, then $\det B = \lambda \det A$.

An immediate consequence of Propositions 2 and 4 is the following useful result.

Proposition 5. If any column of a matrix is a multiple of another column of the matrix (or any row is a multiple of another row), then the value of $\det(A)$ is zero.

Proposition 6. If a multiple of one row (or column) is added to another row (or column), then the value of the determinant is not changed.

Proposition 7. If A and B are square matrices such that the product AB exists, then

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

5.4.3 The efficient numerical evaluation of determinants

Proposition 8. If U is a square row-echelon matrix, then $\det(U)$ is equal to the product of the diagonal entries of U.

Proposition 9. If A is a square matrix and U is an equivalent row-echelon form obtained from A by Gaussian elimination using row interchanges and adding a multiple of one row to another, then $\det(A) = \epsilon \det(U)$, where $\epsilon = +1$ if an even number of row interchanges have been made, and $\epsilon = -1$ if an odd number of row interchanges have been made.

As a summary, we have the following.

1. If $A \longrightarrow B$ by $R_i \leftrightarrow R_j$, then $\det(A) = -\det(B)$.
2. If $A \longrightarrow B$ by $R_i = R_i + aR_j$ and $a \neq 0$, then $\det(A) = \det(B)$.
3. If all the entries of R_i in A has a common factor λ and B is the matrix formed from A by divding R_i by λ, then $\det(A) = \lambda \det(B)$. That is, we factorise λ from R_i.

Using the above, we can evaluate a determinant by row reduction. We illustrate the technique by the following example.

Example 1. Factorise $\begin{vmatrix} 1 & a^2 & a^3 \\ 1 & b^2 & b^3 \\ 1 & c^2 & c^3 \end{vmatrix}$.

Solution.

$$
\begin{aligned}
& \begin{vmatrix} 1 & a^2 & a^3 \\ 1 & b^2 & b^3 \\ 1 & c^2 & c^3 \end{vmatrix} \\
= & \begin{vmatrix} 1 & a^2 & a^3 \\ 0 & b^2-a^2 & b^3-a^3 \\ 0 & c^2-a^2 & c^3-a^3 \end{vmatrix} && \begin{array}{l} R_2 = R_2 - R_1 \\ R_3 = R_3 - R_1 \end{array} \\
= & (b-a)(c-a)\begin{vmatrix} 1 & a^2 & a^3 \\ 0 & a+b & a^2+ab+b^2 \\ 0 & a+c & a^2+ac+c^2 \end{vmatrix} && \begin{array}{l} \text{Proposition 4} \\ \text{factorise } (b-a) \text{ from } R_2 \\ \text{factorise } (c-a) \text{ from } R_3 \end{array} \\
= & (b-a)(c-a)\begin{vmatrix} 1 & a^2 & a^3 \\ 0 & a+b & a^2+ab+b^2 \\ 0 & c-b & a(c-b)+c^2-b^2 \end{vmatrix} && R_3 = R_3 - R_2 \\
= & (b-a)(c-a)(c-b)\begin{vmatrix} 1 & a^2 & a^3 \\ 0 & a+b & a^2+ab+b^2 \\ 0 & 1 & a+b+c \end{vmatrix} && \begin{array}{l} \text{Proposition 4} \\ \text{factorise } (c-b) \text{ from } R_3 \end{array} \\
= & (b-a)(c-a)(c-b)[(a+b)(a+b+c)-(a^2+ab+b^2)] && \text{expand along the first column} \\
= & (a-b)(b-c)(c-a)(ab+bc+ca)
\end{aligned}
$$

$\diamondsuit$

5.4.4 Determinants and solutions of $A\mathbf{x} = \mathbf{b}$

Proposition 10. Let A be an $n \times n$ matrix.

1. *If* $\det(A) \neq 0$*, the equation* $A\mathbf{x} = \mathbf{b}$ *has a solution and the solution is unique for all* $\mathbf{b} \in \mathbb{R}^n$.
2. *If* $\det(A) = 0$*, the equation* $A\mathbf{x} = \mathbf{b}$ *either has no solution or an infinite number of solutions for a given* $\mathbf{b}$.

Proposition 11. For a square matrix A, the homogeneous system of equations $A\mathbf{x} = \mathbf{0}$ has a non-zero solution if and only if $\det(A) = 0$.

Proposition 12. A square matrix A is invertible if and only if $\det(A) \neq 0$.

Proposition 13. If A is an invertible matrix, then $\det(A^{-1}) = \dfrac{1}{\det(A)}$.