MTH4115/4215
Vectors and Matrices

Claudia Garetto
Matthew Lewis
Weini Huang
c.garetto@qmul.ac.uk
w.huang@qmul.ac.uk
matthew.lewis@qmul.ac.uk

Contents

Chapter 1

Introduction

1.1 Why learning Vectors and Matrices

The subject of "Vectors and Matrices" (often called Linear Algebra) is one of the basic disciplines of Mathematics. Together with Calculus it provides a good foundation for more advanced Mathematics and it has several applications outside Mathematics: Physics, Engineering, Computer Science, etc. It also gives you the best starting point to attend Linear Algebra in Y2. We thank the previous lectures of this module for their contribution to these lecture notes.

1.2 How to study for this module

It is extremely important in this module, as in any Mathematics module, to make use of the correct mathematical language and to be able to write sentences that make sense in Mathematics. There are topics in this module that might be familiar from A-levels however here we want to look at them from the point of view of a mathematician not of a secondary school student. We will aim to write down definitions, theorems and also some proofs in abstract terms, learning to think in a logical way as a mathematicians do! Be reassured that there will be plenty of examples and applications so that even the more abstract concepts will become clear. Every chapter will finish with a section called More. The material in this section is NOT examinable however provides some extra interesting topics related to the contents of the chapter. For instance, the profiles of some mathematicians using vectors and matrices in their research, some problems to challenge your brain, some connections to future modules you might want to attend. I hope this will make the module more interesting and enjojyable!

Chapter 2

Elements of Logic and Set theory

We begin by building a common mathematical language. **As any language maths has its own alphabet** made of symbols. It is important to know them to be able to have a meaningful mathematical conversation. These are the building block of our mathematical background that we will construct during this module.

2.1 Symbols

In this section we collect the mathematical symbols that you will encounter all the time in your mathematical studies. We distinguish them into two classes: set theoretic symbols and logical symbols.

2.1.1 Set theoretic symbols

- $\in$ is in, is an element of
- $\subset$ is a subset of, is contained in
- $\subseteq$ subset (possibly equal)
- $\cup$ union
- $\cap$ intersection
- $\emptyset$ the empty set.

We can also reverse these symbols or negate them. For instance, the reverse of $\subset$ is $\supset$ which means contains and the negation of $\in$ is obtained by striking out the symbol, i.e., $\notin$. This last symbol stands for *is not an element of*.

2.1.2 Logical symbols

- $\forall$ for all
- $\exists$ there exists
- $\nexists$ there does not exist
- $\Rightarrow$ implies, is a sufficient condition for
- $\nRightarrow$ does not imply
- $\Leftrightarrow$ if and only if, iff

These symbols are quantifiers ($\forall$, $\exists$, ...) or express logical relations ($\Rightarrow$, $\Leftrightarrow$, ...). We will come back to them later on.

2.2 Set Theory

Every mathematical object belongs to a *set*. It is therefore important to have some foundations of set theory.

Definition 2.2.1. *A set is a collection of objects. The objects are referred to as* elements *of the set.*

We can represent a set by listing its elements or by stating a property that determines membership or in other words defines the set. It is common to use a capital letter to denote a set. In details, the set of numbers 1, 2, 3 is denoted as

$$S = \{1, 2, 3\}.$$

Here we have listed the elements of the set S. The set of all positive numbers (numbers greater than 0) can be represented by stating its defining property, i.e.,

$$T = \{x : x > 0\},$$
$$T = \{x| \, x > 0\}.$$

Note that : and | read as *such that*.

Some more definitions and notations:

- If x is an element of S then we write $x \in S$.
- If x does not belong to S then we write $x \notin S$.
- S is a subset of T if every element of S belongs to T, i.e.,

$$x \in S \Rightarrow x \in T.$$

- If S is a subset of T we write $S \subset T$. If the set might actually be equal then we can use the notation $S \subseteq T$. Analogously, we can write $T \supset S$ or $T \supseteq S$.
- The empty set $\emptyset$ is the set with no elements. It is a subset of every other set.

Definition 2.2.2. *Let A and B be two sets. The union of A and B is the set of all x such that $x \in A$* **or** *$x \in B$. In detail,*

$$A \cup B = \{x : x \in A \mathit{\ or\ } x \in B\}.$$

The intersection of A and B is

$$A \cap B = \{x : x \in A \mathit{\ and\ } x \in B\}.$$

The set of all elements of A which does not belong to B is denoted by $A \setminus B$.

Remark 2.2.3. It is clear from the definition above that $A \cap B$ could be empty. Moreover,

$$\begin{aligned} A \cap B &\subseteq A, \\ A \cap B &\subseteq B, \\ A &\subseteq A \cup B, \\ B &\subseteq A \cup B, \\ A \cap B &\subseteq A \cup B, \\ A \setminus B &\subseteq A. \end{aligned}$$

Example 2.2.4. Let $A = \{1, 2, 3, 4\}$ and $B = \{1, 2, 5\}$. Compute $A \cup B$, $A \cap B$ and $A \setminus B$. What is the relation between $(A \cup B) \setminus (A \cap B)$ and $A \setminus B$?

You are probably already very familiar with the following sets of numbers:

$$\begin{aligned} \mathbb{N} &= \{1, 2, 3, 4, \cdots\} \quad \text{natural numbers} \\ \mathbb{N}_0 &= \{0, 1, 2, 3, \cdots\} \quad \text{natural numbers with 0 included} \\ \mathbb{Z} &= \{0, \pm 1, \pm 2, \pm 3, \cdots\} \quad \text{integers} \\ \mathbb{Q} &= \{a/b : a \in \mathbb{Z}, b \in \mathbb{N}\} \quad \text{rational numbers} \end{aligned}$$

By definition of these sets we have immediately

$$\mathbb{N} \subset \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q}.$$

There exists a bigger set which contains all of these sets of numbers: it is the set $\mathbb{R}$ of real numbers. Intuitively it is the set of all points on a straight line extending indefinitely in both directions. We will give a rigorous definition later on but at the moment we can already write the following chain of inclusions:

$$\mathbb{N} \subset \mathbb{N}_0 \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R}.$$

2.3 Logic

It is important in every Mathematics module to have some notions of logic. We will consider statements (i.e., sentences in Logic also called propositions) which are either true or false but not both. These are all examples of statements:

1. All students in this room have dark hair.
2. $3 \leq 5$
3. 326 is an odd number.
4. 13 is a prime number.
5. 5 is not positive.
6. For all real numbers x, we have $x^2 \geq 0$.
7. There exists a real number a such that $a^2 = 2$.

A statement can contain a *variable*, such as "The number x is positive". Here x is a variable, and the statement can be true or false depending on the value of x. For example, it is true for $x = 1$ and false for $x = -2$.

2.3.1 Combining statements

Using statements with variables one can form new statements in several ways:

(i) Using the quantifier $\forall$: from the statement $b^2 \geq \frac{1}{2}$ one can form the statement

$$\forall b \in \mathbb{N}, \text{one has } b^2 \geq \frac{1}{2}.$$

This statement is true.

(ii) Using the quantifiers $\exists$: $\exists b \in \mathbb{Q}$ such that $b^2 \geq \frac{1}{2}$.

(iii) Using the implication $\Rightarrow$:

$$b^2 \geq \frac{1}{2} \quad \Rightarrow \quad b^2 \geq \frac{1}{3}.$$

(iv) Using the double implication $\Leftrightarrow$:

$$b^2 \geq \frac{1}{2} \quad \Leftrightarrow \quad -b^2 \leq -\frac{1}{2}.$$

Finally, one can combine all the above ways of forming statements.

2.3.2 Negating statements

It is very important to understand how to negate a statement. If a statement is true then its negation is false and vice versa. There is a basic rule that we need to follow:

When negating statements containing symbols $\exists$ and $\forall$, one should replace $\exists$ by $\forall$ and vice versa.

When negating statements containing the symbol $\Rightarrow$ one should replace $\Rightarrow$ by $\not\Rightarrow$.

Example 2.3.1. Let us write down a statement A:

$$\forall x \in \mathbb{N}\ x \geq 2\ \Rightarrow x+1 \geq 3.$$

A is true. Let us negate A. We use the notation $\neg A$. The statement $\neg A$ is given by

$$\exists x \in \mathbb{N}: \quad x \geq 2 \not\Rightarrow x+1 \geq 3.$$

$\neg A$ can be equivalently written as follows:

$$\exists x \in \mathbb{N}, x \geq 2: \quad x+1 < 3.$$

2.3.3 Equivalence and implication

Let A and B be two statements which contain the same variable. Then $A \Leftrightarrow B$ means that B is true if and only if A is true. In equivalences, you need to specify the range of the variable, so such statements will usually include the quantifier $\forall$. Examples:

1. $\forall x \in \mathbb{R}$: $x > 0 \Leftrightarrow x+1 > 1$
2. $\forall n \in \mathbb{Z}$: $n > \frac{1}{2} \Leftrightarrow n \geq 1$
3. $\forall k \in \mathbb{N}$: $k > 2 \Leftrightarrow k^2 > 4$
4. $\forall m \in \mathbb{Z}$: m is even $\Leftrightarrow m^2$ is even.

$A \Leftrightarrow B$ can be rephrased as 'A is a necessary and sufficient condition for B'.

Let A and B be two statements which contain the same variable. $A \Rightarrow B$ means 'if A is true then B is true'. In implications, you need to specify the range of the variable, so such statements will usually include the quantifier $\forall$. Examples:

1. $\forall a \in \mathbb{R}$, one has: $a \geq 1 \Rightarrow a^2 \geq 1$;
2. $\forall x \in \mathbb{R}$: x is a solution of $(x^2-1)=0 \Rightarrow x$ is a solution of $x^2=1$;
3. $\forall a, b \in \mathbb{R}$, one has: $a \in (0,1)$ and $b \in (0,1) \Rightarrow (a+b)/2 \in (0,1)$.

$A \Rightarrow B$ can be equivalently rephrased in one of the following ways:

1. A implies B.
2. If A then B.
3. A is a sufficient condition for B.
4. B is a necessary condition for A.

We say that $A \Rightarrow B$ is a conditional statement. Note that it can happen that $A \Rightarrow B$ but NOT $B \Rightarrow A$. Examples:

1. $k \in \mathbb{N} \Rightarrow k \in \mathbb{Z}$ but $k \in \mathbb{Z} \not\Rightarrow k \in \mathbb{N}$;
2. $\forall a \in \mathbb{R}$, one has $a \geq 1 \Rightarrow a^2 \geq 1$, but $a^2 \geq 1 \not\Rightarrow a \geq 1$.

Warning: It is a very common mistake to (explicitly or implicitly) confuse the statements $A \Rightarrow B$ and $B \Rightarrow A$.

2.4 Some basics mathematical language

It is important to fix some common mathematical language so that we can have a mathematical conversation which makes sense. In the course of this module we will make use of the terminology listed below.

1. Definition: an explanation of the mathematical meaning of a word.
2. Theorem: a statement that has been proven to be true.
3. Proof: the logical argument to prove that a statement is true.
4. Proposition: a less important but nonetheless interesting true statement.
5. Corollary: a true statement that is a simple deduction from a theorem or proposition
6. Lemma: a true statement used in proving other true statements (that is, a less important theorem that is helpful in the proof of other results).

Let's clarify these concepts with some examples.

Definition 2.4.1. *Let A and B be two points in $\mathbb{R}^3$. The vector $\overrightarrow{AB}$ is the segment with starting point A and ending point B.*

This is a definition because it explains a general concept and it is given for arbitrary points A and B. Do not confuse definitions with examples. If you choose two specific points A and B in $\mathbb{R}^3$ and you write down or sketch the corresponding vector you are giving an example NOT a definition.

Theorem 2.4.2. *There is no rational number that satisfies the equation* $x^2 = 2$.

Corollary 2.4.3. $\sqrt{2}$ *is an irrational number.*

This follows immediately from the previous theorem and therefore it is a corollary. Indeed, $\sqrt{2}$ solves the equation $x^2 = 2$ and therefore by the previous theorem cannot be rational.

Proposition 2.4.4. *Let* $S = \{2k,\ k \in \mathbb{N}\}$. *If* $x \in S$ *then* $x^2 \in S$

Note that the proposition above it is still an interesting statement but clearly less important than the theorem above. In Proposition 2.4.4 we see the hypothesis (A: $x \in S$) and the thesis (B: $x^2 \in S$). Proving this proposition means to find a logical argument that shows $A \Rightarrow B$. This is a mathematical proof. This proposition might be useful in proving another more important theorem. In that case we use it as a lemma.

2.5 More

In this section you can find some extra material to challenge your understanding and expand a little your knowledge. Let us begin with the concept of proof by contradiction. There will then be some problems you might want to try. Note that some of the topics in this section will come back to you during Y2, for instance in the module Convergence and Continuity.

2.5.1 Proof by contradiction

A theorem is given by a first statement called hypothesis (a proposition A which is true) and a final statement B. The theorem states that A implies B. We need to prove with a rigorous mathematical argument, which will often requires a chain of implications, that $A \Rightarrow B$ is true. As an explanatory example let us consider the following theorem:

Theorem: There is no rational number that satisfies the equation $x^2 = 2$

Our hypothesis or starting point is

$$A : x^2 = 2.$$

Our final statement of what we want to prove is

$$B : x \notin \mathbb{Q}.$$

Proving the theorem in a direct way means to prove $A \Rightarrow B$. This is equivalent to prove that $\neg B \Rightarrow \neg A$. Indeed, $\neg B \Rightarrow \neg A$ is the contrapositive of $A \Rightarrow B$ and a contrapositive to an implication is true iff the implication itself is true. Proving that the contrapositive is true means to prove that $\neg B$ implies $\neg A$. Doing a **proof by contradiction** means to assume that A is true and that $\neg B$ holds and then to get a contradiction to A. For this reason we use the terminology proof by contradiction. In some cases, as this one, the proof by contradiction is easier than the direct proof.

2.5.2 Converse versus contrapositive

Let A and B be statements depending on a variable. Then a *converse* to $A \Rightarrow B$ is $B \Rightarrow A$ and a *contrapositive* to $A \Rightarrow B$ is $\neg B \Rightarrow \neg A$. We have seen above that a contrapositive to an implication is true iff the implication itself is true. However, a converse may or may not be true regardless of whether the implication itself is true or not. This is illustrated by the following examples:

1. "If you live in London, then you live in England" (true) Converse: "If you live in England, then you live in London" (false). Contrapositive: "If you don't live in England, then you don't live in London" (true)

2. "$x > 1 \Rightarrow x^2 > 1$" (true). Converse: "$x^2 > 1 \Rightarrow x > 1$ (false). Contrapositive: "$x^2 \leq 1 \Rightarrow x \leq 1$" (true).

2.5.3 Problems

1. Let $S = \{2k,\ k \in \mathbb{N}\}$ and $T = \{2k+1,\ k \in \mathbb{N}\}$.

   (i) Compute $S \cup T$ and $S \cap T$.

   (ii) Is $\mathbb{N}_0 \subset S \cup T$? Justify your answer.

   (iii) Prove that if $x \in S$ then $x^2 \in S$.

   (iv) Does (iii) hold for the set T? Justify your answer with a short proof or with a counterexample.

2. Let A, B, C be three sets with $A, B \subseteq C$. Prove that

   (i) $C \setminus B \subseteq C \setminus A$ if $A \subseteq B$;

   (ii) $C \setminus (A \cup B) = (C \setminus A) \cap (C \setminus B)$;

   (iii) $C \setminus (A \cap B) = (C \setminus A) \cup (C \setminus B)$.

   Hint: to prove that two sets X and Y are equal you need to prove the double inclusion: $X \subseteq Y$ and $Y \subseteq X$.

3. Let us consider the following propositions:

   A: For all $x \in \mathbb{Q}$ there exists $k \in \mathbb{N}$ such that $x \leq k$;

   B: For all $x \in \mathbb{Q}$ there exists $k \in \mathbb{N}$ such that $x + 2 \leq k$.

   Prove that

   (i) $A \Rightarrow B$;

   (ii) $\neg B \Rightarrow \neg A$;

   (iii) $A \Leftrightarrow B$.

   How to you call the proof of (ii) in relation to the proof of (i)?

4. Prove by contradiction that there exists a unique natural number x such that $x^2 = x$.

   Hint: Start by proving that $\exists x \in \mathbb{N}$ such that $x^2 = x$. Then, prove the uniqueness of x by contradiction.

Chapter 3

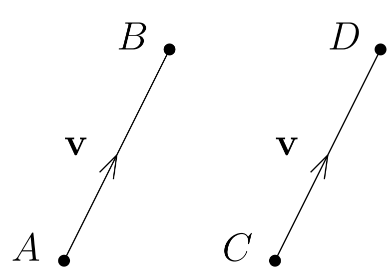
Where all starts: vectors in the plane and the space

In this chapter you will find material that is probably familiar to you from A-levels. Let us try to look at it from a more advanced (university) point of view. We will mainly work in $\mathbb{R}^3$ (space) or $\mathbb{R}^2$ (plane) however when possible we will try to be more general and work in $\mathbb{R}^n$.

3.1 Position vectors and geometrical interpretation

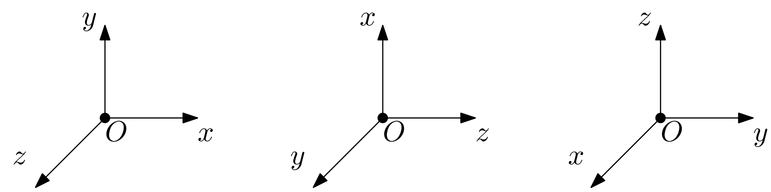
Definition 3.1.1. *Let A and B be two points in $\mathbb{R}^3$. The vector $\overrightarrow{AB}$ is the segment with starting point A and ending point B.*

This segment is reduced to a point if $A = B$. The length of the segment $\overrightarrow{AB}$ is denoted by $|\overrightarrow{AB}|$ and is a non-negative number. In this module we will identify with $\overrightarrow{AB}$ all the vectors having the same length and the same direction and denote them by the same letter $\mathbf{v}$. See explanatory example below:



This means that we can identify $\overrightarrow{AB}$ with $\overrightarrow{CD}$ or in other words they are both represented by the same vector $\mathbf{v}$. It is common to denote vectors with letters in bold, underlined or overlined.

As you know from school, $\mathbb{R}^3$ is the set of all points of the type $P = (a, b, c)$. Let us choose an origin O and 3 mutual perpendicular axes (the $x-$, $y-$ and $z-$ axes) arranged in a right-handed system as in the figures below:



For the sake of simplicity we will fix the third system of axes in the figure above. Let us consider the points $O=(0,0,0)$, $P_1=(1,0,0)$, $P_2=(0,1,0)$ and $P_3=(0,0,1)$. We easily see that P_1, P_2 and P_3 belong to the x-axis, y-axis and z-axis, respectively. Let us consider now the vectors

$$\overrightarrow{OP_1},\quad \overrightarrow{OP_2},\quad \overrightarrow{OP_3},$$

These are important vectors since they determine our axes and have length (or norm) 1. They are called position vectors because they determine the position of the points P_i, $i=1,2,3$. We will make use of the following notations:

$$\begin{aligned}\overrightarrow{OP_1}&=\mathbf{i},\\ \overrightarrow{OP_2}&=\mathbf{j},\\ \overrightarrow{OP_3}&=\mathbf{k}.\end{aligned}$$

Definition 3.1.2. *Let $A=(x_A,y_A,z_A)$ be a point in $\mathbb{R}^3$. The position vector $\overrightarrow{OA}$ is defined by $x_A\mathbf{i}+y_A\mathbf{j}+z_A\mathbf{k}$ and it is geometrically represented by the segment with starting point O and ending point A.*

Note that if we identify $\overrightarrow{OA}$ with the point A we can easily sum two position vectors or multiply them by a scalar $\lambda\in\mathbb{R}$ since we just need to make use of the addition and multiplication by scalar in $\mathbb{R}^3$. These two operations can be more generally defined in $\mathbb{R}^n$ for all $n\geq 1$.

Definition 3.1.3. *Let $\mathbb{R}^n$ the set of all n-ple $z=(z_1,z_2,\cdots,z_n)$[1] of real numbers z_i, $i=1,\cdots,n$. Addition and multiplication by real scalars in $\mathbb{R}^n$ are defined as follows:*

(i) for all $x=(x_1,x_2,\cdots,x_n)$ and $y=(y_1,y_2,\cdots,y_n)$,

$$x+y=(x_1+y_1,x_2+y_2,\cdots,x_n+y_n);$$

[1] One can use here the row notation as well as the column notation. The column notation is more common when talking about vectors rather than points as we will see in the next section.

(ii) for all $x = (x_1, x_2, \cdots, x_n)$ *and* $\lambda \in \mathbb{R}$,

$$\lambda x = (\lambda x_1, \lambda x_2, \cdots, \lambda x_n)$$

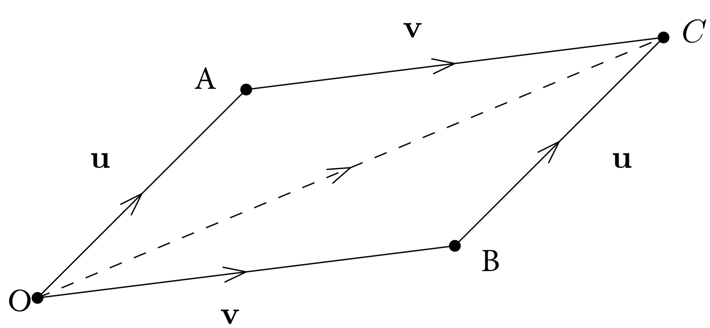
By employing Definition 3.1.3 we therefore have that

$$\overrightarrow{OA} + \overrightarrow{OB} = \overrightarrow{OC},$$

where $C = (x_A + x_B, y_A + y_B, z_A + z_B)$. Analogously, by multiplying $\overrightarrow{OA}$ by -1 we obtain

$$-\overrightarrow{OA} = (-x_A, -y_A, -z_A) = \overrightarrow{AO}.$$

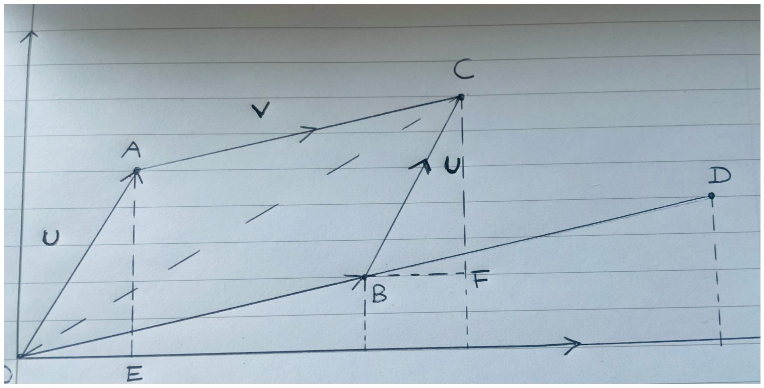
We can give a geometric interpretation to the vector sum $\overrightarrow{OC}$ which is know as parallelogram rule. We begin by constructing the parallelogram with sides $\overrightarrow{OA}$ and $\overrightarrow{OB}$. It will have vertices O, A, B and a fourth one that we will prove to be C. In detail, assuming without loss of generality to be in $\mathbb{R}^2$, we set $\overrightarrow{OA} = \mathbf{u}$, $\overrightarrow{OB} = \mathbf{v}$ and we construct the parallelogram



$\overrightarrow{OC}$ is the diagonal of the parallelogram with start point O. We need to prove that $\overrightarrow{OC} = \mathbf{u} + \mathbf{v}$.

Proposition 3.1.4. *The sum of the vectors* $\overrightarrow{OA} + \overrightarrow{OB}$ *is the vector* $\overrightarrow{OC}$ *which represents the diagonal of the parallelogram constructed on* $\overrightarrow{OA}$ *and* $\overrightarrow{OB}$.

Proof. Let us consider the picture below that we have constructed by shifting the vectors $\mathbf{u}$ and $\mathbf{v}$ from the start point O to the start point B:



By the geometry of the triangles OAE and BCF we easily see that

$$x_C = |\overrightarrow{OC}| + |\overrightarrow{BF}| = x_B + x_A,$$
$$y_C = y_B + |\overrightarrow{FC}| = y_B + y_A.$$

This proves that $\overrightarrow{OC}$ is the sum of the vectors $\overrightarrow{OA}$ and $\overrightarrow{OB}$. $\square$

Since we can identify $\overrightarrow{OA}$ with $\overrightarrow{BC}$ and $\overrightarrow{OB}$ with $\overrightarrow{AC}$ we immediately obtain the following corollary of the parallelogram rule that we will call triangle rule.

Corollary 3.1.5. *Let $OBCA$ be the parallelogram above. Then,*

$$\overrightarrow{OC} = \overrightarrow{OA} + \overrightarrow{AC},$$
$$\overrightarrow{OC} = \overrightarrow{OB} + \overrightarrow{BC}.$$

We are now ready to go back to the definition of the vector $\overrightarrow{AB}$ in $\mathbb{R}^3$ and express it in terms of $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$

Proposition 3.1.6. *Let $A = (x_A, y_A, z_A)$ and $B = (x_B, y_B, z_B)$ be two points in $\mathbb{R}^3$. The vector $\overrightarrow{AB}$ is the sum*

$$\overrightarrow{AO} + \overrightarrow{OB} = (x_B - x_A)\mathbf{i} + (y_B - y_A)\mathbf{j} + (z_B - z_A)\mathbf{k}.$$

Proof. By Corollary 3.1.5 we have that

$$\overrightarrow{AB} = \overrightarrow{AO} + \overrightarrow{OB}.$$

By multiplication by the scalar -1 and definition of the position vectors, we can write

$$\overrightarrow{AB} = -\overrightarrow{OA} + \overrightarrow{OB} = -x_A\mathbf{i} - y_A\mathbf{j} - z_A\mathbf{k} + x_B\mathbf{i} + y_B\mathbf{j} + z_B\mathbf{k}.$$

Hence,

$$\overrightarrow{AB} = (x_B - x_A)\mathbf{i} + (y_B - y_A)\mathbf{j} + (z_B - z_A)\mathbf{k}.$$

$\square$

Conclusion: every vector in $\mathbb{R}^3$ can be regarded as a triple of real numbers, for instance in the case of $\overrightarrow{AB}$ the real numbers $x_B - x_A$, $y_B - y_A$ and $z_B - z_A$. To distinguish vectors from points we could therefore define vectors in $\mathbb{R}^3$ as columns

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

of real numbers. This could be done in general in $\mathbb{R}^n$. Of course, working in dimension $n > 3$ we lose the geometrical interpretation seen so far.

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

In the sequel $\mathbb{R}^n$ is the set of the column vectors

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$$

of real numbers. Addition and multiplication by real scalars are defined as in Definition 3.1.3: for $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$ and $\mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$,

$$v + w = \mathbf{v} = \begin{pmatrix} v_1 + w_1 \\ v_2 + w_2 \\ \vdots \\ v_n + w_n \end{pmatrix};$$

for any $\lambda \in \mathbb{R}$,

$$\lambda\mathbf{v} = \mathbf{v} = \begin{pmatrix} \lambda v_1 \\ \lambda v_2 \\ \vdots \\ \lambda v_n \end{pmatrix}.$$

We recall that addition and multiplication in $\mathbb{R}$ fulfil the following properties:

(i) for all $x, y \in \mathbb{R}$, $x + y = y + x$, (addition in $\mathbb{R}$ is commutative),

(ii) for all $x, y, z \in \mathbb{R}$, $(x + y) + z = x + (y + z)$, (addition in $\mathbb{R}$ is associative),

(iii) for all $x \in \mathbb{R}$, $x + 0 = x$, (0 is the identity for the addition),

(iv) for all $x \in \mathbb{R}$, $x + (-x) = 0$ ($-x$ is the additive inverse of x),

(v) for all $x, y \in \mathbb{R}$, $xy = yx$, (multiplication in $\mathbb{R}$ is commutative),

(vi) for all $x, y, z \in \mathbb{R}$, $(xy)z = x(yz)$, (multiplication in $\mathbb{R}$ associative),

(vii) for all $x \in \mathbb{R}$, $1x = x$, (1 is the identity for the multiplication),

(viii) for all $x \neq 0 \in \mathbb{R}$, $x(\frac{1}{x}) = 1$ ($\frac{1}{x}$ is the multiplicative inverse of x),

(ix) for all $x, y, z \in \mathbb{R}$, $x(y + z) = xy + xz$, (distributive property).

These properties are immediately inherited by the operations defined in $\mathbb{R}^n$ since these operations are defined componentwise. In particular the identity for addition in $\mathbb{R}^n$ is the **0**-vector with all components equal to 0. With summarise this result in the following proposition.

Proposition 3.2.1. *$\mathbb{R}^n$ is a set closed with respect to addition and scalar multiplication. In addition, the following properties hold:*

(i) for all $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$, (addition in $\mathbb{R}^n$ is commutative),

(ii) for all $\mathbf{v}, \mathbf{w}, \mathbf{z} \in \mathbb{R}^n$, $(\mathbf{v} + \mathbf{w}) + \mathbf{z} = \mathbf{v} + (\mathbf{w} + \mathbf{z})$, (addition in $\mathbb{R}^n$ is associative),

(iii) for all $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} + \mathbf{0} = \mathbf{v}$, ($\mathbf{0}$ is the identity for the addition),

(iv) for all $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} + (-\mathbf{v}) = 0$ ($-\mathbf{v}$ is the additive inverse of $\mathbf{v}$),

(v) for all $\alpha, \beta \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, $(\alpha\beta)\mathbf{v} = \alpha(\beta\mathbf{v})$, (multiplication by scalars is associative),

(vi) for all $\mathbf{v} \in \mathbb{R}^n$, $1\mathbf{v} = \mathbf{v}$, (1 is the identity for the multiplication by scalars),

(vi) for all $\alpha \in \mathbb{R}$ and $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, $\alpha(\mathbf{v}, +\mathbf{w}) = \alpha\mathbf{v} + \alpha\mathbf{w}$, (distributive property),

(vii) or all $\alpha\,\beta \in \mathbb{R}$ and $\mathbf{v} \in \mathbb{R}^n$, $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \alpha\mathbf{w}$, (distributive property).

Definition 3.2.2. *A set V with two operations (addition and multiplication by scalars) fulfilling the properties of the proposition above is called vector space.*

If the scalars are real we will talk of real vector spaces however one could choose scalars in $\mathbb{C}$ as well. We therefore have that $\mathbb{R}^n$, $n \geq 1$ is an example of a vector space on $\mathbb{R}$. You will encounter vector spaces in many other modules, for instance in Linear Algebra I and II.

We now introduce the following notion of standard vectors and we see how vectors in $\mathbb{R}^n$ can be easily expressed in terms of them.

Definition 3.2.3. *Let $i = 1, \cdots, n$. The standard vector $\mathbf{e_i}$ is the column vector with the i-th entry equal to 1 and all the others equal to 0.*

We therefore have

$$\mathbf{e_1} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \mathbf{e_2} = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \quad \quad , \mathbf{e_n} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

It is common to use the notations $\mathbf{e_1} = \mathbf{i}$, $\mathbf{e_2} = \mathbf{j}$ in $\mathbb{R}^2$ and $\mathbf{e_1} = \mathbf{i}$, $\mathbf{e_2} = \mathbf{j}$ and $\mathbf{e_3} = \mathbf{k}$ in $\mathbb{R}^3$.

Proposition 3.2.4. *Every vector* $\mathbf{v}$ *in* $\mathbb{R}^n$ *can be written as a unique linear combination of the standard vectors, i.e., there exists a unique choice of* $v_i \in \mathbb{R}$, $i = 1, \cdots, n$, *such that*

$$\mathbf{v} = \sum_{i=1}^{n} v_i \mathbf{e_i}.$$

Proof. We need to prove existence and uniqueness of the components v_i for $i = 1, \cdots, n$. Let

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

By the definition of the sum of vectors and multiplication by scalars seen earlier we have that

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = v_1 \mathbf{e_1} + v_2 \mathbf{e_2} + \cdots + v_n \mathbf{e_n} = \sum_{i=1}^{n} v_i \mathbf{e_i}.$$

This proves the existence part. To prove uniqueness we assume that $\mathbf{v}$ can be written as

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}.$$

and

$$\mathbf{v} = \begin{pmatrix} v_1' \\ v_2' \\ \vdots \\ v_n' \end{pmatrix}.$$

It follows that

$$\begin{pmatrix} v_1 - v_1' \\ v_2 - v_2' \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

This leads to $v_i = v_i'$ for all $i = 1, \cdots, n$.

□

Example 3.2.5. Let $\mathbf{v} = \mathbf{i} - 2\mathbf{j} + 5\mathbf{k}$ and $\mathbf{w} = -4\mathbf{i} + \mathbf{j} - 3\mathbf{k}$. Hence by the properties of the addition of vectors and multiplication by scalars we have that

$$\mathbf{v} + \mathbf{w} = (1-4)\mathbf{i} + (-2+1)\mathbf{j} + (5-3)\mathbf{k} = -3\mathbf{i} - \mathbf{j} + 2\mathbf{k}.$$

Definition 3.2.6. *Let $\mathbf{v} = \sum_{i=1}^n v_i \mathbf{e_i}$ be a vector in $\mathbb{R}^n$. Its length (or norm or module) is defined as*

$$|\mathbf{v}| = \sqrt{\sum_{i=1}^n v_i^2}.$$

Let $\mathbf{v} \neq \mathbf{0}$. By multiplying $\mathbf{v}$ by the scalar $\frac{1}{|\mathbf{v}|}$ we obtain a unit vector, i.e., a vector with norm 1. One can easily check, by using the Pythagoras Theorem that the formula given in Definition 3.2.6 is the actual length of the vector $\mathbf{v}$ in $\mathbb{R}^3$.

Proposition 3.2.7. *Let $\mathbf{v} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ and let P be the point in $\mathbb{R}^3$ with coordinates (a, b, c). Then $|\mathbf{v}|$ is the length of the segment OP.*

Proof. We assume that $P \neq O$ otherwise the statement is trivial. To compute the length of the segment OP we project P on the xy-plane. We get the point $Q = (a, b, 0)$. The length of the segment OQ by Pythagoras theorem on the xy-plane is given by $\sqrt{a^2 + b^2}$. Let us consider the triangle OQP with sides OQ and QP. By applying Pythagoras Theorem again, we have that the length of OP is given by

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

□

We conclude this chapter by remarking that while vectors in $\mathbb{R}^3$ are important for describing physical space, vectors in higher dimension are important as well in many applications. For example, a system of k mass points moving in three-dimensional space can be described by a vector with $n = 3k$ components. In Special Relativity, space and time are combined into a vector with four components. And finally, in Quantum Mechanics, the quantum states of physical systems are described by vectors which, depending on the system, can be basically any size. For this reason, it is important to give a definition in $\mathbb{R}^n$ with arbitrary n. For en extra motivating example please see the More section.

3.3 Problems

1. Let A and B be two points in $\mathbb{R}^2$. Prove that $\overrightarrow{OA} - \overrightarrow{OB}$ is represented by the diagonal in the parallelogram $OACB$ defined by the points A and B.

3.4 More

3.4.1 Application: Google's search algorithm

Modern internet search engines order search results by assigning a page rank to each website. As we will see, this task can be formulated as a problem in linear algebra. Consider an internet with n web sites labeled by an index $k = 1, \cdots, n$. Each site k has n_k links to some of the other sites and is linked to by the sites $L_k \subset \{1, \cdots, n\}$. We would like to assign a page rank x_k to each site k. A first attempt might be to define the page rank of a page k as the number of pages linking to it. However, it is desirable that a page linked to by high-ranked pages has itself a higher rank than a page linked to by low-ranked pages, even if the number of links is the same in each case. So, an improved version might be to define x_k as the sum of all page ranks x_j of the pages linking to k, so as a sum over all x_j , where $j \in L_k$. As a further refinement, a link to page k from a page j with a low number of links n_j might be considered worth more than a link from a page j with a high number of links. Altogether, this leads to the following proposal for the page rank

$$x_k = \sum_{j \in L_k} \frac{x_j}{n_j}.$$

Note that these are n equations (one for each page rank x_k) and that the sum on the RHS runs over all pages j which link to page k. The equation above constitute a system of n linear equations for the variables $x_1, \cdots, x_n$ (while the number of links, n_j , are given constants). The solution of this system is therefore a vector in $\mathbb{R}^n$.

Solving linear systems of equations is an important part of this module and there will be a chapter devoted to them.

3.4.2 The origins of algebra

In the sequel we collect the profiles of some historical mathematicians that can be considered the *founders* of algebra. These profiles are taken from a booklet of Diversifying the Maths Curriculum which is available here and might be of interest for you to read about mathematicians from a diverse background: women, ethnic minorities, queer, disable, etc..

Diophantus

Diophantus, the 'father of algebra', is best known for his *Arithmetica*, a work on the solution of algebraic equations and on the theory of numbers. Essentially nothing is known of his life and the date at which he lived (about 200-284 in Alexandria, Egypt) is also not sure.

The *Arithmetica* is a collection of 130 problems giving numerical solutions of determinate and indeterminate equations. The method for solving the latter is now known as Diophantine analysis.

Even if Diophantus is regarded as the 'father of algebra' there is no doubt that many of the methods for solving linear and quadratic equations go back to Babylonian mathematics. Diophantus' work has become famous in recent years due to its connection with Fermat's Last Theorem.

Read more on MacTutor.

Euclid of Alexandria

Euclid of Alexandria (325 BCE - 265 BCE) is the most prominent mathematician of antiquity. His treatise on mathematics *The Elements* makes him the leading mathematics teacher of all time. However little is known of Euclid's life except that he taught at Alexandria in Egypt.

The Elements became the centre of mathematical teaching for 2000 years. Probably Euclid did not prove the results in *The Elements* but the organisation of the material and its exposition are due to him.

The Elements begins with definitions and five postulates. The first three postulates are postulates of construction, for example the first postulate states that it is possible to draw a straight line between any two points. These postulates also implicitly assume the existence of points, lines and circles and then the existence of other geometric objects are deduced from the fact that these exist.

Barten van der Waerden, a Dutch mathematician famous for his work in topology and history of mathematics, says: *Almost from the time of its writing and lasting almost to the present,* The Elements *has exerted a continuous and major influence on human affairs. It was the primary source of geometric reasoning, theorems, and methods at least until the advent of non-Euclidean geometry in the 19th century. It is sometimes said that, next to the Bible,* The Elements *may be the most translated, published, and studied of all the books produced in the Western world.*

Read more on Britannica

Abu Ja'far Muhammad ibn Musa Al-Khwarizmi

Al-Khwarizmi was an Islamic mathematician and astronomer during the 8th and 9th centries. Al-Khwarizmi is recognised for his contributions to Hindu-Arabic numerals and for his treatise *Hisab al-jabr w'al-muqabala*, which is the first recorded text on the topic of algebra. In the book, Al-Khwarizmi introduced algebraic operations *al-jabr* and *al-muqabala* to simplify quadratic equations, which he then solved with the geometric picture of "completing the square" that is now taught in schools. The Latinised translations of "Al-Khwarizmi" and "al-jabr" in the 12th century gave us the origin of the words "algorithm" and "algebra" respectively.

Little is know about Al-Khwarizmi's life, other than that he worked in the House of Wisdom in Baghdad. Here, he and his colleagues (the Bana Musa brothers) would translate scientific and philosophical manuscripts from Ancient Greek and also publish original research in algebra, geometry and astronomy.

Read more on MacTutor or Encyclopedia Britannica.

Hypatia

Hypatia of Alexandria (370-415), born in Egypt, is the first known woman in mathematics who made major contributions to the field. Hypatia was the daughter of the mathematician and philosopher Theon of Alexandria. She became head of the Platonist school at Alexandria in about 400 CE. Hypatia was accused of paganism by the early Christians, and was eventually murdered for her scientific knowledge.

Despite not having evidence of any original mathematical research, it is known that she assisted her father in writing his commentary on Ptolemy's *Almagest* and in producing a new version of Euclid's *Elements*. Hypatia also wrote commentaries on Diophantus's *Arithmetica*, on Apollonius's *Conics* and on Ptolemy's astronomical works. She provided her advice to Synesius on the construction of an astrolabe and a hydroscope. Throughout her life, Hypatia was a notable compiler, editor and preserver of earlier mathematical works. However, almost all her work is lost.

Read more on MacTutor.

Chapter 4

A little bit of geometry: equations of lines

In this chapter we recall some basic facts about lines in $\mathbb{R}^3$. In particular we will express them via the vector language that we have learn in the previous chapter. Note that we will find different ways to describe the same geometrical object in $\mathbb{R}^3$. The definition seen in this chapter can be easily generalised to the n-dimensional context.

4.1 Equations of Lines

4.1.1 Parametric type equations

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

$$\overrightarrow{PR} = \overrightarrow{PO} + \overrightarrow{OR} = \overrightarrow{OR} - \overrightarrow{OP} = \mathbf{r} - \mathbf{p} = \lambda \mathbf{u}.$$

Definition 4.1.1. *The equation* $\mathbf{r} = \mathbf{p} + \lambda \mathbf{u}$ *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 λ. 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 λ for which $\mathbf{r} = \mathbf{p} + \lambda \mathbf{u}$ then R lies on l; if there is no such λ 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:

$$\begin{cases} x = p_1 + \lambda u_1 \\ y = p_2 + \lambda u_2 \\ z = p_3 + \lambda u_3 \end{cases} \tag{4.1.1}$$

Definition 4.1.2. *The equations* (4.1.1) *are called the* parametric equations *for the line* l.

The variable λ 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*).

4.1.2 Cartesian equations

If $u_1 \neq 0, u_2 \neq 0, u_3 \neq 0$ we can eliminate the parameter λ 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.

4.1.3 Line described by two points

Another natural way of describing a line is by giving two points that lie on it. If P and Q are distinct points 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.

4.2 Lines through the origin and example of sub-vector spaces

Let us consider the equation of the line l passing through the origin O and defined via the vector $\mathbf{u}$, i.e., $\mathbf{r} = \lambda\mathbf{u}$, for $\lambda \in \mathbb{R}$. This gives us the set

$$V = \{\lambda\mathbf{u} : \lambda \in \mathbb{R}\}.$$

V is a vector space according to Definition 3.2.2 seen in the previous chapter. Indeed, it is closed with respect to sum of vectors and multiplication of scalars. These operations have all the properties required by the definition of a vector space. Becuase it is contained in a vector space itself we say that it is a sub-vector space of $\mathbb{R}^n$.

Proposition 4.2.1. *For all $v, v_1, v_2 \in V$ and all $\alpha \in \mathbb{R}$,*

$$\begin{aligned} v_1 + v_2 &\in V, \\ \alpha v &\in V. \end{aligned}$$

Proof. If $v_1, v_2 \in V$ then there exist $\lambda_1, \lambda_2 \in \mathbb{R}$ such that $v_1 = \lambda_1\mathbf{u}$ and $v_2 = \lambda_2\mathbf{u}$. Hence, by the distributive property of the product by scalars we have

$$v_1 + v_2 = \lambda_1\mathbf{u} + \lambda_2\mathbf{u} = (\lambda_1 + \lambda_2)\mathbf{u}.$$

It follows that $v_1 + v_2 \in V$. Analogously, if $v \in V$ then there exists $\lambda \in \mathbb{R}$ such that $v = \lambda\mathbf{u}$. Hence, by the associative property of the product by scalars,

$$\alpha v = \alpha(\lambda\mathbf{u}) = (\alpha\lambda)\mathbf{u}.$$

This proves that αv belongs to V as well. $\square$

Proposition 4.2.2. *Let* $\mathbf{i}$ *and* $\mathbf{j}$ *be the standard vector in* $\mathbb{R}^2$*. The set,*

$$V = \{x\mathbf{i} + y\mathbf{j} : \ x, y \in \mathbb{R}\},$$

is a sub-vector space of $\mathbb{R}^2$*.*

V is the set of all linear combinations of $\mathbf{i}$ and $\mathbf{j}$ and it is denoted by $\mathrm{Span}(\mathbf{i}, \mathbf{j})$. Geometrically speaking it is the (x, y)-plane. We leave the proof of Proposition 4.2.2 as an exercise.

Let now $\mathbf{u}$ and $\mathbf{v}$ be two non-zero vectors in $\mathbb{R}^3$.

$$\mathrm{Span}(\mathbf{u}, \mathbf{v}) = \{\alpha\mathbf{u} + \beta\mathbf{v} : \ \alpha, \beta \in \mathbb{R}\}.$$

It is immediate to see that $\mathrm{Span}(\mathbf{u}, \mathbf{v})$ is a sub-vector space of $\mathbb{R}^3$. However the question is does $\mathrm{Span}(\mathbf{u}, \mathbf{v})$ represent a plane in $\mathbb{R}^3$ or a line? Analogously, if we take three vectors, let's say $\mathbf{u}, \mathbf{v}, \mathbf{w}$ does $\mathrm{Span}(\mathbf{u}, \mathbf{v}, \mathbf{w})$ represents the whole space, or just a plane or a line? It all depends on the relationship between the vectors $\mathbf{u}, \mathbf{v}, \mathbf{w}$ that leads to the concepts of *linear dependence and independence* that you will see in Linear Algebra I. In a nutshell, if $\mathbf{w}$ is a linear combination of $\mathbf{u}$ and $\mathbf{w}$ then

$$\mathrm{Span}(\mathbf{u}, \mathbf{v}, \mathbf{w}) = \mathrm{Span}(\mathbf{u}, \mathbf{v}).$$

If $\mathbf{v}$ is not of the type $\lambda\mathbf{u}$, i.e., $\mathbf{u}$ and $\mathbf{v}$ are linearly independent then $\mathrm{Span}(\mathbf{u}, \mathbf{v})$ represents a plane in $\mathbb{R}^3$. If there exists $\lambda \in \mathbb{R}$ such that $\mathbf{v} = \lambda u$ then

$$\mathrm{Span}(\mathbf{u}, \mathbf{v}) = \{\alpha\mathbf{u} : \ \alpha \in \mathbb{R}\}.$$

This means that $\mathbf{u}$ and $\mathbf{v}$ are linearly dependent and therefore $\mathrm{Span}(\mathbf{u}, \mathbf{v})$ is the line through the origin defined by the vector $\mathbf{u}$.

4.3 Problems

1. Let $\mathbf{p} = \overrightarrow{OP}$ and let $\mathbf{u}$ be a vector in $\mathbb{R}^3$. Let

$$V = \{\mathbf{p} + \lambda\mathbf{u} : \ \lambda \in \mathbb{R}\}.$$

   Is V a sub-vector space? Justify your answer with a short argument.

2. Prove Proposition 4.2.2.

4.4 More

4.4.1 The beginning of Cartesian Geometry

René Descartes (1596-1650) was a French philosopher whose work, La Géométrie includes his application of algebra to geometry from which we now have Cartesian geometry.

Descartes was educated at the Jesuit college of La Flèche in Anjou. He took courses in classics, logic and traditional Aristotelian philosophy. He also learnt mathematics from the books of Clavius, while studying all the branches of mathematics, namely arithmetic, geometry, astronomy and music. School had made Descartes understand how little he knew, the only subject which was satisfactory in his eyes was mathematics. This idea became the foundation for his way of thinking, and was to form the basis for all his works.

[I want to promote a] completely new science by which all questions in general may be solved that can be proposed about any kind of quantity, continuous as well as discrete. But each according to its own nature. ... In arithmetic, for instance, some questions can be solved by rational numbers, some by surd numbers, and others can be imagined but not solved. For continuous quantity I hope to prove that, similarly, certain problems can be solved by using only straight or circular lines, that some problems require other curves for their solution, but still curves which arise from one single motion and which therefore can be traced by the new compasses, which I consider to be no less certain and geometrical than the usual compasses by which circles are traced; and, finally, that other problems can be solved by curved lines generated by separate motions not subordinate to one another.

Around 1628, after some time travelling in Europe, Descartes settled in Holland. Here, encouraged by some intellectual friends, he wrote a treatise on science under the title *Discours de la méthode pour bien conduire sa raison et chercher la vérité dans les sciences.* Three appendices to this work were *La Dioptrique*, *Les Météores*, and *La Géométrie.*

The work describes what Descartes considers is a more satisfactory means of acquiring knowledge than that presented by Aristotle's logic. Only mathematics, Descartes feels, is certain, so all must be based on mathematics.

La Géométrie is by far the most important part of this work. The importance of this work can be summarised in four points:

1. He makes the first step towards a theory of invariants, which removes arbitrariness;
2. Algebra makes it possible to recognise the typical problems in geometry and to bring together problems which in geometrical dress would not appear to be related at all;
3. Algebra imports into geometry the most natural principles of division and the most natural hierarchy of method.
4. Not only can questions of solvability and geometrical possibility be decided elegantly, quickly and fully from the parallel algebra, without it they cannot be decided at all.

Chapter 5

Scalar Product and Vector Product

In this chapter we will see two different types of vector products in $\mathbb{R}^3$ and their geometrical applications. In particular we will see how geometrical concepts can be expressed in terms of operations between vectors.

5.1 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 θ (in radians) between the line segments $\overrightarrow{AB}$ and $\overrightarrow{AC}$ with $0 \leq \theta \leq \pi$.

Definition 5.1.1. *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 θ *is the angle between* $\mathbf{u}$ *and* $\mathbf{v}$.

Definition 5.1.2. *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 5.1.3. *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 $u_1v_1 + u_2v_2 + u_3v_3 = 0$ and so the result is true.

Suppose that $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$ and let θ 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. □

Remark 5.1.4. The scalar product can be defined in general for two vectors in $\mathbb{R}^n$. Indeed, as a consequence of Theorem 5.1.3, given $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ with components $(u_i)_{i=1,\cdots,n}$ and $(v_i)_{i=1,\cdots,n}$, respectively we define $\mathbf{u}\cdot\mathbf{v}$ as

$$\sum_{i=1}^{n} u_i v_i.$$

In $\mathbb{R}^3$, Theorem 5.1.3 can be used to find the angle θ between two non-zero vectors given in coordinates. Rearranging the definition of $\mathbf{u}\cdot\mathbf{v}$ and substituting the formula of Theorem 5.1.3 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 5.1.5 (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 5.1.3. □

5.2 Interesting inequalities for vectors

In this section we prove some important inequalities that are related to the scalar product. We work in $\mathbb{R}^3$ however these inequalities hold in $\mathbb{R}^n$ as well.

Proposition 5.2.1 (Cauchy-Schwarz Inequality)**.** *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be two vectors in* $\mathbb{R}^3$*. The following inequality holds:*

$$|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|.$$

Proof. If either $\mathbf{u}$ or $\mathbf{v}$ is the zero vector then the inequality is trivial since it turns out to be the equality $0 = 0$. Let us assume that $\mathbf{u} \neq \mathbf{0}$ and $\mathbf{v} \neq \mathbf{0}$. For the sake of simplicity, assume first that $|\mathbf{u}| = |\mathbf{v}| = 1$. By the properties of the scalar product we have

$$\begin{aligned} 0 &\leq |\mathbf{u} + \mathbf{v}|^2 = (\mathbf{u} + \mathbf{v}) \cdot (\mathbf{u} + \mathbf{v}) = |\mathbf{u}|^2 + 2\mathbf{u} \cdot \mathbf{v} + |\mathbf{v}|^2 = 2(1 + \mathbf{u} \cdot \mathbf{v}), \\ 0 &\leq |\mathbf{u} - \mathbf{v}|^2 = (\mathbf{u} - \mathbf{v}) \cdot (\mathbf{u} - \mathbf{v}) = |\mathbf{u}|^2 - 2\mathbf{u} \cdot \mathbf{v} + |\mathbf{v}|^2 = 2(1 - \mathbf{u} \cdot \mathbf{v}). \end{aligned}$$

This implies

$$\mathbf{u} \cdot \mathbf{v} \geq -1, \qquad \mathbf{u} \cdot \mathbf{v} \leq 1,$$

i.e.

$$|\mathbf{u} \cdot \mathbf{v}| \leq 1 = |\mathbf{u}||\mathbf{v}|.$$

This proves the Cauchy-Schwarz inequality when the vectors $\mathbf{u}$ and $\mathbf{v}$ have norm 1. For arbitrary non-zero vectors $\mathbf{u}$ and $\mathbf{v}$, we have immediately that the inequality holds for $\frac{\mathbf{u}}{|\mathbf{u}|}$ and $\frac{\mathbf{v}}{|\mathbf{v}|}$. Hence,

$$\left|\frac{\mathbf{u}}{|\mathbf{u}|} \cdot \frac{\mathbf{v}}{|\mathbf{v}|}\right| \leq 1$$

which implies

$$|\mathbf{u} \cdot \mathbf{v}| \leq |\mathbf{u}||\mathbf{v}|.$$

□

Note that the Cauchy Schwarz inequality tell us that

$$-1 \leq \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|} \leq 1.$$

This is indeed meaningful since as we have seen the fraction $\frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{v}|}$ is the cosine of an angle.

The Cauchy-Schwarz inequality can be used to prove the famous triangle inequality.

Proposition 5.2.2 (Triangle inequality)**.** *Let* $\mathbf{u}$ *and* $\mathbf{v}$ *be two vectors in* $\mathbb{R}^3$*. The following inequality holds:*

$$|\mathbf{u}+\mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|.$$

Proof. As for the Cauchy-Schwarz Inequality, the triangle inequality is trivial and becomes an equality when either $\mathbf{u}$ or $\mathbf{v}$ is the zero-vector. In general, by applying the Cauchy-Schwartz inequality we have

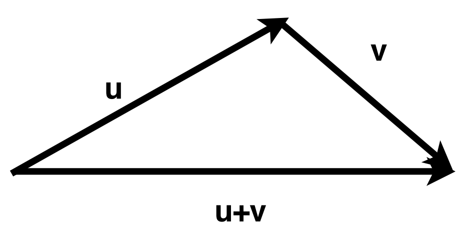
$$|\mathbf{u}+\mathbf{v}|^2 = (\mathbf{u}+\mathbf{v})\cdot(\mathbf{u}+\mathbf{v}) = |\mathbf{u}|^2 + 2\mathbf{u}\cdot\mathbf{v} + |\mathbf{v}|^2 \leq |\mathbf{u}|^2 + |\mathbf{v}|^2 + 2|\mathbf{u}\cdot\mathbf{v}| \leq (|\mathbf{u}| + |\mathbf{v}|)^2.$$

This is equivalent to

$$|\mathbf{u}+\mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|.$$

□

The geometrical interpretation of the triangle inequality is the following:



5.3 The Equation of a Plane

A plane Π in 3-space can be specified by giving

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

By $\mathbf{n}$ being orthogonal to Π we mean that for any two points A and B on Π, 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 Π 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 Π if and only if it satisfies $ax+by+cz=d$.

5.4 Distance from a Point to a Plane

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

Suppose that Π 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 Π closest to Q we need that the vector represented by $\overrightarrow{MQ}$ is orthogonal to the plane Π 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 Π 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 Π.

We could also use this method to find the position vector of M, the point on Π 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 5.4.1. *If the plane Π has equation $\mathbf{r}\cdot\mathbf{n} = d$, and the point Q has position vector $\mathbf{q}$, then the distance between Q and Π is*

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

and the point on Π that is closest to Q has position vector

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

5.5 The vector product

Definition 5.5.1. *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 5.5.2. 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 5.5.3. *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*

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

where θ *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

$$|\mathbf{u}\times\mathbf{v}|^2 = |\mathbf{u}|^2|\mathbf{v}|^2 - (\mathbf{u}\cdot\mathbf{v})^2\,, \tag{5.5.1}$$

because

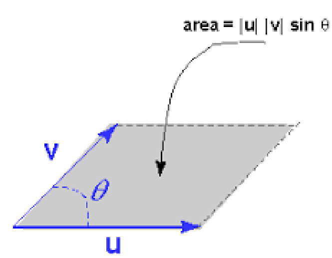
$$\begin{aligned} |\mathbf{u}\times\mathbf{v}|^2 &= (u_2v_3-u_3v_2)^2 + (u_3v_1-u_1v_3)^2 + (u_1v_2-u_2v_1)^2 \\ &= u_2^2v_3^2 + u_3^2v_2^2 + u_3^2v_1^2 + u_1^2v_3^2 + u_1^2v_2^2 + u_2^2v_1^2 - 2(u_2v_3u_3v_2 + u_3v_1u_1v_3 + u_1v_2u_2v_1) \\ &= (u_1^2+u_2^2+u_3^2)(v_1^2+v_2^2+v_3^2) - (u_1v_1+u_2v_2+u_3v_3)^2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

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 (5.5.1) gives $|\mathbf{u}\times\mathbf{v}|^2 = |\mathbf{u}|^2|\mathbf{v}|^2\sin^2\theta$, and hence $|\mathbf{u}\times\mathbf{v}| = |\mathbf{u}||\mathbf{v}|\sin\theta$. ∎

As a consequence of Proposition 5.5.3 we have that $|\mathbf{u}\times\mathbf{v}|$ is equal to the area of the parallelogram indicated in the picture below.



The vector product appears quite often in Physics:

(a) In mechanics the angular momentum of a mass m at position $\mathbf{r}$ and with velocity $\dot{\mathbf{r}}$ is given by $\mathbf{L} = m\,\mathbf{r} \times \dot{\mathbf{r}}$.

(b) The force a magnetic field $\mathbf{B}$ exerts on a particle with charge q and velocity $\dot{\mathbf{r}}$, the so-called Lorentz force, is given by $\mathbf{F} = q\,\dot{\mathbf{r}} \times \mathbf{B}$.

(c) The velocity of a point with coordinate $\mathbf{r}$ in a rotating coordinate system with angular velocity $\mathbf{w}$ is given by $\mathbf{v} = \mathbf{w} \times \mathbf{r}$.

If you want to learn more about applications of the vector products seen in this chapter please refer to the More section at the end of this chapter.

5.6 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 Π 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 Π.

A vector equation for Π is $\mathbf{r} \cdot \mathbf{n} = \mathbf{a} \cdot \mathbf{n}$ (since A is on Π and $\mathbf{n}$ is orthogonal to Π). 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

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

5.7 Equation of the circle in $\mathbb{R}^2$ and the sphere in $\mathbb{R}^3$

We can use the notions seen on vectors so far to deduce the equation of the circle and the sphere. In both cases these are set of points having the same distance from the origin. More precisely, given a radius $r > 0$ the circle with radius r and centre the origin is the set of all position vectors $\mathbf{p} = (x, y)^T$ (the transpose of a row-vector is a column-vector) with length r, i.e.,

$$|\mathbf{p}| = r \quad \Leftrightarrow \quad |\mathbf{p}|^2 = r^2 \quad \Leftrightarrow \quad x^2 + y^2 = r.$$

Analogously, if we argue in $\mathbb{R}^3$ with $\mathbf{p} = (x, y, z)^T$ we find the equation of the sphere:

$$|\mathbf{p}| = r \quad \Leftrightarrow \quad |\mathbf{p}|^2 = r^2 \quad \Leftrightarrow \quad x^2 + y^2 + y^2 = r.$$

If we want to have a centre different from the origin, let's say $C = (a, b)$ or $C = (a, b, c)$ then we set $\mathbf{p} = (x - a, y - b)^T$ and $\mathbf{p} = (x - a, y - b, z - c)^T$, respectively.

Note that we can define *spheres* in any space dimension by using this approach. Indeed, given $\mathbf{x}$ vector in $\mathbb{R}^{n+1}$ with components x_i, $i = 1, \cdots, n+1$ we define the n-sphere with radius 1 as follows:

$$S^n = \{(x_1, \cdots, x_{n+1}) \in \mathbb{R}^{n+1} : \quad |\mathbf{x}| = 1\}.$$

The corresponding equation is

$$\sum_{i=1}^{n+1} x_i^2 = 1.$$

5.8 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 θ 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).

5.9 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

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.

5.10 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 Π be the plane with Cartesian equation $ax + by + cz = d$. The set of points on Π 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.

5.11 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 Π with equation $ax + by + cz = d$ and the line l with parametric equations
$$\begin{cases} x = p_1 + \lambda u_1 \\ y = p_2 + \lambda u_2 \\ z = p_3 + \lambda u_3 \end{cases}$$
we solve
$$a(p_1 + \lambda u_1) + b(p_2 + \lambda u_2) + c(p_3 + \lambda u_3) = d$$
for λ. 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 λ back into the parametric equations for the line then gives the coordinates of the point of intersection.

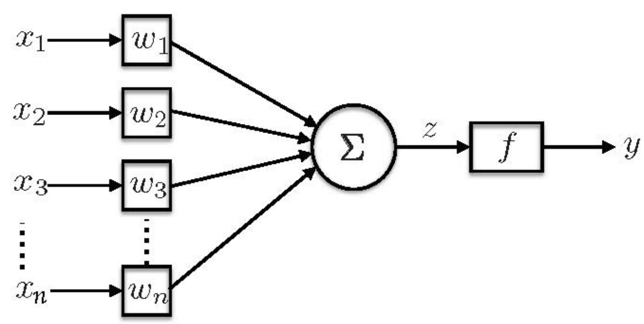
- To find the intersection of the line l_1 with parametric equations
$$\begin{cases} x = p_1 + \lambda u_1 \\ y = p_2 + \lambda u_2 \\ z = p_3 + \lambda u_3 \end{cases}$$
and the line l_2 with parametric equations
$$\begin{cases} x = q_1 + \mu v_1 \\ y = q_2 + \mu v_2 \\ z = q_3 + \mu v_3 \end{cases}$$
we solve
$$\begin{cases} p_1 + \lambda u_1 = q_1 + \mu v_1 \\ p_2 + \lambda u_2 = q_2 + \mu v_2 \\ p_3 + \lambda u_3 = q_3 + \mu v_3 \end{cases}$$
or equivalently solve
$$\begin{cases} \lambda u_1 - \mu v_1 = q_1 - p_1 \\ \lambda u_2 - \mu v_2 = q_2 - p_2 \\ \lambda u_3 - \mu v_3 = q_3 - p_3 \end{cases}$$
for λ and μ. 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).

5.12 Problems

5.13 More

5.13.1 Application: The perceptron - a basic building block of artificial neural networks

Artificial neural networks constitute an important set of methods in modern computing which are motivated by the structure of the human brain. Many of the operating principles of artificial neural networks can be formulated and understood in terms of linear algebra. Here, we would like to introduce one of the basic building blocks of artificial neural networks - the perceptron. The structure of the perceptron is schematically illustrated in the figure below. It receives n real input values



$x_1, x_2, \cdots, x_n$ which can be combined into an n-dimensional input vector $\mathbf{x} = (x_1, x_2, \cdots, x_n)^T$. The internal state of the perceptron is determined by three pieces of data: the real values $w_1, \cdots w_n$ called the *weights* which can be arranged into the n-dimensional weight vector $\mathbf{w} = (w_1, \cdots w_n)^T$, a real number θ, called the *threshold* of the perceptron and a real function f, referred to as the *activation* function. In terms of this data, the perceptron computes the output values y from the input values $\mathbf{x}$ as

$$z = \mathbf{w} \cdot \mathbf{x} - \theta, \qquad y = f(z).$$

The activation function can, for example, be chosen as the sign function

$$f(z) = 1, \quad \text{for } z \geq 0, \qquad f(z) = -1, \quad \text{for } z < 0$$

Given this set-up, the functioning of the perceptron can be phrased in geometrical terms. Consider the equation in $\mathbb{R}^n$ with coordinates $\mathbf{x}$ given by

$$\mathbf{w} \cdot \mathbf{x} = \theta$$

Note that this is simply the equation of a plane (or a hyper-plane in dimensions $n > 3$) in Cartesian form. If a point $\mathbf{x} \in \mathbb{R}^n$ is above (or on) this plane, so that $\mathbf{w} \cdot \mathbf{x} - \theta \geq 0$ then the output of the perceptron is $+1$. On the other hand, for a point $\mathbf{x} \in \mathbb{R}^n$ below this plane, so that $\mathbf{w} \cdot \mathbf{x} - \theta < 0$, the perceptron's output is -1. In other words, the purpose of the perceptron is to decide whether a certain point $\mathbf{x}$ is above or below a given plane

In the context of artificial neural networks the perceptron corresponds to a single neuron. Proper artificial neural networks can be constructed by combining a number of perceptrons into a network, using the output of certain perceptrons within the network as input for others. Such networks of perceptrons correspond to collections of (hyper-) planes and are, for example, applied in the context of pattern recognition.

5.13.2 Brief history of the Cauchy-Schwarz inequality

The Cauchy-Schwarz inequality made its first appearance in the work *Cours d'analyse de l'École Royal Polytechnique* by the French mathematician Augustin-Louis Cauchy (1789-1857). In this work, which was published in 1821, he introduced the inequality in the form of finite sums, although it was only writ- ten as a note. In 1859 a Russian former student of Cauchy, Viktor Yakovlevich Bunyakovsky (1804-1889), published a work on inequalities in the journal *Mémoires de l'Académie Impériale des Sciences de St-Petersbourg.* Here he proved the inequality for infinite sums, written as integrals, for the first time.

(From left) Augustin-Louis Cauchy, Viktor Yakovlevich Bunyakovsky and Karl Hermann Amandus Schwarz.

In 1888, Karl Hermann Amandus Schwarz (1843-1921) published a work on minimal surfaces named *Über ein die flächen kleinsten flächeninhalts betreffendes problem der variationsrechnung* in which he found himself in need of the integral form of Cauchy's inequality, but since he was unaware of the work of Bunyakovsky, he presented the proof as his own. The proofs of Bunyakovsky and Schwarz are not similar and Schwarz's proof is therefore considered independent, although of a later date. A big difference in the methods of Bunyakovsky and Schwarz was in the rigidity of the limiting process, which was of bigger importance for Schwarz. The arguments of Schwarz can also be used in more general settings like in the framework of inner product spaces For this reason the inequality is known without the name of Bunyakovsky.

Chapter 6

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.

6.1 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 $a_1, \dots, a_n$ and b are given real numbers and $x_1, \dots, x_n$ are variables.

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

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

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

It is convenient to rewrite the system above into matrix form. We recall that 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. We will see the basics of matrix theory in the next chapter.

The system

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

is defined via the matrix

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

and can be written as the equation

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

where

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

and

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}.$$

$\mathbf{x}$ is the column of the unknowns while $\mathbf{b}$ is the known right-hand side.

Example 6.1.1.

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

(a) is a 2×2 system, (b) is a 2×3 system, and (c) is a 3×2 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. They are the components of the vector $\mathbf{x}$ which solves the equation $A\mathbf{x} = \mathbf{b}$.

Example 6.1.2. $(1, 2)$ is a solution of Example 6.1.1 (a).
For each $\alpha \in \mathbb{R}$, the 3-tuple $(3, \alpha, \alpha)$ is a solution of Example 6.1.1 (b) (Exercise: check this!).
Example 6.1.1 (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 6.1.3. *Two $m \times n$ systems are said to be **equivalent**, if they have the same solution set.*

Example 6.1.4. 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 6.1.5. *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:

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

Given an $m \times n$ linear system

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

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 6.1.6.

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

A system can be solved by performing operations on the augmented matrix. Corresponding to the three operations given in Lemma 6.1.5 we have the following three operations that can be applied to the augmented matrix, called **elementary row operations**.

Definition 6.1.7 (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.*

6.2 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 6.2.1. *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 6.2.2. 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 6.2.3.

(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 6.2.4. Determine the solution set of the systems given by the following augmented matrices in row echelon form:

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

Solution. (a) The corresponding system is

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

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

(b) The corresponding system is

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

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 α and β 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 $\{\,(2+2\alpha-\beta, \alpha, 1+2\beta, \beta) \mid \alpha, \beta \in \mathbb{R}\,\}$. □

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 6.2.5. 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:

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

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

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

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

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 6.2.6. 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:

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

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

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 6.2.7.

(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 6.2.8. 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.

6.3 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 6.3.1. *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 6.3.2. *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 6.3.3. *A linear system*

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} \tag{6.3.1}$$

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

Given an inhomogeneous system (6.3.1), call the system obtained by setting all b_i's to zero, the ***associated homogeneous system*** *.*

Example 6.3.4.

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

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 6.3.5. *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 6.3.2, 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 6.3.6. *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.

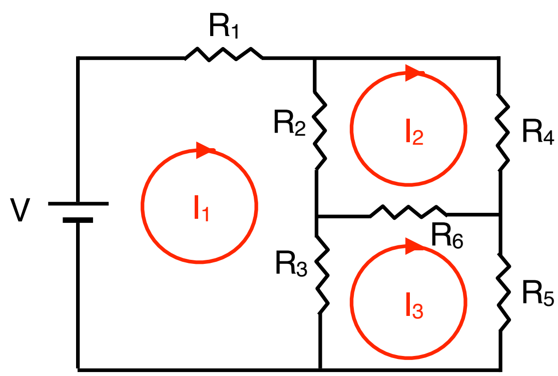
Conversely, if the associated homogeneous system of a given system has the zero solution as its unique solution, then the original inhomogeneous system must have a solution, and this solution must be unique. $\square$

6.4 Problems

6.5 More

6.5.1 Application: linear algebra and circuits

Electrical circuits with batteries and resistors, such as the circuit in the figure below (a simple 3-loop with a battery and resistors) can be described using methods from linear algebra.



To do this, assume that the circuit contains n loops and assign currents I_i, where $i = 1, \cdots, n$, to each loop. Then, applying Ohm's law and Kirchhoff's voltage low (*the voltages along a closed loop must sum to zero*) to each loop leads to the linear system

$$\begin{array}{ccccc} R_{11}I_1 & \cdots & R_{1n}I_n & = & V_1 \\ \vdots & \vdots & \vdots & = & \vdots \\ R_{n1}I_1 & \cdots & R_{nn}I_n & = & V_n \end{array}$$

where R_{ij} describe the various resistors and V_i correspond to the voltages of the batteries. If we introduce the $n \times n$-matrix R with entries R_{ij}, the current vector $\mathbf{I} = (I_1, \cdots, I_n)^T$ and the vector $\mathbf{V} = (V_1, \cdots, V_n)^T$ for the battery voltages this system can, of course, also be written as

$$R\mathbf{I} = \mathbf{V}.$$

This is an $n \times n$ linear system, where we think of the resistors and battery voltages as given, while the currents $I_1, \cdots, I_n$ are the unknowns.

For example, consider the circuit in the picture above. To its three loops we assign the currents I_1, I_2 and I_3 as indicated in the picture. Kirchhoff's voltage law applied to the three loops then leads to

$$\begin{aligned} R_1I_1 + R_2(I_1 - I_2) + R_3(I_1 - I_3) &= V \\ R_2(I_2 - I_1) + R_4I_2 + R_6(I_2 - I_3) &= 0 \\ R_3(I_3 - I_1) + R_6(I_3 - I_2) + R_5I_3 &= 0 \end{aligned}$$

equivalent to

$$
\begin{aligned}
(R_1 + R_2 + R_3)I_1 - R_2 I_2 - R_3 I_3 &= V \\
-R_2 I_1 + (R_2 + R_4 + R_6)I_2 - R_6 I_3 &= 0 \\
-R_3 I_1 - R_6 I_2 + (R_3 + R_5 + R_6)I_3 &= 0.
\end{aligned}
$$

With the current and voltage vectors $\mathbf{I} = (I_1, i_2, I_3)^T$ and $\mathbf{V} = (V, 0, 0)^T$ the matrix R is the equation $R\mathbf{I} = \mathbf{V}$ is then given by

$$
R = \begin{pmatrix} R_1 + R_2 + R_3 & -R_2 & -R_3 \\ -R_2 & R_2 + R_4 + R_6 & -R_6 \\ -R_3 & -R_6 & R_3 + R_5 + R_6 \end{pmatrix}.
$$

Chapter 7

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.

7.1 Matrices and basic properties

We begin by recalling the definition of matrix.

Definition 7.1.1. *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 7.1.2. If

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

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

Definition 7.1.3 (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 7.1.4 (Scalar multiplication)**.** *If $A = (a_{ij})_{m \times n}$ and α is a scalar, then αA is the $m \times n$ matrix whose (i, j)-entry is αa_{ij}.*

Definition 7.1.5 (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 7.1.6. 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 7.1.7 (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 7.1.8. *Let A, B and C be matrices of the same size, and let α and β 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 7.1.9. Simplify $2(A + 3B) - 3(C + 2B)$, where A, B, and C are matrices with the same size.

Solution.

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

□

Remark 7.1.10. From Theorem 7.1.8 we have that the set $\mathcal{A}_{m\times n}$ of all matrices of sixe $m \times n$ together with the operations of addition of matrices and multiplication by real scalar is a vector space.

Definition 7.1.11 (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 7.1.12. 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 7.1.13 (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 7.1.14. *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) $I_mA = AI_n = A$;

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_mA 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 $I_mA = A$.

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

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

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

$\square$

- 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 7.1.15.

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

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 7.1.16. *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 7.1.17. *If A is a square matrix, a matrix B is called an* inverse *of A if*

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

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 2×2 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 7.1.18. *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 7.1.19. *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 $B^{-1}A^{-1}$. □

7.2 Transpose of a matrix

The first new concept we encounter is the following:

Definition 7.2.1. *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 7.2.2.

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

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

Matrix transposition satisfies the following rules:

Theorem 7.2.3. *Assume that α 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 7.2.4. *Let A be invertible. Then A^T is invertible and*

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

Proof. See a Coursework exercise. □

7.3 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 7.3.1. *A matrix is said to be **symmetric** if $A^T = A$.*

Note that a symmetric matrix is necessarily square.

Example 7.3.2.

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

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

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 7.3.3. *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 7.3.4.

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

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

We close this section with the following two observations:

Theorem 7.3.5. *The sum and product of two upper triangular matrices of the same size is upper triangular.*

Proof. See a Coursework exercise. □

7.4 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

7.5 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

$$\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots & & \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array} . \tag{7.5.1}$$

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

$$A\mathbf{x} = \mathbf{b} , \tag{7.5.2}$$

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 7.5.1. 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 (7.5.1) in the form (7.5.2) has a number of other advantages which will become clearer shortly.

7.6 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 7.6.1. *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):*

$$A\mathbf{x} = \mathbf{b} \tag{7.6.3}$$

$$MA\mathbf{x} = M\mathbf{b} \tag{7.6.4}$$

Proof. Note that if $\mathbf{x}$ satisfies (7.6.3), then it clearly satisfies (7.6.4). Conversely, suppose that $\mathbf{x}$ satisfies (7.6.4), 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 (7.6.3). $\square$

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 7.6.2. *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 7.6.3.

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

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

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

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

Example 7.6.4. Let $A = (a_{ij})_{3\times 3}$ and let E_l $(l = 1, 2, 3)$ be defined as in the previous example. Then

$$E_1A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \\ a_{31} & a_{32} & a_{33} \end{pmatrix},$$

$$E_2A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ 4a_{31} & 4a_{32} & 4a_{33} \end{pmatrix},$$

$$E_3A = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} a_{11} + 2a_{31} & a_{12} + 2a_{32} & a_{13} + 2a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}.$$

You should now pause and marvel at the following observation: interchanging rows 1 and 2 of A produces E_1A, multiplying row 3 of A by 4 produces E_2A, and adding 2 times row 3 to row 1 of A produces E_3A.

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 7.6.5. *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 7.6.6. *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 α 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. □

Example 7.6.7. Determine the inverses of the elementary matrices E_1, E_2, and E_3 in Example 7.6.3.

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

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

□

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

Definition 7.6.8. *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:

(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 7.6.6. 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 7.6.9 (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 7.6.6 and since the product of invertible matrices is invertible by Theorem 7.1.19. □

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

Corollary 7.6.10. *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.

7.7 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 7.7.1. 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

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

□

7.8 Problems

7.9 More

7.9.1 Magic squares

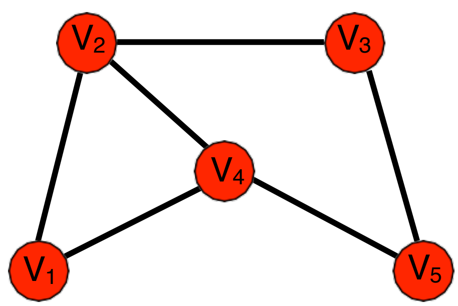
Magic squares are 3×3 matrices such that all rows, all columns and both diagonals sum to the same number. Elementary examples of magic squares are given by

$$M_1 = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}, \quad M_3 = \begin{pmatrix} -1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix}.$$

The set of magic squares is a subspace of the set of all 3×3-matrices. Indeed, it is easy to see that the sum of two magic squares is a magic square and that by multiplying a magic square by a scalar $\lambda \in \mathbb{R}$ we still get a magic square (work out the details as an exercise).

7.9.2 Matrices in graph theory

Graphs are objects consisting of a certain number of vertices, V_1, V_2, ..., V_n and links connecting these vertices. A simple example with five vertices is given below.



This is an undirected graph since links have no direction. Graphs can be related to linear algebra via the *adjacency matrix* which is defined as follows:

$$\begin{aligned} M_{ij} &= 1, \quad \text{if } V_i \text{ and } V_j \text{ are linked} \\ M_{ij} &= 0, \quad \text{otherwise.} \end{aligned}$$

The adjacency matrix for our graph is

$$M = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix} \tag{7.9.5}$$

The matrix M is symmetric due to the underlying graph being undirected. The following fact (which we will not try to prove here) makes the adjacency matrix a useful object.

Fact: The number of possible walks from vertex V_i to vertex V_j over precisely n links in a graph is given by $(M^n)_{ij}$, where M is the adjacency matrix of the graph.

To illustrate this we compute the low powers of the adjacency matrix M in (7.9.5):

$$M^2 = \begin{pmatrix} 2 & 1 & 1 & 1 & 1 \\ 1 & 3 & 0 & 1 & 2 \\ 1 & 0 & 2 & 2 & 0 \\ 1 & 1 & 2 & 3 & 0 \\ 1 & 2 & 0 & 0 & 2 \end{pmatrix}, \qquad M^3 = \begin{pmatrix} 2 & 4 & 2 & 4 & 2 \\ 4 & 2 & 5 & 6 & 1 \\ 2 & 5 & 0 & 1 & 4 \\ 4 & 6 & 1 & 2 & 5 \\ 2 & 1 & 4 & 5 & 0 \end{pmatrix}$$

For example, the number of possible walks from V_1 to V_3 over three links is given by $(M^3)_{13} = 2$. By inspecting the image of our graph it can be seen that these two walks correspond to $V_1 \to V_4 \to V_5 \to V_3$ and $V_1 \to V_4 \to V_2 \to V_3$.

7.9.3 Matrices in cryptography

Matrices can be used for encryption. Here is a basic example for how this works. Suppose we would like to encrypt the text: *linear-algebra-*. First, we translate this text into numerical form using the simple code $- \to 0, a \to 1, b \to 2, \cdots$ and then we split the resulting sequence of numbers into blocks of the same size. Here we use blocks of size three for definiteness. Next, we arrange these numbers into a matrix, with each block forming a column of the matrix. For our sample text this results in

$$T = \begin{pmatrix} 12 & 5 & 0 & 7 & 18 \\ 9 & 1 & 1 & 5 & 1 \\ 14 & 18 & 12 & 2 & 0 \end{pmatrix} \quad \text{for} \quad \begin{matrix} l & e & - & g & r \\ i & a & a & e & a \\ n & r & l & b & - \end{matrix}$$

So far, this is relatively easy to decode, even if we had decided to permute the assignment of letters to numbers. As long as same letters are represented by same numbers, the code can be deciphered by a frequency analysis, at least for a suciently long text. To do this, the relative frequency of each number is determined and compared with the typical frequency with which letters appear in the English language. Matching similar frequencies leads to the key.

For a more sophisticated encryption, define a quadratic *encoding* matrix whose size equals the length of the blocks, so a 3×3-matrix for our case. Basically, the only other restriction on this matrix is that it should be invertible. For our example, let us choose

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

To encode the text, carry out the matrix multiplication

$$T_{\text{enc}} = AT.$$

This produces the matrix

$$\begin{pmatrix} -7 & 12 & 11 & -10 & 19 \\ 10 & -8 & -12 & 12 & 36 \\ -1 & 9 & 13 & -7 & -35 \end{pmatrix}.$$

Note that in T_{enc} same letters are now represented by different numbers. For example, the letter a which appears three times, and corresponds to the three 1's in T, is represented by three different numbers in T_{enc}. Without knowledge of the encoding matrix A it is quite dicult to decypher T_{enc}, particularly for large block sizes. The legitimate receiver of the text should be provided with the inverse A^{-1} of the encoding matrix, for our example given by

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

and can then recover the message by the simply matrix multiplication

$$T = A^{-1} T_{\text{enc}}.$$

7.9.4 Mathematicians from a diverse background working in cryptography

Alan Turing

Alan Turing (1912-1954) was a British mathematician and cryptographer. Nowadays the face of the £50 note and the subject of *The Imitation Game*, his research covered many areas from logic and mathematical foundations, to probability theory, to group theory, to decoding encrypted messages, to mathematical biology and to quantum mechanics. He is best known for developing the *Bombe* machine during the second world war, where he introduced novel statistical approaches to allow the machine to crack the German Enigma codes.

Turing struggled at school, as his strengths in his work lay in his creativity. This would transfer to his research, where he would produce very unconventional (but still correct) solutions compared to the standard methods taught by teachers. Sadly, later in his life, Turing was convicted of homosexuality while it was still illegal in the UK. He was open about his sexuality afterwards, but due to an Official Secrets Act he was forbidden from talking about the depression which plagued him, which lead to him taking his own life.

In academia, Turing is most famous for his theoretical work on modern computers and artificial intelligence. In 1936 he introduced the idea of *Turing machines*, which introduced the notion of "computability" and laid the framework for modern computers decades before they were built. In fact, after the second world war he would introduce code for writing computers. In 1950, he also famously proposed the *Turing test*, which we still use today to determine whether or not a computer can have its own intelligence.

Read more on MacTutor.

Antonia Jones

Antonia Jones (1943-2010) was a British mathematician and computer scientist. Jones contracted polio as a child and loss the ability to walk at age 10, but went on to graduate from Reading University with first-class honors. Later, she earned a Ph.D. in number theory from the University of Cambridge for her work with roots of unity and Diophantine equations.

She is best known, however, for her career in computer science, where she published numerous papers on acoustic pattern recognition, game theory and artificial intelligence (among other areas). Her work lead to the creation of the nearest neighbour (or Gamma) test for neural networks, which is widely used today in nearest neighbour analysis. Jones was also able to use her background in number theory to detect security loopholes, which she used to create high security cryptography. From her work, she also launched her own firm creating random access video controllers to improve the accessibility and ease of use of early computer models.

Read more in this brief summary, this WikiMili article, or in her obituary.

7.10 Ethics in Mathematics

Over the last few decades, we have seen a sociopolitical turn in mathematics education that introduced the study of equity, social justice, identity and power into the apprenticeship of future mathematicians and mathematics teachers, thereby highlighting the impact of existing political and moral discourses and the contingent institutional contexts on both mathematics and its education. We conclude this More section with a problem for Matrices with an ethical component.

Problem

We begin by recalling the following fact.

A square matrix with entries in $\mathbb{R}$ is said to be *column-stochastic* if all of its entries are nonnegative and the entries in each column sum to one. One can prove that every column-stochastic matrix has 1 as an eigenvalue, i.e., for every column-stochastic matrix A there exist $\mathbf{v} \neq \mathbf{0}$ such that

$$A\mathbf{v} = \mathbf{v}.$$

1. The Google PageRank algorithm (simplified) works as follows: Each webpage w_i on the web is assigned a value v_i such that, if L_i is the set of pages that link to w_i, and n_i is the number of outgoing links from page w_i, then

   $$v_i = \sum_{w_j \in L_i} \frac{v_j}{n_j}$$

   That is, each page w_i "donates" $\frac{1}{n_i}$ of its value v_i, uniformly, to each page that it links to, and then the pages are "ranked" by simply ordering the v_i's. Show that, if every page on the web links to at least one other page, then there is at least one way of assigning values to each webpage that satisfies the above relation.

2. What are some effects of choosing this, or indeed any, ranking algorithm on a system as widespread as Google Search?

This question serves two purposes:
First, it shows that a relatively straightforward piece of first-year mathematics can be used to have a huge impact on the way the world functions; the Google pagerank algorithm is (or more accurately, was, as it has now been updated) one of the most influential algorithms in the world. It was often referred to as the \$100,000,000,000 algorithm (the market value of Google at the time). Not every high-impact use of mathematics requires PhD-level content.
Second, there may well be a whole suite of algorithms that could "work" in this scenario;

how does Google decide which one to use? And how does this decision introduce bias into the search results; just because a computer does it, doesn't mean it is impartial!

1. Form a square matrix A where in column i we put the value $1/n_i$ in each entry (i,j) where w_i links to w_j, and 0 in all other entries; there should be n_i such nonzero entries in each column, summing to 1. We now need to find a column vector x of "values" such that $Ax = x$. That is, if there are a total of m webpages on the internet, then we are seeking $\{v_1, \ldots, v_m\}$ such that

$$A \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} \tag{7.10.6}$$

   So we are looking for an eigenvector of A with eigenvalue 1. Now apply the previous result to show that at least one such eigenvector exists.

2. Using a human-designed/chosen algorithm on a search engine as large as Google poses a serious (ethical) question here: which algorithm should be "chosen"? How do we know what the benefits and drawbacks are of such an algorithm when deployed on a search engine that billions of people use every day? This may seem like a purely mathematical choice; it is an algorithm that *somewhat* gives reasonable search results (as perceived by the developers and managers in Google), and is implementable (reduces compute time/resources to levels that are affordable and, more pressingly, lead to a profitable business model). But Google search is used by billions of people every day, so this (hidden) mathematical choice will affect the way a vast proportion of the global population finds (or does not find) webpages. This "choice of algorithm" then becomes a massive, unchecked, proprietary, global censorship object on the most important information tool of the modern age: the internet. There may be no malicious intent here, but that is beside the point. Harm may come to society if the way Google ranks webpages overly-suppresses, or overly-promotes, various content.

Chapter 8

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.

8.1 Determinants of 2×2 and 3×3 matrices

To every 2×2 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 8.1.1. *Let $A = (a_{ij})$ be a 2×2 matrix. The* ***determinant*** *of A, denoted $\det(A)$, is defined by*

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{21}a_{12}\,. \tag{8.1.1}$$

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 8.1.2. *If A and B are 2×2 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*

$$A^{-1} = \frac{1}{\det(A)} \begin{pmatrix} d & -c \\ -b & a \end{pmatrix}. \tag{8.1.2}$$

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 (8.1.2) is well-defined, and we can calculate directly that $CA = I$. From Corollary 7.6.10 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 $C = A^{-1}$. □

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 3×3 determinants:

Definition 8.1.3. *If $A = (a_{ij})$ is a 3×3 matrix, its determinant* $\det(A)$ *is defined by*

$$\det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21}\begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31}\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \tag{8.1.3}$$
$$= a_{11}a_{22}a_{33} - a_{11}a_{32}a_{23} - a_{21}a_{12}a_{33} + a_{21}a_{32}a_{13} + a_{31}a_{12}a_{23} - a_{31}a_{22}a_{13}\,.$$

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

8.2 General definition of determinants

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

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 8.2.1. 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 1×1 matrix $A = (a_{ij})$ by $\det(A) = a_{11}$, we can re-write (8.1.1) and (8.1.3) 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 8.2.2. *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 $a_{11}, a_{21}, \dots, a_{n1}$ 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 8.2.3. 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 8.2.4. *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 8.2.5 (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 8.2.6. 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 8.2.7. 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 4×4 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 3×3 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 8.2.8. *If A is either an upper or a lower triangular matrix, then* $\det(A)$ *is the product of the diagonal entries of A.*

8.3 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 8.3.1. *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 α 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 8.3.11). □

Example 8.3.2.

(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 8.3.3. 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 8.3.4. 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 3×3 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 8.2.8. $\square$

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 2×2 case treated in Theorem 8.1.2(b)).

Theorem 8.3.5. *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 8.3.1 implies that $\det(A) = \gamma \det(U)$ for some $\gamma \neq 0$. If A is invertible, then $\det(U) = 1$ by Theorem 8.2.8, 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 8.2.8, and hence $\det(A) = \gamma \det(U) = 0$. $\square$

Definition 8.3.6. *A square matrix A is called **singular** if* $\det(A) = 0$. *Otherwise it is said to be **nonsingular**.*

Corollary 8.3.7. *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 8.3.8. *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 1×1 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 2×2 matrices (choose $k = 1$); but since we now know that it is true for 2×2 matrices, it must be true for 3×3 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 8.3.1) 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 8.3.9. *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 α 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 8.3.10. 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 8.3.1).

Theorem 8.3.11. *If A is an $n \times n$ matrix and E an elementary $n \times n$ matrix, then*

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

with

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

Proof. By induction on the size of A. The case where A is a 2×2 matrix follows from Theorem 8.1.2(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 2×2 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. $\square$

Using the previous theorem we are now able to prove the second important general result of this chapter (and a generalisation of the 2×2 case treated in Theorem 8.1.2(a)):

Theorem 8.3.12. *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 8.3.5,

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

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 8.3.13. *If A is an invertible matrix then*

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

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 8.3.12 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 8.3.5), and so we can divide both sides of the preceding equation by $\det(A)$ to obtain the required property

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

□

8.4 Cramer's rule

We have already seen how the determinant of a matrix can be used to decide if an $n \times n$ matrix A is invertible, and how to compute the inverse of a matrix. Here we introduce Cramer's Rule, which uses determinants to solve systems of linear equations $A\mathbf{x} = \mathbf{b}$ for the case of quadratic and invertible $n \times n$ matrices A. We recall that the system has a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.

To derive Cramer's rule we first derive the matrices

$$B_{(i)} := (\mathbf{A}^1, \cdots \mathbf{A}^{i-1}, \mathbf{b}, \mathbf{A}^{i+1}, \cdots, \mathbf{A}^n),$$

which are obtained from A by replacing the ith column with $\mathbf{b}$ and keeping all other columns unchanged. We also note that, in terms of the column vectors $\mathbf{A}^j$ of A the linear system $A\mathbf{x} = \mathbf{b}$ can be written as

$$\sum_j x_j \mathbf{A}^j = \mathbf{b},$$

where $\mathbf{x} = (x_1, \cdots, x_n)^T$. By applying the properties of determinants we find

$$\begin{aligned}
\det(B_i) &= \det(\mathbf{A}^1, \cdots \mathbf{A}^{i-1}, \mathbf{b}, \mathbf{A}^{i+1}, \cdots, \mathbf{A}^n) = \det(\mathbf{A}^1, \cdots \mathbf{A}^{i-1}, \sum_j x_j \mathbf{A}^j, \mathbf{A}^{i+1}, \cdots, \mathbf{A}^n) \\
&= \sum_j x_j \det(\mathbf{A}^1, \cdots \mathbf{A}^{i-1}, \mathbf{A}^j, \mathbf{A}^{i+1}, \cdots, \mathbf{A}^n) \\
&= x_i \det(\mathbf{A}^1, \cdots \mathbf{A}^{i-1}, \mathbf{A}^i, \mathbf{A}^{i+1}, \cdots, \mathbf{A}^n) = x_i \det A,
\end{aligned}$$

where in the last equality we used the fact that the determinant of a matrix with two identical columns is equal to 0. Solving for x_i we find the famous Cramer's rule:

$$x_i = \frac{\det(B_i)}{\det(A)} = \frac{\det(\mathbf{A}^1, \cdots \mathbf{A}^{i-1}, \mathbf{A}^i, \mathbf{A}^{i+1}, \cdots, \mathbf{A}^n)}{\det(A)}$$

for the solution $x = (x_1, \cdots, x_n)^T$ of the linear system $A\mathbf{x} = \mathbf{b}$, where A is an invertible $n \times n$ matrix. To solve linear systems explicitly, Cramer's rule is only useful for relatively small systems, due to the $n!$ growth of the determinant. For larger linear systems the Gauss elimination method should be used.

Example 8.4.1. Let us apply Cramer's rule to a linear system $A\mathbf{x} = \mathbf{b}$ with

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

By replacing one column of A with the vector $\mathbf{b}$, we find the three matrices

$$B_{(1)} = \begin{pmatrix} 1 & -1 & 0 \\ 2 & 2 & -2 \\ 0 & 3 & 4 \end{pmatrix}, \quad B_{(2)} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & -2 \\ 0 & 0 & 4 \end{pmatrix}, \quad B_{(3)} = \begin{pmatrix} 2 & -1 & 1 \\ 1 & 2 & 2 \\ 0 & 3 & 0 \end{pmatrix}.$$

By straightforward computations it follows that $\det(A) = 32$, $\det(B_{(1)}) = 22$, $\det(B_{(2)}) = 12$ and $\det(B_{(3)}) = -9$. This leads to

$$\mathbf{x} = \frac{1}{32} \begin{pmatrix} 22 \\ 12 \\ -9 \end{pmatrix}.$$

8.5 Problems

8.6 More

You will encounter determinants many times in your mathematical studies. For instance, a determinant of a very important matrix, called *Jacobian matrix* plays a relevant role when computing integrals and changing coordinates (see Subsection 8.6.1). Analogously matrices and determinants appear when studying systems of differential equations. See some interesting profiles of mathematicians working in differential equations at the end of this section.

8.6.1 Change of coordinates and determinants

The polar coordinates in $\mathbb{R}^2$ are defined as follows:

$$\begin{aligned} x &= r\cos\theta, \\ y &= r\sin\theta, \end{aligned}$$

where $r \geq 0$ and $\theta \in [0, 2\pi)$. We can therefore define the function

$$F : [0,\infty) \times [0, 2\pi) \to \mathbb{R}^2 : (r,\theta) \to (\rho\cos\theta, \rho\sin\theta).$$

The *Jacobian matrix* of F is given by

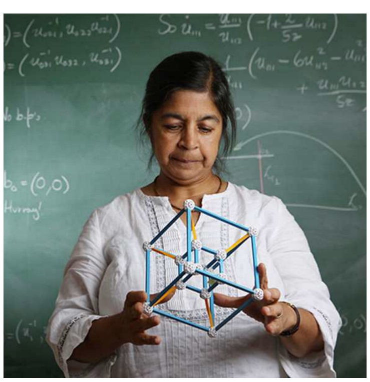
$$J_F(r,\theta) = \begin{pmatrix} \partial_r x & \partial_\theta x \\ \partial_r y & \partial_\theta y \end{pmatrix} = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}.$$

One can easily see that $\det(J_F) = r$. When computing a double integral it could be useful to pass to polar coordinates. This change of coordinates involves the determinant of the Jacobian matrix as follows:

$$\int\int_{F(A)} f(x,y)\,dx\,dy = \int\int_A f(r\cos\theta, r\sin\theta)\det(J_F)\,dr\,d\theta = \int\int_A f(r\cos\theta, r\sin\theta) r\,dr\,d\theta.$$

8.6.2 Mathematicians using linear algebra tools to study differential equations

Nalini Joshi



Nalini Joshi (1958-) is a Burmese-Australian mathematician renowned for her work on non-linear differential equations and integrable systems. Her love of maths began as a child in Myanmar where she loved counting (particularly in the Burmese number system) and games with repetitive patterns, and continued while adjusting moving to Australia at age 12 when she would often contemplate and read about big scientific questions. Joshi wanted to be an astronaut and study astronomy, but in university she found she enjoyed maths much more. *"After trying it for a while, I realized that I could take my time, try alternative beginnings, do one step after another, and get to glimpse all kinds of possibilities along the way."*

Throughout her career, Joshi has been very vocal about promoting diversity and inclusion in academia. She co-founded the Science in Australia Gender Equity initiative, and she is very vocal about challenging the various problems she and others encounter as a woman of colour.

Joshi was formerly head of the Australian Mathematical Society, and the first female chair of applied mathematics at the University of Sydney. Her work focuses on *Painlevé* and *soliton equations*, which are differential equations describing integral systems. She employs tools from analysis, algebra, geometry and the many other interweaving areas of maths to solve these equations, and is renowned for visualising problems in different and unique ways.

"Instead of describing solutions as functions of an independent variable like time, they can be tracked by curves that go through initial values. The first perspective is like pointing a telescope to one point in the sky at night and taking pictures while time is changing. The second perspective is like tracking one star as it follows circular arcs of light in the sky at night."

See more in her university profile or these interviews from 2012 and 2017.

Luis Caffarelli

Luis Caffarelli (1948-) is an Argentine mathematician, regarded as one of the world's leading experts in free boundary problems and nonlinear partial differential equations. In 2023 he was awarded the Abel Prize *"for his seminal contributions to regularity theory for nonlinear partial differential equations including free-boundary problems and the Monge–Ampère equation"*. He is the first South American winner of the award.

Caffarelli earned his Master of Science (1968) and Ph.D. (1972) in mathematics at the University of Buenos Aires. He and his thesis advisor, Calixto Pedro Calderón, went on to write two joint papers: Weak type estimates for the Hardy-Littlewood maximal functions (1974); and On Abel summability of multiple Jacobi series (1974). His proudest achievement, however, was in proving a partial result of the regularity of the Navier-Stokes alongside Louis Nirenberg and Robert Kohn in 1982, which is to this day still the best known result of the (now) Millennium Problem. He currently holds the Sid Richardson Chair at the University of Texas at Austin. He also has been a professor at the University of Minnesota, the University of Chicago, and the Courant Institute of Mathematical Sciences at New York University. From 1986 to 1996 he was a professor at the Institute for Advanced Study in Princeton.

He has received numerous other awards and honors for his work, including the Bocher Memorial Prize, Rolf Schock Prize of the Royal Swedish Academy of Sciences, Leroy P. Steel Prize, the Wolf Prize, and Shaw Prize. He is a member of the National Academy of Sciences, Academy of Medicine, Engineering and Science of Texas, American Mathematical Society, Association for Women in Mathematics, and Society for Industrial and Applied Mathematics.

Read more on MacTutor, this Nature article or on Wikipedia.