ELEMENTARY

LINEAR ALGEBRA

K. R. MATTHEWS

DEPARTMENT OF MATHEMATICS

UNIVERSITY OF QUEENSLAND

Corrected Version, 27th April 2013

Comments to the author at keithmatt@gmail.com

Chapter 1

LINEAR EQUATIONS

1.1 Introduction to linear equations

A *linear equation* in n unknowns $x_1, x_2, \cdots, x_n$ is an equation of the form

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

where $a_1, a_2, \ldots, a_n, b$ are given real numbers.

For example, with x and y instead of x_1 and x_2, the linear equation $2x + 3y = 6$ describes the line passing through the points (3, 0) and (0, 2).

Similarly, with x, y and z instead of x_1, x_2 and x_3, the linear equation $2x + 3y + 4z = 12$ describes the plane passing through the points (6, 0, 0), (0, 4, 0), (0, 0, 3).

A *system* of m linear equations in n unknowns $x_1, x_2, \cdots, x_n$ is a family of linear equations

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

We wish to determine if such a system has a solution, that is to find out if there exist numbers $x_1, x_2, \cdots, x_n$ which satisfy each of the equations simultaneously. We say that the system is *consistent* if it has a solution. Otherwise the system is called *inconsistent*.

Note that the above system can be written concisely as

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad i = 1, 2, \cdots, m.$$

The matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

is called the *coefficient matrix* of the system, while the matrix

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & & & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix}$$

is called the *augmented matrix* of the system.

Geometrically, solving a system of linear equations in two (or three) unknowns is equivalent to determining whether or not a family of lines (or planes) has a common point of intersection.

EXAMPLE 1.1.1 Solve the equation

$$2x + 3y = 6.$$

Solution. The equation $2x + 3y = 6$ is equivalent to $2x = 6 - 3y$ or $x = 3 - \frac{3}{2}y$, where y is arbitrary. So there are infinitely many solutions.

EXAMPLE 1.1.2 Solve the system

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

Solution. We subtract the second equation from the first, to get $2y = 1$ and $y = \frac{1}{2}$. Then $x = y - z = \frac{1}{2} - z$, where z is arbitrary. Again there are infinitely many solutions.

EXAMPLE 1.1.3 Find a polynomial of the form $y = a_0 + a_1x + a_2x^2 + a_3x^3$ which passes through the points $(-3, -2)$, $(-1, 2)$, $(1, 5)$, $(2, 1)$.

Solution. When x has the values -3, -1, 1, 2, then y takes corresponding values -2, 2, 5, 1 and we get four equations in the unknowns a_0, a_1, a_2, a_3:

$$\begin{aligned} a_0 - 3a_1 + 9a_2 - 27a_3 &= -2 \\ a_0 - a_1 + a_2 - a_3 &= 2 \\ a_0 + a_1 + a_2 + a_3 &= 5 \\ a_0 + 2a_1 + 4a_2 + 8a_3 &= 1, \end{aligned}$$

with unique solution $a_0 = 93/20, a_1 = 221/120, a_2 = -23/20, a_3 = -41/120$. So the required polynomial is

$$y = \frac{93}{20} + \frac{221}{120}x - \frac{23}{20}x^2 - \frac{41}{120}x^3.$$

In [26, pages 33–35] there are examples of systems of linear equations which arise from simple electrical networks using Kirchhoff's laws for electrical circuits.

Solving a system consisting of a single linear equation is easy. However if we are dealing with two or more equations, it is desirable to have a systematic method of determining if the system is consistent and to find all solutions.

Instead of restricting ourselves to linear equations with rational or real coefficients, our theory goes over to the more general case where the coefficients belong to an arbitrary *field*. A *field* F is a set F which possesses operations of *addition* and *multiplication* which satisfy the familiar rules of rational arithmetic. There are ten basic properties that a field must have:

THE FIELD AXIOMS.

1. $(a + b) + c = a + (b + c)$ for all a, b, c in F;
2. $(ab)c = a(bc)$ for all a, b, c in F;
3. $a + b = b + a$ for all a, b in F;
4. $ab = ba$ for all a, b in F;
5. there exists an element 0 in F such that $0 + a = a$ for all a in F;
6. there exists an element 1 in F such that $1a = a$ for all a in F;
7. to every a in F, there corresponds an *additive inverse* $-a$ in F, satisfying
$$a + (-a) = 0;$$

8. to every non–zero a in F, there corresponds a *multiplicative inverse* a^{-1} in F, satisfying
$$aa^{-1} = 1;$$

9. $a(b + c) = ab + ac$ for all a, b, c in F;

10. $0 \neq 1$.

With standard definitions such as $a - b = a + (-b)$ and $\frac{a}{b} = ab^{-1}$ for $b \neq 0$, we have the following familiar rules:

$$\begin{aligned}
-(a+b) &= (-a) + (-b), \quad (ab)^{-1} = a^{-1}b^{-1}; \\
-(-a) &= a, \quad (a^{-1})^{-1} = a; \\
-(a-b) &= b - a, \quad (\frac{a}{b})^{-1} = \frac{b}{a}; \\
\frac{a}{b} + \frac{c}{d} &= \frac{ad+bc}{bd}; \\
\frac{a}{b}\frac{c}{d} &= \frac{ac}{bd}; \\
\frac{ab}{ac} &= \frac{b}{c}, \quad \frac{a}{\left(\frac{b}{c}\right)} = \frac{ac}{b}; \\
-(ab) &= (-a)b = a(-b); \\
-\left(\frac{a}{b}\right) &= \frac{-a}{b} = \frac{a}{-b}; \\
0a &= 0; \\
(-a)^{-1} &= -(a^{-1}).
\end{aligned}$$

Fields which have only finitely many elements are of great interest in many parts of mathematics and its applications, for example to coding theory. It is easy to construct fields containing exactly p elements, where p is a prime number. First we must explain the idea of *modular addition* and *modular multiplication*. If a is an integer, we define $a \pmod{p}$ to be the *least remainder on dividing a by p*: That is, if $a = bp + r$, where b and r are integers and $0 \leq r < p$, then $a \pmod{p} = r$.

For example, $-1 \pmod{2} = 1,\ 3 \pmod{3} = 0,\ 5 \pmod{3} = 2$.

Then addition and multiplication mod p are defined by

$$\begin{aligned}
a \oplus b &= (a + b) \,(\mathrm{mod}\, p) \\
a \otimes b &= (ab) \,(\mathrm{mod}\, p).
\end{aligned}$$

For example, with $p = 7$, we have $3 \oplus 4 = 7 \pmod 7 = 0$ and $3 \otimes 5 = 15 \pmod 7 = 1$. Here are the complete addition and multiplication tables mod 7:

$\oplus$	0	1	2	3	4	5	6
0	0	1	2	3	4	5	6
1	1	2	3	4	5	6	0
2	2	3	4	5	6	0	1
3	3	4	5	6	0	1	2
4	4	5	6	0	1	2	3
5	5	6	0	1	2	3	4
6	6	0	1	2	3	4	5

$\otimes$	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6
2	0	2	4	6	1	3	5
3	0	3	6	2	5	1	4
4	0	4	1	5	2	6	3
5	0	5	3	1	6	4	2
6	0	6	5	4	3	2	1

If we now let $\mathbb{Z}_p = \{0, 1, \ldots, p-1\}$, then it can be proved that $\mathbb{Z}_p$ forms a field under the operations of modular addition and multiplication mod p. For example, the additive inverse of 3 in $\mathbb{Z}_7$ is 4, so we write $-3 = 4$ when calculating in $\mathbb{Z}_7$. Also the multiplicative inverse of 3 in $\mathbb{Z}_7$ is 5 , so we write $3^{-1} = 5$ when calculating in $\mathbb{Z}_7$.

In practice, we write $a \oplus b$ and $a \otimes b$ as $a+b$ and ab or $a \times b$ when dealing with linear equations over $\mathbb{Z}_p$.

The simplest field is $\mathbb{Z}_2$, which consists of two elements 0, 1 with addition satisfying $1+1=0$. So in $\mathbb{Z}_2$, $-1 = 1$ and the arithmetic involved in solving equations over $\mathbb{Z}_2$ is very simple.

EXAMPLE 1.1.4 Solve the following system over $\mathbb{Z}_2$:

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

Solution. We add the first equation to the second to get $y = 1$. Then $x = 1 - z = 1 + z$, with z arbitrary. Hence the solutions are $(x, y, z) = (1, 1, 0)$ and $(0, 1, 1)$.

We use $\mathbb{Q}$ and $\mathbb{R}$ to denote the fields of rational and real numbers, respectively. Unless otherwise stated, the field used will be $\mathbb{Q}$.

1.2 Solving linear equations

We show how to solve any system of linear equations over an arbitrary field, using the *GAUSS–JORDAN* algorithm. We first need to define some terms.

DEFINITION 1.2.1 (Row–echelon form) A matrix is in *row–echelon form* if

(i) all zero rows (if any) are at the bottom of the matrix and

(ii) if two successive rows are non–zero, the second row starts with more zeros than the first (moving from left to right).

For example, the matrix

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

is in row–echelon form, whereas the matrix

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

is not in row–echelon form.

The *zero* matrix of any size is always in row–echelon form.

DEFINITION 1.2.2 (Reduced row–echelon form) A matrix is in *reduced row–echelon form* if

1. it is in row–echelon form,
2. the leading (leftmost non–zero) entry in each non–zero row is 1,
3. all other elements of the column in which the leading entry 1 occurs are zeros.

For example the matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 & 2 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

are in reduced row–echelon form, whereas the matrices

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

are not in reduced row–echelon form, but are in row–echelon form.

The *zero* matrix of any size is always in reduced row–echelon form.

Notation. If a matrix is in reduced row–echelon form, it is useful to denote the column numbers in which the leading entries 1 occur, by $c_1, c_2, \ldots, c_r$, with the remaining column numbers being denoted by $c_{r+1}, \ldots, c_n$, where r is the number of non–zero rows. For example, in the 4×6 matrix above, we have $r = 3$, $c_1 = 2$, $c_2 = 4$, $c_3 = 5$, $c_4 = 1$, $c_5 = 3$, $c_6 = 6$.

The following operations are the ones used on systems of linear equations and do not change the solutions.

DEFINITION 1.2.3 (Elementary row operations) Three types of *elementary row operations* can be performed on matrices:

1. Interchanging two rows:

   $R_i \leftrightarrow R_j$ interchanges rows i and j.

2. Multiplying a row by a non–zero scalar:

   $R_i \rightarrow tR_i$ multiplies row i by the non–zero scalar t.

3. Adding a multiple of one row to another row:

   $R_j \rightarrow R_j + tR_i$ adds t times row i to row j.

DEFINITION 1.2.4 (Row equivalence) Matrix A is *row–equivalent* to matrix B if B is obtained from A by a sequence of elementary row operations.

EXAMPLE 1.2.1 Working from left to right,

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 1 \\ 1 & -1 & 2 \end{bmatrix} \quad R_2 \rightarrow R_2 + 2R_3 \quad \begin{bmatrix} 1 & 2 & 0 \\ 4 & -1 & 5 \\ 1 & -1 & 2 \end{bmatrix}$$

$$R_2 \leftrightarrow R_3 \quad \begin{bmatrix} 1 & 2 & 0 \\ 1 & -1 & 2 \\ 4 & -1 & 5 \end{bmatrix} \quad R_1 \rightarrow 2R_1 \quad \begin{bmatrix} 2 & 4 & 0 \\ 1 & -1 & 2 \\ 4 & -1 & 5 \end{bmatrix} = B.$$

Thus A is row–equivalent to B. Clearly B is also row–equivalent to A, by performing the inverse row–operations $R_1 \rightarrow \frac{1}{2}R_1$, $R_2 \leftrightarrow R_3$, $R_2 \rightarrow R_2 - 2R_3$ on B.

It is not difficult to prove that if A and B are row–equivalent augmented matrices of two systems of linear equations, then the two systems have the same solution sets – a solution of the one system is a solution of the other. For example the systems whose augmented matrices are A and B in the above example are respectively

$$\left\{\begin{array}{rcl} x+2y &=& 0\\ 2x+y &=& 1\\ x-y &=& 2 \end{array}\right. \quad \text{and} \quad \left\{\begin{array}{rcl} 2x+4y &=& 0\\ x-y &=& 2\\ 4x-y &=& 5 \end{array}\right.$$

and these systems have precisely the same solutions.

1.3 The Gauss–Jordan algorithm

We now describe the *GAUSS–JORDAN ALGORITHM.* This is a process which starts with a given matrix A and produces a matrix B in reduced row–echelon form, which is row–equivalent to A. If A is the augmented matrix of a system of linear equations, then B will be a much simpler matrix than A from which the consistency or inconsistency of the corresponding system is immediately apparent and in fact the complete solution of the system can be read off.

<u>STEP 1</u>.

Find the first non–zero column moving from left to right, (column c_1) and select a non–zero entry from this column. By interchanging rows, if necessary, ensure that the first entry in this column is non–zero. Multiply row 1 by the multiplicative inverse of a_{1c_1} thereby converting a_{1c_1} to 1. For each non–zero element a_{ic_1}, $i > 1$, (if any) in column c_1, add $-a_{ic_1}$ times row 1 to row i, thereby ensuring that all elements in column c_1, apart from the first, are zero.

<u>STEP 2</u>. If the matrix obtained at Step 1 has its 2nd, $\ldots, m$th rows all zero, the matrix is in reduced row–echelon form. Otherwise suppose that the first column which has a non–zero element in the rows below the first is column c_2. Then $c_1 < c_2$. By interchanging rows below the first, if necessary, ensure that a_{2c_2} is non–zero. Then convert a_{2c_2} to 1 and by adding suitable multiples of row 2 to the remaing rows, where necessary, ensure that all remaining elements in column c_2 are zero.

The process is repeated and will eventually stop after r steps, either because we run out of rows, or because we run out of non–zero columns. In general, the final matrix will be in reduced row–echelon form and will have r non–zero rows, with leading entries 1 in columns $c_1, \ldots, c_r$, respectively.

EXAMPLE 1.3.1

$$\begin{bmatrix} 0 & 0 & 4 & 0 \\ 2 & 2 & -2 & 5 \\ 5 & 5 & -1 & 5 \end{bmatrix} \quad R_1 \leftrightarrow R_2 \quad \begin{bmatrix} 2 & 2 & -2 & 5 \\ 0 & 0 & 4 & 0 \\ 5 & 5 & -1 & 5 \end{bmatrix}$$

$$R_1 \to \tfrac{1}{2}R_1 \quad \begin{bmatrix} 1 & 1 & -1 & \frac{5}{2} \\ 0 & 0 & 4 & 0 \\ 5 & 5 & -1 & 5 \end{bmatrix} \quad R_3 \to R_3 - 5R_1 \quad \begin{bmatrix} 1 & 1 & -1 & \frac{5}{2} \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 4 & -\frac{15}{2} \end{bmatrix}$$

$$R_2 \to \tfrac{1}{4}R_2 \quad \begin{bmatrix} 1 & 1 & -1 & \frac{5}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 4 & -\frac{15}{2} \end{bmatrix} \quad \left\{ \begin{array}{l} R_1 \to R_1 + R_2 \\ R_3 \to R_3 - 4R_2 \end{array} \right. \quad \begin{bmatrix} 1 & 1 & 0 & \frac{5}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -\frac{15}{2} \end{bmatrix}$$

$$R_3 \to \tfrac{-2}{15}R_3 \quad \begin{bmatrix} 1 & 1 & 0 & \frac{5}{2} \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad R_1 \to R_1 - \tfrac{5}{2}R_3 \quad \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The last matrix is in reduced row–echelon form.

REMARK 1.3.1 It is possible to show that a given matrix over an arbitrary field is row–equivalent to *precisely one* matrix which is in reduced row–echelon form.

A flow–chart for the Gauss–Jordan algorithm, based on [1, page 83] is presented in figure 1.1 below.

1.4 Systematic solution of linear systems.

Suppose a system of m linear equations in n unknowns $x_1, \cdots, x_n$ has augmented matrix A and that A is row–equivalent to a matrix B which is in reduced row–echelon form, via the Gauss–Jordan algorithm. Then A and B are $m \times (n+1)$. Suppose that B has r non–zero rows and that the leading entry 1 in row i occurs in column number c_i, for $1 \le i \le r$. Then

$$1 \le c_1 < c_2 < \cdots, < c_r \le n+1.$$

Also assume that the remaining column numbers are $c_{r+1}, \cdots, c_{n+1}$, where

$$1 \le c_{r+1} < c_{r+2} < \cdots < c_n \le n+1.$$

Case 1: $c_r = n+1$. The system is inconsistent. For the last non–zero row of B is $[0, 0, \cdots, 1]$ and the corresponding equation is

$$0x_1 + 0x_2 + \cdots + 0x_n = 1,$$

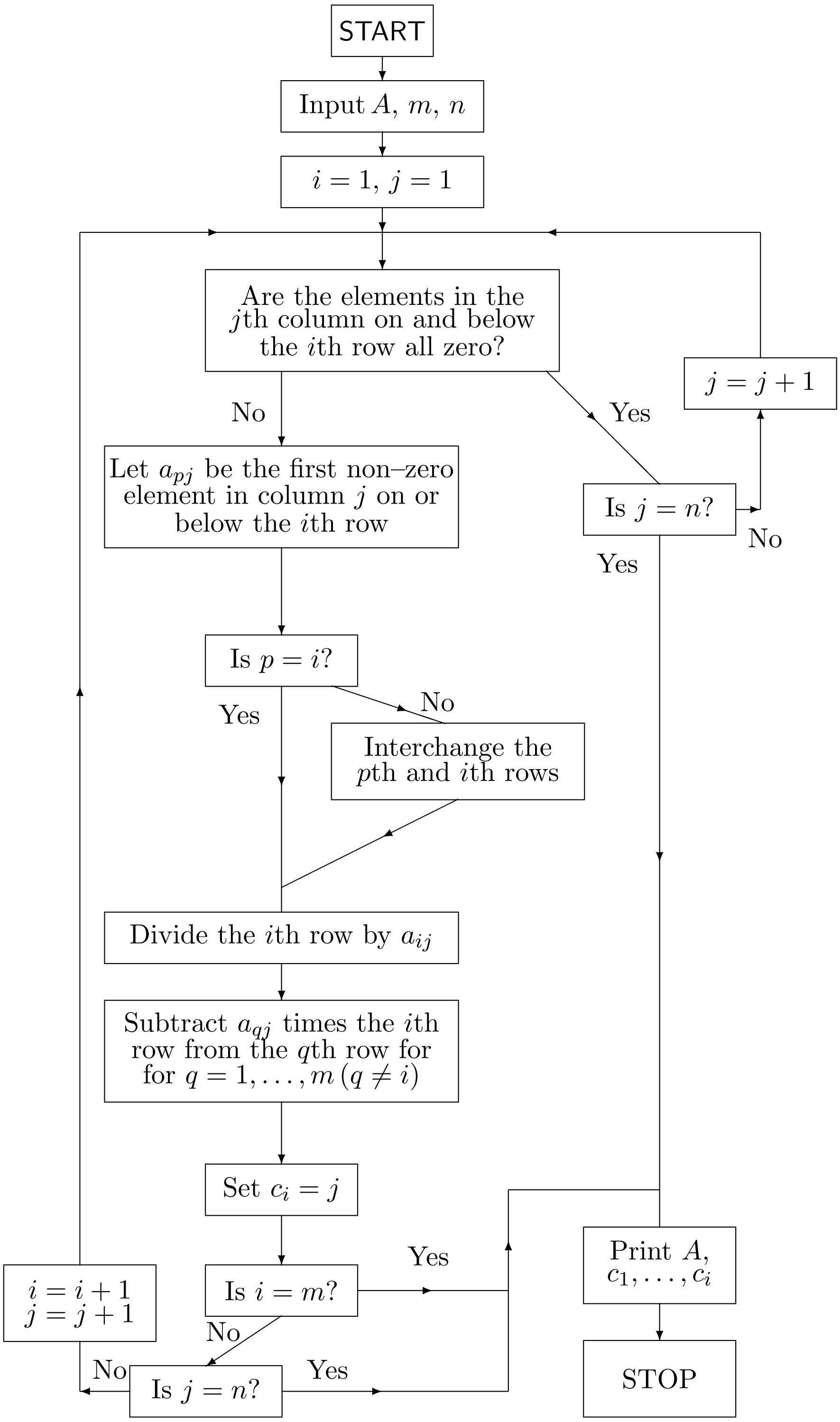

Figure 1.1: Gauss–Jordan algorithm.

which has no solutions. Consequently the original system has no solutions.

Case 2: $c_r \leq n$. The system of equations corresponding to the non–zero rows of B is consistent. First notice that $r \leq n$ here.

If $r = n$, then $c_1 = 1$, $c_2 = 2$, $\cdots$, $c_n = n$ and

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 & d_1 \\ 0 & 1 & \cdots & 0 & d_2 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 1 & d_n \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \cdots & 0 & 0 \end{bmatrix}.$$

There is a unique solution $x_1 = d_1$, $x_2 = d_2$, $\cdots$, $x_n = d_n$.

If $r < n$, there will be more than one solution (infinitely many if the field is infinite). For all solutions are obtained by taking the unknowns $x_{c_1}, \cdots, x_{c_r}$ as *dependent* unknowns and using the r equations corresponding to the non–zero rows of B to express these unknowns in terms of the remaining *independent* unknowns $x_{c_{r+1}}, \ldots, x_{c_n}$, which can take on arbitrary values:

$$\begin{aligned} x_{c_1} &= b_{1\,n+1} - b_{1c_{r+1}}x_{c_{r+1}} - \cdots - b_{1c_n}x_{c_n} \\ &\vdots \\ x_{c_r} &= b_{r\,n+1} - b_{rc_{r+1}}x_{c_{r+1}} - \cdots - b_{rc_n}x_{c_n}. \end{aligned}$$

In particular, taking $x_{c_{r+1}} = 0, \ldots, x_{c_{n-1}} = 0$ and $x_{c_n} = 0, 1$ respectively, produces at least two solutions.

EXAMPLE 1.4.1 Solve the system

$$\begin{aligned} x + y &= 0 \\ x - y &= 1 \\ 4x + 2y &= 1. \end{aligned}$$

Solution. The augmented matrix of the system is

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 1 \\ 4 & 2 & 1 \end{bmatrix}$$

which is row equivalent to

$$B = \begin{bmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & -\frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}.$$

We read off the unique solution $x = \frac{1}{2}$, $y = -\frac{1}{2}$.
(Here $n = 2$, $r = 2$, $c_1 = 1$, $c_2 = 2$. Also $c_r = c_2 = 2 < 3 = n + 1$ and $r = n$.)

EXAMPLE 1.4.2 Solve the system

$$\begin{aligned} 2x_1 + 2x_2 - 2x_3 &= 5 \\ 7x_1 + 7x_2 + x_3 &= 10 \\ 5x_1 + 5x_2 - x_3 &= 5. \end{aligned}$$

Solution. The augmented matrix is

$$A = \begin{bmatrix} 2 & 2 & -2 & 5 \\ 7 & 7 & 1 & 10 \\ 5 & 5 & -1 & 5 \end{bmatrix}$$

which is row equivalent to

$$B = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

We read off inconsistency for the original system.
(Here $n = 3$, $r = 3$, $c_1 = 1$, $c_2 = 3$. Also $c_r = c_3 = 4 = n + 1$.)

EXAMPLE 1.4.3 Solve the system

$$\begin{aligned} x_1 - x_2 + x_3 &= 1 \\ x_1 + x_2 - x_3 &= 2. \end{aligned}$$

Solution. The augmented matrix is

$$A = \begin{bmatrix} 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 2 \end{bmatrix}$$

which is row equivalent to

$$B = \begin{bmatrix} 1 & 0 & 0 & \frac{3}{2} \\ 0 & 1 & -1 & \frac{1}{2} \end{bmatrix}.$$

The complete solution is $x_1 = \frac{3}{2}$, $x_2 = \frac{1}{2} + x_3$, with x_3 arbitrary.
(Here $n = 3$, $r = 2$, $c_1 = 1$, $c_2 = 2$. Also $c_r = c_2 = 2 < 4 = n + 1$ and $r < n$.)

EXAMPLE 1.4.4 Solve the system

$$\begin{aligned} 6x_3 + 2x_4 - 4x_5 - 8x_6 &= 8 \\ 3x_3 + x_4 - 2x_5 - 4x_6 &= 4 \\ 2x_1 - 3x_2 + x_3 + 4x_4 - 7x_5 + x_6 &= 2 \\ 6x_1 - 9x_2 + 11x_4 - 19x_5 + 3x_6 &= 1. \end{aligned}$$

Solution. The augmented matrix is

$$A = \begin{bmatrix} 0 & 0 & 6 & 2 & -4 & -8 & 8 \\ 0 & 0 & 3 & 1 & -2 & -4 & 4 \\ 2 & -3 & 1 & 4 & -7 & 1 & 2 \\ 6 & -9 & 0 & 11 & -19 & 3 & 1 \end{bmatrix}$$

which is row equivalent to

$$B = \begin{bmatrix} 1 & -\frac{3}{2} & 0 & \frac{11}{6} & -\frac{19}{6} & 0 & \frac{1}{24} \\ 0 & 0 & 1 & \frac{1}{3} & -\frac{2}{3} & 0 & \frac{5}{3} \\ 0 & 0 & 0 & 0 & 0 & 1 & \frac{1}{4} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The complete solution is

$$\begin{aligned} x_1 &= \tfrac{1}{24} + \tfrac{3}{2}x_2 - \tfrac{11}{6}x_4 + \tfrac{19}{6}x_5, \\ x_3 &= \tfrac{5}{3} - \tfrac{1}{3}x_4 + \tfrac{2}{3}x_5, \\ x_6 &= \tfrac{1}{4}, \end{aligned}$$

with x_2, x_4, x_5 arbitrary.
(Here $n = 6$, $r = 3$, $c_1 = 1$, $c_2 = 3$, $c_3 = 6$; $c_r = c_3 = 6 < 7 = n + 1$; $r < n$.)

EXAMPLE 1.4.5 Find the rational number t for which the following system is consistent and solve the system for this value of t.

$$\begin{aligned} x+y &= 2 \\ x-y &= 0 \\ 3x-y &= t. \end{aligned}$$

Solution. The augmented matrix of the system is

$$A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & -1 & 0 \\ 3 & -1 & t \end{bmatrix}$$

which is row–equivalent to the simpler matrix

$$B = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & t-2 \end{bmatrix}.$$

Hence if $t \neq 2$ the system is inconsistent. If $t = 2$ the system is consistent and

$$B = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

We read off the solution $x = 1$, $y = 1$.

EXAMPLE 1.4.6 For which rationals a and b does the following system have (i) no solution, (ii) a unique solution, (iii) infinitely many solutions?

$$\begin{aligned} x-2y+3z &= 4 \\ 2x-3y+az &= 5 \\ 3x-4y+5z &= b. \end{aligned}$$

Solution. The augmented matrix of the system is

$$A = \begin{bmatrix} 1 & -2 & 3 & 4 \\ 2 & -3 & a & 5 \\ 3 & -4 & 5 & b \end{bmatrix}$$

$$\left\{\begin{array}{l} R_2 \to R_2 - 2R_1 \\ R_3 \to R_3 - 3R_1 \end{array}\right. \quad \left[\begin{array}{rrrr} 1 & -2 & 3 & 4 \\ 0 & 1 & a-6 & -3 \\ 0 & 2 & -4 & b-12 \end{array}\right]$$

$$R_3 \to R_3 - 2R_2 \quad \left[\begin{array}{rrrr} 1 & -2 & 3 & 4 \\ 0 & 1 & a-6 & -3 \\ 0 & 0 & -2a+8 & b-6 \end{array}\right] = B.$$

Case 1. $a \neq 4$. Then $-2a+8 \neq 0$ and we see that B can be reduced to a matrix of the form

$$\left[\begin{array}{rrrc} 1 & 0 & 0 & u \\ 0 & 1 & 0 & v \\ 0 & 0 & 1 & \frac{b-6}{-2a+8} \end{array}\right]$$

and we have the unique solution $x = u$, $y = v$, $z = (b-6)/(-2a+8)$.

Case 2. $a = 4$. Then

$$B = \left[\begin{array}{rrrr} 1 & -2 & 3 & 4 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & b-6 \end{array}\right].$$

If $b \neq 6$ we get no solution, whereas if $b = 6$ then

$B = \left[\begin{array}{rrrr} 1 & -2 & 3 & 4 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right] \quad R_1 \to R_1 + 2R_2 \quad \left[\begin{array}{rrrr} 1 & 0 & -1 & -2 \\ 0 & 1 & -2 & -3 \\ 0 & 0 & 0 & 0 \end{array}\right]$. We read off the complete solution $x = -2 + z$, $y = -3 + 2z$, with z arbitrary.

EXAMPLE 1.4.7 Find the reduced row–echelon form of the following matrix over $\mathbb{Z}_3$:

$$\left[\begin{array}{rrrr} 2 & 1 & 2 & 1 \\ 2 & 2 & 1 & 0 \end{array}\right].$$

Hence solve the system

$$\begin{aligned} 2x + y + 2z &= 1 \\ 2x + 2y + z &= 0 \end{aligned}$$

over $\mathbb{Z}_3$.

Solution.

$$\begin{bmatrix} 2 & 1 & 2 & 1 \\ 2 & 2 & 1 & 0 \end{bmatrix} \quad R_2 \to R_2 - R_1 \quad \begin{bmatrix} 2 & 1 & 2 & 1 \\ 0 & 1 & -1 & -1 \end{bmatrix} = \begin{bmatrix} 2 & 1 & 2 & 1 \\ 0 & 1 & 2 & 2 \end{bmatrix}$$

$$R_1 \to 2R_1 \quad \begin{bmatrix} 1 & 2 & 1 & 2 \\ 0 & 1 & 2 & 2 \end{bmatrix} \quad R_1 \to R_1 + R_2 \quad \begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 2 & 2 \end{bmatrix}.$$

The last matrix is in reduced row–echelon form.

To solve the system of equations whose augmented matrix is the given matrix over $\mathbb{Z}_3$, we see from the reduced row–echelon form that $x = 1$ and $y = 2 - 2z = 2 + z$, where $z = 0, 1, 2$. Hence there are three solutions to the given system of linear equations: $(x, y, z) = (1, 2, 0)$, $(1, 0, 1)$ and $(1, 1, 2)$.

1.5 Homogeneous systems

A system of homogeneous linear equations is a system of the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0. \end{aligned}$$

Such a system is always consistent as $x_1 = 0, \cdots, x_n = 0$ is a solution. This solution is called the *trivial* solution. Any other solution is called a *non–trivial* solution.

For example the homogeneous system

$$\begin{aligned} x - y &= 0 \\ x + y &= 0 \end{aligned}$$

has only the trivial solution, whereas the homogeneous system

$$\begin{aligned} x - y + z &= 0 \\ x + y + z &= 0 \end{aligned}$$

has the complete solution $x = -z$, $y = 0$, z arbitrary. In particular, taking $z = 1$ gives the non–trivial solution $x = -1$, $y = 0$, $z = 1$.

There is simple but fundamental theorem concerning homogeneous systems.

THEOREM 1.5.1 *A homogeneous system of m linear equations in n unknowns always has a non–trivial solution if $m < n$.*

Proof. Suppose that $m < n$ and that the coefficient matrix of the system is row–equivalent to B, a matrix in reduced row–echelon form. Let r be the number of non–zero rows in B. Then $r \leq m < n$ and hence $n - r > 0$ and so the number $n - r$ of arbitrary unknowns is in fact positive. Taking one of these unknowns to be 1 gives a non–trivial solution.

REMARK 1.5.1 Let two systems of homogeneous equations in n unknowns have coefficient matrices A and B, respectively. If each row of B is a linear combination of the rows of A (i.e. a sum of multiples of the rows of A) and each row of A is a linear combination of the rows of B, then it is easy to prove that the two systems have identical solutions. The converse is true, but is not easy to prove. Similarly if A and B have the same reduced row–echelon form, apart from possibly zero rows, then the two systems have identical solutions and conversely.

There is a similar situation in the case of two systems of linear equations (not necessarily homogeneous), with the proviso that in the statement of the converse, the extra condition that both the systems are consistent, is needed.

1.6 PROBLEMS

1. Which of the following matrices of rationals is in reduced row–echelon form?

$$\text{(a)} \begin{bmatrix} 1 & 0 & 0 & 0 & -3 \\ 0 & 0 & 1 & 0 & 4 \\ 0 & 0 & 0 & 1 & 2 \end{bmatrix} \text{(b)} \begin{bmatrix} 0 & 1 & 0 & 0 & 5 \\ 0 & 0 & 1 & 0 & -4 \\ 0 & 0 & 0 & -1 & 3 \end{bmatrix} \text{(c)} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -2 \end{bmatrix}$$

$$\text{(d)} \begin{bmatrix} 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{(e)} \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \text{(f)} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{(g)} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \text{ [Answers: (a), (e), (g)]}$$

2. Find reduced row–echelon forms which are row–equivalent to the following matrices:

$$\text{(a)} \begin{bmatrix} 0 & 0 & 0 \\ 2 & 4 & 0 \end{bmatrix} \quad \text{(b)} \begin{bmatrix} 0 & 1 & 3 \\ 1 & 2 & 4 \end{bmatrix} \quad \text{(c)} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{(d)} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 0 & 0 \\ -4 & 0 & 0 \end{bmatrix}.$$

[Answers:

(a) $\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ (b) $\begin{bmatrix} 1 & 0 & -2 \\ 0 & 1 & 3 \end{bmatrix}$ (c) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (d) $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$.]

3. Solve the following systems of linear equations by reducing the augmented matrix to reduced row–echelon form:

$$(a) \quad \begin{aligned} x + y + z &= 2 \\ 2x + 3y - z &= 8 \\ x - y - z &= -8 \end{aligned} \qquad (b) \quad \begin{aligned} x_1 + x_2 - x_3 + 2x_4 &= 10 \\ 3x_1 - x_2 + 7x_3 + 4x_4 &= 1 \\ -5x_1 + 3x_2 - 15x_3 - 6x_4 &= 9 \end{aligned}$$

$$(c) \quad \begin{aligned} 3x - y + 7z &= 0 \\ 2x - y + 4z &= \tfrac{1}{2} \\ x - y + z &= 1 \\ 6x - 4y + 10z &= 3 \end{aligned} \qquad (d) \quad \begin{aligned} 2x_2 + 3x_3 - 4x_4 &= 1 \\ 2x_3 + 3x_4 &= 4 \\ 2x_1 + 2x_2 - 5x_3 + 2x_4 &= 4 \\ 2x_1 - 6x_3 + 9x_4 &= 7 \end{aligned}$$

[Answers: (a) $x = -3$, $y = \frac{19}{4}$, $z = \frac{1}{4}$; (b) inconsistent;

(c) $x = -\frac{1}{2} - 3z$, $y = -\frac{3}{2} - 2z$, with z arbitrary;

(d) $x_1 = \frac{19}{2} - 9x_4$, $x_2 = -\frac{5}{2} + \frac{17}{4}x_4$, $x_3 = 2 - \frac{3}{2}x_4$, with x_4 arbitrary.]

4. Show that the following system is consistent if and only if $c = 2a - 3b$ and solve the system in this case.

$$\begin{aligned} 2x - y + 3z &= a \\ 3x + y - 5z &= b \\ -5x - 5y + 21z &= c. \end{aligned}$$

[Answer: $x = \frac{a+b}{5} + \frac{2}{5}z$, $y = \frac{-3a+2b}{5} + \frac{19}{5}z$, with z arbitrary.]

5. Find the value of t for which the following system is consistent and solve the system for this value of t.

$$\begin{aligned} x + y &= 1 \\ tx + y &= t \\ (1 + t)x + 2y &= 3. \end{aligned}$$

[Answer: $t = 2$; $x = 1$, $y = 0$.]

6. Solve the homogeneous system

$$\begin{aligned} -3x_1 + x_2 + x_3 + x_4 &= 0 \\ x_1 - 3x_2 + x_3 + x_4 &= 0 \\ x_1 + x_2 - 3x_3 + x_4 &= 0 \\ x_1 + x_2 + x_3 - 3x_4 &= 0. \end{aligned}$$

[Answer: $x_1 = x_2 = x_3 = x_4$, with x_4 arbitrary.]

7. For which rational numbers λ does the homogeneous system

$$\begin{aligned} x + (\lambda - 3)y &= 0 \\ (\lambda - 3)x + y &= 0 \end{aligned}$$

have a non–trivial solution?

[Answer: $\lambda = 2,\ 4$.]

8. Solve the homogeneous system

$$\begin{aligned} 3x_1 + x_2 + x_3 + x_4 &= 0 \\ 5x_1 - x_2 + x_3 - x_4 &= 0. \end{aligned}$$

[Answer: $x_1 = -\frac{1}{4}x_3$, $x_2 = -\frac{1}{4}x_3 - x_4$, with x_3 and x_4 arbitrary.]

9. Let A be the coefficient matrix of the following homogeneous system of n equations in n unknowns:

$$\begin{aligned} (1-n)x_1 + x_2 + \cdots + x_n &= 0 \\ x_1 + (1-n)x_2 + \cdots + x_n &= 0 \\ \cdots &= 0 \\ x_1 + x_2 + \cdots + (1-n)x_n &= 0. \end{aligned}$$

Find the reduced row–echelon form of A and hence, or otherwise, prove that the solution of the above system is $x_1 = x_2 = \cdots = x_n$, with x_n arbitrary.

10. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a matrix over a field F. Prove that A is row–equivalent to $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ if $ad - bc \neq 0$, but is row–equivalent to a matrix whose second row is zero, if $ad - bc = 0$.

11. For which rational numbers a does the following system have (i) no solutions (ii) exactly one solution (iii) infinitely many solutions?

$$\begin{aligned} x + 2y - 3z &= 4 \\ 3x - y + 5z &= 2 \\ 4x + y + (a^2 - 14)z &= a + 2. \end{aligned}$$

[Answer: $a = -4$, no solution; $a = 4$, infinitely many solutions; $a \neq \pm 4$, exactly one solution.]

12. Solve the following system of homogeneous equations over $\mathbb{Z}_2$:

$$\begin{aligned} x_1 + x_3 + x_5 &= 0 \\ x_2 + x_4 + x_5 &= 0 \\ x_1 + x_2 + x_3 + x_4 &= 0 \\ x_3 + x_4 &= 0. \end{aligned}$$

[Answer: $x_1 = x_2 = x_4 + x_5$, $x_3 = x_4$, with x_4 and x_5 arbitrary elements of $\mathbb{Z}_2$.]

13. Solve the following systems of linear equations over $\mathbb{Z}_5$:

$$(a) \quad \begin{aligned} 2x + y + 3z &= 4 \\ 4x + y + 4z &= 1 \\ 3x + y + 2z &= 0 \end{aligned} \qquad (b) \quad \begin{aligned} 2x + y + 3z &= 4 \\ 4x + y + 4z &= 1 \\ x + y &= 3. \end{aligned}$$

[Answer: (a) $x = 1$, $y = 2$, $z = 0$; (b) $x = 1 + 2z$, $y = 2 + 3z$, with z an arbitrary element of $\mathbb{Z}_5$.]

14. If $(\alpha_1, \ldots, \alpha_n)$ and $(\beta_1, \ldots, \beta_n)$ are solutions of a system of linear equations, prove that

$$((1-t)\alpha_1 + t\beta_1, \ldots, (1-t)\alpha_n + t\beta_n)$$

is also a solution.

15. If $(\alpha_1, \ldots, \alpha_n)$ is a solution of a system of linear equations, prove that the complete solution is given by $x_1 = \alpha_1 + y_1, \ldots, x_n = \alpha_n + y_n$, where $(y_1, \ldots, y_n)$ is the general solution of the associated homogeneous system.

16. Find the values of a and b for which the following system is consistent. Also find the complete solution when $a = b = 2$.

$$\begin{aligned} x + y - z + w &= 1 \\ ax + y + z + w &= b \\ 3x + 2y + \quad aw &= 1 + a. \end{aligned}$$

[Answer: $a \neq 2$ or $a = 2 = b$; $x = 1 - 2z$, $y = 3z - w$, with z, w arbitrary.]

17. Let $F = \{0, 1, a, b\}$ be a field consisting of 4 elements.

(a) Determine the addition and multiplication tables of F. (Hint: Prove that the elements $1+0$, $1+1$, $1+a$, $1+b$ are distinct and deduce that $1 + 1 + 1 + 1 = 0$; then deduce that $1 + 1 = 0$.)

(b) A matrix A, whose elements belong to F, is defined by

$$A = \begin{bmatrix} 1 & a & b & a \\ a & b & b & 1 \\ 1 & 1 & 1 & a \end{bmatrix},$$

prove that the reduced row–echelon form of A is given by the matrix

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & b \\ 0 & 0 & 1 & 1 \end{bmatrix}.$$

Chapter 2

MATRICES

2.1 Matrix arithmetic

A matrix over a field F is a rectangular array of elements from F. The symbol $M_{m\times n}(F)$ denotes the collection of all $m \times n$ matrices over F. Matrices will usually be denoted by capital letters and the equation $A = [a_{ij}]$ means that the element in the i–th row and j–th column of the matrix A equals a_{ij}. It is also occasionally convenient to write $a_{ij} = (A)_{ij}$. For the present, all matrices will have rational entries, unless otherwise stated.

EXAMPLE 2.1.1 The formula $a_{ij} = 1/(i+j)$ for $1 \le i \le 3$, $1 \le j \le 4$ defines a 3×4 matrix $A = [a_{ij}]$, namely

$$A = \begin{bmatrix} \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{bmatrix}.$$

DEFINITION 2.1.1 (Equality of matrices) Matrices A, B are said to be equal if A and B have the same size and corresponding elements are equal; i.e., A and $B \in M_{m\times n}(F)$ and $A = [a_{ij}]$, $B = [b_{ij}]$, with $a_{ij} = b_{ij}$ for $1 \le i \le m$, $1 \le j \le n$.

DEFINITION 2.1.2 (Addition of matrices) Let $A = [a_{ij}]$ and $B = [b_{ij}]$ be of the same size. Then $A + B$ is the matrix obtained by adding corresponding elements of A and B; that is

$$A + B = [a_{ij}] + [b_{ij}] = [a_{ij} + b_{ij}].$$

DEFINITION 2.1.3 (Scalar multiple of a matrix) Let $A = [a_{ij}]$ and $t \in F$ (that is t is a *scalar*). Then tA is the matrix obtained by multiplying all elements of A by t; that is

$$tA = t[a_{ij}] = [ta_{ij}].$$

DEFINITION 2.1.4 (Additive inverse of a matrix) Let $A = [a_{ij}]$. Then $-A$ is the matrix obtained by replacing the elements of A by their additive inverses; that is

$$-A = -[a_{ij}] = [-a_{ij}].$$

DEFINITION 2.1.5 (Subtraction of matrices) Matrix subtraction is defined for two matrices $A = [a_{ij}]$ and $B = [b_{ij}]$ of the same size, in the usual way; that is

$$A - B = [a_{ij}] - [b_{ij}] = [a_{ij} - b_{ij}].$$

DEFINITION 2.1.6 (The zero matrix) For each m, n the matrix in $M_{m\times n}(F)$, all of whose elements are zero, is called the *zero* matrix (of size $m \times n$) and is denoted by the symbol 0.

The matrix operations of addition, scalar multiplication, additive inverse and subtraction satisfy the usual laws of arithmetic. (In what follows, s and t will be arbitrary scalars and A, B, C are matrices of the same size.)

1. $(A + B) + C = A + (B + C)$;
2. $A + B = B + A$;
3. $0 + A = A$;
4. $A + (-A) = 0$;
5. $(s + t)A = sA + tA, \quad (s - t)A = sA - tA$;
6. $t(A + B) = tA + tB, \quad t(A - B) = tA - tB$;
7. $s(tA) = (st)A$;
8. $1A = A, \quad 0A = 0, \quad (-1)A = -A$;
9. $tA = 0 \Rightarrow t = 0$ or $A = 0$.

Other similar properties will be used when needed.

DEFINITION 2.1.7 (Matrix product) Let $A = [a_{ij}]$ be a matrix of size $m \times n$ and $B = [b_{jk}]$ be a matrix of size $n \times p$; (that is the number of columns of A equals the number of rows of B). Then AB is the $m \times p$ matrix $C = [c_{ik}]$ whose $(i,\, k)$–th element is defined by the formula

$$c_{ik} = \sum_{j=1}^{n} a_{ij} b_{jk} = a_{i1} b_{1k} + \cdots + a_{in} b_{nk}.$$

EXAMPLE 2.1.2

1. $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1\times 5 + 2\times 7 & 1\times 6 + 2\times 8 \\ 3\times 5 + 4\times 7 & 3\times 6 + 4\times 8 \end{bmatrix} = \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix};$

2. $\begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix} \neq \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix};$

3. $\begin{bmatrix} 1 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 4 \end{bmatrix} = \begin{bmatrix} 3 & 4 \\ 6 & 8 \end{bmatrix};$

4. $\begin{bmatrix} 3 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 11 \end{bmatrix};$

5. $\begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$

Matrix multiplication obeys many of the familiar laws of arithmetic apart from the commutative law.

1. $(AB)C = A(BC)$ if A, B, C are $m \times n$, $n \times p$, $p \times q$, respectively;
2. $t(AB) = (tA)B = A(tB)$, $\;A(-B) = (-A)B = -(AB)$;
3. $(A + B)C = AC + BC$ if A and B are $m \times n$ and C is $n \times p$;
4. $D(A + B) = DA + DB$ if A and B are $m \times n$ and D is $p \times m$.

We prove the associative law only:
First observe that $(AB)C$ and $A(BC)$ are both of size $m \times q$.
Let $A = [a_{ij}]$, $B = [b_{jk}]$, $C = [c_{kl}]$. Then

$$\begin{aligned} ((AB)C)_{il} &= \sum_{k=1}^{p} (AB)_{ik} c_{kl} = \sum_{k=1}^{p} \left(\sum_{j=1}^{n} a_{ij} b_{jk} \right) c_{kl} \\ &= \sum_{k=1}^{p} \sum_{j=1}^{n} a_{ij} b_{jk} c_{kl}. \end{aligned}$$

Similarly

$$(A(BC))_{il} = \sum_{j=1}^{n}\sum_{k=1}^{p} a_{ij}b_{jk}c_{kl}.$$

However the double summations are equal. For sums of the form

$$\sum_{j=1}^{n}\sum_{k=1}^{p} d_{jk} \quad \text{and} \quad \sum_{k=1}^{p}\sum_{j=1}^{n} d_{jk}$$

represent the sum of the np elements of the rectangular array $[d_{jk}]$, by rows and by columns, respectively. Consequently

$$((AB)C)_{il} = (A(BC))_{il}$$

for $1 \leq i \leq m$, $1 \leq l \leq q$. Hence $(AB)C = A(BC)$.

The system of m linear equations in n unknowns

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

is equivalent to a single matrix equation

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

that is $AX = B$, where $A = [a_{ij}]$ is the *coefficient matrix* of the system, $X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ is the *vector of unknowns* and $B = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$ is the *vector of constants.*

Another useful matrix equation equivalent to the above system of linear equations is

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

EXAMPLE 2.1.3 The system

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

is equivalent to the matrix equation

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

and to the equation

$$x \begin{bmatrix} 1 \\ 1 \end{bmatrix} + y \begin{bmatrix} 1 \\ -1 \end{bmatrix} + z \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}.$$

2.2 Linear transformations

An *n–dimensional column vector* is an $n \times 1$ matrix over F. The collection of all n–dimensional column vectors is denoted by F^n.

Every matrix is associated with an important type of function called a *linear transformation*.

DEFINITION 2.2.1 (Linear transformation) We can associate with $A \in M_{m\times n}(F)$, the function $T_A : F^n \to F^m$, defined by $T_A(X) = AX$ for all $X \in F^n$. More explicitly, using components, the above function takes the form

$$\begin{aligned} y_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ y_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ &\vdots \\ y_m &= a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n, \end{aligned}$$

where $y_1, y_2, \cdots, y_m$ are the components of the column vector $T_A(X)$.

The function just defined has the property that

$$T_A(sX + tY) = sT_A(X) + tT_A(Y) \tag{2.1}$$

for all $s, t \in F$ and all n–dimensional column vectors X, Y. For

$$T_A(sX + tY) = A(sX + tY) = s(AX) + t(AY) = sT_A(X) + tT_A(Y).$$

REMARK 2.2.1 It is easy to prove that if $T : F^n \to F^m$ is a function satisfying equation 2.1, then $T = T_A$, where A is the $m \times n$ matrix whose columns are $T(E_1), \ldots, T(E_n)$, respectively, where $E_1, \ldots, E_n$ are the n–dimensional *unit vectors* defined by

$$E_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \ldots \quad , E_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}.$$

One well–known example of a linear transformation arises from rotating the (x, y)–plane in 2-dimensional Euclidean space, anticlockwise through θ radians. Here a point (x, y) will be transformed into the point (x_1, y_1), where

$$\begin{aligned} x_1 &= x\cos\theta - y\sin\theta \\ y_1 &= x\sin\theta + y\cos\theta. \end{aligned}$$

In 3–dimensional Euclidean space, the equations

$$x_1 = x\cos\theta - y\sin\theta,\ y_1 = x\sin\theta + y\cos\theta,\ z_1 = z;$$
$$x_1 = x,\ y_1 = y\cos\phi - z\sin\phi,\ z_1 = y\sin\phi + z\cos\phi;$$
$$x_1 = x\cos\psi + z\sin\psi,\ y_1 = y,\ z_1 = -x\sin\psi + z\cos\psi;$$

correspond to rotations about the positive z, x and y axes, anticlockwise through θ, ϕ, ψ radians, respectively.

The product of two matrices is related to the product of the corresponding linear transformations:

If A is $m \times n$ and B is $n \times p$, then the function $T_A T_B : F^p \to F^m$, obtained by first performing T_B, then T_A is in fact equal to the linear transformation T_{AB}. For if $X \in F^p$, we have

$$T_A T_B(X) = A(BX) = (AB)X = T_{AB}(X).$$

The following example is useful for producing rotations in 3–dimensional animated design. (See [27, pages 97–112].)

EXAMPLE 2.2.1 The linear transformation resulting from successively rotating 3–dimensional space about the positive z, x, y–axes, anticlockwise through θ, ϕ, ψ radians respectively, is equal to T_{ABC}, where

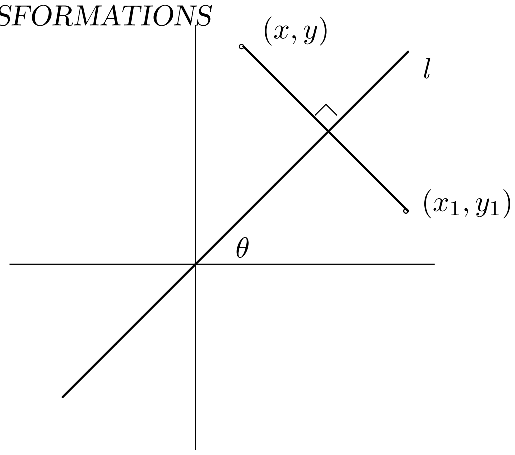

Figure 2.1: Reflection in a line.

$$C = \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}.$$

$$A = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix}.$$

The matrix ABC is quite complicated:

$$A(BC) = \begin{bmatrix} \cos\psi & 0 & \sin\psi \\ 0 & 1 & 0 \\ -\sin\psi & 0 & \cos\psi \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & 0 \\ \cos\phi\sin\theta & \cos\phi\cos\theta & -\sin\phi \\ \sin\phi\sin\theta & \sin\phi\cos\theta & \cos\phi \end{bmatrix}$$

$$= \begin{bmatrix} \cos\psi\cos\theta+\sin\psi\sin\phi\sin\theta & -\cos\psi\sin\theta+\sin\psi\sin\phi\cos\theta & \sin\psi\cos\phi \\ \cos\phi\sin\theta & \cos\phi\cos\theta & -\sin\phi \\ -\sin\psi\cos\theta+\cos\psi\sin\phi\sin\theta & \sin\psi\sin\theta+\cos\psi\sin\phi\cos\theta & \cos\psi\cos\phi \end{bmatrix}.$$

EXAMPLE 2.2.2 Another example from geometry is reflection of the plane in a line l inclined at an angle θ to the positive x–axis.

We reduce the problem to the simpler case $\theta = 0$, where the equations of transformation are $x_1 = x$, $y_1 = -y$. First rotate the plane clockwise through θ radians, thereby taking l into the x–axis; next reflect the plane in the x–axis; then rotate the plane anticlockwise through θ radians, thereby restoring l to its original position.

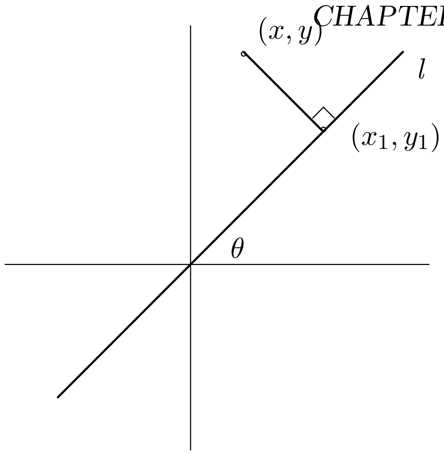

Figure 2.2: Projection on a line.

In terms of matrices, we get transformation equations

$$\begin{aligned}
\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \begin{bmatrix} \cos\theta & \sin\theta \\ \sin\theta & -\cos\theta \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \begin{bmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.
\end{aligned}$$

The more general transformation

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = a \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} u \\ v \end{bmatrix}, \quad a > 0,$$

represents a rotation, followed by a scaling and then by a translation. Such transformations are important in computer graphics. See [23, 24].

EXAMPLE 2.2.3 Our last example of a geometrical linear transformation arises from projecting the plane onto a line l through the origin, inclined at angle θ to the positive x–axis. Again we reduce that problem to the simpler case where l is the x–axis and the equations of transformation are $x_1 = x$, $y_1 = 0$.

In terms of matrices, we get transformation equations

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$$

$$
\begin{aligned}
&= \begin{bmatrix} \cos\theta & 0 \\ \sin\theta & 0 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \\
&= \begin{bmatrix} \cos^2\theta & \cos\theta\sin\theta \\ \sin\theta\cos\theta & \sin^2\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.
\end{aligned}
$$

2.3 Recurrence relations

DEFINITION 2.3.1 (The identity matrix) The $n \times n$ matrix $I_n = [\delta_{ij}]$, defined by $\delta_{ij} = 1$ if $i = j$, $\delta_{ij} = 0$ if $i \neq j$, is called the $n \times n$ *identity* matrix of order n. In other words, the columns of the identity matrix of order n are the unit vectors $E_1, \cdots, E_n$, respectively.

For example, $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

THEOREM 2.3.1 If A is $m \times n$, then $I_m A = A = A I_n$.

DEFINITION 2.3.2 (k–th power of a matrix) If A is an $n \times n$ matrix, we define A^k recursively as follows: $A^0 = I_n$ and $A^{k+1} = A^k A$ for $k \geq 0$.

For example $A^1 = A^0 A = I_n A = A$ and hence $A^2 = A^1 A = AA$.

The usual index laws hold provided $AB = BA$:

1. $A^m A^n = A^{m+n}, \quad (A^m)^n = A^{mn}$;
2. $(AB)^n = A^n B^n$;
3. $A^m B^n = B^n A^m$;
4. $(A + B)^2 = A^2 + 2AB + B^2$;
5. $(A + B)^n = \sum_{i=0}^{n} \binom{n}{i} A^i B^{n-i}$;
6. $(A + B)(A - B) = A^2 - B^2$.

We now state a basic property of the natural numbers.

AXIOM 2.3.1 (MATHEMATICAL INDUCTION) *If $\mathcal{P}_n$ denotes a mathematical statement for each $n \geq 1$, satisfying*

(i) *$\mathcal{P}_1$ is true,*

(ii) *the truth of* $\mathcal{P}_n$ *implies that of* $\mathcal{P}_{n+1}$ *for each* $n \geq 1$,

then $\mathcal{P}_n$ *is true for all* $n \geq 1$.

EXAMPLE 2.3.1 Let $A = \begin{bmatrix} 7 & 4 \\ -9 & -5 \end{bmatrix}$. Prove that

$$A^n = \begin{bmatrix} 1+6n & 4n \\ -9n & 1-6n \end{bmatrix} \quad \text{if } n \geq 1.$$

Solution. We use the principle of mathematical induction.

Take $\mathcal{P}_n$ to be the statement

$$A^n = \begin{bmatrix} 1+6n & 4n \\ -9n & 1-6n \end{bmatrix}.$$

Then $\mathcal{P}_1$ asserts that

$$A^1 = \begin{bmatrix} 1+6\times 1 & 4\times 1 \\ -9\times 1 & 1-6\times 1 \end{bmatrix} = \begin{bmatrix} 7 & 4 \\ -9 & -5 \end{bmatrix},$$

which is true. Now let $n \geq 1$ and assume that $\mathcal{P}_n$ is true. We have to deduce that

$$A^{n+1} = \begin{bmatrix} 1+6(n+1) & 4(n+1) \\ -9(n+1) & 1-6(n+1) \end{bmatrix} = \begin{bmatrix} 7+6n & 4n+4 \\ -9n-9 & -5-6n \end{bmatrix}.$$

Now

$$\begin{aligned} A^{n+1} &= A^nA \\ &= \begin{bmatrix} 1+6n & 4n \\ -9n & 1-6n \end{bmatrix}\begin{bmatrix} 7 & 4 \\ -9 & -5 \end{bmatrix} \\ &= \begin{bmatrix} (1+6n)7+(4n)(-9) & (1+6n)4+(4n)(-5) \\ (-9n)7+(1-6n)(-9) & (-9n)4+(1-6n)(-5) \end{bmatrix} \\ &= \begin{bmatrix} 7+6n & 4n+4 \\ -9n-9 & -5-6n \end{bmatrix}, \end{aligned}$$

and "the induction goes through".

The last example has an application to the solution of a system of *recurrence relations*:

EXAMPLE 2.3.2 The following system of recurrence relations holds for all $n \geq 0$:

$$\begin{aligned} x_{n+1} &= 7x_n + 4y_n \\ y_{n+1} &= -9x_n - 5y_n. \end{aligned}$$

Solve the system for x_n and y_n in terms of x_0 and y_0.

Solution. Combine the above equations into a single matrix equation

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} 7 & 4 \\ -9 & -5 \end{bmatrix} \begin{bmatrix} x_n \\ y_n \end{bmatrix},$$

or $X_{n+1} = AX_n$, where $A = \begin{bmatrix} 7 & 4 \\ -9 & -5 \end{bmatrix}$ and $X_n = \begin{bmatrix} x_n \\ y_n \end{bmatrix}$.
We see that

$$\begin{aligned} X_1 &= AX_0 \\ X_2 &= AX_1 = A(AX_0) = A^2X_0 \\ &\vdots \\ X_n &= A^nX_0. \end{aligned}$$

(The truth of the equation $X_n = A^nX_0$ for $n \geq 1$, strictly speaking follows by mathematical induction; however for simple cases such as the above, it is customary to omit the strict proof and supply instead a few lines of motivation for the inductive statement.)

Hence the previous example gives

$$\begin{aligned} \begin{bmatrix} x_n \\ y_n \end{bmatrix} = X_n &= \begin{bmatrix} 1+6n & 4n \\ -9n & 1-6n \end{bmatrix} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} \\ &= \begin{bmatrix} (1+6n)x_0 + (4n)y_0 \\ (-9n)x_0 + (1-6n)y_0 \end{bmatrix}, \end{aligned}$$

and hence $x_n = (1+6n)x_0 + 4ny_0$ and $y_n = (-9n)x_0 + (1-6n)y_0$, for $n \geq 1$.

2.4 PROBLEMS

1. Let A, B, C, D be matrices defined by

$$A = \begin{bmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 5 & 2 \\ -1 & 1 & 0 \\ -4 & 1 & 3 \end{bmatrix},$$

$$C = \begin{bmatrix} -3 & -1 \\ 2 & 1 \\ 4 & 3 \end{bmatrix}, \quad D = \begin{bmatrix} 4 & -1 \\ 2 & 0 \end{bmatrix}.$$

Which of the following matrices are defined? Compute those matrices which are defined.

$$A+B,\ A+C,\ AB,\ BA,\ CD,\ DC,\ D^2.$$

[Answers: $A+C$, BA, CD, D^2;

$$\begin{bmatrix} 0 & -1 \\ 1 & 3 \\ 5 & 4 \end{bmatrix}, \quad \begin{bmatrix} 0 & 12 \\ -4 & 2 \\ -10 & 5 \end{bmatrix}, \quad \begin{bmatrix} -14 & 3 \\ 10 & -2 \\ 22 & -4 \end{bmatrix}, \quad \begin{bmatrix} 14 & -4 \\ 8 & -2 \end{bmatrix}.]$$

2. Let $A = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$. Show that if B is a 3×2 such that $AB = I_2$, then

$$B = \begin{bmatrix} a & b \\ -a-1 & 1-b \\ a+1 & b \end{bmatrix}$$

for suitable numbers a and b. Use the associative law to show that $(BA)^2 B = B$.

3. If $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, prove that $A^2 - (a+d)A + (ad-bc)I_2 = 0$.

4. If $A = \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}$, use the fact $A^2 = 4A - 3I_2$ and mathematical induction, to prove that

$$A^n = \frac{(3^n - 1)}{2} A + \frac{3 - 3^n}{2} I_2 \quad \text{if } n \geq 1.$$

5. A sequence of numbers $x_1, x_2, \ldots, x_n, \ldots$ satisfies the recurrence relation $x_{n+1} = ax_n + bx_{n-1}$ for $n \geq 1$, where a and b are constants. Prove that

$$\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix} = A \begin{bmatrix} x_n \\ x_{n-1} \end{bmatrix},$$

where $A = \begin{bmatrix} a & b \\ 1 & 0 \end{bmatrix}$ and hence express $\begin{bmatrix} x_{n+1} \\ x_n \end{bmatrix}$ in terms of $\begin{bmatrix} x_1 \\ x_0 \end{bmatrix}$.
If $a = 4$ and $b = -3$, use the previous question to find a formula for x_n in terms of x_1 and x_0.

[Answer:

$$x_n = \frac{3^n - 1}{2} x_1 + \frac{3 - 3^n}{2} x_0.]$$

6. Let $A = \begin{bmatrix} 2a & -a^2 \\ 1 & 0 \end{bmatrix}$.

   (a) Prove that

   $$A^n = \begin{bmatrix} (n+1)a^n & -na^{n+1} \\ na^{n-1} & (1-n)a^n \end{bmatrix} \quad \text{if } n \geq 1.$$

   (b) A sequence $x_0, x_1, \ldots, x_n, \ldots$ satisfies $x_{n+1} = 2ax_n - a^2 x_{n-1}$ for $n \geq 1$. Use part (a) and the previous question to prove that $x_n = na^{n-1}x_1 + (1-n)a^n x_0$ for $n \geq 1$.

7. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and suppose that λ_1 and λ_2 are the roots of the quadratic polynomial $x^2 - (a+d)x + ad - bc$. (λ_1 and λ_2 may be equal.) Let k_n be defined by $k_0 = 0$, $k_1 = 1$ and for $n \geq 2$

$$k_n = \sum_{i=1}^{n} \lambda_1^{n-i} \lambda_2^{i-1}.$$

Prove that

$$k_{n+1} = (\lambda_1 + \lambda_2)k_n - \lambda_1 \lambda_2 k_{n-1},$$

if $n \geq 1$. Also prove that

$$k_n = \begin{cases} (\lambda_1^n - \lambda_2^n)/(\lambda_1 - \lambda_2) & \text{if } \lambda_1 \neq \lambda_2, \\ n\lambda_1^{n-1} & \text{if } \lambda_1 = \lambda_2. \end{cases}$$

Use mathematical induction to prove that if $n \geq 1$,

$$A^n = k_n A - \lambda_1 \lambda_2 k_{n-1} I_2,$$

[Hint: Use the equation $A^2 = (a+d)A - (ad - bc)I_2$.]

8. Use Question 7 to prove that if $A = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, then

$$A^n = \frac{3^n}{2}\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + \frac{(-1)^{n-1}}{2}\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}$$

if $n \geq 1$.

9. The Fibonacci numbers are defined by the equations $F_0 = 0$, $F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$ if $n \geq 1$. Prove that

$$F_n = \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right)$$

if $n \geq 0$.

10. Let $r > 1$ be an integer. Let a and b be arbitrary positive integers. Sequences x_n and y_n of positive integers are defined in terms of a and b by the recurrence relations

$$\begin{aligned} x_{n+1} &= x_n + ry_n \\ y_{n+1} &= x_n + y_n, \end{aligned}$$

for $n \geq 0$, where $x_0 = a$ and $y_0 = b$.

Use Question 7 to prove that

$$\frac{x_n}{y_n} \to \sqrt{r} \quad \text{as } n \to \infty.$$

2.5 Non–singular matrices

DEFINITION 2.5.1 (Non–singular matrix) A matrix $A \in M_{n\times n}(F)$ is called *non–singular* or *invertible* if there exists a matrix $B \in M_{n\times n}(F)$ such that

$$AB = I_n = BA.$$

Any matrix B with the above property is called an *inverse* of A. If A does not have an inverse, A is called *singular*.

THEOREM 2.5.1 (Inverses are unique) If A has inverses B and C, then $B = C$.

Proof. Let B and C be inverses of A. Then $AB = I_n = BA$ and $AC = I_n = CA$. Then $B(AC) = BI_n = B$ and $(BA)C = I_nC = C$. Hence because $B(AC) = (BA)C$, we deduce that $B = C$.

REMARK 2.5.1 If A has an inverse, it is denoted by A^{-1}. So

$$AA^{-1} = I_n = A^{-1}A.$$

Also if A is non–singular, it follows that A^{-1} is also non–singular and

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

THEOREM 2.5.2 If A and B are non–singular matrices of the same size, then so is AB. Moreover

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

Proof.

$$(AB)(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AI_nA^{-1} = AA^{-1} = I_n.$$

Similarly

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

REMARK 2.5.2 The above result generalizes to a product of m non–singular matrices: If $A_1, \ldots, A_m$ are non–singular $n \times n$ matrices, then the product $A_1 \ldots A_m$ is also non–singular. Moreover

$$(A_1 \ldots A_m)^{-1} = A_m^{-1} \ldots A_1^{-1}.$$

(Thus the inverse of the product equals the product of the inverses *in the reverse order*.)

EXAMPLE 2.5.1 If A and B are $n \times n$ matrices satisfying $A^2 = B^2 = (AB)^2 = I_n$, prove that $AB = BA$.

Solution. Assume $A^2 = B^2 = (AB)^2 = I_n$. Then A, B, AB are non–singular and $A^{-1} = A$, $B^{-1} = B$, $(AB)^{-1} = AB$.

But $(AB)^{-1} = B^{-1}A^{-1}$ and hence $AB = BA$.

EXAMPLE 2.5.2 $A = \begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix}$ is singular. For suppose $B = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is an inverse of A. Then the equation $AB = I_2$ gives

$$\begin{bmatrix} 1 & 2 \\ 4 & 8 \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

and equating the corresponding elements of column 1 of both sides gives the system

$$\begin{aligned} a + 2c &= 1 \\ 4a + 8c &= 0 \end{aligned}$$

which is clearly inconsistent.

THEOREM 2.5.3 Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and $\Delta = ad - bc \neq 0$. Then A is non–singular. Also

$$A^{-1} = \Delta^{-1} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

REMARK 2.5.3 The expression $ad - bc$ is called the *determinant* of A and is denoted by the symbols $\det A$ or $\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.

Proof. Verify that the matrix $B = \Delta^{-1} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$ satisfies the equation $AB = I_2 = BA$.

EXAMPLE 2.5.3 Let

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 5 & 0 & 0 \end{bmatrix}.$$

Verify that $A^3 = 5I_3$, deduce that A is non–singular and find A^{-1}.

Solution. After verifying that $A^3 = 5I_3$, we notice that

$$A\left(\frac{1}{5}A^2\right) = I_3 = \left(\frac{1}{5}A^2\right)A.$$

Hence A is non–singular and $A^{-1} = \frac{1}{5}A^2$.

THEOREM 2.5.4 If the coefficient matrix A of a system of n equations in n unknowns is non–singular, then the system $AX = B$ has the unique solution $X = A^{-1}B$.

Proof. Assume that A^{-1} exists.

1. (Uniqueness.) Assume that $AX = B$. Then

$$\begin{aligned} (A^{-1}A)X &= A^{-1}B, \\ I_nX &= A^{-1}B, \\ X &= A^{-1}B. \end{aligned}$$

2. (Existence.) Let $X = A^{-1}B$. Then

$$AX = A(A^{-1}B) = (AA^{-1})B = I_nB = B.$$

THEOREM 2.5.5 (Cramer's rule for 2 equations in 2 unknowns) The system

$$\begin{aligned} ax + by &= e \\ cx + dy &= f \end{aligned}$$

has a unique solution if $\Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0$, namely

$$x = \frac{\Delta_1}{\Delta}, \qquad y = \frac{\Delta_2}{\Delta},$$

where

$$\Delta_1 = \begin{vmatrix} e & b \\ f & d \end{vmatrix} \quad \text{and} \quad \Delta_2 = \begin{vmatrix} a & e \\ c & f \end{vmatrix}.$$

Proof. Suppose $\Delta \neq 0$. Then $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has inverse

$$A^{-1} = \Delta^{-1} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

and we know that the system

$$A \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} e \\ f \end{bmatrix}$$

has the unique solution

$$\begin{aligned} \begin{bmatrix} x \\ y \end{bmatrix} = A^{-1} \begin{bmatrix} e \\ f \end{bmatrix} &= \frac{1}{\Delta} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} \\ &= \frac{1}{\Delta} \begin{bmatrix} de - bf \\ -ce + af \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix} = \begin{bmatrix} \Delta_1/\Delta \\ \Delta_2/\Delta \end{bmatrix}. \end{aligned}$$

Hence $x = \Delta_1/\Delta$, $y = \Delta_2/\Delta$.

COROLLARY 2.5.1 The homogeneous system

$$\begin{aligned} ax + by &= 0 \\ cx + dy &= 0 \end{aligned}$$

has only the trivial solution if $\Delta = \begin{vmatrix} a & b \\ c & d \end{vmatrix} \neq 0$.

EXAMPLE 2.5.4 The system

$$\begin{aligned} 7x + 8y &= 100 \\ 2x - 9y &= 10 \end{aligned}$$

has the unique solution $x = \Delta_1/\Delta$, $y = \Delta_2/\Delta$, where

$$\Delta = \begin{vmatrix} 7 & 8 \\ 2 & -9 \end{vmatrix} = -79, \quad \Delta_1 = \begin{vmatrix} 100 & 8 \\ 10 & -9 \end{vmatrix} = -980, \quad \Delta_2 = \begin{vmatrix} 7 & 100 \\ 2 & 10 \end{vmatrix} = -130.$$

So $x = \frac{980}{79}$ and $y = \frac{130}{79}$.

THEOREM 2.5.6 Let A be a square matrix. If A is non–singular, the homogeneous system $AX = 0$ has only the trivial solution. Equivalently, if the homogenous system $AX = 0$ has a non–trivial solution, then A is singular.

Proof. If A is non–singular and $AX = 0$, then $X = A^{-1}0 = 0$.

REMARK 2.5.4 If $A_{*1}, \ldots, A_{*n}$ denote the columns of A, then the equation

$$AX = x_1 A_{*1} + \ldots + x_n A_{*n}$$

holds. Consequently theorem 2.5.6 tells us that if there exist $x_1, \ldots, x_n$, *not all zero*, such that

$$x_1 A_{*1} + \ldots + x_n A_{*n} = 0,$$

that is, if the columns of A are *linearly dependent*, then A is singular. An equivalent way of saying that the columns of A are linearly dependent is that one of the columns of A is expressible as a sum of certain scalar multiples of the remaining columns of A; that is one column is a *linear combination* of the remaining columns.

EXAMPLE 2.5.5

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 4 & 7 \end{bmatrix}$$

is singular. For it can be verified that A has reduced row–echelon form

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

and consequently $AX = 0$ has a non–trivial solution $x = -1$, $y = -1$, $z = 1$.

REMARK 2.5.5 More generally, if A is row–equivalent to a matrix containing a zero row, then A is singular. For then the homogeneous system $AX = 0$ has a non–trivial solution.

An important class of non–singular matrices is that of the *elementary row matrices*.

DEFINITION 2.5.2 (Elementary row matrices) To each of the three types of elementary row operation, there corresponds an *elementary row matrix*, denoted by E_{ij}, $E_i(t)$, $E_{ij}(t)$:

1. E_{ij}, $(i \neq j)$ is obtained from the identity matrix I_n by interchanging rows i and j.

2. $E_i(t)$, $(t \neq 0)$ is obtained by multiplying the i–th row of I_n by t.

3. $E_{ij}(t)$, $(i \neq j)$ is obtained from I_n by adding t times the j–th row of I_n to the i–th row.

EXAMPLE 2.5.6 $(n = 3.)$

$$E_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \; E_2(-1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \; E_{23}(-1) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}.$$

The elementary row matrices have the following distinguishing property:

THEOREM 2.5.7 If a matrix A is pre–multiplied by an elementary row matrix, the resulting matrix is the one obtained by performing the corresponding elementary row–operation on A.

EXAMPLE 2.5.7

$$E_{23}\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} a & b \\ c & d \\ e & f \end{bmatrix} = \begin{bmatrix} a & b \\ e & f \\ c & d \end{bmatrix}.$$

COROLLARY 2.5.2 Elementary row–matrices are non–singular. Indeed

1. $E_{ij}^{-1} = E_{ij}$;
2. $E_i^{-1}(t) = E_i(t^{-1})$;
3. $(E_{ij}(t))^{-1} = E_{ij}(-t)$.

Proof. Taking $A = I_n$ in the above theorem, we deduce the following equations:

$$\begin{aligned} E_{ij}E_{ij} &= I_n \\ E_i(t)E_i(t^{-1}) &= I_n = E_i(t^{-1})E_i(t) \quad \text{if } t \neq 0 \\ E_{ij}(t)E_{ij}(-t) &= I_n = E_{ij}(-t)E_{ij}(t). \end{aligned}$$

EXAMPLE 2.5.8 Find the 3×3 matrix $A = E_3(5)E_{23}(2)E_{12}$ explicitly. Also find A^{-1}.

Solution.

$$A = E_3(5)E_{23}(2)\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} = E_3(5)\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 0 & 5 \end{bmatrix}.$$

To find A^{-1}, we have

$$\begin{aligned} A^{-1} &= (E_3(5)E_{23}(2)E_{12})^{-1} \\ &= E_{12}^{-1}\,(E_{23}(2))^{-1}\,(E_3(5))^{-1} \\ &= E_{12}E_{23}(-2)E_3(5^{-1}) \\ &= E_{12}E_{23}(-2)\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix} \\ &= E_{12}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -\frac{2}{5} \\ 0 & 0 & \frac{1}{5} \end{bmatrix} = \begin{bmatrix} 0 & 1 & -\frac{2}{5} \\ 1 & 0 & 0 \\ 0 & 0 & \frac{1}{5} \end{bmatrix}. \end{aligned}$$

REMARK 2.5.6 Recall that A and B are row–equivalent if B is obtained from A by a sequence of elementary row operations. If $E_1, \ldots, E_r$ are the respective corresponding elementary row matrices, then

$$B = E_r(\ldots(E_2(E_1A))\ldots) = (E_r \ldots E_1)A = PA,$$

where $P = E_r \ldots E_1$ is non–singular. Conversely if $B = PA$, where P is non–singular, then A is row–equivalent to B. For as we shall now see, P is in fact a product of elementary row matrices.

THEOREM 2.5.8 Let A be non–singular $n \times n$ matrix. Then

(i) A is row–equivalent to I_n,

(ii) A is a product of elementary row matrices.

Proof. Assume that A is non–singular and let B be the reduced row–echelon form of A. Then B has no zero rows, for otherwise the equation $AX = 0$ would have a non–trivial solution. Consequently $B = I_n$.

It follows that there exist elementary row matrices $E_1, \ldots, E_r$ such that $E_r(\ldots(E_1A)\ldots) = B = I_n$ and hence $A = E_1^{-1} \ldots E_r^{-1}$, a product of elementary row matrices.

THEOREM 2.5.9 Let A be $n \times n$ and suppose that A is row–equivalent to I_n. Then A is non–singular and A^{-1} can be found by performing the same sequence of elementary row operations on I_n as were used to convert A to I_n.

Proof. Suppose that $E_r \ldots E_1A = I_n$. In other words $BA = I_n$, where $B = E_r \ldots E_1$ is non–singular. Then $B^{-1}(BA) = B^{-1}I_n$ and so $A = B^{-1}$, which is non–singular.

Also $A^{-1} = \left(B^{-1}\right)^{-1} = B = E_r((\ldots(E_1I_n)\ldots)$, which shows that A^{-1} is obtained from I_n by performing the same sequence of elementary row operations as were used to convert A to I_n.

REMARK 2.5.7 It follows from theorem 2.5.9 that if A is singular, then A is row–equivalent to a matrix whose last row is zero.

EXAMPLE 2.5.9 Show that $A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \end{bmatrix}$ is non–singular, find A^{-1} and express A as a product of elementary row matrices.

Solution. We form the *partitioned* matrix $[A|I_2]$ which consists of A followed by I_2. Then any sequence of elementary row operations which reduces A to I_2 will reduce I_2 to A^{-1}. Here

$$[A|I_2] = \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right]$$

$$R_2 \to R_2 - R_1 \quad \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & -1 & -1 & 1 \end{array}\right]$$

$$R_2 \to (-1)R_2 \quad \left[\begin{array}{cc|cc} 1 & 2 & 1 & 0 \\ 0 & 1 & 1 & -1 \end{array}\right]$$

$$R_1 \to R_1 - 2R_2 \quad \left[\begin{array}{cc|cc} 1 & 0 & -1 & 2 \\ 0 & 1 & 1 & -1 \end{array}\right].$$

Hence A is row–equivalent to I_2 and A is non–singular. Also

$$A^{-1} = \left[\begin{array}{cc} -1 & 2 \\ 1 & -1 \end{array}\right].$$

We also observe that

$$E_{12}(-2)E_2(-1)E_{21}(-1)A = I_2.$$

Hence

$$\begin{aligned} A^{-1} &= E_{12}(-2)E_2(-1)E_{21}(-1) \\ A &= E_{21}(1)E_2(-1)E_{12}(2). \end{aligned}$$

The next result is the converse of Theorem 2.5.6 and is useful for proving the non–singularity of certain types of matrices.

THEOREM 2.5.10 Let A be an $n \times n$ matrix with the property that the homogeneous system $AX = 0$ has only the trivial solution. Then A is non–singular. Equivalently, if A is singular, then the homogeneous system $AX = 0$ has a non–trivial solution.

Proof. If A is $n \times n$ and the homogeneous system $AX = 0$ has only the trivial solution, then it follows that the reduced row–echelon form B of A cannot have zero rows and must therefore be I_n. Hence A is non–singular.

COROLLARY 2.5.3 Suppose that A and B are $n \times n$ and $AB = I_n$. Then $BA = I_n$.

Proof. Let $AB = I_n$, where A and B are $n \times n$. We first show that B is non–singular. Assume $BX = 0$. Then $A(BX) = A0 = 0$, so $(AB)X = 0$, $I_nX = 0$ and hence $X = 0$.

Then from $AB = I_n$ we deduce $(AB)B^{-1} = I_nB^{-1}$ and hence $A = B^{-1}$. The equation $BB^{-1} = I_n$ then gives $BA = I_n$.

Before we give the next example of the above criterion for non-singularity, we introduce an important matrix operation.

DEFINITION 2.5.3 (The transpose of a matrix) Let A be an $m \times n$ matrix. Then A^t, the *transpose* of A, is the matrix obtained by interchanging the rows and columns of A. In other words if $A = [a_{ij}]$, then $\left(A^t\right)_{ji} = a_{ij}$. Consequently A^t is $n \times m$.

The transpose operation has the following properties:

1. $\left(A^t\right)^t = A$;
2. $(A \pm B)^t = A^t \pm B^t$ if A and B are $m \times n$;
3. $(sA)^t = sA^t$ if s is a scalar;
4. $(AB)^t = B^tA^t$ if A is $m \times n$ and B is $n \times p$;
5. If A is non–singular, then A^t is also non–singular and
$$\left(A^t\right)^{-1} = \left(A^{-1}\right)^t;$$
6. $X^tX = x_1^2 + \ldots + x_n^2$ if $X = [x_1, \ldots, x_n]^t$ is a column vector.

We prove only the fourth property. First check that both $(AB)^t$ and B^tA^t have the same size $(p \times m)$. Moreover, corresponding elements of both matrices are equal. For if $A = [a_{ij}]$ and $B = [b_{jk}]$, we have

$$\begin{aligned}
\left((AB)^t\right)_{ki} &= (AB)_{ik} \\
&= \sum_{j=1}^{n} a_{ij}b_{jk} \\
&= \sum_{j=1}^{n} \left(B^t\right)_{kj} \left(A^t\right)_{ji} \\
&= \left(B^tA^t\right)_{ki}.
\end{aligned}$$

There are two important classes of matrices that can be defined concisely in terms of the transpose operation.

DEFINITION 2.5.4 (Symmetric matrix) A matrix A is *symmetric* if $A^t = A$. In other words A is square ($n \times n$ say) and $a_{ji} = a_{ij}$ for all $1 \le i \le n$, $1 \le j \le n$. Hence

$$A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$$

is a general 2×2 symmetric matrix.

DEFINITION 2.5.5 (Skew–symmetric matrix) A matrix A is called *skew–symmetric* if $A^t = -A$. In other words A is square ($n \times n$ say) and $a_{ji} = -a_{ij}$ for all $1 \le i \le n$, $1 \le j \le n$.

REMARK 2.5.8 Taking $i = j$ in the definition of skew–symmetric matrix gives $a_{ii} = -a_{ii}$ and so $a_{ii} = 0$. Hence

$$A = \begin{bmatrix} 0 & b \\ -b & 0 \end{bmatrix}$$

is a general 2×2 skew–symmetric matrix.

We can now state a second application of the above criterion for non–singularity.

COROLLARY 2.5.4 Let B be an $n \times n$ skew–symmetric matrix. Then $A = I_n - B$ is non–singular.

Proof. Let $A = I_n - B$, where $B^t = -B$. By Theorem 2.5.10 it suffices to show that $AX = 0$ implies $X = 0$.

We have $(I_n - B)X = 0$, so $X = BX$. Hence $X^tX = X^tBX$.
Taking transposes of both sides gives

$$\begin{aligned} (X^tBX)^t &= (X^tX)^t \\ X^tB^t(X^t)^t &= X^t(X^t)^t \\ X^t(-B)X &= X^tX \\ -X^tBX &= X^tX = X^tBX. \end{aligned}$$

Hence $X^tX = -X^tX$ and $X^tX = 0$. But if $X = [x_1, \ldots, x_n]^t$, then $X^tX = x_1^2 + \ldots + x_n^2 = 0$ and hence $x_1 = 0, \ldots, x_n = 0$.

2.6 Least squares solution of equations

Suppose $AX = B$ represents a system of linear equations with real coefficients which may be inconsistent, because of the possibility of experimental errors in determining A or B. For example, the system

$$\begin{aligned} x &= 1 \\ y &= 2 \\ x + y &= 3.001 \end{aligned}$$

is inconsistent.

It can be proved that the associated system $A^tAX = A^tB$ is always consistent and that any solution of this system minimizes the sum $r_1^2 + \ldots + r_m^2$, where $r_1, \ldots, r_m$ (the *residuals*) are defined by

$$r_i = a_{i1}x_1 + \ldots + a_{in}x_n - b_i,$$

for $i = 1, \ldots, m$. The equations represented by $A^tAX = A^tB$ are called the *normal equations* corresponding to the system $AX = B$ and any solution of the system of normal equations is called a *least squares* solution of the original system.

EXAMPLE 2.6.1 Find a least squares solution of the above inconsistent system.

Solution. Here $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$, $X = \begin{bmatrix} x \\ y \end{bmatrix}$, $B = \begin{bmatrix} 1 \\ 2 \\ 3.001 \end{bmatrix}$.

Then $A^tA = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

Also $A^tB = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3.001 \end{bmatrix} = \begin{bmatrix} 4.001 \\ 5.001 \end{bmatrix}$.

So the normal equations are

$$\begin{aligned} 2x + y &= 4.001 \\ x + 2y &= 5.001 \end{aligned}$$

which have the unique solution

$$x = \frac{3.001}{3}, \quad y = \frac{6.001}{3}.$$

EXAMPLE 2.6.2 Points $(x_1, y_1), \ldots, (x_n, y_n)$ are experimentally determined and should lie on a line $y = mx + c$. Find a least squares solution to the problem.

Solution. The points have to satisfy

$$\begin{aligned} mx_1 + c &= y_1 \\ &\vdots \\ mx_n + c &= y_n, \end{aligned}$$

or $Ax = B$, where

$$A = \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix}, \; X = \begin{bmatrix} m \\ c \end{bmatrix}, \; B = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

The normal equations are given by $(A^tA)X = A^tB$. Here

$$A^tA = \begin{bmatrix} x_1 & \ldots & x_n \\ 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{bmatrix} = \begin{bmatrix} x_1^2 + \ldots + x_n^2 & x_1 + \ldots + x_n \\ x_1 + \ldots + x_n & n \end{bmatrix}$$

Also

$$A^tB = \begin{bmatrix} x_1 & \ldots & x_n \\ 1 & \ldots & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_1y_1 + \ldots + x_ny_n \\ y_1 + \ldots + y_n \end{bmatrix}.$$

It is not difficult to prove that

$$\Delta = \det(A^tA) = \sum_{1 \le i < j \le n} (x_i - x_j)^2,$$

which is positive unless $x_1 = \ldots = x_n$. Hence if not all of $x_1, \ldots, x_n$ are equal, A^tA is non–singular and the normal equations have a unique solution. This can be shown to be

$$m = \frac{1}{\Delta} \sum_{1 \le i < j \le n} (x_i - x_j)(y_i - y_j), \; c = \frac{1}{\Delta} \sum_{1 \le i < j \le n} (x_iy_j - x_jy_i)(x_i - x_j).$$

REMARK 2.6.1 The matrix A^tA is symmetric.

2.7 PROBLEMS

1. Let $A = \begin{bmatrix} 1 & 4 \\ -3 & 1 \end{bmatrix}$. Prove that A is non–singular, find A^{-1} and express A as a product of elementary row matrices.

   [Answer: $A^{-1} = \begin{bmatrix} \frac{1}{13} & -\frac{4}{13} \\ \frac{3}{13} & \frac{1}{13} \end{bmatrix}$,

   $A = E_{21}(-3)E_2(13)E_{12}(4)$ is one such decomposition.]

2. A square matrix $D = [d_{ij}]$ is called *diagonal* if $d_{ij} = 0$ for $i \neq j$. (That is the *off–diagonal* elements are zero.) Prove that pre–multiplication of a matrix A by a diagonal matrix D results in matrix DA whose rows are the rows of A multiplied by the respective diagonal elements of D. State and prove a similar result for post–multiplication by a diagonal matrix.

   Let diag $(a_1, \ldots, a_n)$ denote the diagonal matrix whose *diagonal* elements d_{ii} are $a_1, \ldots, a_n$, respectively. Show that

   $$\text{diag}\,(a_1, \ldots, a_n)\text{diag}\,(b_1, \ldots, b_n) = \text{diag}\,(a_1 b_1, \ldots, a_n b_n)$$

   and deduce that if $a_1 \ldots a_n \neq 0$, then diag $(a_1, \ldots, a_n)$ is non–singular and

   $$(\text{diag}\,(a_1, \ldots, a_n))^{-1} = \text{diag}\,(a_1^{-1}, \ldots, a_n^{-1}).$$

   Also prove that diag $(a_1, \ldots, a_n)$ is singular if $a_i = 0$ for some i.

3. Let $A = \begin{bmatrix} 0 & 0 & 2 \\ 1 & 2 & 6 \\ 3 & 7 & 9 \end{bmatrix}$. Prove that A is non–singular, find A^{-1} and express A as a product of elementary row matrices.

   [Answers: $A^{-1} = \begin{bmatrix} -12 & 7 & -2 \\ \frac{9}{2} & -3 & 1 \\ \frac{1}{2} & 0 & 0 \end{bmatrix}$,

   $A = E_{12}E_{31}(3)E_{23}E_3(2)E_{12}(2)E_{13}(24)E_{23}(-9)$ is one such decomposition.]

4. Find the rational number k for which the matrix $A = \begin{bmatrix} 1 & 2 & k \\ 3 & -1 & 1 \\ 5 & 3 & -5 \end{bmatrix}$ is singular. [Answer: $k = -3$.]

5. Prove that $A = \begin{bmatrix} 1 & 2 \\ -2 & -4 \end{bmatrix}$ is singular and find a non–singular matrix P such that PA has last row zero.

6. If $A = \begin{bmatrix} 1 & 4 \\ -3 & 1 \end{bmatrix}$, verify that $A^2 - 2A + 13I_2 = 0$ and deduce that $A^{-1} = -\frac{1}{13}(A - 2I_2)$.

7. Let $A = \begin{bmatrix} 1 & 1 & -1 \\ 0 & 0 & 1 \\ 2 & 1 & 2 \end{bmatrix}$.

   (i) Verify that $A^3 = 3A^2 - 3A + I_3$.

   (ii) Express A^4 in terms of A^2, A and I_3 and hence calculate A^4 explicitly.

   (iii) Use (i) to prove that A is non–singular and find A^{-1} explicitly.

   [Answers: (ii) $A^4 = 6A^2 - 8A + 3I_3 = \begin{bmatrix} -11 & -8 & -4 \\ 12 & 9 & 4 \\ 20 & 16 & 5 \end{bmatrix}$;

   (iii) $A^{-1} = A^2 - 3A + 3I_3 = \begin{bmatrix} -1 & -3 & 1 \\ 2 & 4 & -1 \\ 0 & 1 & 0 \end{bmatrix}$.]

8. (i) Let B be an $n \times n$ matrix such that $B^3 = 0$. If $A = I_n - B$, prove that A is non–singular and $A^{-1} = I_n + B + B^2$.
   Show that the system of linear equations $AX = b$ has the solution

$$X = b + Bb + B^2b.$$

   (ii) If $B = \begin{bmatrix} 0 & r & s \\ 0 & 0 & t \\ 0 & 0 & 0 \end{bmatrix}$, verify that $B^3 = 0$ and use (i) to determine $(I_3 - B)^{-1}$ explicitly.

[Answer: $\begin{bmatrix} 1 & r & s+rt \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$.]

9. Let A be $n \times n$.

   (i) If $A^2 = 0$, prove that A is singular.

   (ii) If $A^2 = A$ and $A \neq I_n$, prove that A is singular.

10. Use Question 7 to solve the system of equations

$$\begin{aligned} x + y - z &= a \\ z &= b \\ 2x + y + 2z &= c \end{aligned}$$

   where a, b, c are given rationals. Check your answer using the Gauss–Jordan algorithm.

   [Answer: $x = -a - 3b + c$, $y = 2a + 4b - c$, $z = b$.]

11. Determine explicitly the following products of 3×3 elementary row matrices.

   (i) $E_{12}E_{23}$ (ii) $E_1(5)E_{12}$ (iii) $E_{12}(3)E_{21}(-3)$ (iv) $(E_1(100))^{-1}$
   (v) E_{12}^{-1} (vi) $(E_{12}(7))^{-1}$ (vii) $(E_{12}(7)E_{31}(1))^{-1}$.

   [Answers: (i) $\begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$ (ii) $\begin{bmatrix} 0 & 5 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (iii) $\begin{bmatrix} -8 & 3 & 0 \\ -3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

   (iv) $\begin{bmatrix} 1/100 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (v) $\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (vi) $\begin{bmatrix} 1 & -7 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (vii) $\begin{bmatrix} 1 & -7 & 0 \\ 0 & 1 & 0 \\ -1 & 7 & 1 \end{bmatrix}$.]

12. Let A be the following product of 4×4 elementary row matrices:

$$A = E_3(2)E_{14}E_{42}(3).$$

   Find A and A^{-1} explicitly.

   [Answers: $A = \begin{bmatrix} 0 & 3 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$, $A^{-1} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 1 & -3 & 0 & 0 \end{bmatrix}$.]

13. Determine which of the following matrices over $\mathbb{Z}_2$ are non–singular and find the inverse, where possible.

$$\text{(a)}\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \quad \text{(b)}\begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}.$$ [Answer: $\text{(a)}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix}$.]

14. Determine which of the following matrices are non–singular and find the inverse, where possible.

$$\text{(a)}\begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 0 \\ 2 & 0 & 0 \end{bmatrix} \quad \text{(b)}\begin{bmatrix} 2 & 2 & 4 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{(c)}\begin{bmatrix} 4 & 6 & -3 \\ 0 & 0 & 7 \\ 0 & 0 & 5 \end{bmatrix}$$

$$\text{(d)}\begin{bmatrix} 2 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 7 \end{bmatrix} \quad \text{(e)}\begin{bmatrix} 1 & 2 & 4 & 6 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 \end{bmatrix} \quad \text{(f)}\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{bmatrix}.$$

[Answers: $\text{(a)}\begin{bmatrix} 0 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 1 & -1 & -1 \end{bmatrix}$ $\text{(b)}\begin{bmatrix} -1/2 & 2 & 1 \\ 0 & 0 & 1 \\ 1/2 & -1 & -1 \end{bmatrix}$ $\text{(d)}\begin{bmatrix} 1/2 & 0 & 0 \\ 0 & -1/5 & 0 \\ 0 & 0 & 1/7 \end{bmatrix}$

$\text{(e)}\begin{bmatrix} 1 & -2 & 0 & -3 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1/2 \end{bmatrix}$.]

15. Let A be a non–singular $n \times n$ matrix. Prove that A^t is non–singular and that $(A^t)^{-1} = (A^{-1})^t$.

16. Prove that $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ has no inverse if $ad - bc = 0$.

[Hint: Use the equation $A^2 - (a+d)A + (ad - bc)I_2 = 0$.]

17. Prove that the real matrix $A = \begin{bmatrix} 1 & a & b \\ -a & 1 & c \\ -b & -c & 1 \end{bmatrix}$ is non–singular by proving that A is row–equivalent to I_3.

18. If $P^{-1}AP = B$, prove that $P^{-1}A^nP = B^n$ for $n \geq 1$.

19. Let $A = \begin{bmatrix} 2/3 & 1/4 \\ 1/3 & 3/4 \end{bmatrix}$, $P = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}$. Verify that $P^{-1}AP = \begin{bmatrix} 5/12 & 0 \\ 0 & 1 \end{bmatrix}$ and deduce that

$$A^n = \frac{1}{7}\begin{bmatrix} 3 & 3 \\ 4 & 4 \end{bmatrix} + \frac{1}{7}\left(\frac{5}{12}\right)^n \begin{bmatrix} 4 & -3 \\ -4 & 3 \end{bmatrix}.$$

20. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a *Markov* matrix; that is a matrix whose elements are non–negative and satisfy $a+c = 1 = b+d$. Also let $P = \begin{bmatrix} b & 1 \\ c & -1 \end{bmatrix}$. Prove that if $A \neq I_2$ then

(i) P is non–singular and $P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & a+d-1 \end{bmatrix}$,

(ii) $A^n \rightarrow \dfrac{1}{b+c}\begin{bmatrix} b & b \\ c & c \end{bmatrix}$ as $n \rightarrow \infty$, if $A \neq \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$.

21. If $X = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$ and $Y = \begin{bmatrix} -1 \\ 3 \\ 4 \end{bmatrix}$, find XX^t, X^tX, YY^t, Y^tY.

[Answers: $\begin{bmatrix} 5 & 11 & 17 \\ 11 & 25 & 39 \\ 17 & 39 & 61 \end{bmatrix}$, $\begin{bmatrix} 35 & 44 \\ 44 & 56 \end{bmatrix}$, $\begin{bmatrix} 1 & -3 & -4 \\ -3 & 9 & 12 \\ -4 & 12 & 16 \end{bmatrix}$, 26.]

22. Prove that the system of linear equations

$$\begin{aligned} x + 2y &= 4 \\ x + y &= 5 \\ 3x + 5y &= 12 \end{aligned}$$

is inconsistent and find a least squares solution of the system.

[Answer: $x = 6$, $y = -7/6$.]

23. The points (0, 0), (1, 0), (2, −1), (3, 4), (4, 8) are required to lie on a parabola $y = a + bx + cx^2$. Find a least squares solution for a, b, c. Also prove that no parabola passes through these points.

[Answer: $a = \frac{1}{5}$, $b = -2$, $c = 1$.]

24. If A is a symmetric $n \times n$ real matrix and B is $n \times m$, prove that $B^t AB$ is a symmetric $m \times m$ matrix.

25. If A is $m \times n$ and B is $n \times m$, prove that AB is singular if $m > n$.

26. Let A and B be $n \times n$. If A or B is singular, prove that AB is also singular.

Chapter 3

SUBSPACES

3.1 Introduction

Throughout this chapter, we will be studying F^n, the set of n–dimensional column vectors with components from a field F. We continue our study of matrices by considering an important class of subsets of F^n called *subspaces*. These arise naturally for example, when we solve a system of m linear homogeneous equations in n unknowns.

We also study the concept of linear dependence of a family of vectors. This was introduced briefly in Chapter 2, Remark 2.5.4. Other topics discussed are the *row space, column space* and *null space* of a matrix over F, the *dimension* of a subspace, particular examples of the latter being the *rank* and *nullity* of a matrix.

3.2 Subspaces of F^n

DEFINITION 3.2.1 A subset S of F^n is called a subspace of F^n if

1. The zero vector belongs to S; (that is, $0 \in S$);
2. If $u \in S$ and $v \in S$, then $u + v \in S$; (S is said to be closed under vector addition);
3. If $u \in S$ and $t \in F$, then $tu \in S$; (S is said to be closed under scalar multiplication).

EXAMPLE 3.2.1 Let $A \in M_{m\times n}(F)$. Then the set of vectors $X \in F^n$ satisfying $AX = 0$ is a subspace of F^n called the *null space* of A and is denoted here by $N(A)$. (It is sometimes called the *solution space* of A.)

Proof. (1) $A0 = 0$, so $0 \in N(A)$; (2) If $X, Y \in N(A)$, then $AX = 0$ and $AY = 0$, so $A(X+Y) = AX + AY = 0 + 0 = 0$ and so $X + Y \in N(A)$; (3) If $X \in N(A)$ and $t \in F$, then $A(tX) = t(AX) = t0 = 0$, so $tX \in N(A)$.

For example, if $A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, then $N(A) = \{0\}$, the set consisting of just the zero vector. If $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$, then $N(A)$ is the set of all scalar multiples of $[-2, 1]^t$.

EXAMPLE 3.2.2 Let $X_1, \ldots, X_m \in F^n$. Then the set consisting of all linear combinations $x_1X_1 + \cdots + x_mX_m$, where $x_1, \ldots, x_m \in F$, is a subspace of F^n. This subspace is called the subspace *spanned* or *generated* by $X_1, \ldots, X_m$ and is denoted here by $\langle X_1, \ldots, X_m \rangle$. We also call $X_1, \ldots, X_m$ a spanning family for $S = \langle X_1, \ldots, X_m \rangle$.

Proof. (1) $0 = 0X_1 + \cdots + 0X_m$, so $0 \in \langle X_1, \ldots, X_m \rangle$; (2) If $X, Y \in \langle X_1, \ldots, X_m \rangle$, then $X = x_1X_1 + \cdots + x_mX_m$ and $Y = y_1X_1 + \cdots + y_mX_m$, so

$$\begin{aligned} X + Y &= (x_1X_1 + \cdots + x_mX_m) + (y_1X_1 + \cdots + y_mX_m) \\ &= (x_1 + y_1)X_1 + \cdots + (x_m + y_m)X_m \in \langle X_1, \ldots, X_m \rangle. \end{aligned}$$

(3) If $X \in \langle X_1, \ldots, X_m \rangle$ and $t \in F$, then

$$\begin{aligned} X &= x_1X_1 + \cdots + x_mX_m \\ tX &= t(x_1X_1 + \cdots + x_mX_m) \\ &= (tx_1)X_1 + \cdots + (tx_m)X_m \in \langle X_1, \ldots, X_m \rangle. \end{aligned}$$

For example, if $A \in M_{m \times n}(F)$, the subspace generated by the columns of A is an important subspace of F^m and is called the *column space* of A. The column space of A is denoted here by $C(A)$. Also the subspace generated by the rows of A is a subspace of F^n and is called the *row space* of A and is denoted by $R(A)$.

EXAMPLE 3.2.3 For example $F^n = \langle E_1, \ldots, E_n \rangle$, where $E_1, \ldots, E_n$ are the n–dimensional unit vectors. For if $X = [x_1, \ldots, x_n]^t \in F^n$, then $X = x_1E_1 + \cdots + x_nE_n$.

EXAMPLE 3.2.4 Find a spanning family for the subspace S of $\mathbb{R}^3$ defined by the equation $2x - 3y + 5z = 0$.

Solution. (S is in fact the null space of $[2,\ -3,\ 5]$, so S is indeed a subspace of $\mathbb{R}^3$.)

If $[x,\ y,\ z]^t \in S$, then $x = \frac{3}{2}y - \frac{5}{2}z$. Then

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} \frac{3}{2}y - \frac{5}{2}z \\ y \\ z \end{bmatrix} = y\begin{bmatrix} \frac{3}{2} \\ 1 \\ 0 \end{bmatrix} + z\begin{bmatrix} -\frac{5}{2} \\ 0 \\ 1 \end{bmatrix}$$

and conversely. Hence $[\frac{3}{2},\ 1,\ 0]^t$ and $[-\frac{5}{2},\ 0,\ 1]^t$ form a spanning family for S.

The following result is easy to prove:

LEMMA 3.2.1 Suppose each of $X_1, \ldots, X_r$ is a linear combination of $Y_1, \ldots, Y_s$. Then any linear combination of $X_1, \ldots, X_r$ is a linear combination of $Y_1, \ldots, Y_s$.

As a corollary we have

THEOREM 3.2.1 Subspaces $\langle X_1, \ldots, X_r\rangle$ and $\langle Y_1, \ldots, Y_s\rangle$ are equal if each of $X_1, \ldots, X_r$ is a linear combination of $Y_1, \ldots, Y_s$ and each of $Y_1, \ldots, Y_s$ is a linear combination of $X_1, \ldots, X_r$.

COROLLARY 3.2.1 $\langle X_1, \ldots, X_r,\ Z_1, \ldots, Z_t\rangle$ and $\langle X_1, \ldots, X_r\rangle$ are equal if each of $Z_1, \ldots, Z_t$ is a linear combination of $X_1, \ldots, X_r$.

EXAMPLE 3.2.5 If X and Y are vectors in $\mathbb{R}^n$, then

$$\langle X,\ Y\rangle = \langle X+Y,\ X-Y\rangle.$$

Solution. Each of $X+Y$ and $X-Y$ is a linear combination of X and Y. Also

$$X = \frac{1}{2}(X+Y) + \frac{1}{2}(X-Y) \quad \text{and} \quad Y = \frac{1}{2}(X+Y) - \frac{1}{2}(X-Y),$$

so each of X and Y is a linear combination of $X+Y$ and $X-Y$.

There is an important application of Theorem 3.2.1 to row equivalent matrices (see Definition 1.2.4):

THEOREM 3.2.2 If A is row equivalent to B, then $R(A) = R(B)$.

Proof. Suppose that B is obtained from A by a sequence of elementary row operations. Then it is easy to see that each row of B is a linear combination of the rows of A. But A can be obtained from B by a sequence of elementary operations, so each row of A is a linear combination of the rows of B. Hence by Theorem 3.2.1, $R(A) = R(B)$.

REMARK 3.2.1 If A is row equivalent to B, it is not always true that $C(A) = C(B)$.

For example, if $A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $B = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$, then B is in fact the reduced row–echelon form of A. However we see that

$$C(A) = \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle = \left\langle \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right\rangle$$

and similarly $C(B) = \left\langle \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\rangle$.

Consequently $C(A) \neq C(B)$, as $\begin{bmatrix} 1 \\ 1 \end{bmatrix} \in C(A)$ but $\begin{bmatrix} 1 \\ 1 \end{bmatrix} \notin C(B)$.

3.3 Linear dependence

We now recall the definition of linear dependence and independence of a family of vectors in F^n given in Chapter 2.

DEFINITION 3.3.1 Vectors $X_1, \ldots, X_m$ in F^n are said to be *linearly dependent* if there exist scalars $x_1, \ldots, x_m$, *not all zero*, such that

$$x_1 X_1 + \cdots + x_m X_m = 0.$$

In other words, $X_1, \ldots, X_m$ are linearly dependent if some X_i is expressible as a linear combination of the remaining vectors.

$X_1, \ldots, X_m$ are called *linearly independent* if they are not linearly dependent. Hence $X_1, \ldots, X_m$ are linearly independent if and only if the equation

$$x_1 X_1 + \cdots + x_m X_m = 0$$

has only the trivial solution $x_1 = 0, \ldots, x_m = 0$.

EXAMPLE 3.3.1 The following three vectors in $\mathbb{R}^3$

$$X_1 = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad X_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}, \quad X_3 = \begin{bmatrix} -1 \\ 7 \\ 12 \end{bmatrix}$$

are linearly dependent, as $2X_1 + 3X_2 + (-1)X_3 = 0$.

REMARK 3.3.1 If $X_1, \ldots, X_m$ are linearly independent and

$$x_1X_1 + \cdots + x_mX_m = y_1X_1 + \cdots + y_mX_m,$$

then $x_1 = y_1, \ldots, x_m = y_m$. For the equation can be rewritten as

$$(x_1 - y_1)X_1 + \cdots + (x_m - y_m)X_m = 0$$

and so $x_1 - y_1 = 0, \ldots, x_m - y_m = 0$.

THEOREM 3.3.1 A family of m vectors in F^n will be linearly dependent if $m > n$. Equivalently, any linearly independent family of m vectors in F^n must satisfy $m \leq n$.

Proof. The equation $x_1X_1+\cdots+x_mX_m = 0$ is equivalent to n homogeneous equations in m unknowns. By Theorem 1.5.1, such a system has a non–trivial solution if $m > n$.

The following theorem is an important generalization of the last result and is left as an exercise for the interested student:

THEOREM 3.3.2 A family of s vectors in $\langle X_1, \ldots, X_r\rangle$ will be linearly dependent if $s > r$. Equivalently, a linearly independent family of s vectors in $\langle X_1, \ldots, X_r\rangle$ must have $s \leq r$.

Here is a useful criterion for linear independence which is sometimes called the *left–to–right test*:

THEOREM 3.3.3 Vectors $X_1, \ldots, X_m$ in F^n are linearly independent if

(a) $X_1 \neq 0$;

(b) For each k with $1 < k \leq m$, X_k is not a linear combination of $X_1, \ldots, X_{k-1}$.

One application of this criterion is the following result:

THEOREM 3.3.4 Every subspace S of F^n can be represented in the form $S = \langle X_1, \ldots, X_m\rangle$, where $m \leq n$.

Proof. If $S = \{0\}$, there is nothing to prove – we take $X_1 = 0$ and $m = 1$.

So we assume S contains a non–zero vector X_1; then $\langle X_1\rangle \subseteq S$ as S is a subspace. If $S = \langle X_1\rangle$, we are finished. If not, S will contain a vector X_2, not a linear combination of X_1; then $\langle X_1, X_2\rangle \subseteq S$ as S is a subspace. If

$S = \langle X_1,\, X_2\rangle$, we are finished. If not, S will contain a vector X_3 which is not a linear combination of X_1 and X_2. This process must eventually stop, for at stage k we have constructed a family of k linearly independent vectors $X_1, \ldots, X_k$, all lying in F^n and hence $k \leq n$.

There is an important relationship between the columns of A and B, if A is row–equivalent to B.

THEOREM 3.3.5 Suppose that A is row equivalent to B and let $c_1, \ldots, c_r$ be distinct integers satisfying $1 \leq c_i \leq n$. Then

(a) Columns $A_{*c_1}, \ldots, A_{*c_r}$ of A are linearly dependent if and only if the corresponding columns of B are linearly dependent; indeed more is true:

$$x_1 A_{*c_1} + \cdots + x_r A_{*c_r} = 0 \Leftrightarrow x_1 B_{*c_1} + \cdots + x_r B_{*c_r} = 0.$$

(b) Columns $A_{*c_1}, \ldots, A_{*c_r}$ of A are linearly independent if and only if the corresponding columns of B are linearly independent.

(c) If $1 \leq c_{r+1} \leq n$ and c_{r+1} is distinct from $c_1, \ldots, c_r$, then

$$A_{*c_{r+1}} = z_1 A_{*c_1} + \cdots + z_r A_{*c_r} \Leftrightarrow B_{*c_{r+1}} = z_1 B_{*c_1} + \cdots + z_r B_{*c_r}.$$

Proof. First observe that if $Y = [y_1, \ldots, y_n]^t$ is an n–dimensional column vector and A is $m \times n$, then

$$AY = y_1 A_{*1} + \cdots + y_n A_{*n}.$$

Also $AY = 0 \Leftrightarrow BY = 0$, if B is row equivalent to A. Then (a) follows by taking $y_i = x_{c_j}$ if $i = c_j$ and $y_i = 0$ otherwise.

(b) is logically equivalent to (a), while (c) follows from (a) as

$$\begin{aligned}
A_{*c_{r+1}} &= z_1 A_{*c_1} + \cdots + z_r A_{*c_r} \\
&\Leftrightarrow z_1 A_{*c_1} + \cdots + z_r A_{*c_r} + (-1) A_{*c_{r+1}} = 0 \\
&\Leftrightarrow z_1 B_{*c_1} + \cdots + z_r B_{*c_r} + (-1) B_{*c_{r+1}} = 0 \\
&\Leftrightarrow B_{*c_{r+1}} = z_1 B_{*c_1} + \cdots + z_r B_{*c_r}.
\end{aligned}$$

EXAMPLE 3.3.2 The matrix

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

has reduced row–echelon form equal to

$$B = \begin{bmatrix} 1 & 0 & 2 & 0 & -1 \\ 0 & 1 & 3 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix}.$$

We notice that B_{*1}, B_{*2} and B_{*4} are linearly independent and hence so are A_{*1}, A_{*2} and A_{*4}. Also

$$\begin{aligned} B_{*3} &= 2B_{*1} + 3B_{*2} \\ B_{*5} &= (-1)B_{*1} + 2B_{*2} + 3B_{*4}, \end{aligned}$$

so consequently

$$\begin{aligned} A_{*3} &= 2A_{*1} + 3A_{*2} \\ A_{*5} &= (-1)A_{*1} + 2A_{*2} + 3A_{*4}. \end{aligned}$$

3.4 Basis of a subspace

We now come to the important concept of *basis* of a vector subspace.

DEFINITION 3.4.1 Vectors $X_1, \ldots, X_m$ belonging to a subspace S are said to form a basis of S if

(a) Every vector in S is a linear combination of $X_1, \ldots, X_m$;

(b) $X_1, \ldots, X_m$ are linearly independent.

Note that (a) is equivalent to the statement that $S = \langle X_1, \ldots, X_m \rangle$ as we automatically have $\langle X_1, \ldots, X_m \rangle \subseteq S$. Also, in view of Remark 3.3.1 above, (a) and (b) are equivalent to the statement that every vector in S is *uniquely* expressible as a linear combination of $X_1, \ldots, X_m$.

EXAMPLE 3.4.1 The unit vectors $E_1, \ldots, E_n$ form a basis for F^n.

REMARK 3.4.1 The subspace $\{0\}$, consisting of the zero vector alone, does not have a basis. For every vector in a linearly independent family must necessarily be non–zero. (For example, if $X_1 = 0$, then we have the non–trivial linear relation

$$1X_1 + 0X_2 + \cdots + 0X_m = 0$$

and $X_1, \ldots, X_m$ would be linearly dependent.)

However if we exclude this case, every other subspace of F^n has a basis:

THEOREM 3.4.1 A subspace of the form $\langle X_1, \ldots, X_m \rangle$, where at least one of $X_1, \ldots, X_m$ is non–zero, has a basis $X_{c_1}, \ldots, X_{c_r}$, where $1 \leq c_1 < \cdots < c_r \leq m$.

Proof. (The *left–to–right algorithm*). Let c_1 be the least index k for which X_k is non–zero. If $c_1 = m$ or if all the vectors X_k with $k > c_1$ are linear combinations of X_{c_1}, terminate the algorithm and let $r = 1$. Otherwise let c_2 be the least integer $k > c_1$ such that X_k is not a linear combination of X_{c_1}.

If $c_2 = m$ or if all the vectors X_k with $k > c_2$ are linear combinations of X_{c_1} and X_{c_2}, terminate the algorithm and let $r = 2$. Eventually the algorithm will terminate at the r–th stage, either because $c_r = m$, or because all vectors X_k with $k > c_r$ are linear combinations of $X_{c_1}, \ldots, X_{c_r}$.

Then it is clear by the construction of $X_{c_1}, \ldots, X_{c_r}$, using Corollary 3.2.1 that

(a) $\langle X_{c_1}, \ldots, X_{c_r} \rangle = \langle X_1, \ldots, X_m \rangle$;

(b) the vectors $X_{c_1}, \ldots, X_{c_r}$ are linearly independent by the left–to–right test.

Consequently $X_{c_1}, \ldots, X_{c_r}$ form a basis (called the *left–to–right basis*) for the subspace $\langle X_1, \ldots, X_m \rangle$.

EXAMPLE 3.4.2 Let X and Y be linearly independent vectors in $\mathbb{R}^n$. Then the subspace $\langle 0,\, 2X,\, X,\, -Y,\, X + Y \rangle$ has left–to–right basis consisting of $2X$, $-Y$.

A subspace S will in general have more than one basis. For example, any permutation of the vectors in a basis will yield another basis. Given one particular basis, one can determine all bases for S using a simple formula. This is left as one of the problems at the end of this chapter.

We settle for the following important fact about bases:

THEOREM 3.4.2 Any two bases for a subspace S must contain the same number of elements.

Proof. For if $X_1, \ldots, X_r$ and $Y_1, \ldots, Y_s$ are bases for S, then $Y_1, \ldots, Y_s$ form a linearly independent family in $S = \langle X_1, \ldots, X_r \rangle$ and hence $s \leq r$ by Theorem 3.3.2. Similarly $r \leq s$ and hence $r = s$.

DEFINITION 3.4.2 This number is called the *dimension* of S and is written $\dim S$. Naturally we define $\dim\{0\} = 0$.

It follows from Theorem 3.3.1 that for any subspace S of F^n, we must have $\dim S \le n$.

EXAMPLE 3.4.3 If $E_1, \ldots, E_n$ denote the n–dimensional unit vectors in F^n, then $\dim \langle E_1, \ldots, E_i \rangle = i$ for $1 \le i \le n$.

The following result gives a useful way of exhibiting a basis.

THEOREM 3.4.3 A linearly independent family of m vectors in a subspace S, with $\dim S = m$, must be a basis for S.

Proof. Let $X_1, \ldots, X_m$ be a linearly independent family of vectors in a subspace S, where $\dim S = m$. We have to show that every vector $X \in S$ is expressible as a linear combination of $X_1, \ldots, X_m$. We consider the following family of vectors in S: $X_1, \ldots, X_m, X$. This family contains $m+1$ elements and is consequently linearly dependent by Theorem 3.3.2. Hence we have

$$x_1X_1 + \cdots + x_mX_m + x_{m+1}X = 0, \tag{3.1}$$

where not all of $x_1, \ldots, x_{m+1}$ are zero. Now if $x_{m+1} = 0$, we would have

$$x_1X_1 + \cdots + x_mX_m = 0,$$

with not all of $x_1, \ldots, x_m$ zero, contradicting the assumption that $X_1 \ldots, X_m$ are linearly independent. Hence $x_{m+1} \ne 0$ and we can use equation 3.1 to express X as a linear combination of $X_1, \ldots, X_m$:

$$X = \frac{-x_1}{x_{m+1}}X_1 + \cdots + \frac{-x_m}{x_{m+1}}X_m.$$

3.5 Rank and nullity of a matrix

We can now define three important integers associated with a matrix.

DEFINITION 3.5.1 Let $A \in M_{m\times n}(F)$. Then

(a) column rank $A = \dim C(A)$;

(b) row rank $A = \dim R(A)$;

(c) nullity $A = \dim N(A)$.

We will now see that the reduced row–echelon form B of a matrix A allows us to exhibit bases for the row space, column space and null space of A. Moreover, an examination of the number of elements in each of these bases will immediately result in the following theorem:

THEOREM 3.5.1 Let $A \in M_{m\times n}(F)$. Then

(a) column rank $A =$ row rank A;

(b) column rank $A +$ nullity $A = n$.

Finding a basis for $R(A)$: The r non–zero rows of B form a basis for $R(A)$ and hence row rank $A = r$.

For we have seen earlier that $R(A) = R(B)$. Also

$$\begin{array}{rcl} R(B) & = & \langle B_{1*}, \ldots, B_{m*}\rangle \\ & = & \langle B_{1*}, \ldots, B_{r*}, 0 \ldots, 0\rangle \\ & = & \langle B_{1*}, \ldots, B_{r*}\rangle. \end{array}$$

The linear independence of the non–zero rows of B is proved as follows: Let the leading entries of rows $1, \ldots, r$ of B occur in columns $c_1, \ldots, c_r$. Suppose that

$$x_1 B_{1*} + \cdots + x_r B_{r*} = 0.$$

Then equating components $c_1, \ldots, c_r$ of both sides of the last equation, gives $x_1 = 0, \ldots, x_r = 0$, in view of the fact that B is in reduced row– echelon form.

Finding a basis for $C(A)$: The r columns $A_{*c_1}, \ldots, A_{*c_r}$ form a basis for $C(A)$ and hence column rank $A = r$. For it is clear that columns $c_1, \ldots, c_r$ of B form the left–to–right basis for $C(B)$ and consequently from parts (b) and (c) of Theorem 3.3.5, it follows that columns $c_1, \ldots, c_r$ of A form the left–to–right basis for $C(A)$.

Finding a basis for $N(A)$: For notational simplicity, let us suppose that $c_1 = 1, \ldots, c_r = r$. Then B has the form

$$B = \begin{bmatrix} 1 & 0 & \cdots & 0 & b_{1r+1} & \cdots & b_{1n} \\ 0 & 1 & \cdots & 0 & b_{2r+1} & \cdots & b_{2n} \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 1 & b_{rr+1} & \cdots & b_{rn} \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}.$$

Then $N(B)$ and hence $N(A)$ are determined by the equations

$$\begin{aligned} x_1 &= (-b_{1r+1})x_{r+1} + \cdots + (-b_{1n})x_n \\ &\vdots \\ x_r &= (-b_{rr+1})x_{r+1} + \cdots + (-b_{rn})x_n, \end{aligned}$$

where $x_{r+1}, \ldots, x_n$ are arbitrary elements of F. Hence the general vector X in $N(A)$ is given by

$$\begin{aligned} \begin{bmatrix} x_1 \\ \vdots \\ x_r \\ x_{r+1} \\ \vdots \\ x_n \end{bmatrix} &= x_{r+1} \begin{bmatrix} -b_{1r+1} \\ \vdots \\ -b_{rr+1} \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \cdots + x_n \begin{bmatrix} -b_n \\ \vdots \\ -b_{rn} \\ 0 \\ \vdots \\ 1 \end{bmatrix} \qquad (3.2) \\ &= x_{r+1}X_1 + \cdots + x_n X_{n-r}. \end{aligned}$$

Hence $N(A)$ is spanned by $X_1, \ldots, X_{n-r}$, as $x_{r+1}, \ldots, x_n$ are arbitrary. Also $X_1, \ldots, X_{n-r}$ are linearly independent. For equating the right hand side of equation 3.2 to 0 and then equating components $r+1, \ldots, n$ of both sides of the resulting equation, gives $x_{r+1} = 0, \ldots, x_n = 0$.

Consequently $X_1, \ldots, X_{n-r}$ form a basis for $N(A)$.

Theorem 3.5.1 now follows. For we have

$$\begin{aligned} \text{row rank}\, A &= \dim R(A) = r \\ \text{column rank}\, A &= \dim C(A) = r. \end{aligned}$$

Hence

$$\text{row rank}\, A = \text{column rank}\, A.$$

Also

$$\text{column rank}\, A + \text{nullity}\, A = r + \dim N(A) = r + (n - r) = n.$$

DEFINITION 3.5.2 The common value of column rank A and row rank A is called the *rank* of A and is denoted by rank A.

EXAMPLE 3.5.1 Given that the reduced row–echelon form of

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

equal to

$$B = \begin{bmatrix} 1 & 0 & 2 & 0 & -1 \\ 0 & 1 & 3 & 0 & 2 \\ 0 & 0 & 0 & 1 & 3 \end{bmatrix},$$

find bases for $R(A)$, $C(A)$ and $N(A)$.

Solution. $[1, 0, 2, 0, -1]$, $[0, 1, 3, 0, 2]$ and $[0, 0, 0, 1, 3]$ form a basis for $R(A)$. Also

$$A_{*1} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \ A_{*2} = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \ A_{*4} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

form a basis for $C(A)$.

Finally $N(A)$ is given by

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2x_3 + x_5 \\ -3x_3 - 2x_5 \\ x_3 \\ -3x_5 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 1 \\ -2 \\ 0 \\ -3 \\ 1 \end{bmatrix} = x_3 X_1 + x_5 X_2,$$

where x_3 and x_5 are arbitrary. Hence X_1 and X_2 form a basis for $N(A)$.

Here $\text{rank}\, A = 3$ and $\text{nullity}\, A = 2$.

EXAMPLE 3.5.2 Let $A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$. Then $B = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$ is the reduced row–echelon form of A.

Hence $[1, 2]$ is a basis for $R(A)$ and $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ is a basis for $C(A)$. Also $N(A)$ is given by the equation $x_1 = -2x_2$, where x_2 is arbitrary. Then

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -2x_2 \\ x_2 \end{bmatrix} = x_2 \begin{bmatrix} -2 \\ 1 \end{bmatrix}$$

and hence $\begin{bmatrix} -2 \\ 1 \end{bmatrix}$ is a basis for $N(A)$.

Here $\text{rank}\, A = 1$ and $\text{nullity}\, A = 1$.

EXAMPLE 3.5.3 Let $A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Then $B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ is the reduced row–echelon form of A.

Hence $[1, 0]$, $[0, 1]$ form a basis for $R(A)$ while $[1, 3]$, $[2, 4]$ form a basis for $C(A)$. Also $N(A) = \{0\}$.

Here $\text{rank}\, A = 2$ and $\text{nullity}\, A = 0$.

We conclude this introduction to vector spaces with a result of great theoretical importance.

THEOREM 3.5.2 Every linearly independent family of vectors in a subspace S can be extended to a basis of S.

Proof. Suppose S has basis $X_1, \ldots, X_m$ and that $Y_1, \ldots, Y_r$ is a linearly independent family of vectors in S. Then

$$S = \langle X_1, \ldots, X_m \rangle = \langle Y_1, \ldots, Y_r,\, X_1, \ldots, X_m \rangle,$$

as each of $Y_1, \ldots, Y_r$ is a linear combination of $X_1, \ldots, X_m$.

Then applying the left–to–right algorithm to the second spanning family for S will yield a basis for S which includes $Y_1, \ldots, Y_r$.

3.6 PROBLEMS

1. Which of the following subsets of $\mathbb{R}^2$ are subspaces?

   (a) $[x,\, y]$ satisfying $x = 2y$;
   (b) $[x,\, y]$ satisfying $x = 2y$ and $2x = y$;
   (c) $[x,\, y]$ satisfying $x = 2y + 1$;
   (d) $[x,\, y]$ satisfying $xy = 0$;
   (e) $[x,\, y]$ satisfying $x \geq 0$ and $y \geq 0$.

   [Answer: (a) and (b).]

2. If $X,\, Y,\, Z$ are vectors in $\mathbb{R}^n$, prove that

$$\langle X,\, Y,\, Z \rangle = \langle X + Y,\, X + Z,\, Y + Z \rangle.$$

3. Determine if $X_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 2 \end{bmatrix}$, $X_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 2 \end{bmatrix}$ and $X_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 3 \end{bmatrix}$ are linearly independent in $\mathbb{R}^4$.

4. For which real numbers λ are the following vectors linearly independent in $\mathbb{R}^3$?

$$X_1 = \begin{bmatrix} \lambda \\ -1 \\ -1 \end{bmatrix}, \quad X_2 = \begin{bmatrix} -1 \\ \lambda \\ -1 \end{bmatrix}, \quad X_3 = \begin{bmatrix} -1 \\ -1 \\ \lambda \end{bmatrix}.$$

5. Find bases for the row, column and null spaces of the following matrix over $\mathbb{Q}$:
$$A = \begin{bmatrix} 1 & 1 & 2 & 0 & 1 \\ 2 & 2 & 5 & 0 & 3 \\ 0 & 0 & 0 & 1 & 3 \\ 8 & 11 & 19 & 0 & 11 \end{bmatrix}.$$

6. Find bases for the row, column and null spaces of the following matrix over $\mathbb{Z}_2$:
$$A = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}.$$

7. Find bases for the row, column and null spaces of the following matrix over $\mathbb{Z}_5$:
$$A = \begin{bmatrix} 1 & 1 & 2 & 0 & 1 & 3 \\ 2 & 1 & 4 & 0 & 3 & 2 \\ 0 & 0 & 0 & 1 & 3 & 0 \\ 3 & 0 & 2 & 4 & 3 & 2 \end{bmatrix}.$$

8. Find bases for the row, column and null spaces of the matrix A defined in section 1.6, Problem 17. (Note: In this question, F is a field of four elements.)

9. If $X_1, \ldots, X_m$ form a basis for a subspace S, prove that
$$X_1,\, X_1 + X_2, \ldots, X_1 + \cdots + X_m$$
also form a basis for S.

10. Let $A = \begin{bmatrix} a & b & c \\ 1 & 1 & 1 \end{bmatrix}$. Classify a, b, c such that (a) $\operatorname{rank} A = 1$; (b) $\operatorname{rank} A = 2$.

    [Answer: (a) $a = b = c$; (b) at least two of a, b, c are distinct.]

11. Let S be a subspace of F^n with $\dim S = m$. If $X_1, \ldots, X_m$ are vectors in S with the property that $S = \langle X_1, \ldots, X_m \rangle$, prove that $X_1 \ldots, X_m$ form a basis for S.

12. Find a basis for the subspace S of $\mathbb{R}^3$ defined by the equation
$$x + 2y + 3z = 0.$$

Verify that $Y_1 = [-1,\ -1,\ 1]^t \in S$ and find a basis for S which includes Y_1.

13. Let $X_1, \ldots, X_m$ be vectors in F^n. If $X_i = X_j$, where $i < j$, prove that $X_1, \ldots X_m$ are linearly dependent.

14. Let $X_1, \ldots, X_{m+1}$ be vectors in F^n. Prove that

$$\dim \langle X_1, \ldots, X_{m+1} \rangle = \dim \langle X_1, \ldots, X_m \rangle$$

if X_{m+1} is a linear combination of $X_1, \ldots, X_m$, but

$$\dim \langle X_1, \ldots, X_{m+1} \rangle = \dim \langle X_1, \ldots, X_m \rangle + 1$$

if X_{m+1} is not a linear combination of $X_1, \ldots, X_m$.

Deduce that the system of linear equations $AX = B$ is consistent, if and only if

$$\text{rank}\,[A|B] = \text{rank}\,A.$$

15. Let $a_1, \ldots, a_n$ be elements of F, not all zero. Prove that the set of vectors $[x_1, \ldots, x_n]^t$ where $x_1, \ldots, x_n$ satisfy

$$a_1 x_1 + \cdots + a_n x_n = 0$$

is a subspace of F^n with dimension equal to $n - 1$.

16. Prove Lemma 3.2.1, Theorem 3.2.1, Corollary 3.2.1 and Theorem 3.3.2.

17. Let R and S be subspaces of F^n, with $R \subseteq S$. Prove that

$$\dim R \leq \dim S$$

and that equality implies $R = S$. (This is a very useful way of proving equality of subspaces.)

18. Let R and S be subspaces of F^n. If $R \cup S$ is a subspace of F^n, prove that $R \subseteq S$ or $S \subseteq R$.

19. Let $X_1, \ldots, X_r$ be a basis for a subspace S. Prove that all bases for S are given by the family $Y_1, \ldots, Y_r$, where

$$Y_i = \sum_{j=1}^{r} a_{ij} X_j,$$

and where $A = [a_{ij}] \in M_{r \times r}(F)$ is a non–singular matrix.

Chapter 4

DETERMINANTS

DEFINITION 4.0.1 If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$, we define the *determinant* of A, (also denoted by $\det A$,) to be the scalar

$$\det A = a_{11}a_{22} - a_{12}a_{21}.$$

The notation $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ is also used for the determinant of A.

If A is a real matrix, there is a geometrical interpretation of $\det A$. If $P = (x_1, y_1)$ and $Q = (x_2, y_2)$ are points in the plane, forming a triangle with the origin $O = (0, 0)$, then apart from sign, $\frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$ is the area of the triangle OPQ. For, using polar coordinates, let $x_1 = r_1 \cos\theta_1$ and $y_1 = r_1 \sin\theta_1$, where $r_1 = OP$ and θ_1 is the angle made by the ray $\overrightarrow{OP}$ with the positive x–axis. Then triangle OPQ has area $\frac{1}{2}OP \cdot OQ \sin\alpha$, where $\alpha = \angle POQ$. If triangle OPQ has anti–clockwise orientation, then the ray $\overrightarrow{OQ}$ makes angle $\theta_2 = \theta_1 + \alpha$ with the positive x–axis. (See Figure 4.1.)

Also $x_2 = r_2 \cos\theta_2$ and $y_2 = r_2 \sin\theta_2$. Hence

$$\begin{aligned}
\text{Area}\, OPQ &= \frac{1}{2}OP \cdot OQ \sin\alpha \\
&= \frac{1}{2}OP \cdot OQ \sin(\theta_2 - \theta_1) \\
&= \frac{1}{2}OP \cdot OQ(\sin\theta_2 \cos\theta_1 - \cos\theta_2 \sin\theta_1) \\
&= \frac{1}{2}(OQ \sin\theta_2 \cdot OP \cos\theta_1 - OQ \cos\theta_2 \cdot OP \sin\theta_1)
\end{aligned}$$

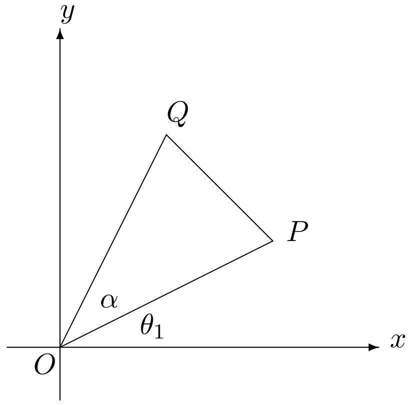

Figure 4.1: Area of triangle OPQ.

$$
\begin{aligned}
&= \frac{1}{2}(y_2x_1 - x_2y_1) \\
&= \frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}.
\end{aligned}
$$

Similarly, if triangle OPQ has clockwise orientation, then its area equals $-\frac{1}{2}\begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix}$.

For a general triangle $P_1P_2P_3$, with $P_i = (x_i, y_i)$, $i = 1, 2, 3$, we can take P_1 as the origin. Then the above formula gives

$$
\frac{1}{2}\begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix} \quad \text{or} \quad -\frac{1}{2}\begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1 \end{vmatrix},
$$

according as vertices $P_1P_2P_3$ are anti–clockwise or clockwise oriented.

We now give a recursive definition of the determinant of an $n \times n$ matrix $A = [a_{ij}]$, $n \geq 3$.

DEFINITION 4.0.2 (Minor) Let $M_{ij}(A)$ (or simply M_{ij} if there is no ambiguity) denote the determinant of the $(n-1) \times (n-1)$ submatrix of A formed by deleting the i–th row and j–th column of A. ($M_{ij}(A)$ is called the (i, j) *minor* of A.)

Assume that the determinant function has been defined for matrices of size $(n-1) \times (n-1)$. Then $\det A$ is defined by the so–called *first–row Laplace* *expansion*:

$$\begin{aligned}\det A &= a_{11}M_{11}(A) - a_{12}M_{12}(A) + \ldots + (-1)^{1+n}M_{1n}(A)\\ &= \sum_{j=1}^{n}(-1)^{1+j}a_{1j}M_{1j}(A).\end{aligned}$$

For example, if $A = [a_{ij}]$ is a 3×3 matrix, the Laplace expansion gives

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

The recursive definition also works for 2×2 determinants, if we define the determinant of a 1×1 matrix $[t]$ to be the scalar t:

$$\det A = a_{11}M_{11}(A) - a_{12}M_{12}(A) = a_{11}a_{22} - a_{12}a_{21}.$$

EXAMPLE 4.0.1 If $P_1P_2P_3$ is a triangle with $P_i = (x_i, y_i)$, $i = 1, 2, 3$, then the area of triangle $P_1P_2P_3$ is

$$\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1\end{vmatrix} \quad \text{or} \quad -\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1\end{vmatrix},$$

according as the orientation of $P_1P_2P_3$ is anti–clockwise or clockwise.

For from the definition of 3×3 determinants, we have

$$\begin{aligned}\frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1\end{vmatrix} &= \frac{1}{2}\left(x_1\begin{vmatrix} y_2 & 1 \\ y_3 & 1\end{vmatrix} - y_1\begin{vmatrix} x_2 & 1 \\ x_3 & 1\end{vmatrix} + \begin{vmatrix} x_2 & y_2 \\ x_3 & y_3\end{vmatrix}\right)\\ &= \frac{1}{2}\begin{vmatrix} x_2 - x_1 & y_2 - y_1 \\ x_3 - x_1 & y_3 - y_1\end{vmatrix}.\end{aligned}$$

One property of determinants that follows immediately from the definition is the following:

THEOREM 4.0.1 If a row of a matrix is zero, then the value of the determinant is zero.

(The corresponding result for columns also holds, but here a simple proof by induction is needed.)

One of the simplest determinants to evaluate is that of a lower triangular matrix.

THEOREM 4.0.2 Let $A = [a_{ij}]$, where $a_{ij} = 0$ if $i < j$. Then

$$\det \mathrm{A} = a_{11}a_{22}\dots a_{nn}. \tag{4.1}$$

An important special case is when A is a diagonal matrix.

If A =diag $(a_1, \dots, a_n)$ then $\det A = a_1 \dots a_n$. In particular, for a scalar matrix tI_n, we have $\det(tI_n) = t^n$.

Proof. Use induction on the size n of the matrix.

The result is true for $n = 2$. Now let $n > 2$ and assume the result true for matrices of size $n - 1$. If A is $n \times n$, then expanding $\det A$ along row 1 gives

$$\begin{aligned}\det A &= a_{11}\begin{vmatrix} a_{22} & 0 & \dots & 0 \\ a_{32} & a_{33} & \dots & 0 \\ \vdots & & & \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \\ &= a_{11}(a_{22}\dots a_{nn})\end{aligned}$$

by the induction hypothesis.

If A is upper triangular, equation 4.1 remains true and the proof is again an exercise in induction, with the slight difference that the column version of theorem 4.0.1 is needed.

REMARK 4.0.1 It can be shown that the expanded form of the determinant of an $n \times n$ matrix A consists of $n!$ signed products $\pm a_{1i_1}a_{2i_2}\dots a_{ni_n}$, where $(i_1, i_2, \dots, i_n)$ is a permutation of $(1, 2, \dots, n)$, the sign being 1 or -1, according as the number of *inversions* of $(i_1, i_2, \dots, i_n)$ is even or odd. An inversion occurs when $i_r > i_s$ but $r < s$. (The proof is not easy and is omitted.)

The definition of the determinant of an $n \times n$ matrix was given in terms of the first–row expansion. The next theorem says that we can expand the determinant along any row or column. (The proof is not easy and is omitted.)

THEOREM 4.0.3

$$\det A = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} M_{ij}(A)$$

for $i = 1, \ldots, n$ (the so–called i–th row expansion) and

$$\det A = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} M_{ij}(A)$$

for $j = 1, \ldots, n$ (the so–called j–th column expansion).

REMARK 4.0.2 The expression $(-1)^{i+j}$ obeys the chess–board pattern of signs:

$$\begin{bmatrix} + & - & + & \ldots \\ - & + & - & \ldots \\ + & - & + & \ldots \\ \vdots & & & \end{bmatrix}.$$

The following theorems can be proved by straightforward inductions on the size of the matrix:

THEOREM 4.0.4 A matrix and its transpose have equal determinants; that is

$$\det A^t = \det A.$$

THEOREM 4.0.5 If two rows of a matrix are equal, the determinant is zero. Similarly for columns.

THEOREM 4.0.6 If two rows of a matrix are interchanged, the determinant changes sign.

EXAMPLE 4.0.2 If $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ are distinct points, then the line through P_1 and P_2 has equation

$$\begin{vmatrix} x & y & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0.$$

For, expanding the determinant along row 1, the equation becomes

$$ax + by + c = 0,$$

where

$$a = \begin{vmatrix} y_1 & 1 \\ y_2 & 1 \end{vmatrix} = y_1 - y_2 \text{ and } b = -\begin{vmatrix} x_1 & 1 \\ x_2 & 1 \end{vmatrix} = x_2 - x_1.$$

This represents a line, as not both a and b can be zero. Also this line passes through P_i, $i = 1, 2$. For the determinant has its first and i–th rows equal if $x = x_i$ and $y = y_i$ and is consequently zero.

There is a corresponding formula in three–dimensional geometry. If P_1, P_2, P_3 are non–collinear points in three–dimensional space, with $P_i = (x_i, y_i, z_i)$, $i = 1, 2, 3$, then the equation

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0$$

represents the plane through P_1, P_2, P_3. For, expanding the determinant along row 1, the equation becomes $ax + by + cz + d = 0$, where

$$a = \begin{vmatrix} y_1 & z_1 & 1 \\ y_2 & z_2 & 1 \\ y_3 & z_3 & 1 \end{vmatrix}, \; b = -\begin{vmatrix} x_1 & z_1 & 1 \\ x_2 & z_2 & 1 \\ x_3 & z_3 & 1 \end{vmatrix}, \; c = \begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}.$$

As we shall see in chapter 6, this represents a plane if at least one of a, b, c is non–zero. However, apart from sign and a factor $\frac{1}{2}$, the determinant expressions for a, b, c give the values of the areas of projections of triangle $P_1P_2P_3$ on the (y, z), (x, z) and (x, y) planes, respectively. Geometrically, it is then clear that at least one of a, b, c is non–zero. It is also possible to give an algebraic proof of this fact.

Finally, the plane passes through P_i, $i = 1, 2, 3$ as the determinant has its first and i–th rows equal if $x = x_i$, $y = y_i$, $z = z_i$ and is consequently zero. We now work towards proving that a matrix is non–singular if its determinant is non–zero.

DEFINITION 4.0.3 (Cofactor) The (i, j) cofactor of A, denoted by $C_{ij}(A)$ (or C_{ij} if there is no ambiguity) is defined by

$$C_{ij}(A) = (-1)^{i+j} M_{ij}(A).$$

REMARK 4.0.3 It is important to notice that $C_{ij}(A)$, like $M_{ij}(A)$, does not depend on a_{ij}. Use will be made of this observation presently.

In terms of the cofactor notation, Theorem 4.0.3 takes the form

THEOREM 4.0.7

$$\det A = \sum_{j=1}^{n} a_{ij} C_{ij}(A)$$

for $i = 1, \ldots, n$ and

$$\det A = \sum_{i=1}^{n} a_{ij} C_{ij}(A)$$

for $j = 1, \ldots, n$.

Another result involving cofactors is

THEOREM 4.0.8 Let A be an $n \times n$ matrix. Then

$$(a) \quad \sum_{j=1}^{n} a_{ij} C_{kj}(A) = 0 \quad \text{if } i \neq k.$$

Also

$$(b) \quad \sum_{i=1}^{n} a_{ij} C_{ik}(A) = 0 \quad \text{if } j \neq k.$$

Proof.

If A is $n \times n$ and $i \neq k$, let B be the matrix obtained from A by replacing row k by row i. Then $\det B = 0$ as B has two identical rows.

Now expand $\det B$ along row k. We get

$$\begin{aligned} 0 = \det B &= \sum_{j=1}^{n} b_{kj} C_{kj}(B) \\ &= \sum_{j=1}^{n} a_{ij} C_{kj}(A), \end{aligned}$$

in view of Remark 4.0.3.

DEFINITION 4.0.4 (Adjoint) If $A = [a_{ij}]$ is an $n \times n$ matrix, the *adjoint* of A, denoted by $\text{adj}\, A$, is the transpose of the matrix of cofactors. Hence

$$\text{adj}\, A = \begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}.$$

Theorems 4.0.7 and 4.0.8 may be combined to give

THEOREM 4.0.9 Let A be an $n \times n$ matrix. Then

$$A(\text{adj}\, A) = (\det A)I_n = (\text{adj}\, A)A.$$

Proof.

$$\begin{aligned} (A\, \text{adj}\, A)_{ik} &= \sum_{j=1}^{n} a_{ij} (\text{adj}\, A)_{jk} \\ &= \sum_{j=1}^{n} a_{ij} C_{kj}(A) \\ &= \delta_{ik} \det A \\ &= ((\det A)I_n)_{ik}. \end{aligned}$$

Hence $A(\text{adj}\, A) = (\det A)I_n$. The other equation is proved similarly.

COROLLARY 4.0.1 (Formula for the inverse) If $\det A \neq 0$, then A is non–singular and

$$A^{-1} = \frac{1}{\det A} \text{adj}\, A.$$

EXAMPLE 4.0.3 The matrix

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

is non–singular. For

$$\begin{aligned} \det A &= \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} - 2 \begin{vmatrix} 4 & 6 \\ 8 & 9 \end{vmatrix} + 3 \begin{vmatrix} 4 & 5 \\ 8 & 8 \end{vmatrix} \\ &= -3 + 24 - 24 \\ &= -3 \neq 0. \end{aligned}$$

Also

$$
\begin{aligned}
A^{-1} &= \frac{1}{-3}\begin{bmatrix} C_{11} & C_{21} & C_{31} \\ C_{12} & C_{22} & C_{32} \\ C_{13} & C_{23} & C_{33} \end{bmatrix} \\
&= -\frac{1}{3}\begin{bmatrix} \begin{vmatrix} 5 & 6 \\ 8 & 9 \end{vmatrix} & -\begin{vmatrix} 2 & 3 \\ 8 & 9 \end{vmatrix} & \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} \\ -\begin{vmatrix} 4 & 6 \\ 8 & 9 \end{vmatrix} & \begin{vmatrix} 1 & 3 \\ 8 & 9 \end{vmatrix} & -\begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} \\ \begin{vmatrix} 4 & 5 \\ 8 & 8 \end{vmatrix} & -\begin{vmatrix} 1 & 2 \\ 8 & 8 \end{vmatrix} & \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} \end{bmatrix} \\
&= -\frac{1}{3}\begin{bmatrix} -3 & 6 & -3 \\ 12 & -15 & 6 \\ -8 & 8 & -3 \end{bmatrix}.
\end{aligned}
$$

The following theorem is useful for simplifying and numerically evaluating a determinant. Proofs are obtained by expanding along the corresponding row or column.

THEOREM 4.0.10 The determinant is a linear function of each row and column. For example

$$
(a)\quad \begin{vmatrix} a_{11}+a'_{11} & a_{12}+a'_{12} & a_{13}+a'_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} a'_{11} & a'_{12} & a'_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}
$$

$$
(b)\quad \begin{vmatrix} ta_{11} & ta_{12} & ta_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = t\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.
$$

COROLLARY 4.0.2 If a multiple of a row is added to *another* row, the value of the determinant is unchanged. Similarly for columns.

Proof. We illustrate with a 3×3 example, but the proof is really quite general.

$$
\begin{vmatrix} a_{11}+ta_{21} & a_{12}+ta_{22} & a_{13}+ta_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + \begin{vmatrix} ta_{21} & ta_{22} & ta_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}
$$

$$
= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + t\begin{vmatrix} a_{21} & a_{22} & a_{23} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} + t \times 0
$$

$$
= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.
$$

To evaluate a determinant numerically, it is advisable to reduce the matrix to row–echelon form, recording any sign changes caused by row interchanges, together with any factors taken out of a row, as in the following examples.

EXAMPLE 4.0.4 Evaluate the determinant

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

Solution. Using row operations $R_2 \to R_2 - 4R_1$ and $R_3 \to R_3 - 8R_1$ and then expanding along the first column, gives

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 8 & 8 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 & 3 \\ 0 & -3 & -6 \\ 0 & -8 & -15 \end{vmatrix} = \begin{vmatrix} -3 & -6 \\ -8 & -15 \end{vmatrix}$$
$$= -3\begin{vmatrix} 1 & 2 \\ -8 & -15 \end{vmatrix} = -3\begin{vmatrix} 1 & 2 \\ 0 & 1 \end{vmatrix} = -3.$$

EXAMPLE 4.0.5 Evaluate the determinant

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

Solution.

$$\begin{vmatrix} 1 & 1 & 2 & 1 \\ 3 & 1 & 4 & 5 \\ 7 & 6 & 1 & 2 \\ 1 & 1 & 3 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 1 & 2 & 1 \\ 0 & -2 & -2 & 2 \\ 0 & -1 & -13 & -5 \\ 0 & 0 & 1 & 3 \end{vmatrix}$$
$$= -2\begin{vmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & -1 & -13 & -5 \\ 0 & 0 & 1 & 3 \end{vmatrix}$$
$$= -2\begin{vmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & -12 & -6 \\ 0 & 0 & 1 & 3 \end{vmatrix}$$
$$= 2\begin{vmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & -12 & -6 \end{vmatrix}$$

$$= 2\begin{vmatrix} 1 & 1 & 2 & 1 \\ 0 & 1 & 1 & -1 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 30 \end{vmatrix} = 60.$$

EXAMPLE 4.0.6 (Vandermonde determinant) Prove that

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b-a)(c-a)(c-b).$$

Solution. Subtracting column 1 from columns 2 and 3 , then expanding along row 1, gives

$$\begin{aligned} \begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} &= \begin{vmatrix} 1 & 0 & 0 \\ a & b-a & c-a \\ a^2 & b^2-a^2 & c^2-a^2 \end{vmatrix} \\ &= \begin{vmatrix} b-a & c-a \\ b^2-a^2 & c^2-a^2 \end{vmatrix} \\ &= (b-a)(c-a)\begin{vmatrix} 1 & 1 \\ b+a & c+a \end{vmatrix} = (b-a)(c-a)(c-b). \end{aligned}$$

REMARK 4.0.4 From theorems 4.0.6, 4.0.10 and corollary 4.0.2, we deduce

(a) $\det(E_{ij}A) = -\det A$,

(b) $\det(E_i(t)A) = t\det A$, if $t \neq 0$,

(c) $\det(E_{ij}(t)A) = \det A$.

It follows that if A is row–equivalent to B, then $\det B = c\det A$, where $c \neq 0$. Hence $\det B \neq 0 \Leftrightarrow \det A \neq 0$ and $\det B = 0 \Leftrightarrow \det A = 0$. Consequently from theorem 2.5.8 and remark 2.5.7, we have the following important result:

THEOREM 4.0.11 Let A be an $n \times n$ matrix. Then

(i) A is non–singular if and only if $\det A \neq 0$;

(ii) A is singular if and only if $\det A = 0$;

(iii) the homogeneous system $AX = 0$ has a non–trivial solution if and only if $\det A = 0$.

EXAMPLE 4.0.7 Find the rational numbers a for which the following homogeneous system has a non–trivial solution and solve the system for these values of a:

$$\begin{aligned} x - 2y + 3z &= 0 \\ ax + 3y + 2z &= 0 \\ 6x + y + az &= 0. \end{aligned}$$

Solution. The coefficient determinant of the system is

$$\begin{aligned} \Delta = \begin{vmatrix} 1 & -2 & 3 \\ a & 3 & 2 \\ 6 & 1 & a \end{vmatrix} &= \begin{vmatrix} 1 & -2 & 3 \\ 0 & 3+2a & 2-3a \\ 0 & 13 & a-18 \end{vmatrix} \\ &= \begin{vmatrix} 3+2a & 2-3a \\ 13 & a-18 \end{vmatrix} \\ &= (3+2a)(a-18) - 13(2-3a) \\ &= 2a^2 + 6a - 80 = 2(a+8)(a-5). \end{aligned}$$

So $\Delta = 0 \Leftrightarrow a = -8$ or $a = 5$ and these values of a are the only values for which the given homogeneous system has a non–trivial solution.

If $a = -8$, the coefficient matrix has reduced row–echelon form equal to

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -2 \\ 0 & 0 & 0 \end{bmatrix}$$

and so the complete solution is $x = z$, $y = 2z$, with z arbitrary. If $a = 5$, the coefficient matrix has reduced row–echelon form equal to

$$\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{bmatrix}$$

and so the complete solution is $x = -z$, $y = z$, with z arbitrary.

EXAMPLE 4.0.8 Find the values of t for which the following system is consistent and solve the system in each case:

$$\begin{aligned} x + y &= 1 \\ tx + y &= t \\ (1+t)x + 2y &= 3. \end{aligned}$$

Solution. Suppose that the given system has a solution (x_0, y_0). Then the following homogeneous system

$$\begin{aligned} x+y+z &= 0 \\ tx+y+tz &= 0 \\ (1+t)x+2y+3z &= 0 \end{aligned}$$

will have a non–trivial solution

$$x = x_0, \quad y = y_0, \quad z = -1.$$

Hence the coefficient determinant Δ is zero. However

$$\Delta = \begin{vmatrix} 1 & 1 & 1 \\ t & 1 & t \\ 1+t & 2 & 3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ t & 1-t & 0 \\ 1+t & 1-t & 2-t \end{vmatrix} = \begin{vmatrix} 1-t & 0 \\ 1-t & 2-t \end{vmatrix} = (1-t)(2-t).$$

Hence $t = 1$ or $t = 2$. If $t = 1$, the given system becomes

$$\begin{aligned} x+y &= 1 \\ x+y &= 1 \\ 2x+2y &= 3 \end{aligned}$$

which is clearly inconsistent. If $t = 2$, the given system becomes

$$\begin{aligned} x+y &= 1 \\ 2x+y &= 2 \\ 3x+2y &= 3 \end{aligned}$$

which has the unique solution $x = 1$, $y = 0$.

To finish this section, we present an old (1750) method of solving a system of n equations in n unknowns called *Cramer's rule* . The method is not used in practice. However it has a theoretical use as it reveals explicitly how the solution depends on the coefficients of the augmented matrix.

THEOREM 4.0.12 (Cramer's rule) The system of n linear equations in n unknowns $x_1, \ldots, x_n$

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n &= b_n \end{aligned}$$

has a unique solution if $\Delta = \det[a_{ij}] \neq 0$, namely

$$x_1 = \frac{\Delta_1}{\Delta},\ x_2 = \frac{\Delta_2}{\Delta}, \ldots, x_n = \frac{\Delta_n}{\Delta},$$

where Δ_i is the determinant of the matrix formed by replacing the i–th column of the coefficient matrix A by the entries $b_1, b_2, \ldots, b_n$.

Proof. Suppose the coefficient determinant $\Delta \neq 0$. Then by corollary 4.0.1, A^{-1} exists and is given by $A^{-1} = \frac{1}{\Delta}\operatorname{adj} A$ and the system has the unique solution

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = A^{-1}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} C_{11} & C_{21} & \cdots & C_{n1} \\ C_{12} & C_{22} & \cdots & C_{n2} \\ \vdots & & & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{bmatrix}\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

$$= \frac{1}{\Delta}\begin{bmatrix} b_1C_{11} + b_2C_{21} + \ldots + b_nC_{n1} \\ b_2C_{12} + b_2C_{22} + \ldots + b_nC_{n2} \\ \vdots \\ b_nC_{1n} + b_2C_{2n} + \ldots + b_nC_{nn} \end{bmatrix}.$$

However the i–th component of the last vector is the expansion of Δ_i along column i. Hence

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \frac{1}{\Delta}\begin{bmatrix} \Delta_1 \\ \Delta_2 \\ \vdots \\ \Delta_n \end{bmatrix} = \begin{bmatrix} \Delta_1/\Delta \\ \Delta_2/\Delta \\ \vdots \\ \Delta_n/\Delta \end{bmatrix}.$$

4.1 PROBLEMS

.

1. If the points $P_i = (x_i, y_i)$, $i = 1, 2, 3, 4$ form a quadrilateral with vertices in anti–clockwise orientation, prove that the area of the quadrilateral equals

$$\frac{1}{2}\left(\begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} + \begin{vmatrix} x_2 & x_3 \\ y_2 & y_3 \end{vmatrix} + \begin{vmatrix} x_3 & x_4 \\ y_3 & y_4 \end{vmatrix} + \begin{vmatrix} x_4 & x_1 \\ y_4 & y_1 \end{vmatrix}\right).$$

(This formula generalizes to a simple polygon and is known as the *Surveyor's formula.*)

2. Prove that the following identity holds by expressing the left–hand side as the sum of 8 determinants:

$$\begin{vmatrix} a+x & b+y & c+z \\ x+u & y+v & z+w \\ u+a & v+b & w+c \end{vmatrix} = 2 \begin{vmatrix} a & b & c \\ x & y & z \\ u & v & w \end{vmatrix}.$$

3. Prove that

$$\begin{vmatrix} n^2 & (n+1)^2 & (n+2)^2 \\ (n+1)^2 & (n+2)^2 & (n+3)^2 \\ (n+2)^2 & (n+3)^2 & (n+4)^2 \end{vmatrix} = -8.$$

4. Evaluate the following determinants:

$$\text{(a)} \begin{vmatrix} 246 & 427 & 327 \\ 1014 & 543 & 443 \\ -342 & 721 & 621 \end{vmatrix} \qquad \text{(b)} \begin{vmatrix} 1 & 2 & 3 & 4 \\ -2 & 1 & -4 & 3 \\ 3 & -4 & -1 & 2 \\ 4 & 3 & -2 & -1 \end{vmatrix}.$$

[Answers: (a) -29400000; (b) 900.]

5. Compute the inverse of the matrix

$$A = \begin{bmatrix} 1 & 0 & -2 \\ 3 & 1 & 4 \\ 5 & 2 & -3 \end{bmatrix}$$

by first computing the adjoint matrix.

[Answer: $A^{-1} = \frac{-1}{13} \begin{bmatrix} -11 & -4 & 2 \\ 29 & 7 & -10 \\ 1 & -2 & 1 \end{bmatrix}$.]

6. Prove that the following identities hold:

$$\text{(i)} \quad \begin{vmatrix} 2a & 2b & b-c \\ 2b & 2a & a+c \\ a+b & a+b & b \end{vmatrix} = -2(a-b)^2(a+b),$$

$$\text{(ii)} \quad \begin{vmatrix} b+c & b & c \\ c & c+a & a \\ b & a & a+b \end{vmatrix} = 2a(b^2+c^2).$$

7. Let $P_i = (x_i, y_i)$, $i = 1, 2, 3$. If x_1, x_2, x_3 are distinct, prove that there is precisely one curve of the form $y = ax^2 + bx + c$ passing through P_1, P_2 and P_3.

8. Let
$$A = \begin{bmatrix} 1 & 1 & -1 \\ 2 & 3 & k \\ 1 & k & 3 \end{bmatrix}.$$
Find the values of k for which $\det A = 0$ and hence, or otherwise, determine the value of k for which the following system has more than one solution:
$$\begin{aligned} x + y - z &= 1 \\ 2x + 3y + kz &= 3 \\ x + ky + 3z &= 2. \end{aligned}$$
Solve the system for this value of k and determine the solution for which $x^2 + y^2 + z^2$ has least value.

[Answer: $k = 2$; $x = 10/21$, $y = 13/21$, $z = 2/21$.]

9. By considering the coefficient determinant, find all rational numbers a and b for which the following system has (i) no solutions, (ii) exactly one solution, (iii) infinitely many solutions:
$$\begin{aligned} x - 2y + bz &= 3 \\ ax + \qquad 2z &= 2 \\ 5x + 2y \qquad &= 1. \end{aligned}$$
Solve the system in case (iii).

[Answer: (i) $ab = 12$ and $a \neq 3$, no solution; $ab \neq 12$, unique solution; $a = 3$, $b = 4$, infinitely many solutions; $x = -\frac{2}{3}z + \frac{2}{3}$, $y = \frac{5}{3}z - \frac{7}{6}$, with z arbitrary.]

10. Express the determinant of the matrix
$$B = \begin{bmatrix} 1 & 1 & 2 & 1 \\ 1 & 2 & 3 & 4 \\ 2 & 4 & 7 & 2t+6 \\ 2 & 2 & 6-t & t \end{bmatrix}$$

as a polynomial in t and hence determine the rational values of t for which B^{-1} exists.

[Answer: $\det B = (t-2)(2t-1)$; $t \neq 2$ and $t \neq \frac{1}{2}$.]

11. If A is a 3×3 matrix over a field and $\det A \neq 0$, prove that

$$\begin{aligned}\text{(i)}\quad \det(\operatorname{adj} A) &= (\det A)^2,\\ \text{(ii)}\quad (\operatorname{adj} A)^{-1} &= \frac{1}{\det A}A = \operatorname{adj}(A^{-1}).\end{aligned}$$

12. Suppose that A is a real 3×3 matrix such that $A^tA = I_3$.
    (i) Prove that $A^t(A - I_3) = -(A - I_3)^t$.
    (ii) Prove that $\det A = \pm 1$.
    (iii) Use (i) to prove that if $\det A = 1$, then $\det(A - I_3) = 0$.

13. If A is a square matrix such that one column is a linear combination of the remaining columns, prove that $\det A = 0$. Prove that the converse also holds.

14. Use Cramer's rule to solve the system

$$\begin{aligned}-2x + 3y - z &= 1\\ x + 2y - z &= 4\\ -2x - y + z &= -3.\end{aligned}$$

[Answer: $x = 2$, $y = 3$, $z = 4$.]

15. Use remark 4.0.4 to deduce that

$$\det E_{ij} = -1, \quad \det E_i(t) = t, \quad \det E_{ij}(t) = 1$$

and use theorem 2.5.8 and induction, to prove that

$$\det(BA) = \det B \det A,$$

if B is non–singular. Also prove that the formula holds when B is singular.

16. Prove that

$$\begin{vmatrix} a+b+c & a+b & a & a\\ a+b & a+b+c & a & a\\ a & a & a+b+c & a+b\\ a & a & a+b & a+b+c \end{vmatrix} = c^2(2b+c)(4a+2b+c).$$

17. Prove that

$$\begin{vmatrix} 1+u_1 & u_1 & u_1 & u_1 \\ u_2 & 1+u_2 & u_2 & u_2 \\ u_3 & u_3 & 1+u_3 & u_3 \\ u_4 & u_4 & u_4 & 1+u_4 \end{vmatrix} = 1 + u_1 + u_2 + u_3 + u_4.$$

18. Let $A \in M_{n\times n}(F)$. If $A^t = -A$, prove that $\det A = 0$ if n is odd and $1 + 1 \neq 0$ in F.

19. Prove that

$$\begin{vmatrix} 1 & 1 & 1 & 1 \\ r & 1 & 1 & 1 \\ r & r & 1 & 1 \\ r & r & r & 1 \end{vmatrix} = (1-r)^3.$$

20. Express the determinant

$$\begin{vmatrix} 1 & a^2 - bc & a^4 \\ 1 & b^2 - ca & b^4 \\ 1 & c^2 - ab & c^4 \end{vmatrix}$$

as the product of one quadratic and four linear factors.

[Answer: $(b-a)(c-a)(c-b)(a+b+c)(b^2+bc+c^2+ac+ab+a^2)$.]

Chapter 5

COMPLEX NUMBERS

5.1 Constructing the complex numbers

One way of introducing the field $\mathbb{C}$ of complex numbers is via the arithmetic of 2×2 matrices.

DEFINITION 5.1.1 A complex number is a matrix of the form

$$\begin{bmatrix} x & -y \\ y & x \end{bmatrix},$$

where x and y are real numbers.

Complex numbers of the form $\begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix}$ are scalar matrices and are called *real* complex numbers and are denoted by the symbol $\{x\}$.

The real complex numbers $\{x\}$ and $\{y\}$ are respectively called the *real part* and *imaginary part* of the complex number $\begin{bmatrix} x & -y \\ y & x \end{bmatrix}$.

The complex number $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ is denoted by the symbol i.

We have the identities

$$\begin{aligned} \begin{bmatrix} x & -y \\ y & x \end{bmatrix} &= \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} + \begin{bmatrix} 0 & -y \\ y & 0 \end{bmatrix} = \begin{bmatrix} x & 0 \\ 0 & x \end{bmatrix} + \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} y & 0 \\ 0 & y \end{bmatrix} \\ &= \{x\} + i\{y\}, \end{aligned}$$

$$i^2 = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} = \{-1\}.$$

Complex numbers of the form $i\{y\}$, where y is a non–zero real number, are called *imaginary numbers*.

If two complex numbers are equal, we can equate their real and imaginary parts:

$$\{x_1\} + i\{y_1\} = \{x_2\} + i\{y_2\} \Rightarrow x_1 = x_2 \text{ and } y_1 = y_2,$$

if x_1, x_2, y_1, y_2 are real numbers. Noting that $\{0\} + i\{0\} = \{0\}$, gives the useful special case is

$$\{x\} + i\{y\} = \{0\} \Rightarrow x = 0 \text{ and } y = 0,$$

if x and y are real numbers.

The sum and product of two real complex numbers are also real complex numbers:

$$\{x\} + \{y\} = \{x + y\}, \quad \{x\}\{y\} = \{xy\}.$$

Also, as real complex numbers are scalar matrices, their arithmetic is very simple. They form a field under the operations of matrix addition and multiplication. The additive identity is $\{0\}$, the additive inverse of $\{x\}$ is $\{-x\}$, the multiplicative identity is $\{1\}$ and the multiplicative inverse of $\{x\}$ is $\{x^{-1}\}$. Consequently

$$\{x\} - \{y\} = \{x\} + (-\{y\}) = \{x\} + \{-y\} = \{x - y\},$$

$$\frac{\{x\}}{\{y\}} = \{x\}\{y\}^{-1} = \{x\}\{y^{-1}\} = \{xy^{-1}\} = \left\{\frac{x}{y}\right\}.$$

It is customary to blur the distinction between the real complex number $\{x\}$ and the real number x and write $\{x\}$ as x. Thus we write the complex number $\{x\} + i\{y\}$ simply as $x + iy$.

More generally, the sum of two complex numbers is a complex number:

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2); \tag{5.1}$$

and (using the fact that scalar matrices commute with all matrices under matrix multiplication and $\{-1\}A = -A$ if A is a matrix), the product of two complex numbers is a complex number:

$$\begin{aligned}
&(x_1 + iy_1)(x_2 + iy_2) = x_1(x_2 + iy_2) + (iy_1)(x_2 + iy_2)\\
&= x_1x_2 + x_1(iy_2) + (iy_1)x_2 + (iy_1)(iy_2)\\
&= x_1x_2 + ix_1y_2 + iy_1x_2 + i^2y_1y_2\\
&= (x_1x_2 + \{-1\}y_1y_2) + i(x_1y_2 + y_1x_2)\\
&= (x_1x_2 - y_1y_2) + i(x_1y_2 + y_1x_2), \qquad (5.2)
\end{aligned}$$

The set $\mathbb{C}$ of complex numbers forms a field under the operations of matrix addition and multiplication. The additive identity is 0, the additive inverse of $x + iy$ is the complex number $(-x) + i(-y)$, the multiplicative identity is 1 and the multiplicative inverse of the non–zero complex number $x + iy$ is the complex number $u + iv$, where

$$u = \frac{x}{x^2 + y^2} \text{ and } v = \frac{-y}{x^2 + y^2}.$$

(If $x + iy \neq 0$, then $x \neq 0$ or $y \neq 0$, so $x^2 + y^2 \neq 0$.)

From equations 5.1 and 5.2, we observe that addition and multiplication of complex numbers is performed just as for real numbers, replacing i^2 by -1, whenever it occurs.

A useful identity satisfied by complex numbers is

$$r^2 + s^2 = (r + is)(r - is).$$

This leads to a method of expressing the ratio of two complex numbers in the form $x + iy$, where x and y are real complex numbers.

$$\begin{aligned}\frac{x_1 + iy_1}{x_2 + iy_2} &= \frac{(x_1 + iy_1)(x_2 - iy_2)}{(x_2 + iy_2)(x_2 - iy_2)} \\ &= \frac{(x_1x_2 + y_1y_2) + i(-x_1y_2 + y_1x_2)}{x_2^2 + y_2^2}.\end{aligned}$$

The process is known as *rationalization of the denominator*.

5.2 Calculating with complex numbers

We can now do all the standard linear algebra calculations over the field of complex numbers – find the reduced row–echelon form of an matrix whose elements are complex numbers, solve systems of linear equations, find inverses and calculate determinants.

For example, solve the system

$$\begin{aligned}(1 + i)z + (2 - i)w &= 2 + 7i \\ 7z + (8 - 2i)w &= 4 - 9i.\end{aligned}$$

The coefficient determinant is

$$\begin{aligned}\begin{vmatrix} 1 + i & 2 - i \\ 7 & 8 - 2i \end{vmatrix} &= (1 + i)(8 - 2i) - 7(2 - i) \\ &= (8 - 2i) + i(8 - 2i) - 14 + 7i \\ &= -4 + 13i \neq 0.\end{aligned}$$

Hence by Cramer's rule, there is a unique solution:

$$\begin{aligned}
z &= \frac{\begin{vmatrix} 2+7i & 2-i \\ 4-9i & 8-2i \end{vmatrix}}{-4+13i} \\
&= \frac{(2+7i)(8-2i)-(4-9i)(2-i)}{-4+13i} \\
&= \frac{2(8-2i)+(7i)(8-2i)-\{(4(2-i)-9i(2-i)\}}{-4+13i} \\
&= \frac{16-4i+56i-14i^2-\{8-4i-18i+9i^2\}}{-4+13i} \\
&= \frac{31+74i}{-4+13i} \\
&= \frac{(31+74i)(-4-13i)}{(-4)^2+13^2} \\
&= \frac{838-699i}{(-4)^2+13^2} \\
&= \frac{838}{185}-\frac{699}{185}i
\end{aligned}$$

and similarly $w = \frac{-698}{185} + \frac{229}{185}i$.

An important property enjoyed by complex numbers is that every complex number has a square root.

THEOREM 5.2.1 If w is a non–zero complex number, then the equation $z^2 = w$ has a solution $z \in \mathbb{C}$.

Proof. Let $w = a + ib$, a, $b \in \mathbb{R}$.

Case 1. Suppose $b = 0$. Then if $a > 0$, $z = \sqrt{a}$ is a solution, while if $a < 0$, $i\sqrt{-a}$ is a solution.

Case 2. Suppose $b \neq 0$. Let $z = x + iy$, x, $y \in \mathbb{R}$. Then the equation $z^2 = w$ becomes

$$(x+iy)^2 = x^2 - y^2 + 2xyi = a + ib,$$

so equating real and imaginary parts gives

$$x^2 - y^2 = a \quad \text{and} \quad 2xy = b.$$

Hence $x \neq 0$ and $y = b/(2x)$. Consequently

$$x^2 - \left(\frac{b}{2x}\right)^2 = a,$$

so $4x^4 - 4ax^2 - b^2 = 0$ and $4(x^2)^2 - 4a(x^2) - b^2 = 0$. Hence

$$x^2 = \frac{4a \pm \sqrt{16a^2 + 16b^2}}{8} = \frac{a \pm \sqrt{a^2 + b^2}}{2}.$$

However $x^2 > 0$, so we must take the + sign, as $a - \sqrt{a^2 + b^2} < 0$. Hence

$$x^2 = \frac{a + \sqrt{a^2 + b^2}}{2}, \quad x = \pm\sqrt{\frac{a + \sqrt{a^2 + b^2}}{2}}.$$

Then y is determined by $y = b/(2x)$.

EXAMPLE 5.2.1 Solve the equation $z^2 = 1 + i$.

Solution. Put $z = x + iy$. Then the equation becomes

$$(x + iy)^2 = x^2 - y^2 + 2xyi = 1 + i,$$

so equating real and imaginary parts gives

$$x^2 - y^2 = 1 \text{ and } 2xy = 1.$$

Hence $x \neq 0$ and $y = 1/(2x)$. Consequently

$$x^2 - \left(\frac{1}{2x}\right)^2 = 1,$$

so $4x^4 - 4x^2 - 1 = 0$. Hence

$$x^2 = \frac{4 \pm \sqrt{16 + 16}}{8} = \frac{1 \pm \sqrt{2}}{2}.$$

Hence

$$x^2 = \frac{1 + \sqrt{2}}{2} \quad \text{and} \quad x = \pm\sqrt{\frac{1 + \sqrt{2}}{2}}.$$

Then

$$y = \frac{1}{2x} = \pm\frac{1}{\sqrt{2}\sqrt{1 + \sqrt{2}}}.$$

Hence the solutions are

$$z = \pm\left(\sqrt{\frac{1 + \sqrt{2}}{2}} + \frac{i}{\sqrt{2}\sqrt{1 + \sqrt{2}}}\right).$$

EXAMPLE 5.2.2 Solve the equation $z^2 + (\sqrt{3}+i)z + 1 = 0$.

Solution. Because every complex number has a square root, the familiar formula
$$z = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$$
for the solution of the general quadratic equation $az^2 + bz + c = 0$ can be used, where now $a(\neq 0)$, b, $c \in \mathbb{C}$. Hence

$$\begin{aligned} z &= \frac{-(\sqrt{3}+i) \pm \sqrt{(\sqrt{3}+i)^2 - 4}}{2} \\ &= \frac{-(\sqrt{3}+i) \pm \sqrt{(3+2\sqrt{3}i-1) - 4}}{2} \\ &= \frac{-(\sqrt{3}+i) \pm \sqrt{-2+2\sqrt{3}i}}{2}. \end{aligned}$$

Now we have to solve $w^2 = -2 + 2\sqrt{3}i$. Put $w = x + iy$. Then $w^2 = x^2 - y^2 + 2xyi = -2 + 2\sqrt{3}i$ and equating real and imaginary parts gives $x^2 - y^2 = -2$ and $2xy = 2\sqrt{3}$. Hence $y = \sqrt{3}/x$ and so $x^2 - 3/x^2 = -2$. So $x^4 + 2x^2 - 3 = 0$ and $(x^2+3)(x^2-1) = 0$. Hence $x^2 - 1 = 0$ and $x = \pm 1$. Then $y = \pm\sqrt{3}$. Hence $(1+\sqrt{3}i)^2 = -2 + 2\sqrt{3}i$ and the formula for z now becomes

$$\begin{aligned} z &= \frac{-\sqrt{3} - i \pm (1+\sqrt{3}i)}{2} \\ &= \frac{1-\sqrt{3}+(1+\sqrt{3})i}{2} \quad \text{or} \quad \frac{-1-\sqrt{3}-(1+\sqrt{3})i}{2}. \end{aligned}$$

EXAMPLE 5.2.3 Find the cube roots of 1.

Solution. We have to solve the equation $z^3 = 1$, or $z^3 - 1 = 0$. Now $z^3 - 1 = (z-1)(z^2+z+1)$. So $z^3 - 1 = 0 \Rightarrow z - 1 = 0$ or $z^2 + z + 1 = 0$. But
$$z^2 + z + 1 = 0 \Rightarrow z = \frac{-1 \pm \sqrt{1^2-4}}{2} = \frac{-1 \pm \sqrt{3}i}{2}.$$
So there are 3 cube roots of 1, namely 1 and $(-1 \pm \sqrt{3}i)/2$.

We state the next theorem without proof. It states that every non–constant polynomial with complex number coefficients has a root in the field of complex numbers.

THEOREM 5.2.2 (Gauss) If $f(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$, where $a_n \neq 0$ and $n \geq 1$, then $f(z) = 0$ for some $z \in \mathbb{C}$.

It follows that in view of the *factor* theorem, which states that if $a \in F$ is a root of a polynomial $f(z)$ with coefficients from a field F, then $z - a$ is a factor of $f(z)$, that is $f(z) = (z - a)g(z)$, where the coefficients of $g(z)$ also belong to F. By repeated application of this result, we can factorize any polynomial with complex coefficients into a product of linear factors with complex coefficients:

$$f(z) = a_n(z - z_1)(z - z_2)\cdots(z - z_n).$$

There are available a number of computational algorithms for finding good approximations to the roots of a polynomial with complex coefficients.

5.3 Geometric representation of $\mathbb{C}$

Complex numbers can be represented as points in the plane, using the correspondence $x + iy \leftrightarrow (x, y)$. The representation is known as the *Argand diagram* or *complex plane*. The real complex numbers lie on the x–axis, which is then called the *real axis*, while the imaginary numbers lie on the y–axis, which is known as the *imaginary axis*. The complex numbers with positive imaginary part lie in the *upper half plane*, while those with negative imaginary part lie in the *lower half plane*.

Because of the equation

$$(x_1 + iy_1) + (x_2 + iy_2) = (x_1 + x_2) + i(y_1 + y_2),$$

complex numbers add vectorially, using the parallellogram law. Similarly, the complex number $z_1 - z_2$ can be represented by the vector from (x_2, y_2) to (x_1, y_1), where $z_1 = x_1 + iy_1$ and $z_2 = x_2 + iy_2$. (See Figure 5.1.)

The geometrical representation of complex numbers can be very useful when complex number methods are used to investigate properties of triangles and circles. It is very important in the branch of calculus known as Complex Function theory, where geometric methods play an important role.

We mention that the line through two distinct points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ has the form $z = (1 - t)z_1 + tz_2$, $t \in \mathbb{R}$, where $z = x + iy$ is any point on the line and $z_i = x_i + iy_i$, $i = 1, 2$. For the line has parametric equations

$$x = (1 - t)x_1 + tx_2, \quad y = (1 - t)y_1 + ty_2$$

and these can be combined into a single equation $z = (1 - t)z_1 + tz_2$.

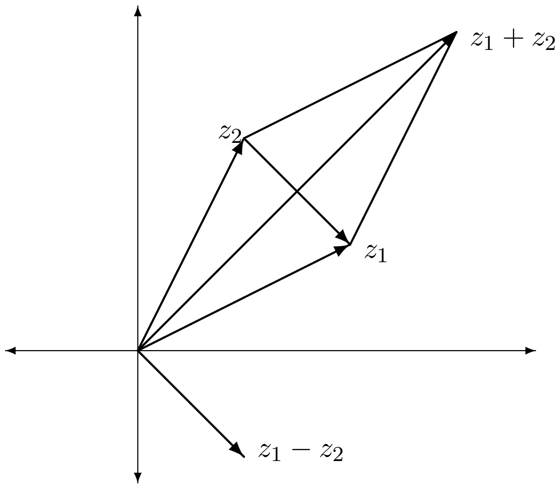

Figure 5.1: Complex addition and subraction.

Circles have various equation representations in terms of complex numbers, as will be seen later.

5.4 Complex conjugate

DEFINITION 5.4.1 (Complex conjugate) If $z = x + iy$, the *complex conjugate* of z is the complex number defined by $\overline{z} = x - iy$. Geometrically, the complex conjugate of z is obtained by reflecting z in the real axis (see Figure 5.2).

The following properties of the complex conjugate are easy to verify:

1. $\overline{z_1 + z_2} = \overline{z_1} + \overline{z_2}$;
2. $\overline{-z} = -\,\overline{z}$.
3. $\overline{z_1 - z_2} = \overline{z_1} - \overline{z_2}$;
4. $\overline{z_1 z_2} = \overline{z_1}\,\overline{z_2}$;
5. $\overline{(1/z)} = 1/\overline{z}$;
6. $\overline{(z_1/z_2)} = \overline{z_1}/\overline{z_2}$;

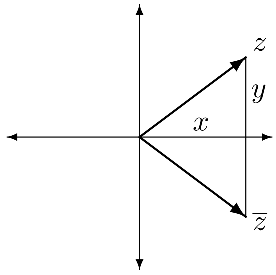

Figure 5.2: The complex conjugate of z: $\overline{z}$.

7. z is real if and only if $\overline{z} = z$;

8. With the standard convention that the real and imaginary parts are denoted by $\operatorname{Re} z$ and $\operatorname{Im} z$, we have

$$\operatorname{Re} z = \frac{z + \overline{z}}{2}, \quad \operatorname{Im} z = \frac{z - \overline{z}}{2i};$$

9. If $z = x + iy$, then $z\overline{z} = x^2 + y^2$.

THEOREM 5.4.1 If $f(z)$ is a polynomial with real coefficients, then its non–real roots occur in complex–conjugate pairs, i.e. if $f(z) = 0$, then $f(\overline{z}) = 0$.

Proof. Suppose $f(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = 0$, where $a_n, \ldots, a_0$ are real. Then

$$\begin{aligned} 0 = \overline{0} = \overline{f(z)} &= \overline{a_n z^n} + \overline{a_{n-1} z^{n-1}} + \cdots + \overline{a_1 z} + \overline{a_0} \\ &= \overline{a_n}\,\overline{z^n} + \overline{a_{n-1}}\,\overline{z^{n-1}} + \cdots + \overline{a_1}\,\overline{z} + \overline{a_0} \\ &= a_n \overline{z}^n + a_{n-1} \overline{z}^{n-1} + \cdots + a_1 \overline{z} + a_0 \\ &= f(\overline{z}). \end{aligned}$$

EXAMPLE 5.4.1 Discuss the position of the roots of the equation

$$z^4 = -1$$

in the complex plane.

Solution. The equation $z^4 = -1$ has real coefficients and so its roots come in complex conjugate pairs. Also if z is a root, so is $-z$. Also there are clearly no real roots and no imaginary roots. So there must be one root w in the first quadrant, with all remaining roots being given by $\overline{w}$, $-w$ and $-\overline{w}$. In fact, as we shall soon see, the roots lie evenly spaced on the unit circle.

The following theorem is useful in deciding if a polynomial $f(z)$ has a multiple root a; that is if $(z-a)^m$ divides $f(z)$ for some $m \geq 2$. (The proof is left as an exercise.)

THEOREM 5.4.2 If $f(z) = (z-a)^m g(z)$, where $m \geq 2$ and $g(z)$ is a polynomial, then $f'(a) = 0$ and the polynomial and its derivative have a common root.

From theorem 5.4.1 we obtain a result which is very useful in the explicit integration of rational functions (i.e. ratios of polynomials) with real coefficients.

THEOREM 5.4.3 If $f(z)$ is a non–constant polynomial with real coefficients, then $f(z)$ can be factorized as a product of real linear factors and real quadratic factors.

Proof. In general $f(z)$ will have r real roots $z_1, \ldots, z_r$ and $2s$ non–real roots z_{r+1}, $\overline{z}_{r+1}, \ldots, z_{r+s}$, $\overline{z}_{r+s}$, occurring in complex–conjugate pairs by theorem 5.4.1. Then if a_n is the coefficient of highest degree in $f(z)$, we have the factorization

$$\begin{aligned} f(z) &= a_n(z-z_1)\cdots(z-z_r)\times \\ &\quad \times(z-z_{r+1})(z-\overline{z}_{r+1})\cdots(z-z_{r+s})(z-\overline{z}_{r+s}). \end{aligned}$$

We then use the following identity for $j = r+1, \ldots, r+s$ which in turn shows that paired terms give rise to real quadratic factors:

$$\begin{aligned} (z-z_j)(z-\overline{z}_j) &= z^2 - (z_j + \overline{z}_j)z + z_j\overline{z}_j \\ &= z^2 - 2\mathrm{Re}\, z_j + (x_j^2 + y_j^2), \end{aligned}$$

where $z_j = x_j + iy_j$.

A well–known example of such a factorization is the following:

EXAMPLE 5.4.2 Factorize $z^4 + 1$ into real linear and quadratic factors.

Solution. Clearly there are no real roots. Also we have the preliminary factorization $z^4 + 1 = (z^2 - i)(z^2 + i)$. Now the roots of $z^2 - i$ are easily verified to be $\pm(1+i)/\sqrt{2}$, so the roots of $z^2 + i$ must be $\pm(1-i)/\sqrt{2}$.

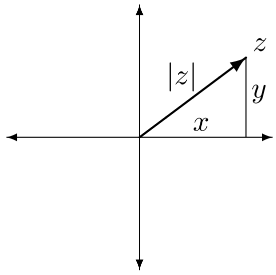

Figure 5.3: The modulus of z: $|z|$.

In other words the roots are $w = (1+i)/\sqrt{2}$ and $\overline{w}$, $-w$, $-\overline{w}$. Grouping conjugate–complex terms gives the factorization

$$\begin{aligned} z^4+1 &= (z-w)(z-\overline{w})(z+w)(z+\overline{w}) \\ &= (z^2-2z\mathrm{Re}\,w+w\overline{w})(z^2+2z\mathrm{Re}\,w+w\overline{w}) \\ &= (z^2-\sqrt{2}z+1)(z^2+\sqrt{2}z+1). \end{aligned}$$

5.5 Modulus of a complex number

DEFINITION 5.5.1 (Modulus) If $z = x+iy$, the *modulus* of z is the non–negative real number $|z|$ defined by $|z| = \sqrt{x^2+y^2}$. Geometrically, the modulus of z is the distance from z to 0 (see Figure 5.3).

More generally, $|z_1 - z_2|$ is the distance between z_1 and z_2 in the complex plane. For

$$\begin{aligned} |z_1-z_2| = |(x_1+iy_1)-(x_2+iy_2)| &= |(x_1-x_2)+i(y_1-y_2)| \\ &= \sqrt{(x_1-x_2)^2+(y_1-y_2)^2}. \end{aligned}$$

The following properties of the modulus are easy to verify, using the identity $|z|^2 = z\overline{z}$:

(i) $|z_1z_2| = |z_1||z_2|$;

(ii) $|z^{-1}| = |z|^{-1}$;

(iii) $\left|\dfrac{z_1}{z_2}\right| = \dfrac{|z_1|}{|z_2|}$.

For example, to prove (i):

$$\begin{aligned} |z_1z_2|^2 &= (z_1z_2)\overline{z_1z_2} = (z_1z_2)\overline{z_1}\,\overline{z_2} \\ &= (z_1\overline{z_1})(z_2\overline{z_2}) = |z_1|^2|z_2|^2 = (|z_1||z_2|)^2. \end{aligned}$$

Hence $|z_1z_2| = |z_1||z_2|$.

EXAMPLE 5.5.1 Find $|z|$ when $z = \dfrac{(1+i)^4}{(1+6i)(2-7i)}$.

Solution.

$$\begin{aligned} |z| &= \frac{|1+i|^4}{|1+6i||2-7i|} \\ &= \frac{(\sqrt{1^2+1^2})^4}{\sqrt{1^2+6^2}\sqrt{2^2+(-7)^2}} \\ &= \frac{4}{\sqrt{37}\sqrt{53}}. \end{aligned}$$

THEOREM 5.5.1 (Ratio formulae) If z lies on the line through z_1 and z_2:

$$z = (1-t)z_1 + tz_2, \quad t \in \mathbb{R},$$

we have the useful *ratio formulae*:

(i) $\left|\dfrac{z-z_1}{z-z_2}\right| = \left|\dfrac{t}{1-t}\right|$ if $z \neq z_2$,

(ii) $\left|\dfrac{z-z_1}{z_1-z_2}\right| = |t|$.

Circle equations. The equation $|z - z_0| = r$, where $z_0 \in \mathbb{C}$ and $r > 0$, represents the circle centre z_0 and radius r. For example the equation $|z-(1+2i)| = 3$ represents the circle $(x-1)^2 + (y-2)^2 = 9$.

Another useful circle equation is the *circle of Apollonius* :

$$\left|\frac{z-a}{z-b}\right| = \lambda,$$

where a and b are distinct complex numbers and λ is a positive real number, $\lambda \neq 1$. (If $\lambda = 1$, the above equation represents the perpendicular bisector of the segment joining a and b.)

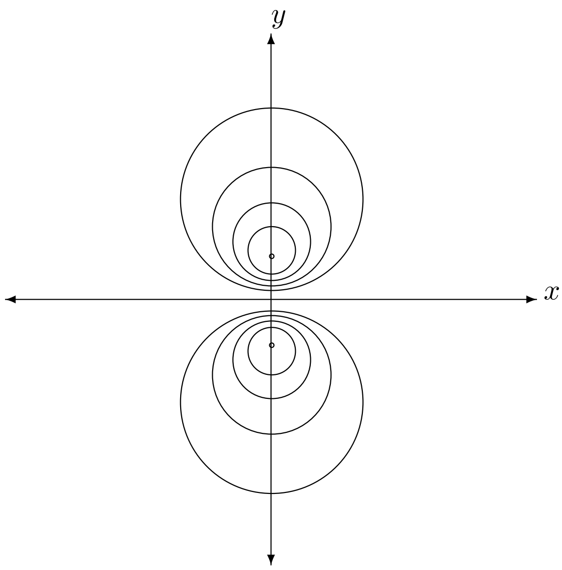

Figure 5.4: Apollonius circles: $\frac{|z+2i|}{|z-2i|} = \frac{1}{4}, \frac{3}{8}, \frac{1}{2}, \frac{5}{8}; \frac{4}{1}, \frac{8}{3}, \frac{2}{1}, \frac{8}{5}$.

An algebraic proof that the above equation represents a circle, runs as follows. We use the following identities:

$$\begin{array}{lrcl} \text{(i)} & |z-a|^2 & = & |z|^2 - 2\mathrm{Re}\,(\overline{z}a) + |a|^2 \\ \text{(ii)} & \mathrm{Re}\,(z_1 \pm z_2) & = & \mathrm{Re}\,z_1 \pm \mathrm{Re}\,z_2 \\ \text{(iii)} & \mathrm{Re}\,(tz) & = & t\mathrm{Re}\,z \text{ if } t \in \mathbb{R}. \end{array}$$

We have

$$\left|\frac{z-a}{z-b}\right| = \lambda \Leftrightarrow |z-a|^2 = \lambda^2|z-b|^2$$

$$\Leftrightarrow |z|^2 - 2\mathrm{Re}\,\{\overline{z}a\} + |a|^2 = \lambda^2(|z|^2 - 2\mathrm{Re}\,\{\overline{z}b\} + |b|^2)$$

$$\Leftrightarrow (1-\lambda^2)|z|^2 - 2\mathrm{Re}\,\{\overline{z}(a-\lambda^2 b)\} = \lambda^2|b|^2 - |a|^2$$

$$\Leftrightarrow |z|^2 - 2\mathrm{Re}\,\left\{\overline{z}\left(\frac{a-\lambda^2 b}{1-\lambda^2}\right)\right\} = \frac{\lambda^2|b|^2-|a|^2}{1-\lambda^2}$$

$$\Leftrightarrow |z|^2 - 2\mathrm{Re}\,\left\{\overline{z}\left(\frac{a-\lambda^2 b}{1-\lambda^2}\right)\right\} + \left|\frac{a-\lambda^2 b}{1-\lambda^2}\right|^2 = \frac{\lambda^2|b|^2-|a|^2}{1-\lambda^2} + \left|\frac{a-\lambda^2 b}{1-\lambda^2}\right|^2.$$

Now it is easily verified that

$$|a-\lambda^2 b|^2 + (1-\lambda^2)(\lambda^2|b|^2 - |a|^2) = \lambda^2|a-b|^2.$$

So we obtain

$$\left|\frac{z-a}{z-b}\right| = \lambda \quad \Leftrightarrow \quad \left|z - \left(\frac{a-\lambda^2 b}{1-\lambda^2}\right)\right|^2 = \frac{\lambda^2|a-b|^2}{|1-\lambda^2|^2}$$
$$\Leftrightarrow \quad \left|z - \left(\frac{a-\lambda^2 b}{1-\lambda^2}\right)\right| = \frac{\lambda|a-b|}{|1-\lambda^2|}.$$

The last equation represents a circle centre z_0, radius r, where

$$z_0 = \frac{a-\lambda^2 b}{1-\lambda^2} \quad \text{and} \quad r = \frac{\lambda|a-b|}{|1-\lambda^2|}.$$

There are two special points on the circle of Apollonius, the points z_1 and z_2 defined by

$$\frac{z_1-a}{z_1-b} = \lambda \quad \text{and} \quad \frac{z_2-a}{z_2-b} = -\lambda,$$

or

$$z_1 = \frac{a-\lambda b}{1-\lambda} \quad \text{and} \quad z_2 = \frac{a+\lambda b}{1+\lambda}. \tag{5.3}$$

It is easy to verify that z_1 and z_2 are distinct points on the line through a and b and that $z_0 = \frac{z_1+z_2}{2}$. Hence the circle of Apollonius is the circle based on the segment z_1, z_2 as diameter.

EXAMPLE 5.5.2 Find the centre and radius of the circle

$$|z-1-i| = 2|z-5-2i|.$$

Solution. Method 1. Proceed algebraically and simplify the equation

$$|x+iy-1-i| = 2|x+iy-5-2i|$$

or

$$|x-1+i(y-1)| = 2|x-5+i(y-2)|.$$

Squaring both sides gives

$$(x-1)^2 + (y-1)^2 = 4((x-5)^2 + (y-2)^2),$$

which reduces to the circle equation

$$x^2 + y^2 - \frac{38}{3}x - \frac{14}{3}y + 38 = 0.$$

Completing the square gives

$$(x - \frac{19}{3})^2 + (y - \frac{7}{3})^2 = \left(\frac{19}{3}\right)^2 + \left(\frac{7}{3}\right)^2 - 38 = \frac{68}{9},$$

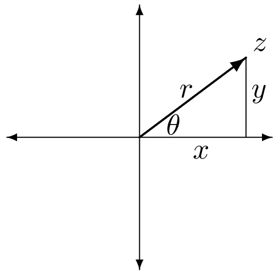

Figure 5.5: The argument of z: $\arg z = \theta$.

so the centre is $(\frac{19}{3}, \frac{7}{3})$ and the radius is $\sqrt{\frac{68}{9}}$.
Method 2. Calculate the diametrical points z_1 and z_2 defined above by equations 5.3:

$$\begin{aligned} z_1 - 1 - i &= 2(z_1 - 5 - 2i) \\ z_2 - 1 - i &= -2(z_2 - 5 - 2i). \end{aligned}$$

We find $z_1 = 9 + 3i$ and $z_2 = (11 + 5i)/3$. Hence the centre z_0 is given by

$$z_0 = \frac{z_1 + z_2}{2} = \frac{19}{3} + \frac{7}{3}i$$

and the radius r is given by

$$r = |z_1 - z_0| = \left|\left(\frac{19}{3} + \frac{7}{3}i\right) - (9 + 3i)\right| = \left|-\frac{8}{3} - \frac{2}{3}i\right| = \frac{\sqrt{68}}{3}.$$

5.6 Argument of a complex number

Let $z = x + iy$ be a non–zero complex number, $r = |z| = \sqrt{x^2 + y^2}$. Then we have $x = r\cos\theta$, $y = r\sin\theta$, where θ is the angle made by z with the positive x–axis. So θ is unique up to addition of a multiple of 2π radians.

DEFINITION 5.6.1 (Argument) Any number θ satisfying the above pair of equations is called an *argument* of z and is denoted by $\arg z$. The particular argument of z lying in the range $-\pi < \theta \leq \pi$ is called the *principal argument* of z and is denoted by $\operatorname{Arg} z$ (see Figure 5.5).

We have $z = r\cos\theta + ir\sin\theta = r(\cos\theta + i\sin\theta)$ and this representation of z is called the *polar representation* or *modulus–argument form* of z.

EXAMPLE 5.6.1 $\text{Arg}\,1 = 0$, $\text{Arg}\,(-1) = \pi$, $\text{Arg}\,i = \frac{\pi}{2}$, $\text{Arg}\,(-i) = -\frac{\pi}{2}$.

We note that $y/x = \tan\theta$ if $x \neq 0$, so θ is determined by this equation up to a multiple of π. In fact

$$\text{Arg}\,z = \tan^{-1}\frac{y}{x} + k\pi,$$

where $k = 0$ if $x > 0$; $k = 1$ if $x < 0$, $y > 0$; $k = -1$ if $x < 0$, $y < 0$.

To determine $\text{Arg}\,z$ graphically, it is simplest to draw the triangle formed by the points 0, x, z on the complex plane, mark in the positive acute angle α between the rays $0, x$ and $0, z$ and determine $\text{Arg}\,z$ geometrically, using the fact that $\alpha = \tan^{-1}(|y|/|x|)$, as in the following examples:

EXAMPLE 5.6.2 Determine the principal argument of z for the followig complex numbers:

$$z = 4 + 3i,\ -4 + 3i,\ -4 - 3i,\ 4 - 3i.$$

Solution. Referring to Figure 5.6, we see that $\text{Arg}\,z$ has the values

$$\alpha,\ \pi - \alpha,\ -\pi + \alpha,\ -\alpha,$$

where $\alpha = \tan^{-1}\frac{3}{4}$.

An important property of the argument of a complex number states that the sum of the arguments of two non–zero complex numbers is an argument of their product:

THEOREM 5.6.1 If θ_1 and θ_2 are arguments of z_1 and z_2, then $\theta_1 + \theta_2$ is an argument of $z_1 z_2$.

Proof. Let z_1 and z_2 have polar representations $z_1 = r_1(\cos\theta_1 + i\sin\theta_1)$ and $z_2 = r_2(\cos\theta_2 + i\sin\theta_2)$. Then

$$\begin{aligned}
z_1 z_2 &= r_1(\cos\theta_1 + i\sin\theta_1)r_2(\cos\theta_2 + i\sin\theta_2) \\
&= r_1 r_2(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2 + i(\cos\theta_1\sin\theta_2 + \sin\theta_1\cos\theta_2)) \\
&= r_1 r_2(\cos(\theta_1 + \theta_2) + i\sin(\theta_1 + \theta_2)),
\end{aligned}$$

which is the polar representation of $z_1 z_2$, as $r_1 r_2 = |z_1||z_2| = |z_1 z_2|$. Hence $\theta_1 + \theta_2$ is an argument of $z_1 z_2$.

An easy induction gives the following generalization to a product of n complex numbers:

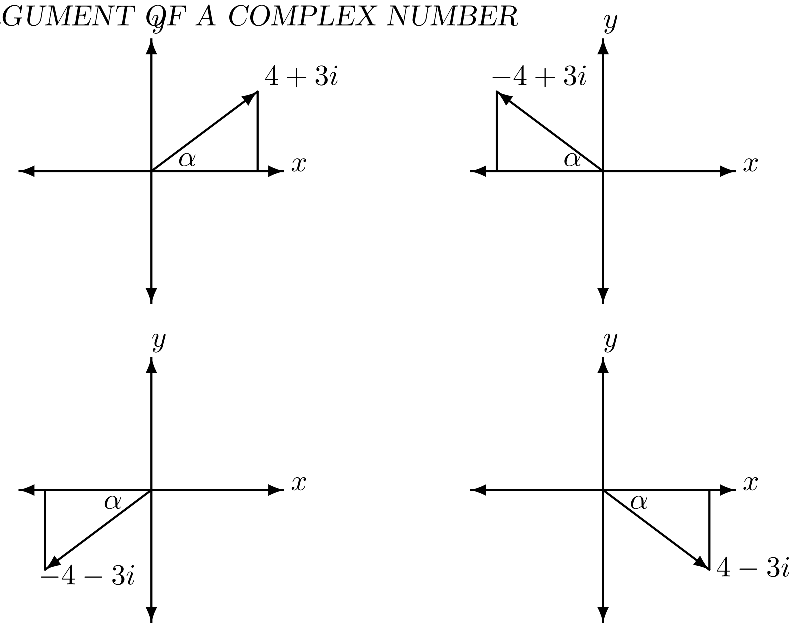

Figure 5.6: Argument examples.

COROLLARY 5.6.1 If $\theta_1, \ldots, \theta_n$ are arguments of $z_1, \ldots, z_n$ respectively, then $\theta_1 + \cdots + \theta_n$ is an argument for $z_1 \cdots z_n$.

Taking $\theta_1 = \cdots = \theta_n = \theta$ in the previous corollary gives

COROLLARY 5.6.2 If θ is an argument of z, then $n\theta$ is an argument for z^n.

THEOREM 5.6.2 If θ is an argument of the non–zero complex number z, then $-\theta$ is an argument of z^{-1}.

Proof. Let θ be an argument of z. Then $z = r(\cos\theta + i\sin\theta)$, where $r = |z|$. Hence

$$\begin{aligned} z^{-1} &= r^{-1}(\cos\theta + i\sin\theta)^{-1} \\ &= r^{-1}(\cos\theta - i\sin\theta) \\ &= r^{-1}(\cos(-\theta) + i\sin(-\theta)). \end{aligned}$$

Now $r^{-1} = |z|^{-1} = |z^{-1}|$, so $-\theta$ is an argument of z^{-1}.

COROLLARY 5.6.3 If θ_1 and θ_2 are arguments of z_1 and z_2, then $\theta_1 - \theta_2$ is an argument of z_1/z_2.

In terms of principal arguments, we have the following equations:

$$\begin{array}{lrcl}
\text{(i)} & \operatorname{Arg}(z_1 z_2) & = & \operatorname{Arg} z_1 + \operatorname{Arg} z_2 + 2k_1\pi, \\
\text{(ii)} & \operatorname{Arg}(z^{-1}) & = & -\operatorname{Arg} z + 2k_2\pi, \\
\text{(iii)} & \operatorname{Arg}(z_1/z_2) & = & \operatorname{Arg} z_1 - \operatorname{Arg} z_2 + 2k_3\pi, \\
\text{(iv)} & \operatorname{Arg}(z_1 \cdots z_n) & = & \operatorname{Arg} z_1 + \cdots + \operatorname{Arg} z_n + 2k_4\pi, \\
\text{(v)} & \operatorname{Arg}(z^n) & = & n \operatorname{Arg} z + 2k_5\pi,
\end{array}$$

where k_1, k_2, k_3, k_4, k_5 are integers.

In numerical examples, we can write (i), for example, as

$$\operatorname{Arg}(z_1 z_2) \equiv \operatorname{Arg} z_1 + \operatorname{Arg} z_2.$$

EXAMPLE 5.6.3 Find the modulus and principal argument of

$$z = \left(\frac{\sqrt{3}+i}{1+i} \right)^{17}$$

and hence express z in modulus–argument form.

Solution. $|z| = \dfrac{|\sqrt{3}+i|^{17}}{|1+i|^{17}} = \dfrac{2^{17}}{(\sqrt{2})^{17}} = 2^{17/2}$.

$$\begin{aligned}
\operatorname{Arg} z &\equiv 17\operatorname{Arg}\left(\frac{\sqrt{3}+i}{1+i} \right) \\
&= 17(\operatorname{Arg}(\sqrt{3}+i) - \operatorname{Arg}(1+i)) \\
&= 17\left(\frac{\pi}{6} - \frac{\pi}{4} \right) = \frac{-17\pi}{12}.
\end{aligned}$$

Hence $\operatorname{Arg} z = \left(\frac{-17\pi}{12}\right) + 2k\pi$, where k is an integer. We see that $k = 1$ and hence $\operatorname{Arg} z = \frac{7\pi}{12}$. Consequently $z = 2^{17/2}\left(\cos\frac{7\pi}{12} + i\sin\frac{7\pi}{12}\right)$.

DEFINITION 5.6.2 If θ is a real number, then we define $e^{i\theta}$ by

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

More generally, if $z = x + iy$, then we define e^z by

$$e^z = e^x e^{iy}.$$

For example,

$$e^{\frac{i\pi}{2}} = i,\ e^{i\pi} = -1,\ e^{-\frac{i\pi}{2}} = -i.$$

The following properties of the complex exponential function are left as exercises:

THEOREM 5.6.3

$$\begin{array}{lrcl}
\text{(i)} & e^{z_1}e^{z_2} & = & e^{z_1+z_2}, \\
\text{(ii)} & e^{z_1}\cdots e^{z_n} & = & e^{z_1+\cdots+z_n}, \\
\text{(iii)} & e^{z} & \neq & 0, \\
\text{(iv)} & (e^{z})^{-1} & = & e^{-z}, \\
\text{(v)} & e^{z_1}/e^{z_2} & = & e^{z_1-z_2}, \\
\text{(vi)} & \overline{e^{z}} & = & e^{\overline{z}}.
\end{array}$$

THEOREM 5.6.4 The equation

$$e^z = 1$$

has the complete solution $z = 2k\pi i$, $k \in \mathbb{Z}$.

Proof. First we observe that

$$e^{2k\pi i} = \cos\,(2k\pi) + i\sin\,(2k\pi) = 1.$$

Conversely, suppose $e^z = 1$, $z = x + iy$. Then $e^x(\cos y + i\sin y) = 1$. Hence $e^x\cos y = 1$ and $e^x\sin y = 0$. Hence $\sin y = 0$ and so $y = n\pi$, $n \in \mathbb{Z}$. Then $e^x\cos\,(n\pi) = 1$, so $e^x(-1)^n = 1$, from which follows $(-1)^n = 1$ as $e^x > 0$. Hence $n = 2k$, $k \in \mathbb{Z}$ and $e^x = 1$. Hence $x = 0$ and $z = 2k\pi i$.

5.7 De Moivre's theorem

The next theorem has many uses and is a special case of theorem 5.6.3(ii). Alternatively it can be proved directly by induction on n.

THEOREM 5.7.1 (De Moivre) If n is a positive integer, then

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

As a first application, we consider the equation $z^n = 1$.

THEOREM 5.7.2 The equation $z^n = 1$ has n distinct solutions, namely the complex numbers $\zeta_k = e^{\frac{2k\pi i}{n}}$, $k = 0, 1, \ldots, n-1$. These lie equally spaced on the unit circle $|z| = 1$ and are obtained by starting at 1, moving round the circle anti–clockwise, incrementing the argument in steps of $\frac{2\pi}{n}$. (See Figure 5.7)

We notice that the roots are the powers of the special root $\zeta = e^{\frac{2\pi i}{n}}$.

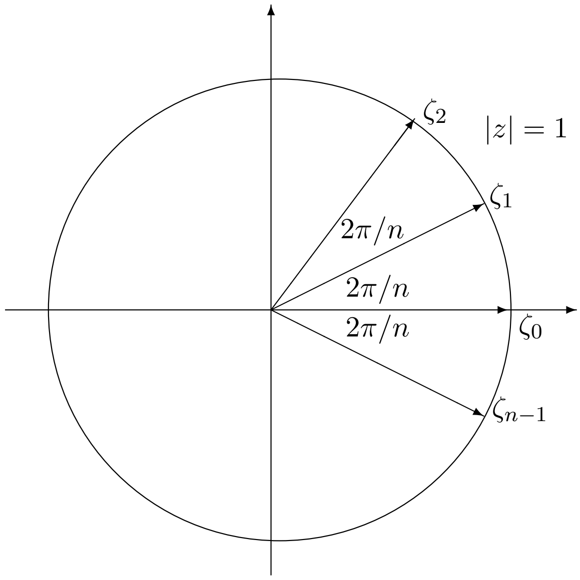

Figure 5.7: The nth roots of unity.

Proof. With ζ_k defined as above,

$$\zeta_k^n = \left(e^{\frac{2k\pi i}{n}}\right)^n = e^{\frac{2k\pi i}{n}n} = 1,$$

by De Moivre's theorem. However $|\zeta_k| = 1$ and $\arg \zeta_k = \frac{2k\pi}{n}$, so the complex numbers ζ_k, $k = 0, 1, \ldots, n-1$, lie equally spaced on the unit circle. Consequently these numbers must be precisely all the roots of $z^n - 1$. For the polynomial $z^n - 1$, being of degree n over a field, can have at most n distinct roots in that field.

The more general equation $z^n = a$, where $a \in, \mathbb{C}, a \neq 0$, can be reduced to the previous case:

Let α be argument of z, so that $a = |a|e^{i\alpha}$. Then if $w = |a|^{1/n}e^{\frac{i\alpha}{n}}$, we have

$$\begin{aligned} w^n &= \left(|a|^{1/n}e^{\frac{i\alpha}{n}}\right)^n \\ &= (|a|^{1/n})^n \left(e^{\frac{i\alpha}{n}}\right)^n \\ &= |a|e^{i\alpha} = a. \end{aligned}$$

So w is a particular solution. Substituting for a in the original equation, we get $z^n = w^n$, or $(z/w)^n = 1$. Hence the complete solution is $z/w =$

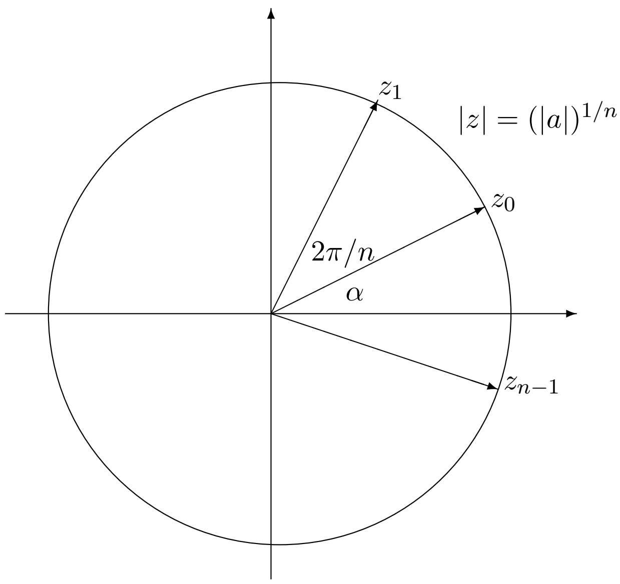

Figure 5.8: The roots of $z^n = a$.

$e^{\frac{2k\pi i}{n}}$, $k = 0,\ 1,\ \ldots, n-1$, or

$$z_k = |a|^{1/n} e^{\frac{i\alpha}{n}} e^{\frac{2k\pi i}{n}} = |a|^{1/n} e^{\frac{i(\alpha+2k\pi)}{n}}, \tag{5.4}$$

$k = 0,\ 1,\ \ldots, n-1$. So the roots are equally spaced on the circle

$$|z| = |a|^{1/n}$$

and are generated from the special solution with argument $(\arg a)/n$, by incrementing the argument in steps of $2\pi/n$. (See Figure 5.8.)

EXAMPLE 5.7.1 Factorize the polynomial $z^5 - 1$ as a product of real linear and quadratic factors.

Solution. The roots are 1, $e^{\frac{2\pi i}{5}}$, $e^{\frac{-2\pi i}{5}}$, $e^{\frac{4\pi i}{5}}$, $e^{\frac{-4\pi i}{5}}$, using the fact that non–real roots come in conjugate–complex pairs. Hence

$$z^5 - 1 = (z-1)(z - e^{\frac{2\pi i}{5}})(z - e^{\frac{-2\pi i}{5}})(z - e^{\frac{4\pi i}{5}})(z - e^{\frac{-4\pi i}{5}}).$$

Now

$$\begin{aligned}(z - e^{\frac{2\pi i}{5}})(z - e^{\frac{-2\pi i}{5}}) &= z^2 - z(e^{\frac{2\pi i}{5}} + e^{\frac{-2\pi i}{5}}) + 1 \\ &= z^2 - 2z\cos\tfrac{2\pi}{5} + 1.\end{aligned}$$

Similarly

$$(z - e^{\frac{4\pi i}{5}})(z - e^{\frac{-4\pi i}{5}}) = z^2 - 2z\cos\tfrac{4\pi}{5} + 1.$$

This gives the desired factorization.

EXAMPLE 5.7.2 Solve $z^3 = i$.

Solution. $|i| = 1$ and $\operatorname{Arg} i = \frac{\pi}{2} = \alpha$. So by equation 5.4, the solutions are

$$z_k = |i|^{1/3} e^{\frac{i(\alpha + 2k\pi)}{3}},\ k = 0,\ 1,\ 2.$$

First, $k = 0$ gives

$$z_0 = e^{\frac{i\pi}{6}} = \cos\frac{\pi}{6} + i\sin\frac{\pi}{6} = \frac{\sqrt{3}}{2} + \frac{i}{2}.$$

Next, $k = 1$ gives

$$z_1 = e^{\frac{5\pi i}{6}} = \cos\frac{5\pi}{6} + i\sin\frac{5\pi}{6} = \frac{-\sqrt{3}}{2} + \frac{i}{2}.$$

Finally, $k = 2$ gives

$$z_1 = e^{\frac{9\pi i}{6}} = \cos\frac{9\pi}{6} + i\sin\frac{9\pi}{6} = -i.$$

We finish this chapter with two more examples of De Moivre's theorem.

EXAMPLE 5.7.3 If

$$\begin{aligned} C &= 1 + \cos\theta + \cdots + \cos\,(n-1)\theta, \\ S &= \sin\theta + \cdots + \sin\,(n-1)\theta, \end{aligned}$$

prove that

$$C = \frac{\sin\frac{n\theta}{2}}{\sin\frac{\theta}{2}}\cos\tfrac{(n-1)\theta}{2} \text{ and } S = \frac{\sin\frac{n\theta}{2}}{\sin\frac{\theta}{2}}\sin\tfrac{(n-1)\theta}{2},$$

if $\theta \neq 2k\pi,\ k \in \mathbb{Z}$.

Solution.

$$\begin{aligned} C+iS &= 1+(\cos\theta+i\sin\theta)+\cdots+(\cos\,(n-1)\theta+i\sin\,(n-1)\theta)\\ &= 1+e^{i\theta}+\cdots+e^{i(n-1)\theta}\\ &= 1+z+\cdots+z^{n-1},\ \text{where } z=e^{i\theta}\\ &= \frac{1-z^n}{1-z},\ \text{if } z\neq 1,\ \text{i.e. } \theta\neq 2k\pi,\\ &= \frac{1-e^{in\theta}}{1-e^{i\theta}}=\frac{e^{\frac{in\theta}{2}}(e^{\frac{-in\theta}{2}}-e^{\frac{in\theta}{2}})}{e^{\frac{i\theta}{2}}(e^{\frac{-i\theta}{2}}-e^{\frac{i\theta}{2}})}\\ &= e^{i(n-1)\frac{\theta}{2}}\frac{\sin\frac{n\theta}{2}}{\sin\frac{\theta}{2}}\\ &= (\cos\,(n-1)\tfrac{\theta}{2}+i\sin\,(n-1)\tfrac{\theta}{2})\frac{\sin\frac{n\theta}{2}}{\sin\frac{\theta}{2}}. \end{aligned}$$

The result follows by equating real and imaginary parts.

EXAMPLE 5.7.4 Express $\cos\,n\theta$ and $\sin\,n\theta$ in terms of $\cos\theta$ and $\sin\theta$, using the equation $\cos\,n\theta+\sin\,n\theta=(\cos\theta+i\sin\theta)^n$.

Solution. The binomial theorem gives

$$(\cos\theta+i\sin\theta)^n=\cos^n\theta+\binom{n}{1}\cos^{n-1}\theta(i\sin\theta)+\binom{n}{2}\cos^{n-2}\theta(i\sin\theta)^2+\cdots + (i\sin\theta)^n.$$

Equating real and imaginary parts gives

$$\begin{aligned} \cos\,n\theta &= \cos^n\theta-\binom{n}{2}\cos^{n-2}\theta\sin^2\theta+\cdots\\ \sin\,n\theta &= \binom{n}{1}\cos^{n-1}\theta\sin\theta-\binom{n}{3}\cos^{n-3}\theta\sin^3\theta+\cdots . \end{aligned}$$

5.8 PROBLEMS

1. Express the following complex numbers in the form $x+iy$, x, y real:

   (i) $(-3+i)(14-2i)$; (ii) $\dfrac{2+3i}{1-4i}$; (iii) $\dfrac{(1+2i)^2}{1-i}$.

   [Answers: (i) $-40+20i$; (ii) $-\frac{10}{17}+\frac{11}{17}i$; (iii) $-\frac{7}{2}+\frac{i}{2}$.]

2. Solve the following equations:

$$\text{(i)} \qquad iz + (2 - 10i)z = 3z + 2i,$$

$$\text{(ii)} \qquad \begin{aligned} (1+i)z + (2-i)w &= -3i \\ (1+2i)z + (3+i)w &= 2+2i. \end{aligned}$$

[Answers:(i) $z = -\frac{9}{41} - \frac{i}{41}$; (ii) $z = -1 + 5i$, $w = \frac{19}{5} - \frac{8i}{5}$.]

3. Express $1 + (1+i) + (1+i)^2 + \ldots + (1+i)^{99}$ in the form $x + iy$, x, y real. [Answer: $(1 + 2^{50})i$.]

4. Solve the equations: (i) $z^2 = -8 - 6i$; (ii) $z^2 - (3+i)z + 4 + 3i = 0$.

   [Answers: (i) $z = \pm(1 - 3i)$; (ii) $z = 2 - i,\ 1 + 2i$.]

5. Find the modulus and principal argument of each of the following complex numbers:

   (i) $4 + i$; (ii) $-\frac{3}{2} - \frac{i}{2}$; (iii) $-1 + 2i$; (iv) $\frac{1}{2}(-1 + i\sqrt{3})$.

   [Answers: (i) $\sqrt{17}$, $\tan^{-1}\frac{1}{4}$; (ii) $\frac{\sqrt{10}}{2}$, $-\pi + \tan^{-1}\frac{1}{3}$; (iii) $\sqrt{5}$, $\pi - \tan^{-1} 2$.]

6. Express the following complex numbers in modulus-argument form:

   (i) $z = (1+i)(1+i\sqrt{3})(\sqrt{3} - i)$.

   (ii) $z = \dfrac{(1+i)^5(1-i\sqrt{3})^5}{(\sqrt{3}+i)^4}$.

   [Answers:
   (i) $z = 4\sqrt{2}(\cos\frac{5\pi}{12} + i\sin\frac{5\pi}{12})$; (ii) $z = 2^{7/2}(\cos\frac{11\pi}{12} + i\sin\frac{11\pi}{12})$.]

7. (i) If $z = 2(\cos\frac{\pi}{4} + i\sin\frac{\pi}{4})$ and $w = 3(\cos\frac{\pi}{6} + i\sin\frac{\pi}{6})$, find the polar form of
   (a) zw; (b) $\frac{z}{w}$; (c) $\frac{w}{z}$; (d) $\frac{z^5}{w^2}$.

   (ii) Express the following complex numbers in the form $x + iy$:
   (a) $(1+i)^{12}$; (b) $\left(\frac{1-i}{\sqrt{2}}\right)^{-6}$.

   [Answers: (i): (a) $6(\cos\frac{5\pi}{12} + i\sin\frac{5\pi}{12})$; (b) $\frac{2}{3}(\cos\frac{\pi}{12} + i\sin\frac{\pi}{12})$;

   (c) $\frac{3}{2}(\cos -\frac{\pi}{12} + i\sin -\frac{\pi}{12})$; (d) $\frac{32}{9}(\cos\frac{11\pi}{12} + i\sin\frac{11\pi}{12})$;

   (ii): (a) -64; (b) $-i$.]

8. Solve the equations:

$$\text{(i) } z^2 = 1 + i\sqrt{3}; \text{ (ii) } z^4 = i; \text{ (iii) } z^3 = -8i; \text{ (iv) } z^4 = 2 - 2i.$$

[Answers: (i) $z = \pm\frac{(\sqrt{3}+i)}{\sqrt{2}}$; (ii) $i^k(\cos\frac{\pi}{8} + i\sin\frac{\pi}{8}), k = 0, 1, 2, 3$; (iii) $z = 2i, -\sqrt{3}-i, \sqrt{3}-i$; (iv) $z = i^k 2^{\frac{3}{8}}(\cos\frac{\pi}{16} - i\sin\frac{\pi}{16}),\ k = 0, 1, 2, 3$.]

9. Find the reduced row–echelon form of the complex matrix

$$\begin{bmatrix} 2+i & -1+2i & 2 \\ 1+i & -1+i & 1 \\ 1+2i & -2+i & 1+i \end{bmatrix}.$$

[Answer: $\begin{bmatrix} 1 & i & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$.]

10. (i) Prove that the line equation $lx + my = n$ is equivalent to

$$\overline{p}z + p\overline{z} = 2n,$$

where $p = l + im$.

(ii) Use (ii) to deduce that reflection in the straight line

$$\overline{p}z + p\overline{z} = n$$

is described by the equation

$$\overline{p}w + p\overline{z} = n.$$

[Hint: The complex number $l + im$ is perpendicular to the given line.]

(iii) Prove that the line $|z-a| = |z-b|$ may be written as $\overline{p}z + p\overline{z} = n$, where $p = b - a$ and $n = |b|^2 - |a|^2$. Deduce that if z lies on the Apollonius circle $\frac{|z-a|}{|z-b|} = \lambda$, then w, the reflection of z in the line $|z-a| = |z-b|$, lies on the Apollonius circle $\frac{|z-a|}{|z-b|} = \frac{1}{\lambda}$.

11. Let a and b be distinct complex numbers and $0 < \alpha < \pi$.

(i) Prove that each of the following sets in the complex plane represents a circular arc and sketch the circular arcs on the same diagram:

$$\operatorname{Arg}\frac{z-a}{z-b} = \alpha,\ -\alpha,\ \pi-\alpha,\ \alpha-\pi.$$

Also show that $\operatorname{Arg}\dfrac{z-a}{z-b} = \pi$ represents the line segment joining a and b, while $\operatorname{Arg}\dfrac{z-a}{z-b} = 0$ represents the remaining portion of the line through a and b.

(ii) Use (i) to prove that four distinct points z_1, z_2, z_3, z_4 are concyclic or collinear, if and only if the *cross–ratio*

$$\frac{z_4-z_1}{z_4-z_2}\bigg/\frac{z_3-z_1}{z_3-z_2}$$

is real.

(iii) Use (ii) to derive *Ptolemy's* Theorem: Distinct points A, B, C, D are concyclic or collinear, if and only if one of the following holds:

$$\begin{aligned} AB\cdot CD + BC\cdot AD &= AC\cdot BD \\ BD\cdot AC + AD\cdot BC &= AB\cdot CD \\ BD\cdot AC + AB\cdot CD &= AD\cdot BC. \end{aligned}$$

Chapter 6

EIGENVALUES AND EIGENVECTORS

6.1 Motivation

We motivate the chapter on eigenvalues by discussing the equation

$$ax^2 + 2hxy + by^2 = c,$$

where not all of a, h, b are zero. The expression $ax^2 + 2hxy + by^2$ is called a *quadratic form* in x and y and we have the identity

$$ax^2 + 2hxy + by^2 = \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} a & h \\ h & b \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = X^tAX,$$

where $X = \begin{bmatrix} x \\ y \end{bmatrix}$ and $A = \begin{bmatrix} a & h \\ h & b \end{bmatrix}$. A is called the matrix of the quadratic form.

We now rotate the x, y axes anticlockwise through θ radians to new x_1, y_1 axes. The equations describing the rotation of axes are derived as follows:

Let P have coordinates (x, y) relative to the x, y axes and coordinates (x_1, y_1) relative to the x_1, y_1 axes. Then referring to Figure 6.1:

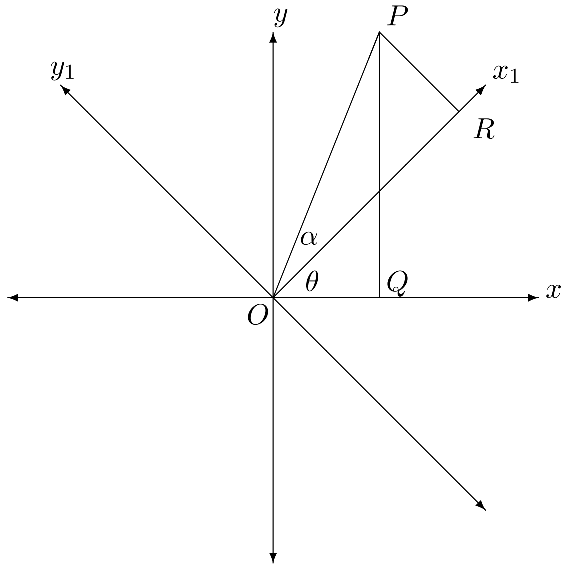

Figure 6.1: Rotating the axes.

$$\begin{aligned} x &= OQ = OP\cos\,(\theta+\alpha) \\ &= OP(\cos\theta\cos\alpha - \sin\theta\sin\alpha) \\ &= (OP\cos\alpha)\cos\theta - (OP\sin\alpha)\sin\theta \\ &= OR\cos\theta - PR\sin\theta \\ &= x_1\cos\theta - y_1\sin\theta. \end{aligned}$$

Similarly $y = x_1\sin\theta + y_1\cos\theta$.

We can combine these transformation equations into the single matrix equation:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix},$$

or $X = PY$, where $X = \begin{bmatrix} x \\ y \end{bmatrix}$, $Y = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$. We note that the columns of P give the directions of the positive x_1 and y_1 axes. Also P is an orthogonal matrix – we have $PP^t = I_2$ and so $P^{-1} = P^t$. The matrix P has the special property that $\det P = 1$.

A matrix of the type $P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$ is called a *rotation* matrix. We shall show soon that any 2×2 real orthogonal matrix with determinant equal to 1 is a rotation matrix.

We can also solve for the new coordinates in terms of the old ones:

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = Y = P^t X = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix},$$

so $x_1 = x\cos\theta + y\sin\theta$ and $y_1 = -x\sin\theta + y\cos\theta$. Then

$$X^t AX = (PY)^t A(PY) = Y^t(P^t AP)Y.$$

Now suppose, as we later show, that it is possible to choose an angle θ so that $P^t AP$ is a diagonal matrix, say $\text{diag}(\lambda_1, \lambda_2)$. Then

$$X^t AX = \begin{bmatrix} x_1 & y_1 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \lambda_1 x_1^2 + \lambda_2 y_1^2 \qquad (6.1)$$

and relative to the new axes, the equation $ax^2 + 2hxy + by^2 = c$ becomes $\lambda_1 x_1^2 + \lambda_2 y_1^2 = c$, which is quite easy to sketch. This curve is symmetrical about the x_1 and y_1 axes, with P_1 and P_2, the respective columns of P, giving the directions of the axes of symmetry.

Also it can be verified that P_1 and P_2 satisfy the equations

$$AP_1 = \lambda_1 P_1 \text{ and } AP_2 = \lambda_2 P_2.$$

These equations force a restriction on λ_1 and λ_2. For if $P_1 = \begin{bmatrix} u_1 \\ v_1 \end{bmatrix}$, the first equation becomes

$$\begin{bmatrix} a & h \\ h & b \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \lambda_1 \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} \text{ or } \begin{bmatrix} a-\lambda_1 & h \\ h & b-\lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ v_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Hence we are dealing with a homogeneous system of two linear equations in two unknowns, having a non–trivial solution (u_1, v_1). Hence

$$\begin{vmatrix} a-\lambda_1 & h \\ h & b-\lambda_1 \end{vmatrix} = 0.$$

Similarly, λ_2 satisfies the same equation. In expanded form, λ_1 and λ_2 satisfy

$$\lambda^2 - (a+b)\lambda + ab - h^2 = 0.$$

This equation has real roots

$$\lambda = \frac{a+b \pm \sqrt{(a+b)^2 - 4(ab-h^2)}}{2} = \frac{a+b \pm \sqrt{(a-b)^2 + 4h^2}}{2} \qquad (6.2)$$

(The roots are distinct if $a \neq b$ or $h \neq 0$. The case $a = b$ and $h = 0$ needs no investigation, as it gives an equation of a circle.)

The equation $\lambda^2 - (a+b)\lambda + ab - h^2 = 0$ is called the *eigenvalue equation* of the matrix A.

6.2 Definitions and examples

DEFINITION 6.2.1 (Eigenvalue, eigenvector) Let A be a complex square matrix. Then if λ is a complex number and X a *non–zero* complex column vector satisfying $AX = \lambda X$, we call X an *eigenvector* of A, while λ is called an *eigenvalue* of A. We also say that X is an eigenvector corresponding to the eigenvalue λ.

So in the above example P_1 and P_2 are eigenvectors corresponding to λ_1 and λ_2, respectively. We shall give an algorithm which starts from the eigenvalues of $A = \begin{bmatrix} a & h \\ h & b \end{bmatrix}$ and constructs a rotation matrix P such that P^tAP is diagonal.

As noted above, if λ is an eigenvalue of an $n \times n$ matrix A, with corresponding eigenvector X, then $(A - \lambda I_n)X = 0$, with $X \neq 0$, so $\det(A - \lambda I_n) = 0$ and there are at most n distinct eigenvalues of A.

Conversely if $\det(A - \lambda I_n) = 0$, then $(A - \lambda I_n)X = 0$ has a non–trivial solution X and so λ is an eigenvalue of A with X a corresponding eigenvector.

DEFINITION 6.2.2 (Characteristic polynomial, equation) The polynomial $\det(A - \lambda I_n)$ is called the *characteristic polynomial* of A and is often denoted by $\mathrm{ch}_A(\lambda)$. The equation $\det(A - \lambda I_n) = 0$ is called the *characteristic equation* of A. Hence the eigenvalues of A are the roots of the characteristic polynomial of A.

For a 2×2 matrix $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, it is easily verified that the characteristic polynomial is $\lambda^2 - (\mathrm{trace}\, A)\lambda + \det A$, where $\mathrm{trace}\, A = a + d$ is the sum of the diagonal elements of A.

EXAMPLE 6.2.1 Find the eigenvalues of $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ and find all eigenvectors.

Solution. The characteristic equation of A is $\lambda^2 - 4\lambda + 3 = 0$, or

$$(\lambda - 1)(\lambda - 3) = 0.$$

Hence $\lambda = 1$ or 3. The eigenvector equation $(A - \lambda I_n)X = 0$ reduces to

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

or

$$\begin{aligned} (2-\lambda)x + y &= 0 \\ x + (2-\lambda)y &= 0. \end{aligned}$$

Taking $\lambda = 1$ gives

$$\begin{aligned} x + y &= 0 \\ x + y &= 0, \end{aligned}$$

which has solution $x = -y$, y arbitrary. Consequently the eigenvectors corresponding to $\lambda = 1$ are the vectors $\begin{bmatrix} -y \\ y \end{bmatrix}$, with $y \neq 0$.

Taking $\lambda = 3$ gives

$$\begin{aligned} -x + y &= 0 \\ x - y &= 0, \end{aligned}$$

which has solution $x = y$, y arbitrary. Consequently the eigenvectors corresponding to $\lambda = 3$ are the vectors $\begin{bmatrix} y \\ y \end{bmatrix}$, with $y \neq 0$.

Our next result has wide applicability:

THEOREM 6.2.1 Let A be a 2×2 matrix having distinct eigenvalues λ_1 and λ_2 and corresponding eigenvectors X_1 and X_2. Let P be the matrix whose columns are X_1 and X_2, respectively. Then P is non–singular and

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Proof. Suppose $AX_1 = \lambda_1 X_1$ and $AX_2 = \lambda_2 X_2$. We show that the system of homogeneous equations

$$xX_1 + yX_2 = 0$$

has only the trivial solution. Then by theorem 2.5.10 the matrix $P = [X_1|X_2]$ is non–singular. So assume

$$xX_1 + yX_2 = 0. \tag{6.3}$$

Then $A(xX_1 + yX_2) = A0 = 0$, so $x(AX_1) + y(AX_2) = 0$. Hence

$$x\lambda_1 X_1 + y\lambda_2 X_2 = 0. \tag{6.4}$$

Multiplying equation 6.3 by λ_1 and subtracting from equation 6.4 gives

$$(\lambda_2 - \lambda_1)yX_2 = 0.$$

Hence $y = 0$, as $(\lambda_2 - \lambda_1) \neq 0$ and $X_2 \neq 0$. Then from equation 6.3, $xX_1 = 0$ and hence $x = 0$.

Then the equations $AX_1 = \lambda_1 X_1$ and $AX_2 = \lambda_2 X_2$ give

$$\begin{aligned} AP = A[X_1|X_2] = [AX_1|AX_2] &= [\lambda_1 X_1|\lambda_2 X_2] \\ &= [X_1|X_2]\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}, \end{aligned}$$

so

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

EXAMPLE 6.2.2 Let $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ be the matrix of example 6.2.1. Then $X_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ are eigenvectors corresponding to eigenvalues 1 and 3, respectively. Hence if $P = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$, we have

$$P^{-1}AP = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

There are two immediate applications of theorem 6.2.1. The first is to the calculation of A^n: If $P^{-1}AP = \text{diag}\,(\lambda_1, \lambda_2)$, then $A = P\text{diag}\,(\lambda_1, \lambda_2)P^{-1}$ and

$$A^n = \left(P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}P^{-1}\right)^n = P\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}^n P^{-1} = P\begin{bmatrix} \lambda_1^n & 0 \\ 0 & \lambda_2^n \end{bmatrix}P^{-1}.$$

The second application is to solving a system of linear differential equations

$$\begin{aligned} \frac{dx}{dt} &= ax + by \\ \frac{dy}{dt} &= cx + dy, \end{aligned}$$

where $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is a matrix of real or complex numbers and x and y are functions of t. The system can be written in matrix form as $\dot{X} = AX$, where

$$X = \begin{bmatrix} x \\ y \end{bmatrix} \text{ and } \dot{X} = \begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{bmatrix}.$$

We make the substitution $X = PY$, where $Y = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$. Then x_1 and y_1 are also functions of t and

$$\dot{X} = P\dot{Y} = AX = A(PY), \text{ so } \dot{Y} = (P^{-1}AP)Y = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} Y.$$

Hence $\dot{x_1} = \lambda_1 x_1$ and $\dot{y_1} = \lambda_2 y_1$.

These differential equations are well–known to have the solutions $x_1 = x_1(0)e^{\lambda_1 t}$ and $y_1 = y_1(0)e^{\lambda_2 t}$, where $x_1(0)$ is the value of x_1 when $t = 0$.

[If $\frac{dx}{dt} = kx$, where k is a constant, then

$$\frac{d}{dt}\left(e^{-kt}x\right) = -ke^{-kt}x + e^{-kt}\frac{dx}{dt} = -ke^{-kt}x + e^{-kt}kx = 0.$$

Hence $e^{-kt}x$ is constant, so $e^{-kt}x = e^{-k0}x(0) = x(0)$. Hence $x = x(0)e^{kt}$.]

However $\begin{bmatrix} x_1(0) \\ y_1(0) \end{bmatrix} = P^{-1}\begin{bmatrix} x(0) \\ y(0) \end{bmatrix}$, so this determines $x_1(0)$ and $y_1(0)$ in terms of $x(0)$ and $y(0)$. Hence ultimately x and y are determined as explicit functions of t, using the equation $X = PY$.

EXAMPLE 6.2.3 Let $A = \begin{bmatrix} 2 & -3 \\ 4 & -5 \end{bmatrix}$. Use the eigenvalue method to derive an explicit formula for A^n and also solve the system of differential equations

$$\begin{aligned} \frac{dx}{dt} &= 2x - 3y \\ \frac{dy}{dt} &= 4x - 5y, \end{aligned}$$

given $x = 7$ and $y = 13$ when $t = 0$.

Solution. The characteristic polynomial of A is $\lambda^2+3\lambda+2$ which has distinct roots $\lambda_1 = -1$ and $\lambda_2 = -2$. We find corresponding eigenvectors $X_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $X_2 = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$. Hence if $P = \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}$, we have $P^{-1}AP = \text{diag}(-1, -2)$. Hence

$$\begin{aligned} A^n &= \left(P\text{diag}(-1, -2)P^{-1}\right)^n = P\text{diag}((-1)^n, (-2)^n)P^{-1} \\ &= \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} (-1)^n & 0 \\ 0 & (-2)^n \end{bmatrix}\begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix} \end{aligned}$$

$$\begin{aligned} &= (-1)^n \begin{bmatrix} 1 & 3 \\ 1 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 2^n \end{bmatrix} \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix} \\ &= (-1)^n \begin{bmatrix} 1 & 3 \times 2^n \\ 1 & 4 \times 2^n \end{bmatrix} \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix} \\ &= (-1)^n \begin{bmatrix} 4 - 3 \times 2^n & -3 + 3 \times 2^n \\ 4 - 4 \times 2^n & -3 + 4 \times 2^n \end{bmatrix}. \end{aligned}$$

To solve the differential equation system, make the substitution $X = PY$. Then $x = x_1 + 3y_1$, $y = x_1 + 4y_1$. The system then becomes

$$\begin{aligned} \dot{x}_1 &= -x_1 \\ \dot{y}_1 &= -2y_1, \end{aligned}$$

so $x_1 = x_1(0)e^{-t}$, $y_1 = y_1(0)e^{-2t}$. Now

$$\begin{bmatrix} x_1(0) \\ y_1(0) \end{bmatrix} = P^{-1} \begin{bmatrix} x(0) \\ y(0) \end{bmatrix} = \begin{bmatrix} 4 & -3 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 7 \\ 13 \end{bmatrix} = \begin{bmatrix} -11 \\ 6 \end{bmatrix},$$

so $x_1 = -11e^{-t}$ and $y_1 = 6e^{-2t}$. Hence $x = -11e^{-t} + 3(6e^{-2t}) = -11e^{-t} + 18e^{-2t}$, $y = -11e^{-t} + 4(6e^{-2t}) = -11e^{-t} + 24e^{-2t}$.

For a more complicated example we solve a system of *inhomogeneous* recurrence relations.

EXAMPLE 6.2.4 Solve the system of recurrence relations

$$\begin{aligned} x_{n+1} &= 2x_n - y_n - 1 \\ y_{n+1} &= -x_n + 2y_n + 2, \end{aligned}$$

given that $x_0 = 0$ and $y_0 = -1$.

Solution. The system can be written in matrix form as

$$X_{n+1} = AX_n + B,$$

where

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \text{ and } B = \begin{bmatrix} -1 \\ 2 \end{bmatrix}.$$

It is then an easy induction to prove that

$$X_n = A^n X_0 + (A^{n-1} + \cdots + A + I_2)B. \qquad (6.5)$$

Also it is easy to verify by the eigenvalue method that

$$A^n = \frac{1}{2}\begin{bmatrix} 1+3^n & 1-3^n \\ 1-3^n & 1+3^n \end{bmatrix} = \frac{1}{2}U + \frac{3^n}{2}V,$$

where $U = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $V = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$. Hence

$$\begin{aligned} A^{n-1} + \cdots + A + I_2 &= \frac{n}{2}U + \frac{(3^{n-1} + \cdots + 3 + 1)}{2}V \\ &= \frac{n}{2}U + \frac{(3^n - 1)}{4}V. \end{aligned}$$

Then equation 6.5 gives

$$X_n = \left(\frac{1}{2}U + \frac{3^n}{2}V\right)\begin{bmatrix} 0 \\ -1 \end{bmatrix} + \left(\frac{n}{2}U + \frac{(3^n-1)}{4}V\right)\begin{bmatrix} -1 \\ 2 \end{bmatrix},$$

which simplifies to

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \begin{bmatrix} (2n+1-3^n)/4 \\ (2n-5+3^n)/4 \end{bmatrix}.$$

Hence $x_n = (2n+1-3^n)/4$ and $y_n = (2n-5+3^n)/4$.

REMARK 6.2.1 If $(A - I_2)^{-1}$ existed (that is, if $\det(A - I_2) \neq 0$, or equivalently, if 1 is not an eigenvalue of A), then we could have used the formula

$$A^{n-1} + \cdots + A + I_2 = (A^n - I_2)(A - I_2)^{-1}. \tag{6.6}$$

However the eigenvalues of A are 1 and 3 in the above problem, so formula 6.6 cannot be used there.

Our discussion of eigenvalues and eigenvectors has been limited to 2×2 matrices. The discussion is more complicated for matrices of size greater than two and is best left to a second course in linear algebra. Nevertheless the following result is a useful generalization of theorem 6.2.1. The reader is referred to [28, page 350] for a proof.

THEOREM 6.2.2 Let A be an $n \times n$ matrix having distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ and corresponding eigenvectors $X_1, \ldots, X_n$. Let P be the matrix whose columns are respectively $X_1, \ldots, X_n$. Then P is non–singular and

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Another useful result which covers the case where there are multiple eigenvalues is the following (The reader is referred to [28, pages 351–352] for a proof):

THEOREM 6.2.3 Suppose the characteristic polynomial of A has the factorization

$$\det(\lambda I_n - A) = (\lambda - c_1)^{n_1} \cdots (\lambda - c_t)^{n_t},$$

where $c_1, \ldots, c_t$ are the distinct eigenvalues of A. Suppose that for $i = 1, \ldots, t$, we have $\text{nullity}\,(c_i I_n - A) = n_i$. For each such i, choose a basis $X_{i1}, \ldots, X_{in_i}$ for the *eigenspace* $N(c_i I_n - A)$. Then the matrix

$$P = [X_{11}|\cdots|X_{1n_1}|\cdots|X_{t1}|\cdots|X_{tn_t}]$$

is non–singular and $P^{-1}AP$ is the following diagonal matrix

$$P^{-1}AP = \begin{bmatrix} c_1 I_{n_1} & 0 & \cdots & 0 \\ 0 & c_2 I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & c_t I_{n_t} \end{bmatrix}.$$

(The notation means that on the diagonal there are n_1 elements c_1, followed by n_2 elements c_2,..., n_t elements c_t.)

6.3 PROBLEMS

1. Let $A = \begin{bmatrix} 4 & -3 \\ 1 & 0 \end{bmatrix}$. Find an invertible matrix P such that $P^{-1}AP = \text{diag}\,(1,\,3)$ and hence prove that

$$A^n = \frac{3^n - 1}{2}A + \frac{3 - 3^n}{2}I_2.$$

2. If $A = \begin{bmatrix} 0.6 & 0.8 \\ 0.4 & 0.2 \end{bmatrix}$, prove that A^n tends to a limiting matrix

$$\begin{bmatrix} 2/3 & 2/3 \\ 1/3 & 1/3 \end{bmatrix}$$

as $n \to \infty$.

3. Solve the system of differential equations

$$\frac{dx}{dt} = 3x - 2y$$
$$\frac{dy}{dt} = 5x - 4y,$$

given $x = 13$ and $y = 22$ when $t = 0$.
[Answer: $x = 7e^t + 6e^{-2t}$, $y = 7e^t + 15e^{-2t}$.]

4. Solve the system of recurrence relations

$$x_{n+1} = 3x_n - y_n$$
$$y_{n+1} = -x_n + 3y_n,$$

given that $x_0 = 1$ and $y_0 = 2$.
[Answer: $x_n = 2^{n-1}(3 - 2^n)$, $y_n = 2^{n-1}(3 + 2^n)$.]

5. Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ be a real or complex matrix with distinct eigenvalues λ_1, λ_2 and corresponding eigenvectors X_1, X_2. Also let $P = [X_1|X_2]$.

(a) Prove that the system of recurrence relations

$$x_{n+1} = ax_n + by_n$$
$$y_{n+1} = cx_n + dy_n$$

has the solution

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = \alpha\lambda_1^n X_1 + \beta\lambda_2^n X_2,$$

where α and β are determined by the equation

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = P^{-1} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}.$$

(b) Prove that the system of differential equations

$$\frac{dx}{dt} = ax + by$$
$$\frac{dy}{dt} = cx + dy$$

has the solution

$$\begin{bmatrix} x \\ y \end{bmatrix} = \alpha e^{\lambda_1 t} X_1 + \beta e^{\lambda_2 t} X_2,$$

where α and β are determined by the equation

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = P^{-1} \begin{bmatrix} x(0) \\ y(0) \end{bmatrix}.$$

6. Let $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ be a real matrix with non–real eigenvalues $\lambda = a + ib$ and $\overline{\lambda} = a - ib$, with corresponding eigenvectors $X = U + iV$ and $\overline{X} = U - iV$, where U and V are real vectors. Also let P be the real matrix defined by $P = [U|V]$. Finally let $a + ib = re^{i\theta}$, where $r > 0$ and θ is real.

   (a) Prove that

$$\begin{aligned} AU &= aU - bV \\ AV &= bU + aV. \end{aligned}$$

   (b) Deduce that

$$P^{-1}AP = \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

   (c) Prove that the system of recurrence relations

$$\begin{aligned} x_{n+1} &= a_{11}x_n + a_{12}y_n \\ y_{n+1} &= a_{21}x_n + a_{22}y_n \end{aligned}$$

   has the solution

$$\begin{bmatrix} x_n \\ y_n \end{bmatrix} = r^n\{(\alpha U + \beta V)\cos n\theta + (\beta U - \alpha V)\sin n\theta\},$$

   where α and β are determined by the equation

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = P^{-1} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix}.$$

   (d) Prove that the system of differential equations

$$\begin{aligned} \frac{dx}{dt} &= ax + by \\ \frac{dy}{dt} &= cx + dy \end{aligned}$$

has the solution

$$\begin{bmatrix} x \\ y \end{bmatrix} = e^{at}\{(\alpha U + \beta V)\cos bt + (\beta U - \alpha V)\sin bt\},$$

where α and β are determined by the equation

$$\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = P^{-1}\begin{bmatrix} x(0) \\ y(0) \end{bmatrix}.$$

[Hint: Let $\begin{bmatrix} x \\ y \end{bmatrix} = P\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$. Also let $z = x_1 + iy_1$. Prove that

$$\dot{z} = (a - ib)z$$

and deduce that

$$x_1 + iy_1 = e^{at}(\alpha + i\beta)(\cos bt + i\sin bt).$$

Then equate real and imaginary parts to solve for x_1, y_1 and hence x, y.]

7. (The case of repeated eigenvalues.) Let $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ and suppose that the characteristic polynomial of A, $\lambda^2 - (a+d)\lambda + (ad - bc)$, has a repeated root α. Also assume that $A \neq \alpha I_2$. Let $B = A - \alpha I_2$.

   (i) Prove that $(a-d)^2 + 4bc = 0$.

   (ii) Prove that $B^2 = 0$.

   (iii) Prove that $BX_2 \neq 0$ for some vector X_2; indeed, show that X_2 can be taken to be $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

   (iv) Let $X_1 = BX_2$. Prove that $P = [X_1|X_2]$ is non–singular,

   $$AX_1 = \alpha X_1 \text{ and } AX_2 = \alpha X_2 + X_1$$

   and deduce that

   $$P^{-1}AP = \begin{bmatrix} \alpha & 1 \\ 0 & \alpha \end{bmatrix}.$$

8. Use the previous result to solve system of the differential equations

$$\begin{aligned} \frac{dx}{dt} &= 4x - y \\ \frac{dy}{dt} &= 4x + 8y, \end{aligned}$$

given that $x = 1 = y$ when $t = 0$.

[To solve the differential equation

$$\frac{dx}{dt} - kx = f(t), \quad k \text{ a constant},$$

multiply throughout by e^{-kt}, thereby converting the left–hand side to $\frac{dx}{dt}(e^{-kt}x)$.]

[Answer: $x = (1 - 3t)e^{6t}$, $y = (1 + 6t)e^{6t}$.]

9. Let

$$A = \begin{bmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/4 & 1/2 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}.$$

   (a) Verify that $\det(\lambda I_3 - A)$, the characteristic polynomial of A, is given by

$$(\lambda - 1)\lambda(\lambda - \frac{1}{4}).$$

   (b) Find a non–singular matrix P such that $P^{-1}AP = \operatorname{diag}(1, 0, \frac{1}{4})$.

   (c) Prove that

$$A^n = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} + \frac{1}{3 \cdot 4^n}\begin{bmatrix} 2 & 2 & -4 \\ -1 & -1 & 2 \\ -1 & -1 & 2 \end{bmatrix}$$

   if $n \geq 1$.

10. Let

$$A = \begin{bmatrix} 5 & 2 & -2 \\ 2 & 5 & -2 \\ -2 & -2 & 5 \end{bmatrix}.$$

   (a) Verify that $\det(\lambda I_3 - A)$, the characteristic polynomial of A, is given by

$$(\lambda - 3)^2(\lambda - 9).$$

   (b) Find a non–singular matrix P such that $P^{-1}AP = \operatorname{diag}(3, 3, 9)$.

Chapter 7

Identifying second degree equations

7.1 The eigenvalue method

In this section we apply eigenvalue methods to determine the geometrical nature of the second degree equation

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0, \tag{7.1}$$

where not all of a, h, b are zero.

Let $A = \begin{bmatrix} a & h \\ h & b \end{bmatrix}$ be the matrix of the quadratic form $ax^2 + 2hxy + by^2$. We saw in section 6.1, equation 6.2 that A has real eigenvalues λ_1 and λ_2, given by

$$\lambda_1 = \frac{a + b - \sqrt{(a-b)^2 + 4h^2}}{2}, \ \lambda_2 = \frac{a + b + \sqrt{(a-b)^2 + 4h^2}}{2}.$$

We show that it is always possible to rotate the x, y axes to x_1, y_1 axes whose positive directions are determined by eigenvectors X_1 and X_2 corresponding to $\lambda_1$and λ_2 in such a way that relative to the x_1, y_1 axes, equation 7.1 takes the form

$$a'x^2 + b'y^2 + 2g'x + 2f'y + c = 0. \tag{7.2}$$

Then by completing the square and suitably translating the x_1, y_1 axes, to new x_2, y_2 axes, equation 7.2 can be reduced to one of several standard forms, each of which is easy to sketch. We need some preliminary definitions.

DEFINITION 7.1.1 (Orthogonal matrix) An $n \times n$ real matrix P is called *orthogonal* if

$$P^t P = I_n.$$

It follows that if P is orthogonal, then $\det P = \pm 1$. For

$$\det(P^t P) = \det P^t \det P = (\det P)^2,$$

so $(\det P)^2 = \det I_n = 1$. Hence $\det P = \pm 1$.

If P is an orthogonal matrix with $\det P = 1$, then P is called a *proper* orthogonal matrix.

THEOREM 7.1.1 If P is a 2×2 orthogonal matrix with $\det P = 1$, then

$$P = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

for some θ.

REMARK 7.1.1 Hence, by the discusssion at the beginning of Chapter 6, if P is a proper orthogonal matrix, the coordinate transformation

$$\begin{bmatrix} x \\ y \end{bmatrix} = P \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

represents a rotation of the axes, with new x_1 and y_1 axes given by the repective columns of P.

Proof. Suppose that $P^t P = I_2$, where $\Delta = \det P = 1$. Let

$$P = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

Then the equation

$$P^t = P^{-1} = \frac{1}{\Delta} \operatorname{adj} P$$

gives

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

Hence $a = d$, $b = -c$ and so

$$P = \begin{bmatrix} a & -c \\ c & a \end{bmatrix},$$

where $a^2 + c^2 = 1$. But then the point (a, c) lies on the unit circle, so $a = \cos\theta$ and $c = \sin\theta$, where θ is uniquely determined up to multiples of 2π.

DEFINITION 7.1.2 (Dot product). If $X = \begin{bmatrix} a \\ b \end{bmatrix}$ and $Y = \begin{bmatrix} c \\ d \end{bmatrix}$, then $X \cdot Y$, the *dot product* of X and Y, is defined by

$$X \cdot Y = ac + bd.$$

The dot product has the following properties:

(i) $X \cdot (Y + Z) = X \cdot Y + X \cdot Z$;

(ii) $X \cdot Y = Y \cdot X$;

(iii) $(tX) \cdot Y = t(X \cdot Y)$;

(iv) $X \cdot X = a^2 + b^2$ if $X = \begin{bmatrix} a \\ b \end{bmatrix}$;

(v) $X \cdot Y = X^t Y$.

The *length* of X is defined by

$$||X|| = \sqrt{a^2 + b^2} = (X \cdot X)^{1/2}.$$

We see that $||X||$ is the distance between the origin $O = (0, 0)$ and the point (a, b).

THEOREM 7.1.2 (Geometrical meaning of the dot product) Let $A = (x_1, y_1)$ and $B = (x_2, y_2)$ be points, each distinct from the origin $O = (0, 0)$. Then if $X = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $Y = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, we have

$$X \cdot Y = OA \cdot OB \cos\theta,$$

where θ is the angle between the rays OA and OB.

Proof. By the cosine law applied to triangle OAB, we have

$$AB^2 = OA^2 + OB^2 - 2OA \cdot OB \cos\theta. \tag{7.3}$$

Now $AB^2 = (x_2 - x_1)^2 + (y_2 - y_1)^2$, $OA^2 = x_1^2 + y_1^2$, $OB^2 = x_2^2 + y_2^2$.

Substituting in equation 7.3 then gives

$$(x_2 - x_1)^2 + (y_2 - y_1)^2 = (x_1^2 + y_1^2) + (x_2^2 + y_2^2) - 2OA \cdot OB \cos\theta,$$

which simplifies to give

$$OA \cdot OB \cos\theta = x_1x_2 + y_1y_2 = X \cdot Y.$$

It follows from theorem 7.1.2 that if $A = (x_1, y_1)$ and $B = (x_2, y_2)$ are points distinct from $O = (0, 0)$ and $X = \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$ and $Y = \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}$, then $X \cdot Y = 0$ means that the rays OA and OB are perpendicular. This is the reason for the following definition:

DEFINITION 7.1.3 (Orthogonal vectors) Vectors X and Y are called orthogonal if

$$X \cdot Y = 0.$$

There is also a connection with orthogonal matrices:

THEOREM 7.1.3 Let P be a 2×2 real matrix. Then P is an orthogonal matrix if and only if the columns of P are orthogonal and have unit length.

Proof. P is orthogonal if and only if $P^tP = I_2$. Now if $P = [X_1|X_2]$, the matrix P^tP is an important matrix called the *Gram* matrix of the column vectors X_1 and X_2. It is easy to prove that

$$P^tP = [X_i \cdot X_j] = \begin{bmatrix} X_1 \cdot X_1 & X_1 \cdot X_2 \\ X_2 \cdot X_1 & X_2 \cdot X_2 \end{bmatrix}.$$

Hence the equation $P^tP = I_2$ is equivalent to

$$\begin{bmatrix} X_1 \cdot X_1 & X_1 \cdot X_2 \\ X_2 \cdot X_1 & X_2 \cdot X_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

or, equating corresponding elements of both sides:

$$X_1 \cdot X_1 = 1,\ X_1 \cdot X_2 = 0,\ X_2 \cdot X_2 = 1,$$

which says that the columns of P are orthogonal and of unit length.

The next theorem describes a fundamental property of real symmetric matrices and the proof generalizes to symmetric matrices of any size.

THEOREM 7.1.4 If X_1 and X_2 are eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2 of a real symmetric matrix A, then X_1 and X_2 are orthogonal vectors.

Proof. Suppose

$$AX_1 = \lambda_1 X_1,\ AX_2 = \lambda_2 X_2, \tag{7.4}$$

where X_1 and X_2 are non–zero column vectors, $A^t = A$ and $\lambda_1 \neq \lambda_2$.

We have to prove that $X_1^t X_2 = 0$. From equation 7.4,

$$X_2^t A X_1 = \lambda_1 X_2^t X_1 \tag{7.5}$$

and

$$X_1^t A X_2 = \lambda_2 X_1^t X_2. \tag{7.6}$$

From equation 7.5, taking transposes,

$$(X_2^t A X_1)^t = (\lambda_1 X_2^t X_1)^t$$

so

$$X_1^t A^t X_2 = \lambda_1 X_1^t X_2.$$

Hence

$$X_1^t A X_2 = \lambda_1 X_1^t X_2. \tag{7.7}$$

Finally, subtracting equation 7.6 from equation 7.7, we have

$$(\lambda_1 - \lambda_2) X_1^t X_2 = 0$$

and hence, since $\lambda_1 \neq \lambda_2$,

$$X_1^t X_2 = 0.$$

THEOREM 7.1.5 Let A be a real 2×2 symmetric matrix with distinct eigenvalues λ_1 and λ_2. Then a proper orthogonal 2×2 matrix P exists such that

$$P^t A P = \operatorname{diag}(\lambda_1,\ \lambda_2).$$

Also the rotation of axes

$$\begin{bmatrix} x \\ y \end{bmatrix} = P \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

"diagonalizes" the quadratic form corresponding to A:

$$X^t A X = \lambda_1 x_1^2 + \lambda_2 y_1^2.$$

Proof. Let X_1 and X_2 be eigenvectors corresponding to λ_1 and λ_2. Then by theorem 7.1.4, X_1 and X_2 are orthogonal. By dividing X_1 and X_2 by their lengths (i.e. *normalizing* X_1 and X_2) if necessary, we can assume that X_1 and X_2 have unit length. Then by theorem 7.1.1, $P = [X_1|X_2]$ is an orthogonal matrix. By replacing X_1 by $-X_1$, if necessary, we can assume that $\det P = 1$. Then by theorem 6.2.1, we have

$$P^tAP = P^{-1}AP = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

Also under the rotation $X = PY$,

$$\begin{aligned} X^tAX &= (PY)^tA(PY) = Y^t(P^tAP)Y = Y^t \operatorname{diag}(\lambda_1, \lambda_2)Y \\ &= \lambda_1 x_1^2 + \lambda_2 y_1^2. \end{aligned}$$

EXAMPLE 7.1.1 Let A be the symmetric matrix

$$A = \begin{bmatrix} 12 & -6 \\ -6 & 7 \end{bmatrix}.$$

Find a proper orthogonal matrix P such that P^tAP is diagonal.

Solution. The characteristic equation of A is $\lambda^2 - 19\lambda + 48 = 0$, or

$$(\lambda - 16)(\lambda - 3) = 0.$$

Hence A has distinct eigenvalues $\lambda_1 = 16$ and $\lambda_2 = 3$. We find corresponding eigenvectors

$$X_1 = \begin{bmatrix} -3 \\ 2 \end{bmatrix} \text{ and } X_2 = \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Now $||X_1|| = ||X_2|| = \sqrt{13}$. So we take

$$X_1 = \frac{1}{\sqrt{13}} \begin{bmatrix} -3 \\ 2 \end{bmatrix} \text{ and } X_2 = \frac{1}{\sqrt{13}} \begin{bmatrix} 2 \\ 3 \end{bmatrix}.$$

Then if $P = [X_1|X_2]$, the proof of theorem 7.1.5 shows that

$$P^tAP = \begin{bmatrix} 16 & 0 \\ 0 & 3 \end{bmatrix}.$$

However $\det P = -1$, so replacing X_1 by $-X_1$ will give $\det P = 1$.

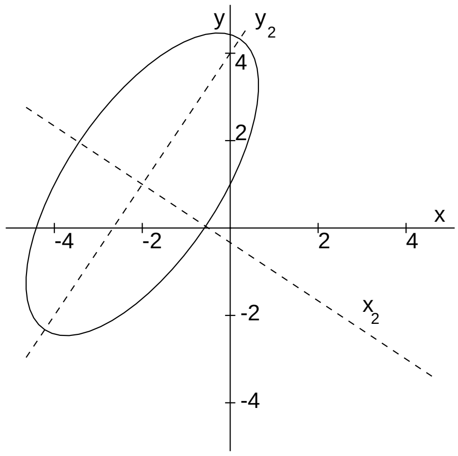

Figure 7.1: $12x^2 - 12xy + 7y^2 + 60x - 38y + 31 = 0$.

REMARK 7.1.2 (A shortcut) Once we have determined one eigenvector $X_1 = \begin{bmatrix} a \\ b \end{bmatrix}$, the other can be taken to be $\begin{bmatrix} -b \\ a \end{bmatrix}$, as these vectors are always orthogonal. Also $P = [X_1|X_2]$ will have $\det P = a^2 + b^2 > 0$.

We now apply the above ideas to determine the geometric nature of second degree equations in x and y.

EXAMPLE 7.1.2 Sketch the curve determined by the equation

$$12x^2 - 12xy + 7y^2 + 60x - 38y + 31 = 0.$$

Solution. With P taken to be the proper orthogonal matrix defined in the previous example by

$$P = \begin{bmatrix} 3/\sqrt{13} & 2/\sqrt{13} \\ -2/\sqrt{13} & 3/\sqrt{13} \end{bmatrix},$$

then as theorem 7.1.1 predicts, P is a rotation matrix and the transformation

$$X = \begin{bmatrix} x \\ y \end{bmatrix} = PY = P \begin{bmatrix} x_1 \\ y_1 \end{bmatrix}$$

or more explicitly

$$x = \frac{3x_1 + 2y_1}{\sqrt{13}},\ y = \frac{-2x_1 + 3y_1}{\sqrt{13}}, \tag{7.8}$$

will rotate the x, y axes to positions given by the respective columns of P. (More generally, we can always arrange for the x_1 axis to point either into the first or fourth quadrant.)

Now $A = \begin{bmatrix} 12 & -6 \\ -6 & 7 \end{bmatrix}$ is the matrix of the quadratic form

$$12x^2 - 12xy + 7y^2,$$

so we have, by Theorem 7.1.5

$$12x^2 - 12xy + 7y^2 = 16x_1^2 + 3y_1^2.$$

Then under the rotation $X = PY$, our original quadratic equation becomes

$$16x_1^2 + 3y_1^2 + \frac{60}{\sqrt{13}}(3x_1 + 2y_1) - \frac{38}{\sqrt{13}}(-2x_1 + 3y_1) + 31 = 0,$$

or

$$16x_1^2 + 3y_1^2 + \frac{256}{\sqrt{13}}x_1 + \frac{6}{\sqrt{13}}y_1 + 31 = 0.$$

Now complete the square in x_1 and y_1:

$$16\left(x_1^2 + \frac{16}{\sqrt{13}}x_1\right) + 3\left(y_1^2 + \frac{2}{\sqrt{13}}y_1\right) + 31 = 0,$$

$$\begin{aligned} 16\left(x_1 + \frac{8}{\sqrt{13}}\right)^2 + 3\left(y_1 + \frac{1}{\sqrt{13}}\right)^2 &= 16\left(\frac{8}{\sqrt{13}}\right)^2 + 3\left(\frac{1}{\sqrt{13}}\right)^2 - 31 \\ &= 48. \end{aligned} \tag{7.9}$$

Then if we perform a translation of axes to the new origin $(x_1, y_1) = (-\frac{8}{\sqrt{13}}, -\frac{1}{\sqrt{13}})$:

$$x_2 = x_1 + \frac{8}{\sqrt{13}},\ y_2 = y_1 + \frac{1}{\sqrt{13}},$$

equation 7.9 reduces to

$$16x_2^2 + 3y_2^2 = 48,$$

or

$$\frac{x_2^2}{3} + \frac{y_2^2}{16} = 1.$$

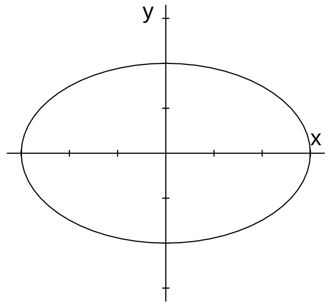

Figure 7.2: $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$, $0 < b < a$: an ellipse.

This equation is now in one of the standard forms listed below as Figure 7.2 and is that of a whose centre is at $(x_2, y_2) = (0, 0)$ and whose axes of symmetry lie along the x_2, y_2 axes. In terms of the original x, y coordinates, we find that the centre is $(x, y) = (-2, 1)$. Also $Y = P^tX$, so equations 7.8 can be solved to give

$$x_1 = \frac{3x - 2y}{\sqrt{13}}, \; y_1 = \frac{2x + 3y}{\sqrt{13}}.$$

Hence the y_2–axis is given by

$$\begin{aligned} 0 = x_2 &= x_1 + \frac{8}{\sqrt{13}} \\ &= \frac{3x - 2y}{\sqrt{13}} + \frac{8}{\sqrt{13}}, \end{aligned}$$

or $3x - 2y + 8 = 0$. Similarly the x_2–axis is given by $2x + 3y + 1 = 0$.

This ellipse is sketched in Figure 7.1.

Figures 7.2, 7.3, 7.4 and 7.5 are a collection of standard second degree equations: Figure 7.2 is an ellipse; Figures 7.3 are hyperbolas (in both these examples, the asymptotes are the lines $y = \pm\frac{b}{a}x$); Figures 7.4 and 7.5 represent parabolas.

EXAMPLE 7.1.3 Sketch $y^2 - 4x - 10y - 7 = 0$.

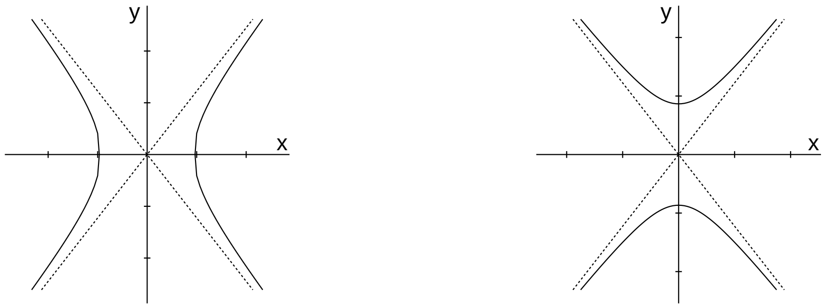

Figure 7.3: (i) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$; (ii) $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = -1,\ 0 < b,\ 0 < a$.

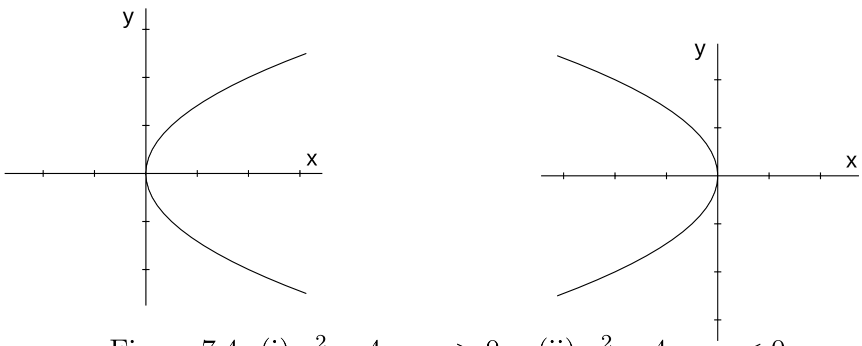

Figure 7.4: (i) $y^2 = 4ax,\ a > 0$; (ii) $y^2 = 4ax,\ a < 0$.

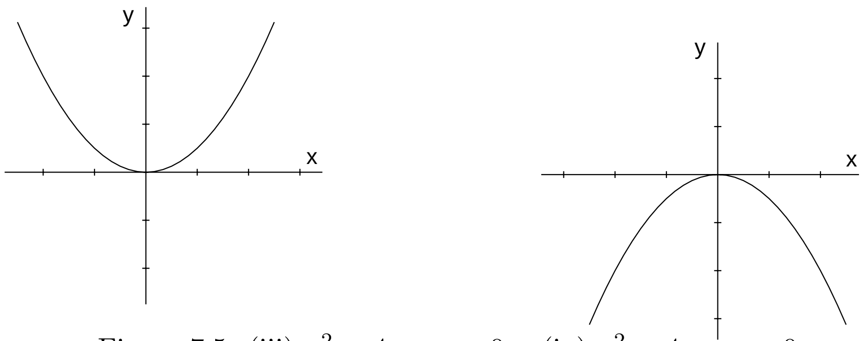

Figure 7.5: (iii) $x^2 = 4ay$, $a > 0$;   (iv) $x^2 = 4ay$, $a < 0$.

Solution. Complete the square:

$$\begin{aligned} y^2 - 10y + 25 - 4x - 32 &= 0 \\ (y-5)^2 = 4x + 32 &= 4(x+8), \end{aligned}$$

or $y_1^2 = 4x_1$, under the translation of axes $x_1 = x + 8$, $y_1 = y - 5$. Hence we get a parabola with vertex at the new origin $(x_1, y_1) = (0, 0)$, i.e. $(x, y) = (-8, 5)$.

The parabola is sketched in Figure 7.6.

EXAMPLE 7.1.4 Sketch the curve $x^2 - 4xy + 4y^2 + 5y - 9 = 0$.

Solution. We have $x^2 - 4xy + 4y^2 = X^tAX$, where

$$A = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix}.$$

The characteristic equation of A is $\lambda^2 - 5\lambda = 0$, so A has distinct eigenvalues $\lambda_1 = 5$ and $\lambda_2 = 0$. We find corresponding unit length eigenvectors

$$X_1 = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 \\ -2 \end{bmatrix}, \; X_2 = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 \\ 1 \end{bmatrix}.$$

Then $P = [X_1|X_2]$ is a proper orthogonal matrix and under the rotation of axes $X = PY$, or

$$\begin{aligned} x &= \frac{x_1 + 2y_1}{\sqrt{5}} \\ y &= \frac{-2x_1 + y_1}{\sqrt{5}}, \end{aligned}$$

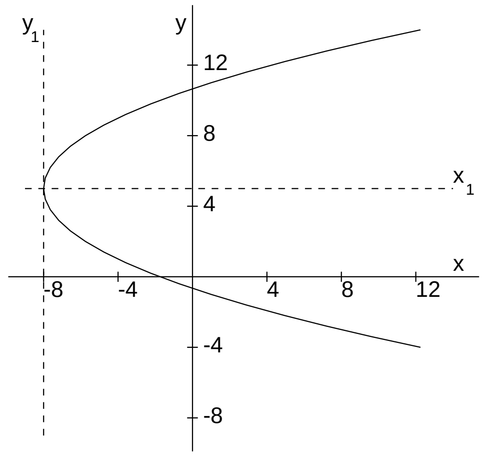

Figure 7.6: $y^2 - 4x - 10y - 7 = 0$.

we have

$$x^2 - 4xy + 4y^2 = \lambda_1 x_1^2 + \lambda_2 y_1^2 = 5x_1^2.$$

The original quadratic equation becomes

$$\begin{aligned} 5x_1^2 + \frac{5}{\sqrt{5}}(-2x_1 + y_1) - 9 &= 0 \\ 5(x_1^2 - \frac{2}{\sqrt{5}}x_1) + \sqrt{5}y_1 - 9 &= 0 \\ 5(x_1 - \frac{1}{\sqrt{5}})^2 = 10 - \sqrt{5}y_1 &= -\sqrt{5}(y_1 - 2\sqrt{5}), \end{aligned}$$

or $5x_2^2 = -\frac{1}{\sqrt{5}}y_2$, where the x_1, y_1 axes have been translated to x_2, y_2 axes using the transformation

$$x_2 = x_1 - \frac{1}{\sqrt{5}}, \quad y_2 = y_1 - 2\sqrt{5}.$$

Hence the vertex of the parabola is at $(x_2, y_2) = (0, 0)$, i.e. $(x_1, y_1) = (\frac{1}{\sqrt{5}}, 2\sqrt{5})$, or $(x, y) = (\frac{21}{5}, \frac{8}{5})$. The axis of symmetry of the parabola is the line $x_2 = 0$, i.e. $x_1 = 1/\sqrt{5}$. Using the rotation equations in the form

$$x_1 = \frac{x - 2y}{\sqrt{5}}$$

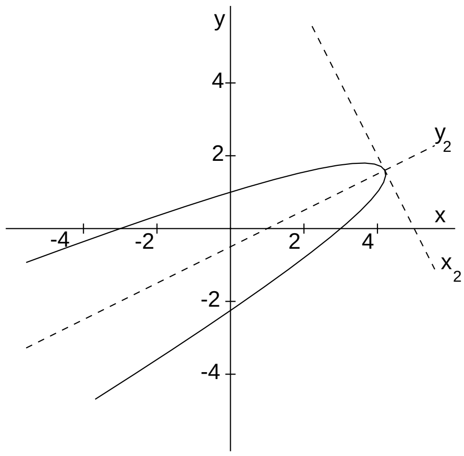

Figure 7.7: $x^2 - 4xy + 4y^2 + 5y - 9 = 0$.

$$y_1 = \frac{2x + y}{\sqrt{5}},$$

we have

$$\frac{x - 2y}{\sqrt{5}} = \frac{1}{\sqrt{5}}, \qquad \text{or} \qquad x - 2y = 1.$$

The parabola is sketched in Figure 7.7.

7.2 A classification algorithm

There are several possible degenerate cases that can arise from the general second degree equation. For example $x^2 + y^2 = 0$ represents the point $(0, 0)$; $x^2 + y^2 = -1$ defines the empty set, as does $x^2 = -1$ or $y^2 = -1$; $x^2 = 0$ defines the line $x = 0$; $(x + y)^2 = 0$ defines the line $x + y = 0$; $x^2 - y^2 = 0$ defines the lines $x - y = 0$, $x + y = 0$; $x^2 = 1$ defines the parallel lines $x = \pm 1$; $(x + y)^2 = 1$ likewise defines two parallel lines $x + y = \pm 1$.

We state without proof a complete classification [1] of the various cases

[1] This classification forms the basis of a computer program which was used to produce the diagrams in this chapter. I am grateful to Peter Adams for his programming assistance.

that can possibly arise for the general second degree equation

$$ax^2 + 2hxy + by^2 + 2gx + 2fy + c = 0. \tag{7.10}$$

It turns out to be more convenient to first perform a suitable translation of axes, before rotating the axes. Let

$$\Delta = \begin{vmatrix} a & h & g \\ h & b & f \\ g & f & c \end{vmatrix}, \quad C = ab - h^2,\ A = bc - f^2,\ B = ca - g^2.$$

If $C \neq 0$, let

$$\alpha = \frac{-\begin{vmatrix} g & h \\ f & b \end{vmatrix}}{C}, \qquad \beta = \frac{-\begin{vmatrix} a & g \\ h & f \end{vmatrix}}{C}. \tag{7.11}$$

CASE 1. $\Delta = 0$.

(1.1) $C \neq 0$. Translate axes to the new origin (α, β), where α and β are given by equations 7.11:

$$x = x_1 + \alpha, \quad y = y_1 + \beta.$$

Then equation 7.10 reduces to

$$ax_1^2 + 2hx_1y_1 + by_1^2 = 0.$$

(a) $C > 0$: **Single point** $(x, y) = (\alpha, \beta)$.

(b) $C < 0$: **Two non–parallel lines** intersecting in $(x, y) = (\alpha, \beta)$. The lines are

$$\frac{y-\beta}{x-\alpha} = \frac{-h \pm \sqrt{-C}}{b} \quad \text{if } b \neq 0,$$

$$x = \alpha \quad \text{and} \quad \frac{y-\beta}{x-\alpha} = -\frac{a}{2h}, \quad \text{if } b = 0.$$

(1.2) $C = 0$.

(a) $h = 0$.

(i) $a = g = 0$.

(A) $A > 0$: **Empty set**.

(B) $A = 0$: **Single line** $y = -f/b$.

(C) $A < 0$: **Two parallel lines**

$$y = \frac{-f \pm \sqrt{-A}}{b}$$

(ii) $b = f = 0$.

(A) $B > 0$: **Empty set**.

(B) $B = 0$: **Single line** $x = -g/a$.

(C) $B < 0$: **Two parallel lines**

$$x = \frac{-g \pm \sqrt{-B}}{a}$$

(b) $h \neq 0$.

(i) $B > 0$: **Empty set**.

(ii) $B = 0$: **Single line** $ax + hy = -g$.

(iii) $B < 0$: **Two parallel lines**

$$ax + hy = -g \pm \sqrt{-B}.$$

CASE 2. $\Delta \neq 0$.

(2.1) $C \neq 0$. Translate axes to the new origin (α, β), where α and β are given by equations 7.11:

$$x = x_1 + \alpha, \quad y = y_1 + \beta.$$

Equation 7.10 becomes

$$ax_1^2 + 2hx_1y_1 + by_1^2 = -\frac{\Delta}{C}. \tag{7.12}$$

CASE 2.1(i) $h = 0$. Equation 7.12 becomes $ax_1^2 + by_1^2 = \frac{-\Delta}{C}$.

(a) $C < 0$: **Hyperbola**.

(b) $C > 0$ and $a\Delta > 0$: **Empty set**.

(c) $C > 0$ and $a\Delta < 0$.

(i) $a = b$: **Circle**, centre (α, β), radius $\sqrt{\frac{g^2+f^2-ac}{a}}$.

(ii) $a \neq b$: **Ellipse**.

CASE 2.1(ii) $h \neq 0$.

Rotate the (x_1, y_1) axes with the new positive x_2–axis in the direction of

$$[(b-a+R)/2, -h],$$

where $R = \sqrt{(a-b)^2 + 4h^2}$.

Then equation 7.12 becomes

$$\lambda_1 x_2^2 + \lambda_2 y_2^2 = -\frac{\Delta}{C}. \tag{7.13}$$

where

$$\lambda_1 = (a+b-R)/2, \ \lambda_2 = (a+b+R)/2,$$

Here $\lambda_1 \lambda_2 = C$.

(a) $C < 0$: **Hyperbola**.
Here $\lambda_2 > 0 > \lambda_1$ and equation 7.13 becomes

$$\frac{x_2^2}{u^2} - \frac{y_2^2}{v^2} = \frac{-\Delta}{|\Delta|},$$

where

$$u = \sqrt{\frac{|\Delta|}{C\lambda_1}}, \ v = \sqrt{\frac{|\Delta|}{-C\lambda_2}}.$$

(b) $C > 0$ and $a\Delta > 0$: **Empty set**.

(c) $C > 0$ and $a\Delta < 0$: **Ellipse**.
Here $\lambda_1, \lambda_2, a, b$ have the same sign and $\lambda_1 \neq \lambda_2$ and equation 7.13 becomes

$$\frac{x_2^2}{u^2} + \frac{y_2^2}{v^2} = 1,$$

where

$$u = \sqrt{\frac{\Delta}{-C\lambda_1}}, \ v = \sqrt{\frac{\Delta}{-C\lambda_2}}.$$

(2.1) $C = 0$.

(a) $h = 0$.

(i) $a = 0$: Then $b \neq 0$ and $g \neq 0$. **Parabola** with vertex

$$\left(\frac{-A}{2gb}, -\frac{f}{b}\right).$$

Translate axes to $(x_1,\, y_1)$ axes:

$$y_1^2 = -\frac{2g}{b}x_1.$$

(ii) $b = 0$: Then $a \neq 0$ and $f \neq 0$. **Parabola** with vertex

$$\left(-\frac{g}{a},\, \frac{-B}{2fa}\right).$$

Translate axes to $(x_1,\, y_1)$ axes:

$$x_1^2 = -\frac{2f}{a}y_1.$$

(b) $h \neq 0$: **Parabola**. Let

$$k = \frac{ga + hf}{a + b}.$$

The vertex of the parabola is

$$\left(\frac{(2akf - hk^2 - hac)}{d},\, \frac{a(k^2 + ac - 2kg)}{d}\right),$$

where $d = 2a(gh - af)$. Now translate to the vertex as the new origin, then rotate to $(x_2,\, y_2)$ axes with the positive x_2–axis along $[sa,\, -sh]$, where $s = \text{sign}\,(a)$.
(The positive x_2–axis points into the first or fourth quadrant.) Then the parabola has equation

$$x_2^2 = \frac{-2st}{\sqrt{a^2 + h^2}}y_2,$$

where $t = (af - gh)/(a + b)$.

REMARK 7.2.1 If $\Delta = 0$, it is not necessary to rotate the axes. Instead it is always possible to translate the axes suitably so that the coefficients of the terms of the first degree vanish.

EXAMPLE 7.2.1 Identify the curve

$$2x^2 + xy - y^2 + 6y - 8 = 0. \tag{7.14}$$

Solution. Here

$$\Delta = \begin{vmatrix} 2 & \frac{1}{2} & 0 \\ \frac{1}{2} & -1 & 3 \\ 0 & 3 & -8 \end{vmatrix} = 0.$$

Let $x = x_1 + \alpha$, $y = y_1 + \beta$ and substitute in equation 7.14 to get

$$2(x_1 + \alpha)^2 + (x_1 + \alpha)(y_1 + \beta) - (y_1 + \beta)^2 + 4(y_1 + \beta) - 8 = 0. \qquad (7.15)$$

Then equating the coefficients of x_1 and y_1 to 0 gives

$$\begin{aligned} 4\alpha + \beta &= 0 \\ \alpha + 2\beta + 4 &= 0, \end{aligned}$$

which has the unique solution $\alpha = -\frac{2}{3}$, $\beta = \frac{8}{3}$. Then equation 7.15 simplifies to

$$2x_1^2 + x_1y_1 - y_1^2 = 0 = (2x_1 - y_1)(x_1 + y_1),$$

so relative to the x_1, y_1 coordinates, equation 7.14 describes two lines: $2x_1 - y_1 = 0$ or $x_1 + y_1 = 0$. In terms of the original x, y coordinates, these lines become $2(x + \frac{2}{3}) - (y - \frac{8}{3}) = 0$ and $(x + \frac{2}{3}) + (y - \frac{8}{3}) = 0$, i.e. $2x - y + 4 = 0$ and $x + y - 2 = 0$, which intersect in the point

$$(x, y) = (\alpha, \beta) = (-\frac{2}{3}, \frac{8}{3}).$$

EXAMPLE 7.2.2 Identify the curve

$$x^2 + 2xy + y^2 + 2x + 2y + 1 = 0. \qquad (7.16)$$

Solution. Here

$$\Delta = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Let $x = x_1 + \alpha$, $y = y_1 + \beta$ and substitute in equation 7.16 to get

$$(x_1+\alpha)^2+2(x_1+\alpha)(y_1+\beta)+(y_1+\beta)^2+2(x_1+\alpha)+2(y_1+\beta)+1 = 0. \quad (7.17)$$

Then equating the coefficients of x_1 and y_1 to 0 gives the same equation

$$2\alpha + 2\beta + 2 = 0.$$

Take $\alpha = 0$, $\beta = -1$. Then equation 7.17 simplifies to

$$x_1^2 + 2x_1y_1 + y_1^2 = 0 = (x_1 + y_1)^2,$$

and in terms of x, y coordinates, equation 7.16 becomes

$$(x + y + 1)^2 = 0, \text{ or } x + y + 1 = 0.$$

7.3 PROBLEMS

1. Sketch the curves

   (i) $x^2 - 8x + 8y + 8 = 0$;

   (ii) $y^2 - 12x + 2y + 25 = 0$.

2. Sketch the hyperbola
   $$4xy - 3y^2 = 8$$
   and find the equations of the asymptotes.

   [Answer: $y = 0$ and $y = \frac{4}{3}x$.]

3. Sketch the ellipse
   $$8x^2 - 4xy + 5y^2 = 36$$
   and find the equations of the axes of symmetry.

   [Answer: $y = 2x$ and $x = -2y$.]

4. Sketch the conics defined by the following equations. Find the centre when the conic is an ellipse or hyperbola, asymptotes if an hyperbola, the vertex and axis of symmetry if a parabola:

   (i) $4x^2 - 9y^2 - 24x - 36y - 36 = 0$;

   (ii) $5x^2 - 4xy + 8y^2 + 4\sqrt{5}x - 16\sqrt{5}y + 4 = 0$;

   (iii) $4x^2 + y^2 - 4xy - 10y - 19 = 0$;

   (iv) $77x^2 + 78xy - 27y^2 + 70x - 30y + 29 = 0$.

   [Answers: (i) hyperbola, centre $(3, -2)$, asymptotes $2x - 3y - 12 = 0$, $2x + 3y = 0$;
   (ii) ellipse, centre $(0, \sqrt{5})$;
   (iii) parabola, vertex $(-\frac{7}{5}, -\frac{9}{5})$, axis of symmetry $2x - y + 1 = 0$;
   (iv) hyperbola, centre $(-\frac{1}{10}, -\frac{7}{10})$, asymptotes $7x + 9y + 7 = 0$ and $11x - 3y - 1 = 0$.]

5. Identify the lines determined by the equations:

   (i) $2x^2 + y^2 + 3xy - 5x - 4y + 3 = 0$;

(ii) $9x^2 + y^2 - 6xy + 6x - 2y + 1 = 0$;

(iii) $x^2 + 4xy + 4y^2 - x - 2y - 2 = 0$.

[Answers: (i) $2x + y - 3 = 0$ and $x + y - 1 = 0$; (ii) $3x - y + 1 = 0$; (iii) $x + 2y + 1 = 0$ and $x + 2y - 2 = 0$.]

Chapter 8

THREE–DIMENSIONAL GEOMETRY

8.1 Introduction

In this chapter we present a vector–algebra approach to three–dimensional geometry. The aim is to present standard properties of lines and planes, with minimum use of complicated three–dimensional diagrams such as those involving similar triangles. We summarize the chapter:

Points are defined as ordered triples of real numbers and the *distance* between points $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ is defined by the formula

$$P_1P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Directed line segments $\overrightarrow{AB}$ are introduced as three–dimensional column vectors: If $A = (x_1, y_1, z_1)$ and $B = (x_2, y_2, z_2)$, then

$$\overrightarrow{AB} = \left[\begin{array}{c} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{array}\right].$$

If P is a point, we let $\mathbf{P} = \overrightarrow{OP}$ and call $\mathbf{P}$ the *position vector* of P.

With suitable definitions of *lines*, *parallel lines*, there are important geometrical interpretations of equality, addition and scalar multiplication of vectors.

(i) Equality of vectors: Suppose A, B, C, D are distinct points such that no three are collinear. Then $\overrightarrow{AB} = \overrightarrow{CD}$ if and only if $\overrightarrow{AB} \parallel \overrightarrow{CD}$ and $\overrightarrow{AC} \parallel \overrightarrow{BD}$ (See Figure 8.1.)

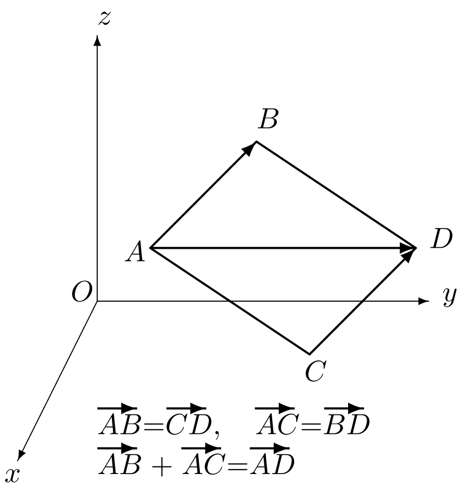

Figure 8.1: Equality and addition of vectors.

(ii) Addition of vectors obeys the *parallelogram law*: Let A, B, C be non–collinear. Then

$$\overrightarrow{AB}+\overrightarrow{AC}=\overrightarrow{AD},$$

where D is the point such that $\overrightarrow{AB} \parallel \overrightarrow{CD}$ and $\overrightarrow{AC} \parallel \overrightarrow{BD}$. (See Figure 8.1.)

(iii) Scalar multiplication of vectors: Let $\overrightarrow{AP}= t\ \overrightarrow{AB}$, where A and B are distinct points. Then P is on the line AB,

$$\frac{AP}{AB} = |t|$$

and

(a) $P = A$ if $t = 0$, $P = B$ if $t = 1$;

(b) P is between A and B if $0 < t < 1$;

(c) B is between A and P if $1 < t$;

(d) A is between P and B if $t < 0$.

(See Figure 8.2.)

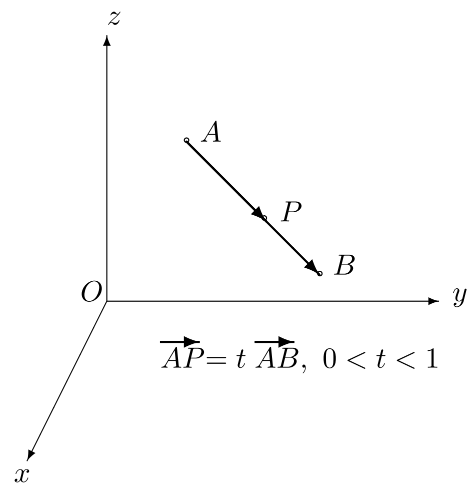

Figure 8.2: Scalar multiplication of vectors.

The *dot product* $X \cdot Y$ of vectors $X = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}$ and $Y = \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix}$, is defined by

$$X \cdot Y = a_1 a_2 + b_1 b_2 + c_1 c_2.$$

The *length* $||X||$ of a vector X is defined by

$$||X|| = (X \cdot X)^{1/2}$$

and the *Cauchy–Schwarz inequality* holds:

$$|X \cdot Y| \le ||X|| \cdot ||Y||.$$

The *triangle inequality* for vector length now follows as a simple deduction:

$$||X + Y|| \le ||X|| + ||Y||.$$

Using the equation

$$AB = ||\ \overrightarrow{AB}\ ||,$$

we deduce the corresponding familiar *triangle inequality* for distance:

$$AB \le AC + CB.$$

The *angle* θ between two non–zero vectors X and Y is then defined by

$$\cos\theta = \frac{X \cdot Y}{||X|| \cdot ||Y||}, \quad 0 \leq \theta \leq \pi.$$

This definition makes sense. For by the Cauchy–Schwarz inequality,

$$-1 \leq \frac{X \cdot Y}{||X|| \cdot ||Y||} \leq 1.$$

Vectors X and Y are said to be *perpendicular* or *orthogonal* if $X \cdot Y = 0$.
Vectors of unit length are called *unit* vectors. The vectors

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

are unit vectors and every vector is a linear combination of **i**, **j** and **k**:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

Non–zero vectors X and Y are *parallel* or *proportional* if the angle between X and Y equals 0 or π; equivalently if $X = tY$ for some real number t. Vectors X and Y are then said to have the same or opposite direction, according as $t > 0$ or $t < 0$.

We are then led to study straight lines. If A and B are distinct points, it is easy to show that $AP + PB = AB$ holds if and only if

$$\overrightarrow{AP} = t\,\overrightarrow{AB}, \text{ where } 0 \leq t \leq 1.$$

A *line* is defined as a set consisting of all points P satisfying

$$\mathbf{P} = \mathbf{P_0} + tX, \quad t \in \mathbb{R} \quad \text{or equivalently} \quad \overrightarrow{P_0P} = tX,$$

for some fixed point P_0 and fixed non–zero vector X called a *direction vector* for the line.

Equivalently, in terms of coordinates,

$$x = x_0 + ta, \; y = y_0 + tb, \; z = z_0 + tc,$$

where $P_0 = (x_0, y_0, z_0)$ and not all of a, b, c are zero.

There is then one and only one line passing passing through two distinct points A and B. It consists of the points P satisfying

$$\overrightarrow{AP} = t\,\overrightarrow{AB},$$

where t is a real number.

The *cross–product* $X \times Y$ provides us with a vector which is perpendicular to both X and Y. It is defined in terms of the components of X and Y:

Let $X = a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$ and $Y = a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}$. Then

$$X \times Y = a\mathbf{i} + b\mathbf{j} + c\mathbf{k},$$

where

$$a = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \quad b = -\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}, \quad c = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

The cross–product enables us to derive elegant formulae for the distance from a point to a line, the area of a triangle and the distance between two skew lines.

Finally we turn to the concept of a plane in three–dimensional space.

A *plane* is a set of points P satisfying an equation of the form

$$\mathbf{P} = \mathbf{P}_0 + sX + tY, \; s, t \in \mathbb{R}, \tag{8.1}$$

where X and Y are non–zero, non–parallel vectors.

In terms of coordinates, equation 8.1 takes the form

$$\begin{aligned} x &= x_0 + sa_1 + ta_2 \\ y &= y_0 + sb_1 + tb_2 \\ z &= z_0 + sc_1 + tc_2, \end{aligned}$$

where $P_0 = (x_0, y_0, z_0)$.

There is then one and only one plane passing passing through three non–collinear points A, B, C. It consists of the points P satisfying

$$\overrightarrow{AP} = s\,\overrightarrow{AB} + t\,\overrightarrow{AC},$$

where s and t are real numbers.

The cross–product enables us to derive a concise equation for the plane through three non–collinear points A, B, C, namely

$$\overrightarrow{AP} \cdot (\overrightarrow{AB} \times \overrightarrow{AC}) = 0.$$

When expanded, this equation has the form

$$ax + by + cz = d,$$

where $a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ is a non–zero vector which is perpendicular to $\overrightarrow{P_1P_2}$ for all points P_1, P_2 lying in the plane. Any vector with this property is said to be a *normal* to the plane.

It is then easy to prove that two planes with non–parallel normal vectors must intersect in a line.

We conclude the chapter by deriving a formula for the distance from a point to a plane.

8.2 Three–dimensional space

DEFINITION 8.2.1 *Three–dimensional space* is the set E^3 of ordered triples $(x,\, y,\, z)$, where $x,\, y,\, z$ are real numbers. The triple $(x,\, y,\, z)$ is called a point P in E^3 and we write $P = (x,\, y,\, z)$. The numbers $x,\, y,\, z$ are called, respectively, the $x,\, y,\, z$ coordinates of P.

The *coordinate axes* are the sets of points:

$\{(x, 0,\, 0)\}$ (x–axis), $\{(0,\, y,\, 0)\}$ (y–axis), $\{(0,\, 0,\, z)\}$ (z–axis).

The only point common to all three axes is the *origin* $O = (0,\, 0,\, 0)$.

The *coordinate planes* are the sets of points:

$\{(x,\, y,\, 0)\}$ (xy–plane), $\{(0,\, y,\, z)\}$ (yz–plane), $\{(x,\, 0,\, z)\}$ (xz–plane).

The *positive octant* consists of the points $(x,\, y,\, z)$, where $x > 0,\, y > 0,\, z > 0$.

We think of the points $(x,\, y,\, z)$ with $z > 0$ as lying above the xy–plane, and those with $z < 0$ as lying beneath the xy–plane. A point $P = (x,\, y,\, z)$ will be represented as in Figure 8.3. The point illustrated lies in the positive octant.

DEFINITION 8.2.2 The *distance* P_1P_2 between points $P_1 = (x_1,\, y_1,\, z_1)$ and $P_2 = (x_2,\, y_2,\, z_2)$ is defined by the formula

$$P_1P_2 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

For example, if $P = (x,\, y,\, z)$,

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

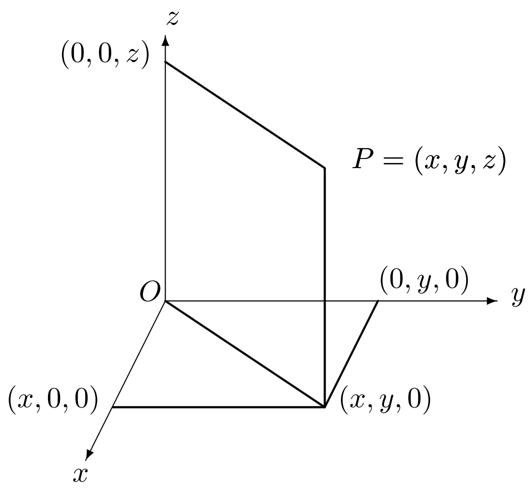

Figure 8.3: Representation of three-dimensional space.

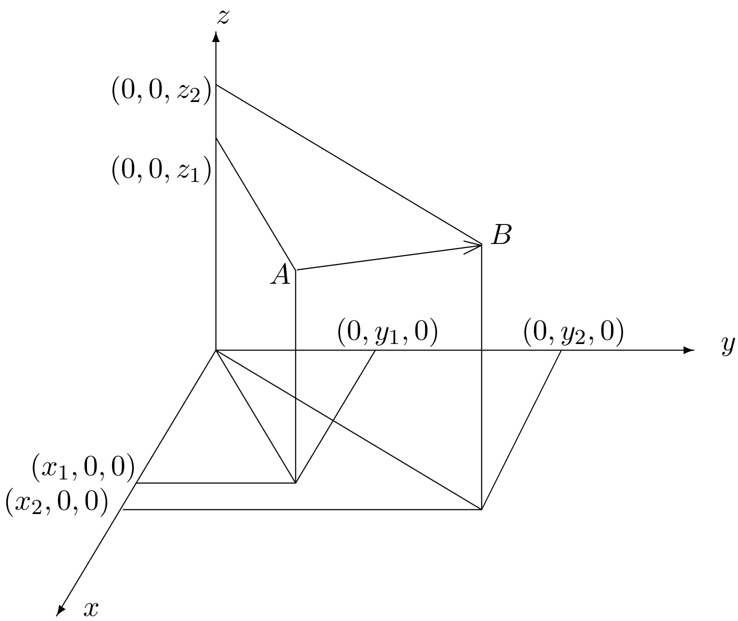

Figure 8.4: The vector $\overrightarrow{AB}$.

DEFINITION 8.2.3 If $A = (x_1, y_1, z_1)$ and $B = (x_2, y_2, z_2)$ we define the symbol $\overrightarrow{AB}$ to be the column vector

$$\overrightarrow{AB} = \begin{bmatrix} x_2 - x_1 \\ y_2 - y_1 \\ z_2 - z_1 \end{bmatrix}.$$

We let $\mathbf{P} = \overrightarrow{OP}$ and call $\mathbf{P}$ the *position vector* of P.

The components of $\overrightarrow{AB}$ are the coordinates of B when the axes are translated to A as origin of coordinates.

We think of $\overrightarrow{AB}$ as being represented by the directed line segment from A to B and think of it as an arrow whose tail is at A and whose head is at B. (See Figure 8.4.)

Some mathematicians think of $\overrightarrow{AB}$ as representing the translation of space which takes A into B.

The following simple properties of $\overrightarrow{AB}$ are easily verified and correspond to how we intuitively think of directed line segments:

(i) $\overrightarrow{AB} = 0 \Leftrightarrow A = B$;

(ii) $\overrightarrow{BA} = -\overrightarrow{AB}$;

(iii) $\overrightarrow{AB} + \overrightarrow{BC} = \overrightarrow{AC}$ (the triangle law);

(iv) $\overrightarrow{BC} = \overrightarrow{AC} - \overrightarrow{AB} = \mathbf{C} - \mathbf{B}$;

(v) if X is a vector and A a point, there is exactly one point B such that $\overrightarrow{AB} = X$, namely that defined by $\mathbf{B} = \mathbf{A} + X$.

To derive properties of the distance function and the vector function $\overrightarrow{P_1P_2}$, we need to introduce the *dot product* of two vectors in $\mathbb{R}^3$.

8.3 Dot product

DEFINITION 8.3.1 If $X = \begin{bmatrix} a_1 \\ b_1 \\ c_1 \end{bmatrix}$ and $Y = \begin{bmatrix} a_2 \\ b_2 \\ c_2 \end{bmatrix}$, then $X \cdot Y$, the *dot product* of X and Y, is defined by

$$X \cdot Y = a_1a_2 + b_1b_2 + c_1c_2.$$

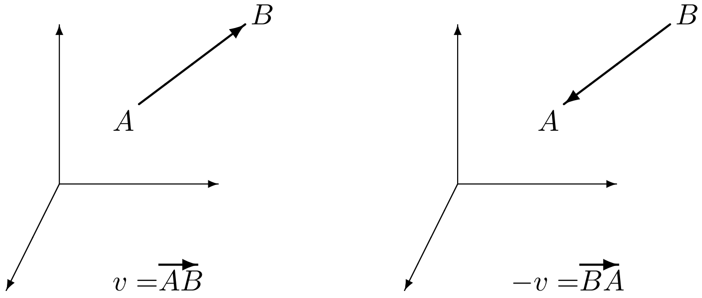

Figure 8.5: The negative of a vector.

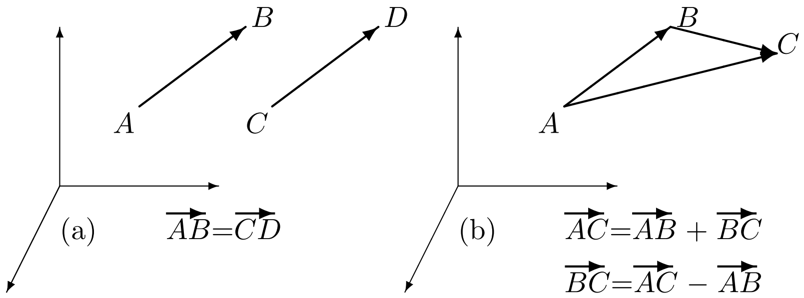

Figure 8.6: (a) Equality of vectors; (b) Addition and subtraction of vectors.

The dot product has the following properties:

(i) $X \cdot (Y + Z) = X \cdot Y + X \cdot Z$;

(ii) $X \cdot Y = Y \cdot X$;

(iii) $(tX) \cdot Y = t(X \cdot Y)$;

(iv) $X \cdot X = a^2 + b^2 + c^2$ if $X = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$;

(v) $X \cdot Y = X^t Y$;

(vi) $X \cdot X = 0$ if and only if $X = 0$.

The *length* of X is defined by

$$||X|| = \sqrt{a^2 + b^2 + c^2} = (X \cdot X)^{1/2}.$$

We see that $||\mathbf{P}|| = OP$ and more generally $||\ \overrightarrow{P_1P_2}\ || = P_1P_2$, the distance between P_1 and P_2.

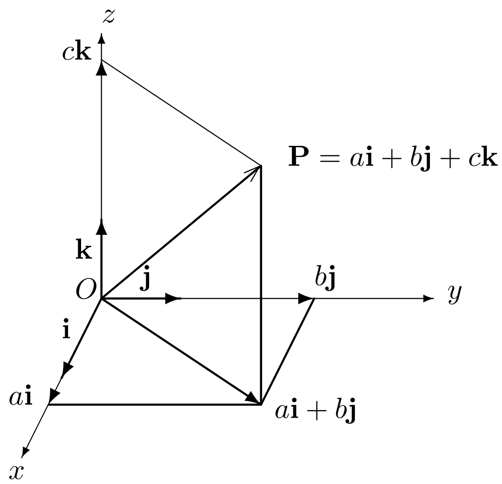

Figure 8.7: Position vector as a linear combination of **i**, **j** and **k**.

Vectors having unit length are called *unit* vectors.

The vectors

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

are unit vectors. Every vector is a linear combination of **i**, **j** and **k**:

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}.$$

(See Figure 8.7.)

It is easy to prove that

$$||tX|| = |t| \cdot ||X||,$$

if t is a real number. Hence if X is a non–zero vector, the vectors

$$\pm\frac{1}{||X||}X$$

are unit vectors.

A useful property of the length of a vector is

$$||X \pm Y||^2 = ||X||^2 \pm 2X \cdot Y + ||Y||^2. \tag{8.2}$$

The following important property of the dot product is widely used in mathematics:

THEOREM 8.3.1 The Cauchy–Schwarz inequality
If X and Y are vectors in $\mathbb{R}^3$, then

$$|X \cdot Y| \leq ||X|| \cdot ||Y||. \tag{8.3}$$

Moreover if $X \neq 0$ and $Y \neq 0$, then

$$\begin{aligned} X \cdot Y = ||X|| \cdot ||Y|| \quad &\Leftrightarrow \quad Y = tX,\ t > 0, \\ X \cdot Y = -||X|| \cdot ||Y|| \quad &\Leftrightarrow \quad Y = tX,\ t < 0. \end{aligned}$$

Proof. If $X = 0$, then inequality 8.3 is trivially true. So assume $X \neq 0$. Now if t is any real number, by equation 8.2,

$$\begin{aligned} 0 \leq ||tX - Y||^2 &= ||tX||^2 - 2(tX) \cdot Y + ||Y||^2 \\ &= t^2||X||^2 - 2(X \cdot Y)t + ||Y||^2 \\ &= at^2 - 2bt + c, \end{aligned}$$

where $a = ||X||^2 > 0$, $b = X \cdot Y$, $c = ||Y||^2$.

Hence

$$\begin{aligned} a(t^2 - \frac{2b}{a}t + \frac{c}{a}) &\geq 0 \\ \left(t - \frac{b}{a}\right)^2 + \frac{ca - b^2}{a^2} &\geq 0 \quad . \end{aligned}$$

Substituting $t = b/a$ in the last inequality then gives

$$\frac{ac - b^2}{a^2} \geq 0,$$

so

$$|b| \leq \sqrt{ac} = \sqrt{a}\sqrt{c}$$

and hence inequality 8.3 follows.

To discuss equality in the Cauchy–Schwarz inequality, assume $X \neq 0$ and $Y \neq 0$.

Then if $X \cdot Y = ||X|| \cdot ||Y||$, we have for all t

$$\begin{aligned} ||tX - Y||^2 &= t^2||X||^2 - 2tX \cdot Y + ||Y||^2 \\ &= t^2||X||^2 - 2t||X|| \cdot ||Y|| + ||Y||^2 \\ &= ||tX - Y||^2. \end{aligned}$$

Taking $t = ||X||/||Y||$ then gives $||tX - Y||^2 = 0$ and hence $tX - Y = 0$. Hence $Y = tX$, where $t > 0$. The case $X \cdot Y = -||X|| \cdot ||Y||$ is proved similarly.

COROLLARY 8.3.1 (The triangle inequality for vectors)
If X and Y are vectors, then

$$||X + Y|| \leq ||X|| + ||Y||. \tag{8.4}$$

Moreover if $X \neq 0$ and $Y \neq 0$, then equality occurs in inequality 8.4 if and only if $Y = tX$, where $t > 0$.

Proof.

$$\begin{aligned} ||X+Y||^2 &= ||X||^2 + 2X \cdot Y + ||Y||^2 \\ &\leq ||X||^2 + 2||X|| \cdot ||Y|| + ||Y||^2 \\ &= (||X|| + ||Y||)^2 \end{aligned}$$

and inequality 8.4 follows.

If $||X + Y|| = ||X|| + ||Y||$, then the above proof shows that

$$X \cdot Y = ||X|| \cdot ||Y||.$$

Hence if $X \neq 0$ and $Y \neq 0$, the first case of equality in the Cauchy–Schwarz inequality shows that $Y = tX$ with $t > 0$.

The triangle inequality for vectors gives rise to a corresponding inequality for the distance function:

THEOREM 8.3.2 (The triangle inequality for distance)
If A, B, C are points, then

$$AC \leq AB + BC. \tag{8.5}$$

Moreover if $B \neq A$ and $B \neq C$, then equality occurs in inequality 8.5 if and only if $\overrightarrow{AB} = r\,\overrightarrow{AC}$, where $0 < r < 1$.

Proof.

$$\begin{aligned} AC = ||\,\overrightarrow{AC}\,|| &= ||\,\overrightarrow{AB} + \overrightarrow{BC}\,|| \\ &\leq ||\,\overrightarrow{AB}\,|| + ||\,\overrightarrow{BC}\,|| \\ &= AB + BC. \end{aligned}$$

Moreover if equality occurs in inequality 8.5 and $B \neq A$ and $B \neq C$, then $X = \overrightarrow{AB} \neq 0$ and $Y = \overrightarrow{BC} \neq 0$ and the equation $AC = AB + BC$ becomes $||X + Y|| = ||X|| + ||Y||$. Hence the case of equality in the vector triangle inequality gives

$$Y = \overrightarrow{BC} = tX = t\,\overrightarrow{AB}, \text{ where } t > 0.$$

Then

$$\begin{aligned} \overrightarrow{BC} &= \overrightarrow{AC} - \overrightarrow{AB} = t\,\overrightarrow{AB} \\ \overrightarrow{AC} &= (1+t)\,\overrightarrow{AB} \\ \overrightarrow{AB} &= r\,\overrightarrow{AC}, \end{aligned}$$

where $r = 1/(t+1)$ satisfies $0 < r < 1$.

8.4 Lines

DEFINITION 8.4.1 A line in E^3 is the set $\mathcal{L}(P_0, X)$ consisting of all points P satisfying

$$\mathbf{P} = \mathbf{P_0} + tX, \quad t \in \mathbb{R} \quad \text{or equivalently} \quad \overrightarrow{P_0P} = tX, \tag{8.6}$$

for some fixed point P_0 and fixed non–zero vector X. (See Figure 8.8.)

Equivalently, in terms of coordinates, equation 8.6 becomes

$$x = x_0 + ta,\ y = y_0 + tb,\ z = z_0 + tc,$$

where not all of a, b, c are zero.

The following familiar property of straight lines is easily verified.

THEOREM 8.4.1 If A and B are distinct points, there is one and only one line containing A and B, namely $\mathcal{L}(A, \overrightarrow{AB})$ or more explicitly the line defined by $\overrightarrow{AP} = t\,\overrightarrow{AB}$, or equivalently, in terms of position vectors:

$$\mathbf{P} = (1-t)\mathbf{A} + t\mathbf{B} \quad \text{or} \quad \mathbf{P} = \mathbf{A} + t\,\overrightarrow{AB}\,. \tag{8.7}$$

Equations 8.7 may be expressed in terms of coordinates:
if $A = (x_1, y_1, z_1)$ and $B = (x_2, y_2, z_2)$, then

$$x = (1-t)x_1 + tx_2,\ y = (1-t)y_1 + ty_2,\ z = (1-t)z_1 + tz_2.$$

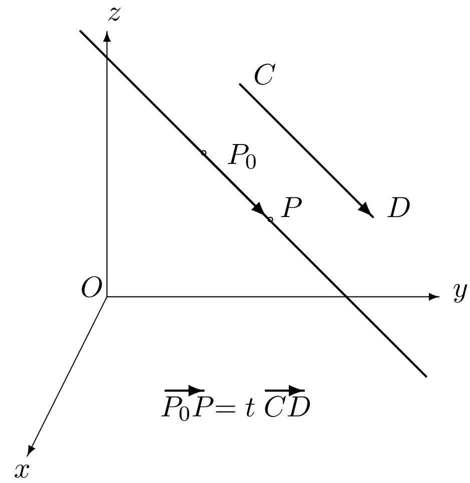

Figure 8.8: Representation of a line.

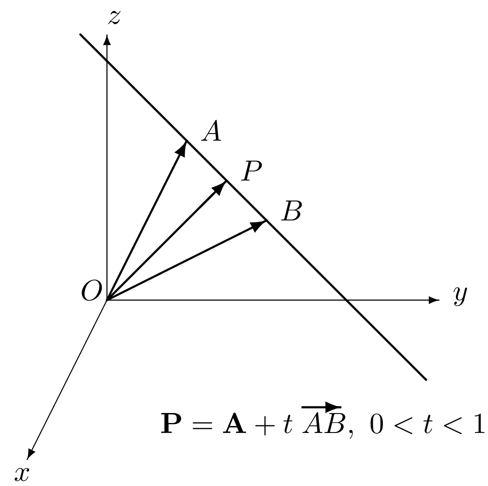

Figure 8.9: The line segment AB.

There is an important geometric significance in the number t of the above equation of the line through A and B. The proof is left as an exercise:

THEOREM 8.4.2 (Joachimsthal's ratio formulae)
If t is the parameter occurring in theorem 8.4.1, then

$$\text{(i)}\quad |t| = \frac{AP}{AB}; \quad \text{(ii)}\quad \left|\frac{t}{1-t}\right| = \frac{AP}{PB} \quad \text{if } P \neq B.$$

Also

(iii) P is between A and B if $0 < t < 1$;

(iv) B is between A and P if $1 < t$;

(v) A is between P and B if $t < 0$.

(See Figure 8.9.)

For example, $t = \frac{1}{2}$ gives the mid–point P of the segment AB:

$$\mathbf{P} = \frac{1}{2}(\mathbf{A} + \mathbf{B}).$$

EXAMPLE 8.4.1 $\mathcal{L}$ is the line AB, where $A = (-4,\ 3,\ 1)$, $B = (1,\ 1,\ 0)$; $\mathcal{M}$ is the line CD, where $C = (2,\ 0,\ 2)$, $D = (-1,\ 3,\ -2)$; $\mathcal{N}$ is the line EF, where $E = (1,\ 4,\ 7)$, $F = (-4,\ -3,\ -13)$. Find which pairs of lines intersect and also the points of intersection.

Solution. In fact only $\mathcal{L}$ and $\mathcal{N}$ intersect, in the point $(-\frac{2}{3},\ \frac{5}{3},\ \frac{1}{3})$. For example, to determine if $\mathcal{L}$ and $\mathcal{N}$ meet, we start with vector equations for $\mathcal{L}$ and $\mathcal{N}$:

$$\mathbf{P} = \mathbf{A} + t\,\overrightarrow{AB}, \quad \mathbf{Q} = \mathbf{E} + s\,\overrightarrow{EF},$$

equate $\mathbf{P}$ and $\mathbf{Q}$ and solve for s and t:

$$(-4\mathbf{i} + 3\mathbf{j} + k) + t(5\mathbf{i} - 2\mathbf{j} - \mathbf{k}) = (\mathbf{i} + 4\mathbf{j} + 7\mathbf{k}) + s(-5\mathbf{i} - 7\mathbf{j} - 20\mathbf{k}),$$

which on simplifying, gives

$$\begin{aligned} 5t + 5s &= 5 \\ -2t + 7s &= 1 \\ -t + 20s &= 6 \end{aligned}$$

This system has the unique solution $t = \frac{2}{3}$, $s = \frac{1}{3}$ and this determines a corresponding point P where the lines meet, namely $P = (-\frac{2}{3},\ \frac{5}{3},\ \frac{1}{3})$.

The same method yields inconsistent systems when applied to the other pairs of lines.

EXAMPLE 8.4.2 If $A = (5,\ 0,\ 7)$ and $B = (2,\ -3,\ 6)$, find the points P on the line AB which satisfy $AP/PB = 3$.

Solution. Use the formulae

$$\mathbf{P} = \mathbf{A} + t\,\overrightarrow{AB} \quad \text{and} \quad \left|\frac{t}{1-t}\right| = \frac{AP}{PB} = 3.$$

Then

$$\frac{t}{1-t} = 3 \text{ or } -3,$$

so $t = \frac{3}{4}$ or $t = \frac{3}{2}$. The corresponding points are $(\frac{11}{4},\ \frac{9}{4},\ \frac{25}{4})$ and $(\frac{1}{2},\ \frac{9}{2},\ \frac{11}{2})$.

DEFINITION 8.4.2 Let X and Y be non–zero vectors. Then X is *parallel* or *proportional* to Y if $X = tY$ for some $t \in \mathbb{R}$. We write $X \| Y$ if X is parallel to Y. If $X = tY$, we say that X and Y have the *same* or *opposite direction*, according as $t > 0$ or $t < 0$.

DEFINITION 8.4.3 if A and B are distinct points on a line $\mathcal{L}$, the non–zero vector $\overrightarrow{AB}$ is called a *direction vector* for $\mathcal{L}$.

It is easy to prove that any two direction vectors for a line are parallel.

DEFINITION 8.4.4 Let $\mathcal{L}$ and $\mathcal{M}$ be lines having direction vectors X and Y, respectively. Then $\mathcal{L}$ is *parallel* to $\mathcal{M}$ if X is parallel to Y. Clearly any line is parallel to itself.

It is easy to prove that the line through a given point A and parallel to a given line CD has an equation $\mathbf{P} = \mathbf{A} + t\,\overrightarrow{CD}$.

THEOREM 8.4.3 Let $X = a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$ and $Y = a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}$ be non–zero vectors. Then X is parallel to Y if and only if

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix} = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} = 0. \tag{8.8}$$

Proof. The case equality in the Cauchy–Schwarz inequality (Theorem 8.3.1) shows that X and Y are parallel if and only if

$$|X \cdot Y| = ||X|| \cdot ||Y||.$$

Squaring gives the equivalent equality

$$(a_1a_2 + b_1b_2 + c_1c_2)^2 = (a_1^2 + b_1^2 + c_1^2)(a_2^2 + b_2^2 + c_2^2),$$

which simplifies to

$$(a_1b_2 - a_2b_1)^2 + (b_1c_2 - b_2c_1)^2 + (a_1c_2 - a_2c_1)^2 = 0,$$

which is equivalent to

$$a_1b_2 - a_2b_1 = 0,\ b_1c_2 - b_2c_1 = 0,\ a_1c_2 - a_2c_1 = 0,$$

which is equation 8.8.

Equality of geometrical vectors has a fundamental geometrical interpretation:

THEOREM 8.4.4 Suppose A, B, C, D are distinct points such that no three are collinear. Then $\overrightarrow{AB}=\overrightarrow{CD}$ if and only if $\overrightarrow{AB} \parallel \overrightarrow{CD}$ and $\overrightarrow{AC} \parallel \overrightarrow{BD}$ (See Figure 8.1.)

Proof. If $\overrightarrow{AB}=\overrightarrow{CD}$ then

$$\begin{aligned} \mathbf{B}-\mathbf{A} &= \mathbf{D}-\mathbf{C}, \\ \mathbf{C}-\mathbf{A} &= \mathbf{D}-\mathbf{B} \end{aligned}$$

and so $\overrightarrow{AC}=\overrightarrow{BD}$. Hence $\overrightarrow{AB} \parallel \overrightarrow{CD}$ and $\overrightarrow{AC} \parallel \overrightarrow{BD}$.

Conversely, suppose that $\overrightarrow{AB} \parallel \overrightarrow{CD}$ and $\overrightarrow{AC} \parallel \overrightarrow{BD}$. Then

$$\overrightarrow{AB}= s\,\overrightarrow{CD} \quad \text{and} \quad \overrightarrow{AC}= t\,\overrightarrow{BD},$$

or

$$\mathbf{B}-\mathbf{A} = s(\mathbf{D}-\mathbf{C}) \quad \text{and} \quad \mathbf{C}-\mathbf{A} = t\mathbf{D}-\mathbf{B}.$$

We have to prove $s = 1$ or equivalently, $t = 1$.

Now subtracting the second equation above from the first, gives

$$\mathbf{B}-\mathbf{C} = s(\mathbf{D}-\mathbf{C}) - t(\mathbf{D}-\mathbf{B}),$$

so

$$(1-t)\mathbf{B} = (1-s)\mathbf{C} + (s-t)\mathbf{D}.$$

If $t \neq 1$, then

$$\mathbf{B} = \frac{1-s}{1-t}\mathbf{C} + \frac{s-t}{1-t}\mathbf{D}$$

and B would lie on the line CD. Hence $t = 1$.

8.5 The angle between two vectors

DEFINITION 8.5.1 Let X and Y be non–zero vectors. Then the *angle* between X and Y is the unique value of θ defined by

$$\cos\theta = \frac{X\cdot Y}{||X||\cdot||Y||}, \quad 0\le\theta\le\pi.$$

REMARK 8.5.1 By Cauchy's inequality, we have

$$-1 \le \frac{X\cdot Y}{||X||\cdot||Y||} \le 1,$$

so the above equation does define an angle θ.

In terms of components, if $X = [a_1,\, b_1,\, c_1]^t$ and $Y = [a_2,\, b_2,\, c_2]^t$, then

$$\cos\theta = \frac{a_1a_2 + b_1b_2 + c_1c_2}{\sqrt{a_1^2+b_1^2+c_1^2}\sqrt{a_2^2+b_2^2+c_2^2}}. \tag{8.9}$$

The next result is the well-known *cosine rule* for a triangle.

THEOREM 8.5.1 (Cosine rule) If A, B, C are points with $A\ne B$ and $A\ne C$, then the angle θ between vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$ satifies

$$\cos\theta = \frac{AB^2 + AC^2 - BC^2}{2AB\cdot AC}, \tag{8.10}$$

or equivalently

$$BC^2 = AB^2 + AC^2 - 2AB\cdot AC\cos\theta.$$

(See Figure 8.10.)

Proof. Let $A = (x_1,\, y_1,\, z_1)$, $B = (x_2,\, y_2,\, z_2)$, $C = (x_3,\, y_3,\, z_3)$. Then

$$\begin{aligned}
\overrightarrow{AB} &= a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}\\
\overrightarrow{AC} &= a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}\\
\overrightarrow{BC} &= (a_2-a_1)\mathbf{i} + (b_2-b_1)\mathbf{j} + (c_2-c_1)\mathbf{k},
\end{aligned}$$

where

$$a_i = x_{i+1} - x_1,\ b_i = y_{i+1} - y_1,\ c_i = z_{i+1} - z_1,\ i = 1,\, 2.$$

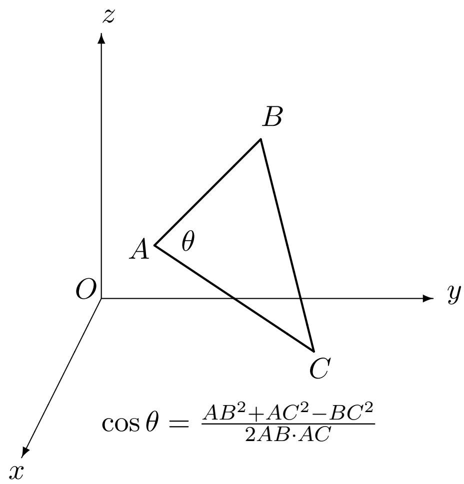

Figure 8.10: The cosine rule for a triangle.

Now by equation 8.9,

$$\cos\theta = \frac{a_1a_2 + b_1b_2 + c_1c_2}{AB \cdot AC}.$$

Also

$$\begin{aligned} AB^2 + AC^2 - BC^2 &= (a_1^2 + b_1^2 + c_1^2) + (a_2^2 + b_2^2 + c_2^2) \\ &\quad - ((a_2 - a_1)^2 + (b_2 - b_1)^2 + (c_2 - c_1)^2) \\ &= 2a_1a_2 + 2b_1b_2 + c_1c_2. \end{aligned}$$

Equation 8.10 now follows, since

$$\overrightarrow{AB} \cdot \overrightarrow{AC} = a_1a_2 + b_1b_2 + c_1c_2.$$

EXAMPLE 8.5.1 Let $A = (2, 1, 0)$, $B = (3, 2, 0)$, $C = (5, 0, 1)$. Find the angle θ between vectors $\overrightarrow{AB}$ and $\overrightarrow{AC}$.

Solution.

$$\cos\theta = \frac{\overrightarrow{AB} \cdot \overrightarrow{AC}}{AB \cdot AC}.$$

Now

$$\overrightarrow{AB} = \mathbf{i} + \mathbf{j} \quad \text{and} \quad \overrightarrow{AC} = 3\mathbf{i} - \mathbf{j} + \mathbf{k}.$$

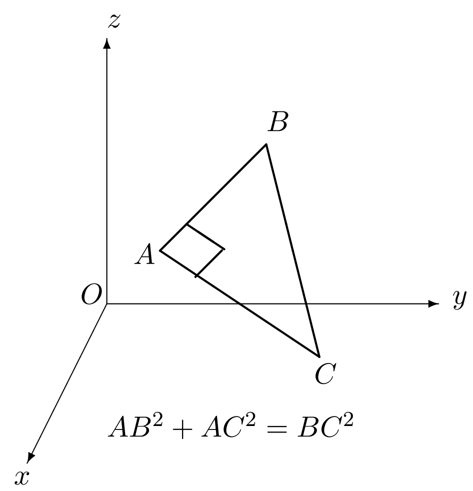

Figure 8.11: Pythagoras' theorem for a right–angled triangle.

Hence

$$\cos\theta = \frac{1\times 3 + 1\times(-1) + 0\times 1}{\sqrt{1^2+1^2+0^2}\sqrt{3^2+(-1)^2+1^2}} = \frac{2}{\sqrt{2}\sqrt{11}} = \frac{\sqrt{2}}{\sqrt{11}}.$$

Hence $\theta = \cos^{-1}\frac{\sqrt{2}}{\sqrt{11}}$.

DEFINITION 8.5.2 If X and Y are vectors satisfying $X\cdot Y = 0$, we say X is *orthogonal* or *perpendicular* to Y.

REMARK 8.5.2 If A, B, C are points forming a triangle and $\overrightarrow{AB}$ is orthogonal to $\overrightarrow{AC}$, then the angle θ between $\overrightarrow{AB}$ and $\overrightarrow{AC}$ satisfies $\cos\theta = 0$ and hence $\theta = \frac{\pi}{2}$ and the triangle is *right–angled* at A.

Then we have *Pythagoras' theorem*:

$$BC^2 = AB^2 + AC^2. \tag{8.11}$$

We also note that $BC \geq AB$ and $BC \geq AC$ follow from equation 8.11. (See Figure 8.11.)

EXAMPLE 8.5.2 Let $A = (2, 9, 8)$, $B = (6, 4, -2)$, $C = (7, 15, 7)$.

Show that $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are perpendicular and find the point D such that $ABDC$ forms a rectangle.

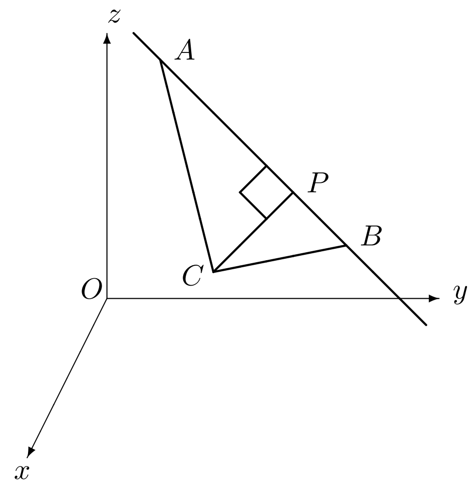

Figure 8.12: Distance from a point to a line.

Solution.

$$\overrightarrow{AB} \cdot \overrightarrow{AC} = (4\mathbf{i} - 5\mathbf{j} - 10\mathbf{k}) \cdot (5\mathbf{i} + 6\mathbf{j} - \mathbf{k}) = 20 - 30 + 10 = 0.$$

Hence $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are perpendicular. Also, the required fourth point D clearly has to satisfy the equation

$$\overrightarrow{BD} = \overrightarrow{AC}, \text{ or equivalently } \mathbf{D} - \mathbf{B} = \overrightarrow{AC}.$$

Hence

$$\mathbf{D} = \mathbf{B} + \overrightarrow{AC} = (6\mathbf{i} + 4\mathbf{j} - 2\mathbf{k}) + (5\mathbf{i} + 6\mathbf{j} - \mathbf{k}) = 11\mathbf{i} + 10\mathbf{j} - 3\mathbf{k},$$

so $D = (11,\ 10,\ -3)$.

THEOREM 8.5.2 (Distance from a point to a line) If C is a point and $\mathcal{L}$ is the line through A and B, then there is exactly one point P on $\mathcal{L}$ such that $\overrightarrow{CP}$ is perpendicular to $\overrightarrow{AB}$, namely

$$\mathbf{P} = \mathbf{A} + t\,\overrightarrow{AB}, \quad t = \frac{\overrightarrow{AC} \cdot \overrightarrow{AB}}{AB^2}. \tag{8.12}$$

Moreover if Q is any point on $\mathcal{L}$, then $CQ \geq CP$ and hence P is the point on $\mathcal{L}$ closest to C.

The shortest distance CP is given by

$$CP = \frac{\sqrt{AC^2AB^2 - (\overrightarrow{AC} \cdot \overrightarrow{AB})^2}}{AB}. \tag{8.13}$$

(See Figure 8.12.)

Proof. Let $\mathbf{P} = \mathbf{A} + t\,\overrightarrow{AB}$ and assume that $\overrightarrow{CP}$ is perpendicular to $\overrightarrow{AB}$. Then

$$\begin{aligned}
\overrightarrow{CP} \cdot \overrightarrow{AB} &= 0\\
(\mathbf{P} - \mathbf{C})\cdot \overrightarrow{AB} &= 0\\
(\mathbf{A} + t\,\overrightarrow{AB} - \mathbf{C})\cdot \overrightarrow{AB} &= 0\\
(\overrightarrow{CA} + t\,\overrightarrow{AB})\cdot \overrightarrow{AB} &= 0\\
\overrightarrow{CA} \cdot \overrightarrow{AB} + t(\overrightarrow{AB} \cdot \overrightarrow{AB}) &= 0\\
-\overrightarrow{AC} \cdot \overrightarrow{AB} + t(\overrightarrow{AB} \cdot \overrightarrow{AB}) &= 0,
\end{aligned}$$

so equation 8.12 follows.

The inequality $CQ \geq CP$, where Q is any point on $\mathcal{L}$, is a consequence of Pythagoras' theorem.

Finally, as $\overrightarrow{CP}$ and $\overrightarrow{PA}$ are perpendicular, Pythagoras' theorem gives

$$\begin{aligned}
CP^2 &= AC^2 - PA^2\\
&= AC^2 - \|t\,\overrightarrow{AB}\|^2\\
&= AC^2 - t^2AB^2\\
&= AC^2 - \left(\frac{\overrightarrow{AC} \cdot \overrightarrow{AB}}{AB^2}\right)^2 AB^2\\
&= \frac{AC^2AB^2 - (\overrightarrow{AC} \cdot \overrightarrow{AB})^2}{AB^2},
\end{aligned}$$

as required.

EXAMPLE 8.5.3 The closest point on the line through $A = (1, 2, 1)$ and $B = (2, -1, 3)$ to the origin is $P = (\frac{17}{14}, \frac{19}{14}, \frac{20}{14})$ and the corresponding shortest distance equals $\frac{5}{14}\sqrt{42}$.

Another application of theorem 8.5.2 is to the projection of a line segment on another line:

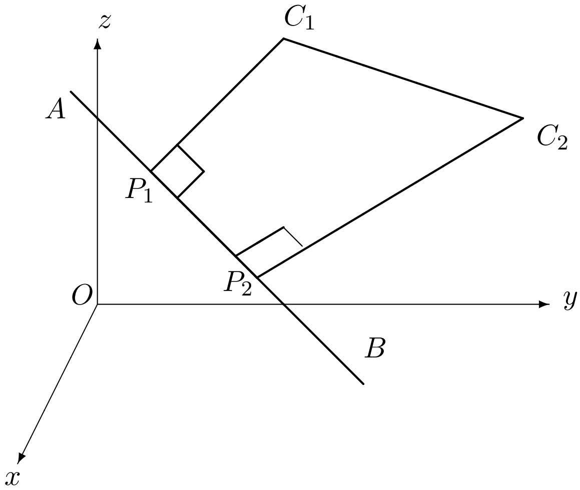

Figure 8.13: Projecting the segment C_1C_2 onto the line AB.

THEOREM 8.5.3 (The projection of a line segment onto a line) Let C_1, C_2 be points and P_1, P_2 be the feet of the perpendiculars from C_1 and C_2 to the line AB. Then

$$P_1P_2 = |\overrightarrow{C_1C_2} \cdot \hat{n}|,$$

where

$$\hat{n} = \frac{1}{AB}\overrightarrow{AB}.$$

Also

$$C_1C_2 \geq P_1P_2. \tag{8.14}$$

(See Figure 8.13.)

Proof. Using equations 8.12, we have

$$\mathbf{P_1} = \mathbf{A} + t_1\overrightarrow{AB}, \quad \mathbf{P_2} = \mathbf{A} + t_2\overrightarrow{AB},$$

where

$$t_1 = \frac{\overrightarrow{AC_1} \cdot \overrightarrow{AB}}{AB^2}, \quad t_2 = \frac{\overrightarrow{AC_2} \cdot \overrightarrow{AB}}{AB^2}.$$

Hence

$$\begin{aligned}\overrightarrow{P_1P_2} &= (\mathbf{A} + t_2\overrightarrow{AB}) - (\mathbf{A} + t_1\overrightarrow{AB}) \\ &= (t_2 - t_1)\overrightarrow{AB},\end{aligned}$$

so

$$\begin{aligned} P_1P_2 &= \|\, \overrightarrow{P_1P_2} \,\| = |t_2 - t_1| AB \\ &= \left| \frac{\overrightarrow{AC_2} \cdot \overrightarrow{AB}}{AB^2} - \frac{\overrightarrow{AC_1} \cdot \overrightarrow{AB}}{AB^2} \right| AB \\ &= \frac{\left| \overrightarrow{C_1C_2} \cdot \overrightarrow{AB} \right|}{AB^2} AB \\ &= \left| \overrightarrow{C_1C_2} \cdot \hat{n} \right|, \end{aligned}$$

where $\hat{n}$ is the unit vector

$$\hat{n} = \frac{1}{AB} \overrightarrow{AB} .$$

Inequality 8.14 then follows from the Cauchy–Schwarz inequality 8.3.

DEFINITION 8.5.3 Two non–intersecting lines are called *skew* if they have non–parallel direction vectors.

Theorem 8.5.3 has an application to the problem of showing that two skew lines have a shortest distance between them. (The reader is referred to problem 16 at the end of the chapter.)

Before we turn to the study of planes, it is convenient to introduce the cross–product of two vectors.

8.6 The cross–product of two vectors

DEFINITION 8.6.1 Let $X = a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$ and $Y = a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}$. Then $X \times Y$, the *cross–product* of X and Y, is defined by

$$X \times Y = a\mathbf{i} + b\mathbf{j} + c\mathbf{k},$$

where

$$a = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \quad b = -\begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix}, \quad c = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}.$$

The vector cross–product has the following properties which follow from properties of 2×2 and 3×3 determinants:

(i) $\mathbf{i} \times \mathbf{j} = \mathbf{k}, \quad \mathbf{j} \times \mathbf{k} = \mathbf{i}, \quad \mathbf{k} \times \mathbf{i} = \mathbf{j};$

(ii) $X \times X = 0$;

(iii) $Y \times X = -X \times Y$;

(iv) $X \times (Y + Z) = X \times Y + X \times Z$;

(v) $(tX) \times Y = t(X \times Y)$;

(vi) (Scalar triple product formula) if $Z = a_3\mathbf{i} + b_3\mathbf{j} + c_3\mathbf{k}$, then

$$X \cdot (Y \times Z) = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = (X \times Y) \cdot Z;$$

(vii) $X \cdot (X \times Y) = 0 = Y \cdot (X \times Y)$;

(viii) $||X \times Y|| = \sqrt{||X||^2||Y||^2 - (X \cdot Y)^2}$;

(ix) if X and Y are non–zero vectors and θ is the angle between X and Y, then

$$||X \times Y|| = ||X|| \cdot ||Y|| \sin\theta.$$

(See Figure 8.14.)

From theorem 8.4.3 and the definition of cross–product, it follows that non–zero vectors X and Y are parallel if and only if $X \times Y = 0$; hence by (vii), the cross–product of two non–parallel, non–zero vectors X and Y, is a non–zero vector perpendicular to both X and Y.

LEMMA 8.6.1 Let X and Y be non–zero, non–parallel vectors.

(i) Z is a linear combination of X and Y, if and only if Z is perpendicular to $X \times Y$;

(ii) Z is perpendicular to X and Y, if and only if Z is parallel to $X \times Y$.

Proof. Let X and Y be non–zero, non–parallel vectors. Then

$$X \times Y \neq 0.$$

Then if $X \times Y = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$, we have

$$\det\,[X \times Y|X|Y]^t = \begin{vmatrix} a & b & c \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = (X \times Y) \cdot (X \times Y) > 0.$$

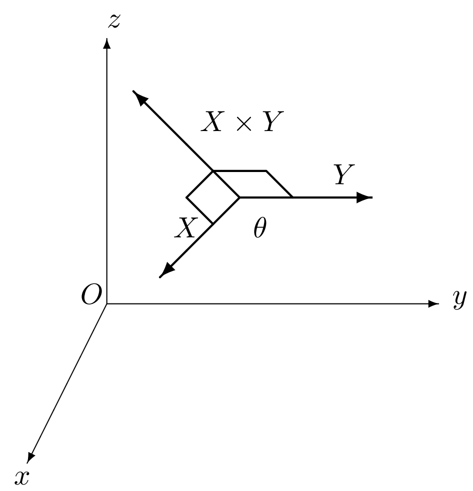

Figure 8.14: The vector cross–product.

Hence the matrix $[X \times Y|X|Y]$ is non–singular. Consequently the linear system

$$r(X \times Y) + sX + tY = Z \tag{8.15}$$

has a unique solution r, s, t.
(i) Suppose $Z = sX + tY$. Then

$$Z \cdot (X \times Y) = sX \cdot (X \times Y) + tY \cdot (X \times Y) = s0 + t0 = 0.$$

Conversely, suppose that

$$Z \cdot (X \times Y) = 0. \tag{8.16}$$

Now from equation 8.15, r, s, t exist satisfying

$$Z = r(X \times Y) + sX + tY.$$

Then equation 8.16 gives

$$\begin{aligned} 0 &= (r(X \times Y) + sX + tY) \cdot (X \times Y) \\ &= r||X \times Y||^2 + sX \cdot (X \times Y) + tY \cdot (Y \times X) \\ &= r||X \times Y||^2. \end{aligned}$$

Hence $r = 0$ and $Z = sX + tY$, as required.
(ii) Suppose $Z = \lambda(X \times Y)$. Then clearly Z is perpendicular to X and Y.

Conversely suppose that Z is perpendicular to X and Y.
Now from equation 8.15, r, s, t exist satisfying

$$Z = r(X \times Y) + sX + tY.$$

Then

$$\begin{aligned} sX \cdot X + tX \cdot Y &= X \cdot Z = 0 \\ sY \cdot X + tY \cdot Y &= Y \cdot Z = 0, \end{aligned}$$

from which it follows that

$$(sX + tY) \cdot (sX + tY) = 0.$$

Hence $sX + tY = 0$ and so $s = 0$, $t = 0$. Consequently $Z = r(X \times Y)$, as required.

The cross–product gives a compact formula for the distance from a point to a line, as well as the area of a triangle.

THEOREM 8.6.1 (Area of a triangle)
If A, B, C are distinct non–collinear points, then

(i) the distance d from C to the line AB is given by

$$d = \frac{||\ \overrightarrow{AB} \times \overrightarrow{AC}\ ||}{AB}, \tag{8.17}$$

(ii) the area of the triangle ABC equals

$$\frac{||\ \overrightarrow{AB} \times \overrightarrow{AC}\ ||}{2} = \frac{||\mathbf{A} \times \mathbf{B} + \mathbf{B} \times \mathbf{C} + \mathbf{C} \times \mathbf{A}||}{2}. \tag{8.18}$$

Proof. The area Δ of triangle ABC is given by

$$\Delta = \frac{AB \cdot CP}{2},$$

where P is the foot of the perpendicular from C to the line AB. Now by formula 8.13, we have

$$\begin{aligned} CP &= \frac{\sqrt{AC^2 \cdot AB^2 - (\overrightarrow{AC} \cdot \overrightarrow{AB})^2}}{AB} \\ &= \frac{||\ \overrightarrow{AB} \times \overrightarrow{AC}\ ||}{AB}, \end{aligned}$$

which, by property (viii) of the cross–product, gives formula 8.17. The second formula of equation 8.18 follows from the equations

$$\begin{aligned}\overrightarrow{AB} \times \overrightarrow{AC} &= (\mathbf{B} - \mathbf{A}) \times (\mathbf{C} - \mathbf{A}) \\ &= \{(\mathbf{B} - \mathbf{A}) \times \mathbf{C}\} - \{(\mathbf{C} - \mathbf{A}) \times \mathbf{A}\} \\ &= \{(\mathbf{B} \times \mathbf{C} - \mathbf{A} \times \mathbf{C})\} - \{(\mathbf{B} \times \mathbf{A} - \mathbf{A} \times \mathbf{A})\} \\ &= \mathbf{B} \times \mathbf{C} - \mathbf{A} \times \mathbf{C} - \mathbf{B} \times \mathbf{A} \\ &= \mathbf{B} \times \mathbf{C} + \mathbf{C} \times \mathbf{A} + \mathbf{A} \times \mathbf{B},\end{aligned}$$

as required.

8.7 Planes

DEFINITION 8.7.1 A *plane* is a set of points P satisfying an equation of the form

$$\mathbf{P} = \mathbf{P}_0 + sX + tY, \quad s,\, t \in \mathbb{R}, \tag{8.19}$$

where X and Y are non–zero, non–parallel vectors.

For example, the xy–plane consists of the points $P = (x,\, y,\, 0)$ and corresponds to the plane equation

$$\mathbf{P} = x\mathbf{i} + y\mathbf{j} = \mathbf{O} + x\mathbf{i} + y\mathbf{j}.$$

In terms of coordinates, equation 8.19 takes the form

$$\begin{aligned}x &= x_0 + sa_1 + ta_2 \\ y &= y_0 + sb_1 + tb_2 \\ z &= z_0 + sc_1 + tc_2,\end{aligned}$$

where $P_0 = (x_0,\, y_0,\, z_0)$ and $(a_1,\, b_1,\, c_1)$ and $(a_2,\, b_2,\, c_2)$ are non–zero and non–proportional.

THEOREM 8.7.1 Let A, B, C be three non–collinear points. Then there is one and only one plane through these points, namely the plane given by the equation

$$\mathbf{P} = \mathbf{A} + s\,\overrightarrow{AB} + t\,\overrightarrow{AC}, \tag{8.20}$$

or equivalently

$$\overrightarrow{AP} = s\,\overrightarrow{AB} + t\,\overrightarrow{AC}\,. \tag{8.21}$$

(See Figure 8.15.)

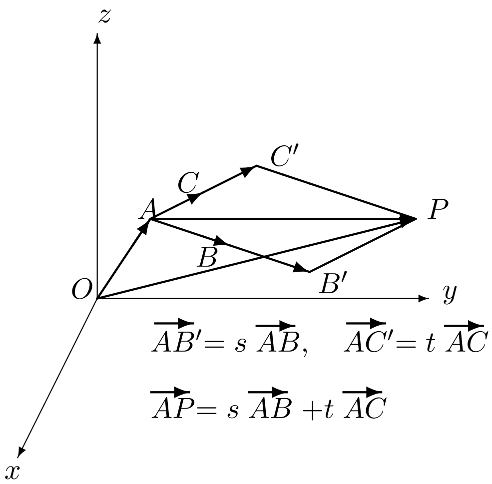

Figure 8.15: Vector equation for the plane ABC.

Proof. First note that equation 8.20 is indeed the equation of a plane through A, B and C, as $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are non–zero and non–parallel and $(s, t) = (0, 0)$, $(1, 0)$ and $(0, 1)$ give $P = A$, B and C, respectively. Call this plane $\mathcal{P}$.

Conversely, suppose $\mathbf{P} = \mathbf{P}_0 + sX + tY$ is the equation of a plane $\mathcal{Q}$ passing through A, B, C. Then $\mathbf{A} = \mathbf{P}_0 + s_0X + t_0Y$, so the equation for $\mathcal{Q}$ may be written

$$\mathbf{P} = \mathbf{A} + (s - s_0)X + (t - t_0)Y = \mathbf{A} + s'X + t'Y;$$

so in effect we can take $P_0 = A$ in the equation of $\mathcal{Q}$. Then the fact that B and C lie on $\mathcal{Q}$ gives equations

$$\mathbf{B} = \mathbf{A} + s_1X + t_1Y, \quad \mathbf{C} = \mathbf{A} + s_2X + t_2Y,$$

or

$$\overrightarrow{AB} = s_1X + t_1Y, \quad \overrightarrow{AC} = s_2X + t_2Y. \tag{8.22}$$

Then equations 8.22 and equation 8.20 show that

$$\mathcal{P} \subseteq \mathcal{Q}.$$

Conversely, it is straightforward to show that because $\overrightarrow{AB}$ and $\overrightarrow{AC}$ are not parallel, we have

$$\begin{vmatrix} s_1 & t_1 \\ s_2 & t_2 \end{vmatrix} \neq 0.$$

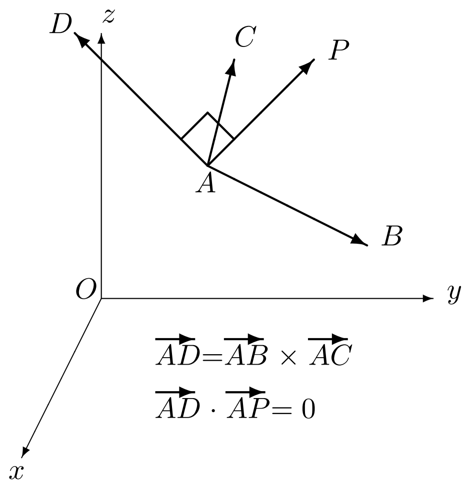

Figure 8.16: Normal equation of the plane ABC.

Hence equations 8.22 can be solved for X and Y as linear combinations of $\overrightarrow{AB}$ and $\overrightarrow{AC}$, allowing us to deduce that

$$\mathcal{Q} \subseteq \mathcal{P}.$$

Hence

$$\mathcal{Q} = \mathcal{P}.$$

THEOREM 8.7.2 (Normal equation for a plane) Let

$$A = (x_1,\, y_1,\, z_1),\ B = (x_2,\, y_2,\, z_2),\ C = (x_3,\, y_3,\, z_3)$$

be three non–collinear points. Then the plane through A, B, C is given by

$$\overrightarrow{AP} \cdot (\overrightarrow{AB} \times \overrightarrow{AC}) = 0, \tag{8.23}$$

or equivalently,

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0, \tag{8.24}$$

where $P = (x,\, y,\, z)$. (See Figure 8.16.)

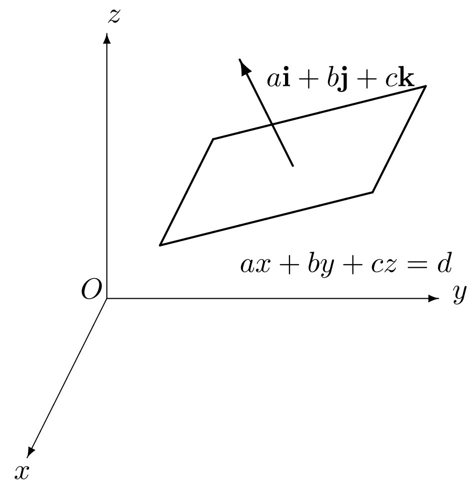

Figure 8.17: The plane $ax + by + cz = d$.

REMARK 8.7.1 Equation 8.24 can be written in more symmetrical form as

$$\begin{vmatrix} x & y & z & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0. \tag{8.25}$$

Proof. Let $\mathcal{P}$ be the plane through A, B, C. Then by equation 8.21, we have $P \in \mathcal{P}$ if and only if $\overrightarrow{AP}$ is a linear combination of $\overrightarrow{AB}$ and $\overrightarrow{AC}$ and so by lemma 8.6.1(i), using the fact that $\overrightarrow{AB} \times \overrightarrow{AC} \neq 0$ here, if and only if $\overrightarrow{AP}$ is perpendicular to $\overrightarrow{AB} \times \overrightarrow{AC}$. This gives equation 8.23.

Equation 8.24 is the scalar triple product version of equation 8.23, taking into account the equations

$$\begin{aligned} \overrightarrow{AP} &= (x - x_1)\mathbf{i} + (y - y_1)\mathbf{j} + (z - z_1)\mathbf{k}, \\ \overrightarrow{AB} &= (x_2 - x_1)\mathbf{i} + (y_2 - y_1)\mathbf{j} + (z_2 - z_1)\mathbf{k}, \\ \overrightarrow{AC} &= (x_3 - x_1)\mathbf{i} + (y_3 - y_1)\mathbf{j} + (z_3 - z_1)\mathbf{k}. \end{aligned}$$

REMARK 8.7.2 Equation 8.24 gives rise to a linear equation in x, y and z:

$$ax + by + cz = d,$$

where $a\mathbf{i}+b\mathbf{j}+c\mathbf{k} \neq 0$. For

$$\begin{vmatrix} x-x_1 & y-y_1 & z-z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix} =$$

$$\begin{vmatrix} x & y & z \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix} - \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2-x_1 & y_2-y_1 & z_2-z_1 \\ x_3-x_1 & y_3-y_1 & z_3-z_1 \end{vmatrix} \quad (8.26)$$

and expanding the first determinant on the right–hand side of equation 8.26 along row 1 gives an expression

$$ax+by+cz$$

where

$$a = \begin{vmatrix} y_2-y_1 & z_2-z_1 \\ y_3-y_1 & z_3-z_1 \end{vmatrix}, \ b = -\begin{vmatrix} x_2-x_1 & z_2-z_1 \\ x_3-x_1 & z_3-z_1 \end{vmatrix}, \ c = \begin{vmatrix} x_2-x_1 & y_2-y_1 \\ x_3-x_1 & y_3-y_1 \end{vmatrix}.$$

But a, b, c are the components of $\overrightarrow{AB} \times \overrightarrow{AC}$, which in turn is non–zero, as A, B, C are non–collinear here.

Conversely if $a\mathbf{i}+b\mathbf{j}+c\mathbf{k} \neq 0$, the equation

$$ax+by+cz=d$$

does indeed represent a plane. For if say $a \neq 0$, the equation can be solved for x in terms of y and z:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\frac{d}{a} \\ 0 \\ 0 \end{bmatrix} + y \begin{bmatrix} -\frac{b}{a} \\ 1 \\ 0 \end{bmatrix} + z \begin{bmatrix} -\frac{c}{a} \\ 0 \\ 1 \end{bmatrix},$$

which gives the plane

$$\mathbf{P} = \mathbf{P}_0 + yX + zY,$$

where $P_0 = (-\frac{d}{a}, 0, 0)$ and $X = -\frac{b}{a}\mathbf{i}+\mathbf{j}$ and $Y = -\frac{c}{a}\mathbf{i}+\mathbf{k}$ are evidently non–parallel vectors.

REMARK 8.7.3 The plane equation $ax+by+cz=d$ is called the *normal form*, as it is easy to prove that if P_1 and P_2 are two points in the plane, then $a\mathbf{i}+b\mathbf{j}+c\mathbf{k}$ is perpendicular to $\overrightarrow{P_1P_2}$. Any non–zero vector with this property is called a *normal* to the plane. (See Figure 8.17.)

By lemma 8.6.1(ii), it follows that every vector X normal to a plane through three non–collinear points A, B, C is parallel to $\overrightarrow{AB} \times \overrightarrow{AC}$, since X is perpendicular to $\overrightarrow{AB}$ and $\overrightarrow{AC}$.

EXAMPLE 8.7.1 Show that the planes

$$x + y - 2z = 1 \quad \text{and} \quad x + 3y - z = 4$$

intersect in a line and find the distance from the point $C = (1,\, 0,\, 1)$ to this line.

Solution. Solving the two equations simultaneously gives

$$x = -\frac{1}{2} + \frac{5}{2}z, \quad y = \frac{3}{2} - \frac{1}{2}z, \tag{8.27}$$

where z is arbitrary. Hence

$$x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = -\frac{1}{2}\mathbf{i} - \frac{3}{2}\mathbf{j} + z(\frac{5}{2}\mathbf{i} - \frac{1}{2}\mathbf{j} + \mathbf{k}),$$

which is the equation of a line $\mathcal{L}$ through $A = (-\frac{1}{2},\, -\frac{3}{2},\, 0)$ and having direction vector $\frac{5}{2}\mathbf{i} - \frac{1}{2}\mathbf{j} + \mathbf{k}$.

We can now proceed in one of three ways to find the closest point on $\mathcal{L}$ to A.

One way is to use equation 8.17 with B defined by

$$\overrightarrow{AB} = \frac{5}{2}\mathbf{i} - \frac{1}{2}\mathbf{j} + \mathbf{k}.$$

Another method minimizes the distance CP, where P ranges over $\mathcal{L}$.

A third way is to find an equation for the plane through C, having $\frac{5}{2}\mathbf{i} - \frac{1}{2}\mathbf{j} + \mathbf{k}$ as a normal. Such a plane has equation

$$5x - y + 2z = d,$$

where d is found by substituting the coordinates of C in the last equation.

$$d = 5 \times 1 - 0 + 2 \times 1 = 7.$$

We now find the point P where the plane intersects the line $\mathcal{L}$. Then $\overrightarrow{CP}$ will be perpendicular to $\mathcal{L}$ and CP will be the required shortest distance from C to $\mathcal{L}$. We find using equations 8.27 that

$$5(-\frac{1}{2} + \frac{5}{2}z) - (\frac{3}{2} - \frac{1}{2}z) + 2z = 7,$$

so $z = \frac{11}{15}$. Hence $P = (\frac{4}{3},\, \frac{17}{15},\, \frac{11}{15})$.

It is clear that through a given line and a point not on that line, there passes exactly one plane. If the line is given as the intersection of two planes, each in normal form, there is a simple way of finding an equation for this plane. More explicitly we have the following result:

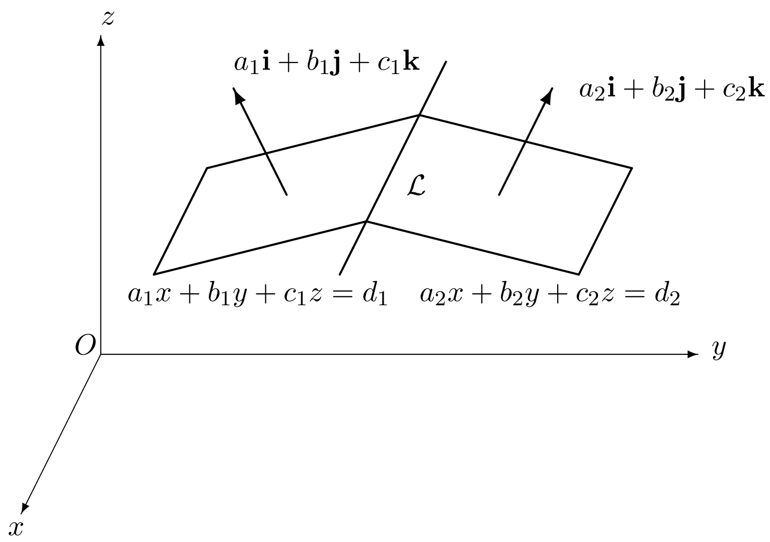

Figure 8.18: Line of intersection of two planes.

THEOREM 8.7.3 Suppose the planes

$$a_1x + b_1y + c_1z = d_1 \tag{8.28}$$
$$a_2x + b_2y + c_2z = d_2 \tag{8.29}$$

have non–parallel normals. Then the planes intersect in a line $\mathcal{L}$.

Moreover the equation

$$\lambda(a_1x + b_1y + c_1z - d_1) + \mu(a_2x + b_2y + c_2z - d_2) = 0, \tag{8.30}$$

where λ and μ are not both zero, gives all planes through $\mathcal{L}$.

(See Figure 8.18.)

Proof. Assume that the normals $a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$ and $a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}$ are non–parallel. Then by theorem 8.4.3, not all of

$$\Delta_1 = \begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}, \quad \Delta_2 = \begin{vmatrix} b_1 & c_1 \\ b_2 & c_2 \end{vmatrix}, \quad \Delta_3 = \begin{vmatrix} a_1 & c_1 \\ a_2 & c_2 \end{vmatrix} \tag{8.31}$$

are zero. If say $\Delta_1 \neq 0$, we can solve equations 8.28 and 8.29 for x and y in terms of z, as we did in the previous example, to show that the intersection forms a line $\mathcal{L}$.

We next have to check that if λ and μ are not both zero, then equation 8.30 represents a plane. (Whatever set of points equation 8.30 represents, this set certainly contains $\mathcal{L}$.)

$$(\lambda a_1 + \mu a_2)x + (\lambda b_1 + \mu b_2)y + (\lambda c_1 + \mu c_2)z - (\lambda d_1 + \mu d_2) = 0.$$

Then we clearly cannot have all the coefficients

$$\lambda a_1 + \mu a_2, \quad \lambda b_1 + \mu b_2, \quad \lambda c_1 + \mu c_2$$

zero, as otherwise the vectors $a_1\mathbf{i} + b_1\mathbf{j} + c_1\mathbf{k}$ and $a_2\mathbf{i} + b_2\mathbf{j} + c_2\mathbf{k}$ would be parallel.

Finally, if $\mathcal{P}$ is a plane containing $\mathcal{L}$, let $P_0 = (x_0, y_0, z_0)$ be a point not on $\mathcal{L}$. Then if we define λ and μ by

$$\lambda = -(a_2x_0 + b_2y_0 + c_2z_0 - d_2), \quad \mu = a_1x_0 + b_1y_0 + c_1z_0 - d_1,$$

then at least one of λ and μ is non–zero. Then the coordinates of P_0 satisfy equation 8.30, which therefore represents a plane passing through $\mathcal{L}$ and P_0 and hence identical with $\mathcal{P}$.

EXAMPLE 8.7.2 Find an equation for the plane through $P_0 = (1, 0, 1)$ and passing through the line of intersection of the planes

$$x + y - 2z = 1 \quad \text{and} \quad x + 3y - z = 4.$$

Solution. The required plane has the form

$$\lambda(x + y - 2z - 1) + \mu(x + 3y - z - 4) = 0,$$

where not both of λ and μ are zero. Substituting the coordinates of P_0 into this equation gives

$$-2\lambda + \mu(-4) = 0, \quad \lambda = -2\mu.$$

So the required equation is

$$-2\mu(x + y - 2z - 1) + \mu(x + 3y - z - 4) = 0,$$

or

$$-x + y + 3z - 2 = 0.$$

Our final result is a formula for the distance from a point to a plane.

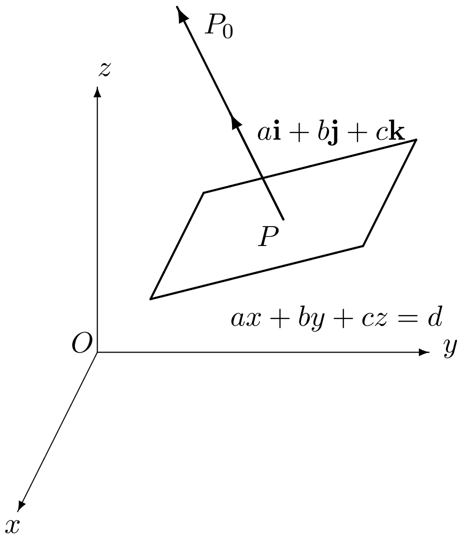

Figure 8.19: Distance from a point P_0 to the plane $ax + by + cz = d$.

THEOREM 8.7.4 (Distance from a point to a plane)
Let $P_0 = (x_0, y_0, z_0)$ and $\mathcal{P}$ be the plane

$$ax + by + cz = d. \tag{8.32}$$

Then there is a unique point P on $\mathcal{P}$ such that $\overrightarrow{P_0P}$ is normal to $\mathcal{P}$. Morever

$$P_0P = \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2 + b^2 + c^2}}$$

(See Figure 8.19.)
Proof. The line through P_0 normal to $\mathcal{P}$ is given by

$$\mathbf{P} = \mathbf{P}_0 + t(a\mathbf{i} + b\mathbf{j} + c\mathbf{k}),$$

or in terms of coordinates

$$x = x_0 + at, \quad y = y_0 + bt, \quad z = z_0 + ct.$$

Substituting these formulae in equation 8.32 gives

$$\begin{aligned} a(x_0 + at) + b(y_0 + bt) + c(z_0 + ct) &= d \\ t(a^2 + b^2 + c^2) &= -(ax_0 + by_0 + cz_0 - d), \end{aligned}$$

so

$$t = -\left(\frac{ax_0 + by_0 + cz_0 - d}{a^2 + b^2 + c^2}\right).$$

Then

$$\begin{aligned} P_0P = ||\ \overrightarrow{P_0P}\ || &= ||t(a\mathbf{i} + b\mathbf{j} + c\mathbf{k})|| \\ &= |t|\sqrt{a^2+b^2+c^2} \\ &= \frac{|ax_0 + by_0 + cz_0 - d|}{a^2+b^2+c^2}\sqrt{a^2+b^2+c^2} \\ &= \frac{|ax_0 + by_0 + cz_0 - d|}{\sqrt{a^2+b^2+c^2}}. \end{aligned}$$

Other interesting geometrical facts about lines and planes are left to the problems at the end of this chapter.

8.8 PROBLEMS

.

1. Find the point where the line through $A = (3,\ -2,\ 7)$ and $B = (13,\ 3,\ -8)$ meets the xz–plane.

   [Ans: (7, 0, 1).]

2. Let A, B, C be non–collinear points. If E is the mid–point of the segment BC and F is the point on the segment EA satisfying $\frac{AF}{EF} = 2$, prove that
$$\mathbf{F} = \frac{1}{3}(\mathbf{A} + \mathbf{B} + \mathbf{C}).$$
   (F is called the *centroid* of triangle ABC.)

3. Prove that the points $(2,\ 1,\ 4),\ (1,\ -1,\ 2),\ (3,\ 3,\ 6)$ are collinear.

4. If $A = (2,\ 3,\ -1)$ and $B = (3,\ 7,\ 4)$, find the points P on the line AB satisfying $PA/PB = 2/5$.

   [Ans: $\left(\frac{16}{7},\ \frac{29}{7},\ \frac{3}{7}\right)$ and $\left(\frac{4}{3},\ \frac{1}{3},\ -\frac{13}{3}\right)$.]

5. Let $\mathcal{M}$ be the line through $A = (1,\ 2,\ 3)$ parallel to the line joining $B = (-2,\ 2,\ 0)$ and $C = (4,\ -1,\ 7)$. Also $\mathcal{N}$ is the line joining $E = (1,\ -1,\ 8)$ and $F = (10,\ -1,\ 11)$. Prove that $\mathcal{M}$ and $\mathcal{N}$ intersect and find the point of intersection.

   [Ans: (7, −1, 10).]

6. Prove that the triangle formed by the points $(-3,\ 5,\ 6)$, $(-2,\ 7,\ 9)$ and $(2,\ 1,\ 7)$ is a 30^o, 60^o, 90^o triangle.

7. Find the point on the line AB closest to the origin, where $A = (-2,\ 1,\ 3)$ and $B = (1,\ 2,\ 4)$. Also find this shortest distance.

   [Ans: $\left(-\frac{16}{11},\ \frac{13}{11},\ \frac{35}{11}\right)$ and $\sqrt{\frac{150}{11}}$.]

8. A line $\mathcal{N}$ is determined by the two planes
$$x + y - 2z = 1, \qquad \text{and} \qquad x + 3y - z = 4.$$
   Find the point P on $\mathcal{N}$ closest to the point $C = (1,\ 0,\ 1)$ and find the distance PC.

   [Ans: $\left(\frac{4}{3},\ \frac{17}{15},\ \frac{11}{15}\right)$ and $\frac{\sqrt{330}}{15}$.]

9. Find a linear equation describing the plane perpendicular to the line of intersection of the planes $x + y - 2z = 4$ and $3x - 2y + z = 1$ and which passes through $(6,\ 0,\ 2)$.

   [Ans: $3x + 7y + 5z = 28$.]

10. Find the length of the projection of the segment AB on the line $\mathcal{L}$, where $A = (1,\ 2,\ 3)$, $B = (5,\ -2,\ 6)$ and $\mathcal{L}$ is the line CD, where $C = (7,\ 1,\ 9)$ and $D = (-1,\ 5,\ 8)$.

    [Ans: $\frac{17}{3}$.]

11. Find a linear equation for the plane through $A = (3,\ -1,\ 2)$, perpendicular to the line $\mathcal{L}$ joining $B = (2,\ 1,\ 4)$ and $C = (-3,\ -1,\ 7)$. Also find the point of intersection of $\mathcal{L}$ and the plane and hence determine the distance from A to $\mathcal{L}$. [Ans: $5x+2y-3z = 7$, $\left(\frac{111}{38},\ \frac{52}{38},\ \frac{131}{38}\right)$, $\sqrt{\frac{293}{38}}$.]

12. If P is a point inside the triangle ABC, prove that
$$\mathbf{P} = r\mathbf{A} + s\mathbf{B} + t\mathbf{C},$$
    where $r + s + t = 1$ and $r > 0$, $s > 0$, $t > 0$.

13. If B is the point where the perpendicular from $A = (6,\ -1,\ 11)$ meets the plane $3x + 4y + 5z = 10$, find B and the distance AB.

    [Ans: $B = \left(\frac{123}{50},\ \frac{-286}{50},\ \frac{255}{50}\right)$ and $AB = \frac{59}{\sqrt{50}}$.]

14. Prove that the triangle with vertices $(-3, 0, 2)$, $(6, 1, 4)$, $(-5, 1, 0)$ has area $\frac{1}{2}\sqrt{333}$.

15. Find an equation for the plane through $(2, 1, 4)$, $(1, -1, 2)$, $(4, -1, 1)$.
[Ans: $2x - 7y + 6z = 21$.]

16. Lines $\mathcal{L}$ and $\mathcal{M}$ are non–parallel in 3–dimensional space and are given by equations
$$\mathbf{P} = \mathbf{A} + sX, \quad \mathbf{Q} = \mathbf{B} + tY.$$
   (i) Prove that there is precisely one pair of points P and Q such that $\overrightarrow{PQ}$ is perpendicular to X and Y.
   (ii) Explain why PQ is the shortest distance between lines $\mathcal{L}$ and $\mathcal{M}$. Also prove that
$$PQ = \frac{|\,(X \times Y)\cdot \overrightarrow{AB}|}{\|X \times Y\|}.$$

17. If $\mathcal{L}$ is the line through $A = (1, 2, 1)$ and $C = (3, -1, 2)$, while $\mathcal{M}$ is the line through $B = (1, 0, 2)$ and $D = (2, 1, 3)$, prove that the shortest distance between $\mathcal{L}$ and $\mathcal{M}$ equals $\frac{13}{\sqrt{62}}$.

18. Prove that the volume of the tetrahedron formed by four non–coplanar points $A_i = (x_i, y_i, z_i)$, $1 \le i \le 4$, is equal to
$$\frac{1}{6}\,|\,(\overrightarrow{A_1A_2} \times \overrightarrow{A_1A_3})\cdot \overrightarrow{A_1A_4}|,$$
which in turn equals the absolute value of the determinant
$$\frac{1}{6}\begin{vmatrix} 1 & x_1 & y_1 & z_1 \\ 1 & x_2 & y_2 & z_2 \\ 1 & x_3 & y_3 & z_3 \\ 1 & x_4 & y_4 & z_4 \end{vmatrix}.$$

19. The points $A = (1, 1, 5)$, $B = (2, 2, 1)$, $C = (1, -2, 2)$ and $D = (-2, 1, 2)$ are the vertices of a tetrahedron. Find the equation of the line through A perpendicular to the face BCD and the distance of A from this face. Also find the shortest distance between the skew lines AD and BC.
[Ans: $\mathbf{P} = (1 + t)(\mathbf{i} + \mathbf{j} + 5\mathbf{k})$; $2\sqrt{3}$; 3.]

Chapter 9

FURTHER READING

Matrix theory has many applications to science, mathematics, economics and engineering. Some of these applications can be found in the books [2, 3, 4, 5, 11, 13, 16, 20, 26, 28].

For the numerical side of matrix theory, [6] is recommended. Its bibliography is also useful as a source of further references.

For applications to:

1. Graph theory, see [7, 13];
2. Coding theory, see [8, 15];
3. Game theory, see [13];
4. Statistics, see [9];
5. Economics, see [10];
6. Biological systems, see [12];
7. Markov non–negative matrices, see [11, 13, 14, 17];
8. The general equation of the second degree in three variables, see [18];
9. Affine and projective geometry, see [19, 21, 22];
10. Computer graphics, see [23, 24].

Bibliography

[1] B. Noble. *Applied Linear Algebra, 1969.* Prentice Hall, NJ.

[2] B. Noble and J.W. Daniel. *Applied Linear Algebra, third edition, 1988.* Prentice Hall, NJ.

[3] R.P. Yantis and R.J. Painter. *Elementary Matrix Algebra with Application, second edition, 1977.* Prindle, Weber and Schmidt, Inc. Boston, Massachusetts.

[4] T.J. Fletcher. *Linear Algebra through its Applications, 1972.* Van Nostrand Reinhold Company, New York.

[5] A.R. Magid. *Applied Matrix Models, 1984.* John Wiley and Sons, New York.

[6] D.R. Hill and C.B. Moler. *Experiments in Computational Matrix Algebra, 1988.* Random House, New York.

[7] N. Deo. *Graph Theory with Applications to Engineering and Computer Science, 1974.* Prentice–Hall, N. J.

[8] V. Pless. *Introduction to the Theory of Error–Correcting Codes, 1982.* John Wiley and Sons, New York.

[9] F.A. Graybill. *Matrices with Applications in Statistics, 1983.* Wadsworth, Belmont Ca.

[10] A.C. Chiang. *Fundamental Methods of Mathematical Economics, second edition, 1974.* McGraw–Hill Book Company, New York.

[11] N.J. Pullman. *Matrix Theory and its Applications, 1976.* Marcel Dekker Inc. New York.

[12] J.M. Geramita and N.J. Pullman. *An Introduction to the Application of Nonnegative Matrices to Biological Systems, 1984.* Queen's Papers in Pure and Applied Mathematics 68. Queen's University, Kingston, Canada.

[13] M. Pearl. *Matrix Theory and Finite Mathematics, 1973.* McGraw–Hill Book Company, New York.

[14] J.G. Kemeny and J.L. Snell. *Finite Markov Chains, 1967.* Van Nostrand Reinhold, N.J.

[15] E.R. Berlekamp. *Algebraic Coding Theory, 1968.* McGraw–Hill Book Company, New York.

[16] G. Strang. *Linear Algebra and its Applications, 1988.* Harcourt Brace Jovanovich, San Diego.

[17] H. Minc. *Nonnegative Matrices, 1988.* John Wiley and Sons, New York.

[18] G.C. Preston and A.R. Lovaglia. *Modern Analytic Geometry, 1971.* Harper and Row, New York.

[19] J.A. Murtha and E.R. Willard. *Linear Algebra and Geometry, 1969.* Holt, Rinehart and Winston, Inc. New York.

[20] L.A. Pipes. *Matrix Methods for Engineering, 1963.* Prentice–Hall, Inc. N. J.

[21] D. Gans. *Transformations and Geometries, 1969.* Appleton–Century–Crofts, New York.

[22] J.N. Kapur. *Transformation Geometry, 1976.* Affiliated East–West Press, New Delhi.

[23] G.C. Reid. *Postscript Language Tutorial and Cookbook, 1988.* Addison–Wesley Publishing Company, New York.

[24] D. Hearn and M.P. Baker. *Computer Graphics, 1989.* Prentice–Hall, Inc. N. J.

[25] C.G. Cullen. *Linear Algebra with Applications, 1988.* Scott, Foresman and Company, Glenview, Illinois.

[26] R.E. Larson and B.H. Edwards. *Elementary Linear Algebra, 1988.* D.C. Heath and Company, Lexington, Massachusetts Toronto.

[27] N. Magnenat–Thalman and D. Thalmann. *State–of–the–art–in Computer Animation, 1989.* Springer–Verlag Tokyo.

[28] W.K. Nicholson. *Elementary Linear Algebra, 1990.* PWS–Kent, Boston.

Index