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:

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

This system has the 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 $\dfrac{a}{b} = ab^{-1}$ 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

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

For example, with $p = 7$, we have $3 \oplus 4 = 7\,(\operatorname{mod} 7) = 0$ and $3 \otimes 5 = 15\,(\operatorname{mod} 7) = 1$. 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) There are three types of *elementary row operations* that 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:

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

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

   $R_j \to R_j + tR_i$ 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,

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

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

Thus A is row–equivalent to B. Clearly B is also row–equivalent to A, by performing the inverse row–operations $R_1 \to \frac{1}{2}R_1$, $R_2 \leftrightarrow R_3$, $R_2 \to R_2 - 2R_3$ 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 $2\text{nd}, \ldots, m\text{th}$ 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 $1 \leq i \leq r$. Then

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

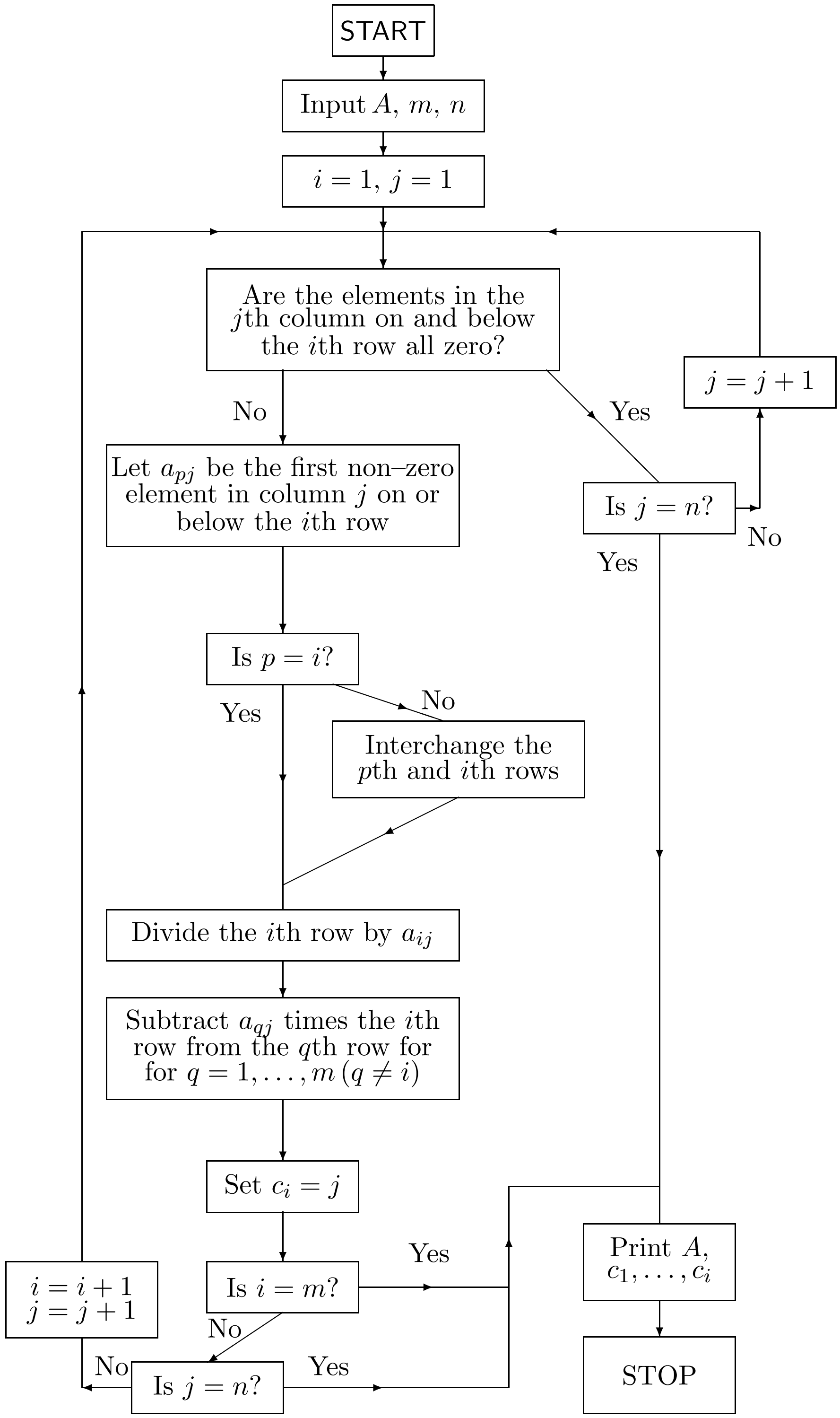

Figure 1.1: Gauss–Jordan algorithm.

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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 $\mathbb{Z}_3$:

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

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?

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

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

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

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)
$$\begin{aligned} x+y+z &= 2 \\ 2x+3y-z &= 8 \\ x-y-z &= -8 \end{aligned}$$

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

(c)
$$\begin{aligned} 3x-y+7z &= 0 \\ 2x-y+4z &= \tfrac{1}{2} \\ x-y+z &= 1 \\ 6x-4y+10z &= 3 \end{aligned}$$

(d)
$$\begin{aligned} 2x_2+3x_3-4x_4 &= 1 \\ 2x_3+3x_4 &= 4 \\ 2x_1+2x_2-5x_3+2x_4 &= 4 \\ 2x_1-6x_3+9x_4 &= 7 \end{aligned}$$

[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 and B are said to be equal if A and B have the same size and corresponding elements are equal; that is 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. $t(AB) = (tA)B = A(tB), \quad A(-B) = (-A)B = -(AB);$

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

The system of m linear equations in n unknowns

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

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) With $A \in M_{m\times n}(F)$, we associate 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, 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

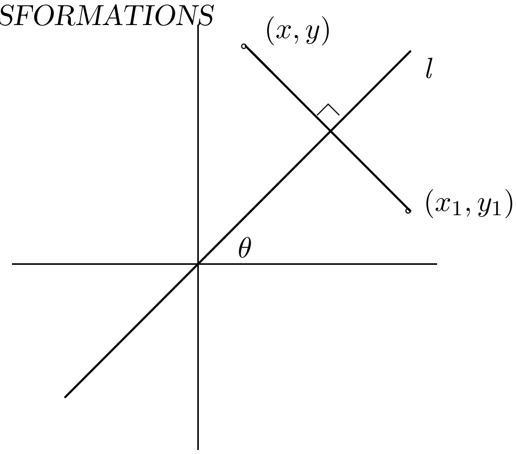

Figure 2.1: Reflection in a line.

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

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

The matrix ABC is quite complicated:

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

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

EXAMPLE 2.2.2 Another example of a linear transformation arising 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.

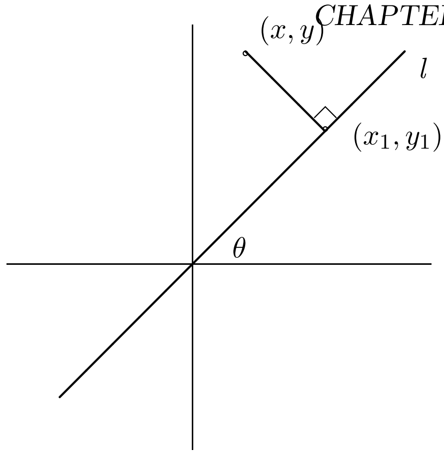

Figure 2.2: Projection on a line.

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

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

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

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

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 $I_m A = A = AI_n$.

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. $A^m A^n = A^{m+n}, \;\; (A^m)^n = A^{mn}$;
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 (PRINCIPLE OF MATHEMATICAL INDUCTION) *If for each* $n \geq 1$, $\mathcal{P}_n$ *denotes a mathematical statement and*

(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

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

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 the recurrence relation $x_{n+1} = 2ax_n - a^2x_{n-1}$ for $n \geq 1$. Use part (a) and the previous question to prove that $x_n = na^{n-1}x_1 + (1-n)a^nx_0$ 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

$$k_{n+1} = (\lambda_1 + \lambda_2)k_n - \lambda_1\lambda_2k_{n-1},$$

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 $n \geq 1$,

$$A^n = k_nA - \lambda_1\lambda_2k_{n-1}I_2,$$

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

8. Use Question 6 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
$$\begin{aligned} x_{n+1} &= x_n + r y_n \\ y_{n+1} &= x_n + y_n, \end{aligned}$$
for $n \geq 0$, where $x_0 = a$ and $y_0 = b$.

Use Question 6 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 square 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

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

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 scalars $x_1, \ldots, x_n$, *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) There are three types, E_{ij}, $E_i(t)$, $E_{ij}(t)$, corresponding to the three kinds of elementary row operation:

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 The three types of 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

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

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

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

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

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 $E_r\left(\ldots(E_1A)\ldots\right) = B = I_n$ 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 $E_r \ldots E_1 A = I_n$. 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 $A^{-1} = \left(B^{-1}\right)^{-1} = B = E_r\left((\ldots(E_1I_n)\ldots\right)$, 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

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

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

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

$$\begin{aligned} &= \sum_{j=1}^{n} \left(B^t\right)_{kj} \left(A^t\right)_{ji} \\ &= \left(B^t A^t\right)_{ki}. \end{aligned}$$

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 real matrix A is called *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 real 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 $X^t X = X^t BX$.
Taking transposes of both sides gives

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

Hence $X^t X = -X^t X$ and $X^t X = 0$. But if $X = [x_1, \ldots, x_n]^t$, then $X^t X = x_1^2 + \ldots + x_n^2 = 0$ 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 $\mathrm{diag}\,(a_1, \ldots, a_n)$ denote the diagonal matrix whose *diagonal* elements d_{ii} are $a_1, \ldots, a_n$, respectively. Show that

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

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

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

   Also prove that $\mathrm{diag}\,(a_1, \ldots, a_n)$ 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.

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

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) $(E_{12}(7)E_{31}(1))^{-1}$.

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

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

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.

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

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

$$\text{(a)} \left[\begin{array}{cccc} 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{array}\right] \quad \text{(b)} \left[\begin{array}{cccc} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{array}\right].$$

[Answer: (a) $\left[\begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 \end{array}\right]$.]

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

$$\text{(a)} \left[\begin{array}{rrr} 1 & 1 & 1 \\ -1 & 1 & 0 \\ 2 & 0 & 0 \end{array}\right] \quad \text{(b)} \left[\begin{array}{rrr} 2 & 2 & 4 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{array}\right] \quad \text{(c)} \left[\begin{array}{rrr} 4 & 6 & -3 \\ 0 & 0 & 7 \\ 0 & 0 & 5 \end{array}\right]$$

$$\text{(d)} \left[\begin{array}{rrr} 2 & 0 & 0 \\ 0 & -5 & 0 \\ 0 & 0 & 7 \end{array}\right] \text{(e)} \left[\begin{array}{rrrr} 1 & 2 & 4 & 6 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 2 \end{array}\right] \text{(f)} \left[\begin{array}{rrr} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 5 & 7 & 9 \end{array}\right].$$

[Answers: (a) $\left[\begin{array}{rrr} 0 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 1 & -1 & -1 \end{array}\right]$ (b) $\left[\begin{array}{rrr} -\frac{1}{2} & 2 & 1 \\ 0 & 0 & 1 \\ \frac{1}{2} & -1 & -1 \end{array}\right]$ (d) $\left[\begin{array}{rrr} \frac{1}{2} & 0 & 0 \\ 0 & -\frac{1}{5} & 0 \\ 0 & 0 & \frac{1}{7} \end{array}\right]$

(e) $\left[\begin{array}{rrrr} 1 & -2 & 0 & -3 \\ 0 & 1 & -2 & 2 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & \frac{1}{2} \end{array}\right]$.]

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 $A = \left[\begin{array}{cc} a & b \\ c & d \end{array}\right]$ 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 $A = \begin{bmatrix} \frac{2}{3} & \frac{1}{4} \\ \frac{1}{3} & \frac{3}{4} \end{bmatrix}$, $P = \begin{bmatrix} 1 & 3 \\ -1 & 4 \end{bmatrix}$. Verify that $P^{-1}AP = \begin{bmatrix} \frac{5}{12} & 0 \\ 0 & 1 \end{bmatrix}$ 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) $A^n \to \dfrac{1}{b+c}\begin{bmatrix} b & b \\ c & c \end{bmatrix}$ as $n \to \infty$, 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^tAB 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 all 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 Subspaces $\langle X_1, \ldots, X_r,\ Z_1, \ldots, Z_t\rangle$ 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

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

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

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

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

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

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 $\dim S \leq n$.

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 $1 \leq i \leq n$.

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, contradictiong the assumption that $X_1 \ldots, X_m$ are linearly independent. Hence $x_{m+1} \neq 0$ 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 $A = n$.

<u>Finding a basis for $R(A)$</u>: 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

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

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

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

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

<u>Finding a basis for $C(A)$</u>: The r columns $A_{*c_1}, \dots, A_{*c_r}$ form a basis for $C(A)$ and hence column rank $A = r$. For it is clear that columns $c_1, \dots, c_r$ 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 $c_1, \dots, c_r$ 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

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

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

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -2x_3 + x_5 \\ -3x_3 - 2x_5 \\ x_3 \\ -3x_5 \\ x_5 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ -3 \\ 1 \\ 0 \\ 0 \end{bmatrix} + x_5 \begin{bmatrix} 1 \\ -2 \\ 0 \\ -3 \\ 1 \end{bmatrix} = x_3X_1 + x_5X_2,$$

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

Here rank $A = 3$ and 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 $\operatorname{rank} A = 1$ and $\operatorname{nullity} A = 1$.

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 $\operatorname{rank} A = 2$ and $\operatorname{nullity} A = 0$.

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 $A = \begin{bmatrix} a & b & c \\ 1 & 1 & 1 \end{bmatrix}$. Find conditions on $a,\ b,\ c$ such that (a) $\text{rank}\, A = 1$; (b) $\text{rank}\, A = 2$.

    [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
$$a_1x_1 + \cdots + a_nx_n = 0$$
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
$$\dim R \le \dim S$$
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.