1
The Algebra of Matrices

If m and n are positive integers then by a **matrix of size m by n**, or an $m \times n$ **matrix**, we shall mean a rectangular array consisting of mn numbers in a boxed display consisting of m rows and n columns. Simple examples of such objects are the following:

$$\text{size } 1 \times 5 : \begin{bmatrix} 10 & 9 & 8 & 7 & 6 \end{bmatrix} \qquad \text{size } 3 \times 2 : \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

$$\text{size } 4 \times 4 : \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 7 \end{bmatrix} \qquad \text{size } 3 \times 1 : \begin{bmatrix} 0 \\ 2 \\ 4 \end{bmatrix}$$

In general we shall display an $m \times n$ matrix as

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & x_{m3} & \dots & x_{mn} \end{bmatrix}.$$

- Note that the *first* suffix gives the number of the **row** and the *second* suffix that of the **column**, so that x_{ij} appears at the intersection of the i-th row and the j-th column.

We shall often find it convenient to abbreviate the above display to simply

$$[x_{ij}]_{m \times n}$$

and refer to x_{ij} as the **(i,j)-th element** or the **(i,j)-th entry** of the matrix.

- Thus the expression $X = [x_{ij}]_{m \times n}$ will be taken to mean that 'X is the $m \times n$ matrix whose (i,j)-th element is x_{ij}'.

Example 1.1

The 3×3 matrix $X = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 2^2 & 2^3 \\ 3 & 3^2 & 3^3 \end{bmatrix}$ can be expressed as $X = [x_{ij}]_{3\times 3}$ where $x_{ij} = i^j$.

Example 1.2

The 3×3 matrix $X = \begin{bmatrix} a & a & a \\ 0 & a & a \\ 0 & 0 & a \end{bmatrix}$ can be expressed as $X = [x_{ij}]_{3\times 3}$ where

$$x_{ij} = \begin{cases} a & \text{if } i \leqslant j; \\ 0 & \text{otherwise.} \end{cases}$$

Example 1.3

The $n \times n$ matrix

$$X = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ e & 1 & 0 & \dots & 0 \\ e^2 & e & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e^{n-1} & e^{n-2} & e^{n-3} & \dots & 1 \end{bmatrix}$$

can be expressed as $X = [x_{ij}]_{n\times n}$ where

$$x_{ij} = \begin{cases} e^{i-j} & \text{if } i \geqslant j; \\ 0 & \text{otherwise.} \end{cases}$$

EXERCISES

1.1 Write out the 3×3 matrix whose entries are given by $x_{ij} = i + j$.

1.2 Write out the 3×3 matrix whose entries are given by

$$x_{ij} = \begin{cases} 1 & \text{if } i + j \text{ is even;} \\ 0 & \text{otherwise.} \end{cases}$$

1.3 Write out the 3×3 matrix whose entries are given by $x_{ij} = (-1)^{i-j}$.

1.4 Write out the $n \times n$ matrix whose entries are given by

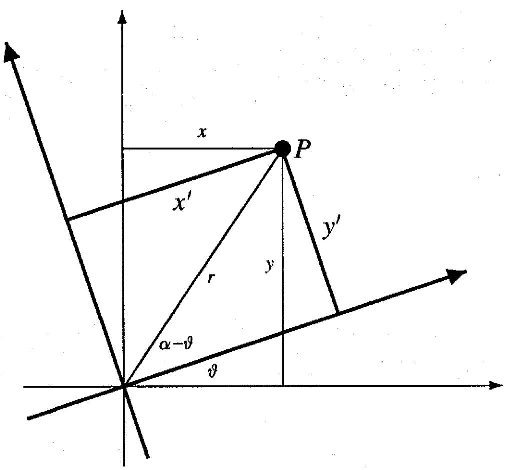
$$x_{ij} = \begin{cases} -1 & \text{if } i > j; \\ 0 & \text{if } i = j; \\ 1 & \text{if } i < j. \end{cases}$$

2
Some Applications of Matrices

We shall now give brief descriptions of some situations to which matrix theory finds a natural application, and some problems to which the solutions are determined by the algebra that we have developed. Some of these applications will be dealt with in greater detail in later chapters.

1. Analytic geometry

In analytic geometry, various transformations of the coordinate axes may be described using matrices. By way of example, suppose that in the two-dimensional cartesian plane we rotate the coordinate axes in an anti-clockwise direction through an angle ϑ, as illustrated in the following diagram:



Let us compute the new coordinates (x', y') of the point P whose old coordinates were (x, y).

From the diagram we have $x = r\cos\alpha$ and $y = r\sin\alpha$ so

$$\begin{aligned} x' = r\cos(\alpha - \vartheta) &= r\cos\alpha\cos\vartheta + r\sin\alpha\sin\vartheta \\ &= x\cos\vartheta + y\sin\vartheta; \\ y' = r\sin(\alpha - \vartheta) &= r\sin\alpha\cos\vartheta - r\cos\alpha\sin\vartheta \\ &= y\cos\vartheta - x\sin\vartheta. \end{aligned}$$

These equations give x', y' in terms of x, y and ϑ. They can be expressed in the matrix form

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}.$$

The 2×2 matrix

$$R_\vartheta = \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix}$$

is called the **rotation matrix** associated with ϑ. It has the following property:

$$R_\vartheta R'_\vartheta = I_2 = R'_\vartheta R_\vartheta.$$

In fact, we have

$$\begin{aligned} R_\vartheta R'_\vartheta &= \begin{bmatrix} \cos\vartheta & \sin\vartheta \\ -\sin\vartheta & \cos\vartheta \end{bmatrix} \begin{bmatrix} \cos\vartheta & -\sin\vartheta \\ \sin\vartheta & \cos\vartheta \end{bmatrix} \\ &= \begin{bmatrix} \cos^2\vartheta + \sin^2\vartheta & -\cos\vartheta\sin\vartheta + \sin\vartheta\cos\vartheta \\ -\sin\vartheta\cos\vartheta + \cos\vartheta\sin\vartheta & \sin^2\vartheta + \cos^2\vartheta \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned}$$

and similarly, as the reader can verify, $R'_\vartheta R_\vartheta = I_2$.

This leads us more generally to the following notion.

Definition

An $n \times n$ matrix A is said to be **orthogonal** if

$$AA' = I_n = A'A.$$

It follows from the above that to every rotation of axes in two dimensions we can associate a real orthogonal matrix ('real' in the sense that its elements are real numbers).

EXERCISES

2.1 If A is an orthogonal $n \times n$ matrix prove that A' is also orthogonal.

2.2 If A and B are orthogonal $n \times n$ matrices prove that AB is also orthogonal.

3
Systems of Linear Equations

We shall now consider in some detail a systematic method of solving systems of linear equations. In working with such systems, there are three basic operations involved:

(1) interchanging two equations (usually for convenience);
(2) multiplying an equation by a non-zero scalar;
(3) forming a new equation by adding one equation to another.

The operation of adding a multiple of one equation to another can be achieved by a combination of (2) and (3).

We begin by considering the following three examples.

Example 3.1

To solve the system

$$\begin{aligned} y + 2z &= 1 \quad (1)\\ x - 2y + z &= 0 \quad (2)\\ 3y - 4z &= 23 \quad (3) \end{aligned}$$

we multiply equation (1) by 3 and subtract the new equation from equation (3) to obtain $-10z = 20$, whence we see that $z = -2$. It then follows from equation (1) that $y = 5$, and then by equation (2) that $x = 2y - z = 12$.

Example 3.2

Consider the system

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

If we add together equations (1) and (2), we obtain equation (3), which is therefore superfluous. Thus we have only two equations in three unknowns. What do we mean by a solution in this case?

Example 3.3

Consider the system

$$\begin{aligned} x + y + z + t &= 1 \quad (1)\\ x - y - z + t &= 3 \quad (2)\\ -x - y + z - t &= 1 \quad (3)\\ -3x + y - 3z - 3t &= 4 \quad (4) \end{aligned}$$

Adding equations (1) and (2), we obtain $x + t = 2$, whence it follows that $y + z = -1$. Adding equations (1) and (3), we obtain $z = 1$ and consequently $y = -2$. Substituting in equation (4), we obtain $-3x - 3t = 9$ so that $x + t = -3$, which is not consistent with $x + t = 2$.

This system therefore does not have a solution. Expressed in another way, given the 4×4 matrix

$$A = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & -1 & 1 \\ -1 & -1 & 1 & -1 \\ -3 & 1 & -3 & -3 \end{bmatrix},$$

there are no numbers x, y, z, t that satisfy the matrix equation

$$A \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \\ 1 \\ 4 \end{bmatrix}.$$

The above three examples were chosen to provoke the question: is there a *systematic* method of tackling systems of linear equations that

(a) avoids a haphazard manipulation of the equations;
(b) yields all the solutions when they exist;
(c) makes it clear when no solution exists?

In what follows our objective will be to obtain a complete answer to this question.

We note first that in dealing with systems of linear equations the 'unknowns' play a secondary rôle. It is in fact the coefficients (which are usually integers) that are important. Indeed, each such system is completely determined by its augmented matrix. In order to work solely with this, we consider the following **elementary row operations** on this matrix:

(1) interchange two rows;
(2) multiply a row by a non-zero scalar;
(3) add one row to another.

These elementary row operations clearly correspond to the basic operations on equations listed above.

4
Invertible Matrices

In Theorem 1.3 we showed that every $m \times n$ matrix A has an additive inverse, denoted by $-A$, which is the unique $m \times n$ matrix X that satisfies the equation $A + X = 0$. We shall now consider the multiplicative analogue of this.

Definition

Let A be an $m \times n$ matrix. Then an $n \times m$ matrix X is said to be a **left inverse** of A if it satisfies the equation $XA = I_n$; and a **right inverse** of A if it satisfies the equation $AX = I_m$.

Example 4.1

Consider the matrices

$$A = \begin{bmatrix} 1 & 1 \\ 4 & 3 \\ 3 & 4 \\ 0 & 0 \end{bmatrix}, \qquad X_{a,b} = \begin{bmatrix} -3 & 1 & 0 & a \\ -3 & 0 & 1 & b \end{bmatrix}.$$

A simple computation shows that $X_{a,b}A = I_2$, and so *A has infinitely many left inverses*. In contrast, *A has no right inverse*. To see this, it suffices to observe that if Y were a 2×4 matrix such that $AY = I_4$ then we would require $[AY]_{4,4} = 1$ which is not possible since all the entries in the fourth row of A are 0.

Example 4.2

The matrix

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

has a common unique left inverse and unique right inverse, namely

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

Theorem 4.1

Let A be an $m \times n$ matrix. Then

(1) *A has a right inverse if and only if* $\text{rank } A = m$;

(2) *A has a left inverse if and only if* $\text{rank } A = n$.

Proof

(1) Suppose that the $n \times m$ matrix X is a right inverse of A, so that we have $AX = I_m$. If $\mathbf{x}_i$ denotes the i-th column of X then this equation can be expanded to the m equations

$$(i = 1, \dots, m) \qquad A\mathbf{x}_i = \Delta_i$$

where Δ_i denotes the i-th column of I_m.

Now each of the matrix equations $A\mathbf{x}_i = \Delta_i$ represents a consistent system of m equations in n unknowns and so, by Theorem 3.17, for each i we have

$$\text{rank } A = \text{rank } A|\Delta_i.$$

Since $\Delta_1, \dots, \Delta_m$ are linearly independent, it follows by considering column ranks that

$$\begin{aligned} \text{rank } A &= \text{rank } A|\Delta_1 \\ &= \text{rank } A|\Delta_1|\Delta_2 \\ &= \cdots \\ &= \text{rank } A|\Delta_1|\Delta_2|\dots|\Delta_n = \text{rank } A|I_m = m. \end{aligned}$$

Conversely, suppose that the rank of A is m. Then necessarily we have that $n \geqslant m$. Consider the Hermite form of A'. Since $H(A')$ is an $n \times m$ matrix and

$$\text{rank } H(A') = \text{rank } A' = \text{rank } A = m,$$

we see that $H(A')$ is of the form

$$H(A') = \begin{bmatrix} I_m \\ 0_{n-m,m} \end{bmatrix}.$$

As this is row-equivalent to A', there exists an $n \times n$ matrix Y such that

$$YA' = \begin{bmatrix} I_m \\ 0_{n-m,m} \end{bmatrix}.$$

Taking transposes, we obtain

$$AY' = \begin{bmatrix} I_m & 0_{m,n-m} \end{bmatrix}.$$

Now let Z be the $n \times m$ matrix consisting of the first m columns of Y'. Then from the form of the immediately preceding equation we see that $AZ = I_m$, whence Z is a right inverse of A.

(2) It is an immediate consequence of Theorem 1.9 that A has a left inverse if and only if its transpose has a right inverse. The result therefore follows by applying (1) to the transpose of A. □

5
Vector Spaces

In order to proceed further with matrices we have to take a wider view of matters. This we do through the following important notion.

Definition

By a **vector space** we shall mean a set V on which there are defined two operations, one called 'addition' and the other called 'multiplication by scalars', such that the following properties hold:

(V_1) $x + y = y + x$ for all $x, y \in V$;
(V_2) $(x + y) + z = x + (y + z)$ for all $x, y, z \in V$;
(V_3) there exists an element $0 \in V$ such that $x + 0 = x$ for every $x \in V$;
(V_4) for every $x \in V$ there exists $-x \in V$ such that $x + (-x) = 0$;
(V_5) $\lambda(x + y) = \lambda x + \lambda y$ for all $x, y \in V$ and all scalars λ;
(V_6) $(\lambda + \mu)x = \lambda x + \mu x$ for all $x \in V$ and all scalars λ, μ;
(V_7) $(\lambda\mu)x = \lambda(\mu x)$ for all $x \in V$ and all scalars λ, μ;
(V_8) $1x = x$ for all $x \in V$.

When the scalars are all real numbers we shall often talk of a **real** vector space; and when the scalars are all complex numbers we shall talk of a **complex** vector space.

- It should be noted that in the definition of a vector space the scalars need not be restricted to be real or complex numbers. They can in fact belong to any 'field' F (which may be regarded informally as a number system in which every non-zero element has a multiplicative inverse). Although in what follows we shall find it convenient to say that 'V is a vector space over a field F' to indicate that the scalars come from a field F, we shall in fact normally assume (i.e. unless explicitly mentioned otherwise) that F is either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers.
- Axioms (V_1) to (V_4) above can be summarised by saying that the algebraic structure $(V; +)$ is an **abelian group**. If we denote by F the field of scalars

(usually IR or $\mathbb{C}$) then multiplication by scalars can be considered as an **action** by F on V, described by $(\lambda, x) \mapsto \lambda x$, which relates the operations in F (addition and multiplication) to that of V (addition) in the way described by the axioms (V_5) to (V_8).

Example 5.1

Let $\mathrm{Mat}_{m\times n}\,\mathrm{IR}$ be the set of all $m \times n$ matrices with real entries. Then Theorems 1.1 to 1.4 collectively show that $\mathrm{Mat}_{m\times n}\,\mathrm{IR}$ is a real vector space under the usual operations of addition of matrices and multiplication by scalars.

Example 5.2

The set IR^n of n-tuples $(x_1, \dots, x_n)$ of real numbers is a real vector space under the following component-wise definitions of addition and mutiplication by scalars:

$$\begin{aligned}(x_1, \dots, x_n) + (y_1, \dots, y_n) &= (x_1 + y_1, \dots, x_n + y_n),\\ \lambda(x_1, \dots, x_n) &= (\lambda x_1, \dots, \lambda x_n).\end{aligned}$$

Geometrically, IR^2 represents the cartesian plane, whereas IR^3 represents three-dimensional space.

Similarly, the set $\mathbb{C}^n$ of n-tuples of complex numbers can be made into both a real vector space (with the scalars real numbers) or a complex vector space (with the scalars complex numbers).

Example 5.3

Let $\mathrm{Map}(\mathrm{IR}, \mathrm{IR})$ be the set of all mappings $f : \mathrm{IR} \to \mathrm{IR}$. For two such mappings f, g define $f + g : \mathrm{IR} \to \mathrm{IR}$ to be the mapping given by the prescription

$$(f + g)(x) = f(x) + g(x),$$

and for every scalar $\lambda \in \mathrm{IR}$ define $\lambda f : \mathrm{IR} \to \mathrm{IR}$ to be the mapping given by the prescription

$$(\lambda f)(x) = \lambda f(x).$$

Then it is readily verified that (V_1) to (V_8) are satisfied, with the rôle of the vector 0 taken by the zero mapping (i.e. the mapping ϑ such that $\vartheta(x) = 0$ for every $x \in \mathrm{IR}$) and $-f$ the mapping given by $(-f)(x) = -f(x)$ for every $x \in \mathrm{IR}$. These operations therefore make $\mathrm{Map}(\mathrm{IR}, \mathrm{IR})$ into a real vector space.

Example 5.4

Let $\mathrm{IR}_n[X]$ be the set of polynomials of degree at most n with real coefficients. The reader will recognise this as the set of objects of the form

$$a_0 + a_1X + a_2X^2 + \cdots + a_nX^n$$

6
Linear Mappings

In the study of any algebraic structure there are two concepts that are of paramount importance. The first is that of a **substructure** (i.e. a subset with the same type of structure), and the second is that of a **morphism** (i.e. a mapping from one structure to another of the same kind that is 'structure-preserving').

So far, we have encountered the notion of a substructure for a vector space; this is called a *subspace*. In this chapter we shall consider the notion of a morphism between vector spaces, i.e. a mapping from one vector space to another that is 'structure-preserving' in the following sense.

Definition

If V and W are vector spaces over the same field F then by a **linear mapping** (or **linear transformation**) from V to W we shall mean a mapping $f : V \to W$ such that

(1) $(\forall x, y \in V) \qquad f(x+y) = f(x) + f(y);$

(2) $(\forall x \in V)(\forall \lambda \in F) \qquad f(\lambda x) = \lambda f(x).$

- If $f : V \to W$ is linear then V is sometimes called the **departure space** and W the **arrival space** of f.

Example 6.1

The mapping $f : \mathrm{IR}^2 \to \mathrm{IR}^3$ given by

$$f(a,b) = (a+b, a-b, b)$$

is linear. In fact, for all (a,b) and (a',b') in IR^2 we have

$$\begin{aligned} f\big((a,b)+(a',b')\big) &= f(a+a', b+b') \\ &= (a+a'+b+b', a+a'-b-b', b+b') \\ &= (a+b, a-b, b) + (a'+b', a'-b', b') \\ &= f(a,b) + f(a',b') \end{aligned}$$

and, for all $(a,b) \in \mathrm{IR}^2$ and all $\lambda \in \mathrm{IR}$,

$$\begin{aligned} f(\lambda(a,b)) &= f(\lambda a, \lambda b) \\ &= (\lambda a + \lambda b, \lambda a - \lambda b, \lambda b) \\ &= \lambda(a+b, a-b, b) \\ &= \lambda f(a,b). \end{aligned}$$

Example 6.2

The mapping $\mathrm{pr}_i : \mathrm{IR}^n \longrightarrow \mathrm{IR}$ described by

$$\mathrm{pr}_i(x_1, \ldots, x_n) = x_i$$

(i.e. the mapping that picks out the i-th coordinate) is called the **i-th projection** of IR^n onto IR. It is readily seen that (1) and (2) above are satisfied, so that pr_i is linear.

Example 6.3

Consider the differentiation map $\mathrm{D} : \mathrm{IR}_n[X] \longrightarrow \mathrm{IR}_n[X]$ given by

$$\mathrm{D}(a_0 + a_1X + \cdots + a_nX^n) = a_1 + 2a_2X + \cdots + na_nX^{n-1}.$$

This mapping is linear; for if $p(X)$ and $q(X)$ are polynomials then we know from analysis that $\mathrm{D}(p(X)+q(X)) = \mathrm{D}p(X) + \mathrm{D}q(X)$ and that, for every scalar λ, $\mathrm{D}(\lambda p(X)) = \lambda\, \mathrm{D}p(X)$.

EXERCISES

6.1 Decide which of the following mappings $f : \mathrm{IR}^3 \longrightarrow \mathrm{IR}^3$ are linear:
(1) $f(x,y,z) = (y,z,0)$;
(2) $f(x,y,z) = (z,-y,x)$;
(3) $f(x,y,z) = (|x|,-z,0)$;
(4) $f(x,y,z) = (x-1,x,y)$;
(5) $f(x,y,z) = (x+y,z,0)$;
(6) $f(x,y,z) = (2x,y-2,4y)$.

6.2 Let $B \in \mathrm{Mat}_{n\times n}\,\mathrm{IR}$ be fixed and non-zero. Which of the following mappings $T_B : \mathrm{Mat}_{n\times n}\,\mathrm{IR} \longrightarrow \mathrm{Mat}_{n\times n}\,\mathrm{IR}$ are linear?
(1) $T_B(X) = XB - BX$;
(2) $T_B(X) = XB^2 + BX$;
(3) $T_B(X) = XB^2 - BX^2$.

6.3 Which of the following mappings are linear?
(1) $f : \mathrm{IR}_n[X] \longrightarrow \mathrm{IR}_3[X]$ given by $f(p(X)) = p(0)X^2 + \mathrm{D}p(0)X^3$;
(2) $f : \mathrm{IR}_n[X] \longrightarrow \mathrm{IR}_{n+1}[X]$ given by $f(p(X)) = p(0) + Xp(X)$;
(3) $f : \mathrm{IR}_n[X] \longrightarrow \mathrm{IR}_{n+1}[X]$ given by $f(p(X)) = 1 + Xp(X)$.

7
The Matrix Connection

We shall now proceed to show how a linear mapping from one finite-dimensional vector space to another can be represented by a matrix. For this purpose, we require the following notion.

Definition

Let V be a finite-dimensional vector space over a field F. By an **ordered basis** of V we shall mean a finite sequence $(v_i)_{1 \leqslant i \leqslant n}$ of elements of V such that $\{v_1, \dots, v_n\}$ is a basis of V.

Note that every basis of n elements gives rise to $n!$ distinct ordered bases, for there are $n!$ permutations on a set of n elements, and therefore $n!$ distinct ways of ordering the elements $v_1, \dots, v_n$.

In what follows we shall find it convenient to abbreviate $(v_i)_{1 \leqslant i \leqslant n}$ to simply $(v_i)_n$.

Suppose now that V and W are vector spaces of dimensions m and n respectively over a field F. Let $(v_i)_m$, $(w_i)_n$ be given ordered bases of V, W and let $f : V \to W$ be linear. We know from Corollary 1 of Theorem 6.7 that f is completely and uniquely determined by its action on the basis $(v_i)_m$. This action is described by expressing each $f(v_i)$ as a linear combination of elements from the basis $(w_i)_n$:

$$\begin{aligned} f(v_1) &= x_{11}w_1 + x_{12}w_2 + \cdots + x_{1n}w_n; \\ f(v_2) &= x_{21}w_1 + x_{22}w_2 + \cdots + x_{2n}w_n; \\ &\ \vdots \\ f(v_m) &= x_{m1}w_1 + x_{m2}w_2 + \cdots + x_{mn}w_n. \end{aligned}$$

The action of f on $(v_i)_m$ is therefore determined by the mn scalars x_{ij} appearing in the above equations. Put another way, *the action of f is completely determined by a knowledge of the $m \times n$ matrix $X = [x_{ij}]$.*

For technical reasons that will be explained later, the *transpose* of this matrix X

is called the **matrix of f relative to the fixed ordered bases** $(v_i)_m, (w_i)_n$. When it is clear what these fixed ordered bases are, we denote the matrix in question by Mat f.

- The reader should note carefully that it is an $\mathbf{n} \times \mathbf{m}$ matrix that represents a linear mapping from an **m**-dimensional vector space to an **n**-dimensional vector space.

Example 7.1

Consider the linear mapping $f : \mathrm{IR}^3 \to \mathrm{IR}^2$ given by

$$f(x,y,z) = (2x - 3y + z,\ 3x - 2y).$$

The action of f on the natural basis of IR^3 is described in terms of the natural basis of IR^2 as follows :

$$\begin{aligned} f(1,0,0) &= \quad (2,3) = \ \ 2(1,0) + 3(0,1) \\ f(0,1,0) &= (-3,-2) = -3(1,0) - 2(0,1) \\ f(0,0,1) &= \quad (1,0) = \ \ 1(1,0) + 0(0,1) \end{aligned}$$

and so we see that the matrix of f relative to the natural ordered bases of IR^3 and IR^2 is the transpose of the above coefficient matrix, namely the 2×3 matrix

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

- Note how the rows of this matrix relate to the definition of f.

Example 7.2

The vector space $\mathrm{IR}_n[X]$ is of dimension $n + 1$ and has the natural ordered basis

$$\{1, X, X^2, \dots, X^n\}.$$

The differentiation mapping $\mathrm{D} : \mathrm{IR}_n[X] \to \mathrm{IR}_n[X]$ is linear, and

$$\begin{aligned} \mathrm{D}1 &= 0 \cdot 1 + 0 \cdot X + \cdots + 0 \cdot X^{n-1} + 0 \cdot X^n \\ \mathrm{D}X &= 1 \cdot 1 + 0 \cdot X + \cdots + 0 \cdot X^{n-1} + 0 \cdot X^n \\ \mathrm{D}X^2 &= 0 \cdot 1 + 2 \cdot X + \cdots + 0 \cdot X^{n-1} + 0 \cdot X^n \\ &\ \ \vdots \\ \mathrm{D}X^n &= 0 \cdot 1 + 0 \cdot X + \cdots + n \cdot X^{n-1} + 0 \cdot X^n \end{aligned}$$

so the matrix of D relative to the natural ordered basis of $\mathrm{IR}_n[X]$ is the $(n+1) \times (n+1)$ matrix

$$\begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & n \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

8
Determinants

In what follows it will be convenient to write an $n \times n$ matrix A in the form

$$A = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n]$$

where, as before, $\mathbf{a}_i$ represents the i-th column of A. Also, the letter F will signify either the field $\mathbb{R}$ of real numbers or the field $\mathbb{C}$ of complex numbers.

Definition

A mapping $D : \mathrm{Mat}_{n\times n} F \longrightarrow F$ is **determinantal** if it is

(a) **multilinear** (or *a linear function of each column*) in the sense that

(D_1) $D[\ldots, \mathbf{b}_i + \mathbf{c}_i, \ldots] = D[\ldots, \mathbf{b}_i, \ldots] + D[\ldots, \mathbf{c}_i, \ldots]$;
(D_2) $D[\ldots, \lambda\mathbf{a}_i, \ldots] = \lambda D[\ldots, \mathbf{a}_i, \ldots]$;

(b) **alternating** in the sense that

(D_3) $D[\ldots, \mathbf{a}_i, \ldots, \mathbf{a}_j, \ldots] = -D[\ldots, \mathbf{a}_j, \ldots, \mathbf{a}_i, \ldots]$;

(c) **1-preserving** in the sense that

(D_4) $D(I_n) = 1_F$.

We first observe that, in the presence of property (D_1), property (D_3) can be expressed in another way.

Theorem 8.1

If D satisfies property (D_1) *then D satisfies property* (D_3) *if and only if it satisfies the property*

(D_3') $D(A) = 0$ *whenever A has two identical columns.*

Proof

$\Rightarrow$: Suppose that A has two identical columns, say $\mathbf{a}_i = \mathbf{a}_j$ with $i \neq j$. Then by (D_3) we have $D(A) = -D(A)$ whence $D(A) = 0$.

$\Leftarrow$: Suppose now that D satisfies (D_1) and (D_3'). Then we have

$$\begin{aligned}
0 &\overset{(D_3')}{=} D[\ldots,\mathbf{a}_i+\mathbf{a}_j,\ldots,\mathbf{a}_i+\mathbf{a}_j,\ldots] \\
&\overset{(D_1)}{=} D[\ldots,\mathbf{a}_i,\ldots,\mathbf{a}_i+\mathbf{a}_j,\ldots]+D[\ldots,\mathbf{a}_j,\ldots,\mathbf{a}_i+\mathbf{a}_j,\ldots] \\
&\overset{(D_1)}{=} D[\ldots,\mathbf{a}_i,\ldots,\mathbf{a}_i,\ldots]+D[\ldots,\mathbf{a}_i,\ldots,\mathbf{a}_j,\ldots] \\
&\qquad +D[\ldots,\mathbf{a}_j,\ldots,\mathbf{a}_i,\ldots]+D[\ldots,\mathbf{a}_j,\ldots,\mathbf{a}_j,\ldots] \\
&\overset{(D_3')}{=} D[\ldots,\mathbf{a}_i,\ldots,\mathbf{a}_j,\ldots]+D[\ldots,\mathbf{a}_j,\ldots,\mathbf{a}_i,\ldots]
\end{aligned}$$

whence (D_3) follows. □

Corollary

D is determinantal if and only if it satisfies $(D_1),(D_2),(D_3'),(D_4)$. □

Example 8.1

Let $D : \mathrm{Mat}_{2\times 2} F \to F$ be given by

$$D\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

Then it is an easy exercise to show that D satisfies the properties $(D_1),(D_2),(D_3'),(D_4)$ and so is determinantal.

In fact, as we shall now show, *this is the only determinantal mapping definable on* $\mathrm{Mat}_{2\times 2} F$.

For this purpose, let

$$\delta_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \delta_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

so that every $A \in \mathrm{Mat}_{2\times 2} F$ can be written in the form

$$A = [a_{11}\delta_1 + a_{21}\delta_2,\ a_{12}\delta_1 + a_{22}\delta_2].$$

Suppose that $f : \mathrm{Mat}_{2\times 2} F \to F$ is determinantal. Then, by (D_1) we have

$$f(A) = f[a_{11}\delta_1,\ a_{12}\delta_1 + a_{22}\delta_2] + f[a_{21}\delta_2,\ a_{12}\delta_1 + a_{22}\delta_2].$$

Applying (D_1) again, the first summand can be expanded to

$$f[a_{11}\delta_1,\ a_{12}\delta_1] + f[a_{11}\delta_1,\ a_{22}\delta_2]$$

which, by (D_2), is

$$a_{11}a_{12}f[\delta_1,\delta_1] + a_{11}a_{22}f[\delta_1,\delta_2].$$

By (D_3') and (D_4), this reduces to $a_{11}a_{22}$.

As for the second summand, by (D_1) this can be expanded to

$$f[a_{21}\delta_2,\ a_{12}\delta_1] + f[a_{21}\delta_2,\ a_{22}\delta_2]$$

9
Eigenvalues and Eigenvectors

Recall that an $n \times n$ matrix B is **similar** to an $n \times n$ matrix A if there is an invertible $n \times n$ matrix P such that $B = P^{-1}AP$. Our objective now is to determine under what conditions an $n \times n$ matrix is similar to a diagonal matrix. In so doing we shall draw together all of the notions that have been previously developed. Unless otherwise specified, A will denote an $n \times n$ matrix over $\mathbb{R}$ or $\mathbb{C}$.

Definition

By an **eigenvalue** (or **latent root**) of A we shall mean a scalar λ for which there exists a *non-zero* $n \times 1$ matrix $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$. Such a (column) matrix $\mathbf{x}$ is called an **eigenvector** (or **latent vector**) associated with λ.

- Note that eigenvectors are by definition *non-zero.*

Theorem 9.1

A scalar λ is an eigenvalue of A if and only if

$$\det(A - \lambda I_n) = 0.$$

Proof

Observe that $A\mathbf{x} = \lambda\mathbf{x}$ can be written in the form

$$(A - \lambda I_n)\mathbf{x} = \mathbf{0}.$$

Then λ is an eigenvalue of A if and only if the homogeneous system of equations

$$(A - \lambda I_n)\mathbf{x} = \mathbf{0}$$

has a *non-zero* solution. By Theorems 3.16 and 4.3, this is the case if and only if the matrix $A - \lambda I_n$ is *not* invertible, and by Theorem 8.9 this is equivalent to $\det(A - \lambda I_n)$ being zero. $\square$

Corollary

Similar matrices have the same eigenvalues.

Proof

It suffices to observe that, by Theorem 8.7,

$$\begin{aligned}\det(P^{-1}AP - \lambda I_n) &= \det[P^{-1}(A-\lambda I_n)P]\\ &= \det P^{-1}\cdot\det(A-\lambda I_n)\cdot\det P\\ &= \det(A-\lambda I_n). \quad \square\end{aligned}$$

Note that with $A = [a_{ij}]_{n\times n}$ we have

$$\det(A-\lambda I_n) = \det\begin{bmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n}\\ a_{21} & a_{22}-\lambda & \dots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda\end{bmatrix}$$

and, recalling that the product of the diagonal elements is a term in the $\sum\limits_{\sigma}$-expansion, we see that this is a polynomial of degree n in λ. We call this the **characteristic polynomial** of A. By the **characteristic equation** of A we mean the equation

$$\det(A-\lambda I_n) = 0.$$

Thus Theorem 9.1 can be expressed by saying that the eigenvalues of A are the roots of the characteristic equation.

Recall that over the field $\mathbb{C}$ of complex numbers this equation has n roots, some of which may be repeated.

If $\lambda_1,\dots,\lambda_k$ are the distinct roots (= eigenvalues) then the characteristic polynomial factorises in the form

$$(-1)^n(\lambda-\lambda_1)^{r_1}(\lambda-\lambda_2)^{r_2}\cdots(\lambda-\lambda_k)^{r_k}.$$

We call $r_1,\dots,r_k$ the **algebraic multiplicities** of $\lambda_1,\dots,\lambda_k$.

Example 9.1

Consider the matrix

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

We have

$$\det(A-\lambda I_2) = \det\begin{bmatrix} -\lambda & 1\\ -1 & -\lambda\end{bmatrix} = \lambda^2+1.$$

Since λ^2+1 has no real roots, we see that A has no real eigenvalues. However, if we regard A as a matrix over $\mathbb{C}$ then A has two eigenvalues, namely i and $-i$, each being of algebraic multiplicity 1.

10

The Minimum Polynomial

In Chapter 9 we introduced the notions of *eigenvalue* and *eigenvector* of a matrix or of a linear mapping. There we concentrated our attention on showing the importance of these notions in solving particular problems. Here we shall take a closer algebraic look.

We begin by considering again the vector space $\text{Mat}_{n\times n} F$ which, as we know, has the natural basis $\{E_{ij}\ ;\ i,j = 1,\dots,n\}$ and so is of dimension n^2. Thus, recalling Corollary 3 of Theorem 5.8, we have that every set of n^2+1 elements of $\text{Mat}_{n\times n} F$ must be linearly dependent. In particular, given any $A \in \text{Mat}_{n\times n} F$, the n^2+1 powers

$$A^0 = I_n,\ A,\ A^2,\ A^3,\ \dots, A^{n^2}$$

are linearly dependent and so there is a non-zero polynomial

$$p(X) = a_0 + a_1X + a_2X^2 + \cdots + a_{n^2}X^{n^2} \in F[X]$$

such that $p(A) = 0$. The same is of course true for any $f \in \text{Lin}(V, V)$ where V is of dimension n; for, by Theorem 7.2, we have $\text{Lin}(V, V) \simeq \text{Mat}_{n\times n} F$.

But we can do better than this: there is in fact a polynomial $p(X)$ which is *of degree at most n* and such that $p(A) = 0$.

This is the celebrated **Cayley–Hamilton Theorem** which we shall now establish. Since we choose to work in $\text{Mat}_{n\times n} F$, the proof that we shall give is considered 'elementary'. There are other (much more 'elegant') proofs which use $\text{Lin}(V, V)$.

Recall that if $A \in \text{Mat}_{n\times n} F$ then the characteristic polynomial of A is

$$c_A(\lambda) = \det(A - \lambda I_n)$$

and that $c_A(\lambda)$ is of degree n in the indeterminate λ.

Theorem 10.1

[Cayley–Hamilton] $c_A(A) = 0$.

Proof

Let $B = A - \lambda I_n$ and

$$c_A(\lambda) = \det B = b_0 + b_1\lambda + \cdots + b_n\lambda^n.$$

Consider the adjugate matrix adj B. By definition, this is an $n \times n$ matrix whose entries are polynomials in λ of degree at most $n-1$, and so we have

$$\operatorname{adj} B = B_0 + B_1\lambda + \cdots + B_{n-1}\lambda^{n-1}$$

for some $n \times n$ matrices $B_0, \dots, B_{n-1}$. Recalling from Theorem 8.8 that $B \cdot \operatorname{adj} B = (\det B)I_n$, we have

$$(\det B)I_n = B \cdot \operatorname{adj} B = (A - \lambda I_n) \operatorname{adj} B = A \operatorname{adj} B - \lambda \operatorname{adj} B,$$

i.e. we have the polynomial identity

$$b_0 I_n + b_1 I_n \lambda + \cdots + b_n I_n \lambda^n = AB_0 + \cdots + AB_{n-1}\lambda^{n-1} - B_0\lambda - \cdots - B_{n-1}\lambda^n.$$

Equating coefficients of like powers, we obtain

$$\begin{aligned} b_0 I_n &= AB_0 \\ b_1 I_n &= AB_1 - B_0 \\ &\vdots \\ b_{n-1} I_n &= AB_{n-1} - B_{n-2} \\ b_n I_n &= -B_{n-1}. \end{aligned}$$

Multiplying the first equation on the left by $A^0 = I_n$, the second by A, the third by A^2, and so on, we obtain

$$\begin{aligned} b_0 I_n &= AB_0 \\ b_1 A &= A^2 B_1 - AB_0 \\ &\vdots \\ b_{n-1} A^{n-1} &= A^n B_{n-1} - A^{n-1} B_{n-2} \\ b_n A^n &= -A^n B_{n-1}. \end{aligned}$$

Adding these equations together, we obtain $c_A(A) = 0$. □

The Cayley–Hamilton Theorem is really quite remarkable, it being far from obvious that an $n \times n$ matrix should satisfy a polynomial equation of degree n.

Suppose now that k is the lowest degree for which a polynomial $p(X)$ exists such that $p(A) = 0$. Dividing $p(X)$ by its leading coefficient, we obtain a monic polynomial $m(X)$ of degree k which has A as a zero. Suppose that $m'(X)$ is another monic polynomial of degree k such that $m'(A) = 0$. Then $m(X) - m'(X)$ is a non-zero polynomial of degree less than k which has A as a zero. This contradicts the above assumption on k. Consequently, $m(X)$ is the unique monic polynomial of least degree having A as a zero. This leads to the following:

Definition

If $A \in \operatorname{Mat}_{n\times n} F$ then the **minimum polynomial** of A is the monic polynomial $m_A(X)$ of least degree such that $m_A(A) = 0$.

Theorem 10.2

If $p(X)$ is a polynomial such that $p(A) = 0$ then the minimum polynomial $m_A(X)$ divides $p(X)$.

11
Computer Assistance

Many applications of linear algebra require careful, and sometimes rather tedious, calculations by hand. As the reader will be aware, these can often be subject to error. The use of a computer is therefore called for. As far as computation in algebra is concerned, there are several packages that have been developed specifically for this purpose. In this chapter we give a brief introduction, by way of a tutorial, to the package 'LinearAlgebra' in MAPLE 7. Having mastered the techniques, the reader may freely check some of the answers to previous questions!

Having opened MAPLE, begin with the following input:

```
> with(LinearAlgebra):
```

(1) *Matrices*

There are several different ways to input a matrix. Here is the first, which merely gives the matrix as a list of its rows (the matrix palette may also be used to do this). At each stage the MAPLE output is generated immediately following the semi-colon on pressing the ENTER key.

For example, we can input the matrices from Exercise 1.12 as follows: for the first, we do

input:

```
> m1:=Matrix([[3,1,-2],[2,-2,0],[-1,1,2]]);
```

output:

$$m1 := \begin{bmatrix} 3 & 1 & -2 \\ 2 & -2 & 0 \\ -1 & 1 & 2 \end{bmatrix}$$

In order to illustrate how to do matrix algebra with MAPLE, let us input the second matrix of Exercise 1.12:

```
> m2:=Matrix([[1,1,1],[1,-1,1],[0,1,2]]);
```

$$m2 := \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

One way of adding these matrices is by using the 'Add' command:

```
> m3:=Add(m1,m2);
```

$$m3 := \begin{bmatrix} 4 & 2 & -1 \\ 3 & -3 & 1 \\ -1 & 2 & 4 \end{bmatrix}$$

As for multiplying matrices, this can be achieved by using the 'Multiply' command. To multiply the above matrices, for example, input:

```
> m4:=Multiply(m1,m2);
```

$$m4 := \begin{bmatrix} 4 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Now 'Add' also allows linear combinations to be computed. Here, for example, is how to obtain $3m1 + 4m2$:

http://www.springer.com/978-1-85233-662-2

Basic Linear Algebra
Blyth, T.S.; Robertson, E.F.
2002, XI, 232 p., Softcover
ISBN: 978-1-85233-662-2