1
Vector Spaces, Subspaces, and Linear Transformations

The study of matrices is based on the concept of linear transformations between two vector spaces. It is therefore necessary to define what this concept means in order to understand the setup of a matrix. In this chapter, as well as in the remainder of the book, the set of all real numbers is denoted by R, and its elements are referred to as scalars. The set of all n-tuples of real numbers will be denoted by R^n ($n \geq 1$).

1.1 Vector Spaces

This section introduces the reader to ideas that are used extensively in many books on linear and matrix algebra. They involve extensions of the Euclidean geometry which are important in the current mathematical literature and are described here as a convenient introductory reference for the reader. We confine ourselves to real numbers and to vectors whose elements are real numbers.

1.1.1 Euclidean Space

A vector $(x_0, y_0)'$ of two elements can be thought of as representing a point in a two-dimensional Euclidean space using the familiar Cartesian x, y coordinates, as in Figure 1.1. Similarly, a vector $(x_0, y_0, z_0)'$ of three elements can represent a point in a three-dimensional Euclidean space, also shown in Figure 1.1. In general, a vector of n elements can be said to represent a point (an n-tuple) in what is called an n-dimensional Euclidean space. This is a special case of a wider concept called a vector space, which we now define.

Definition 1.1 (Vector Spaces) *A vector space over R is a set of elements, denoted by V, which can be added or multiplied by scalars, in such a way that the sum of two elements of V is an element of V, and the product of an element of V by a scalar is an element of V. Furthermore, the following properties must be satisfied:*

1. $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$ for all $\boldsymbol{u}$, $\boldsymbol{v}$ in V.
2. $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$ for all $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ in V.
3. There exists an element in V, called the zero element and is denoted by $\mathbf{0}$, such that $\mathbf{0} + \boldsymbol{u} = \boldsymbol{u} + \mathbf{0} = \boldsymbol{u}$ for every $\boldsymbol{u}$ in V.
4. For each $\boldsymbol{u}$ in V, there exists a unique element $-\boldsymbol{u}$ in V such that $\boldsymbol{u} + (-\boldsymbol{u}) = (-\boldsymbol{u}) + \boldsymbol{u} = \mathbf{0}$.
5. For every $\boldsymbol{u}$ and $\boldsymbol{v}$ in V and any scalar α, $\alpha(\boldsymbol{u} + \boldsymbol{v}) = \alpha\boldsymbol{u} + \alpha\boldsymbol{v}$.
6. $(\alpha + \beta)\boldsymbol{u} = \alpha\boldsymbol{u} + \beta\boldsymbol{u}$ for any scalars α and β and any $\boldsymbol{u}$ in V.
7. $\alpha(\beta\boldsymbol{u}) = (\alpha\beta)\boldsymbol{u}$ for any scalars α and β and any $\boldsymbol{u}$ in V.
8. For every $\boldsymbol{u}$ in V, $\mathbf{1}\boldsymbol{u} = \boldsymbol{u}$, where $\mathbf{1}$ is the number one, and $0\boldsymbol{u} = \mathbf{0}$, where 0 is the number zero.

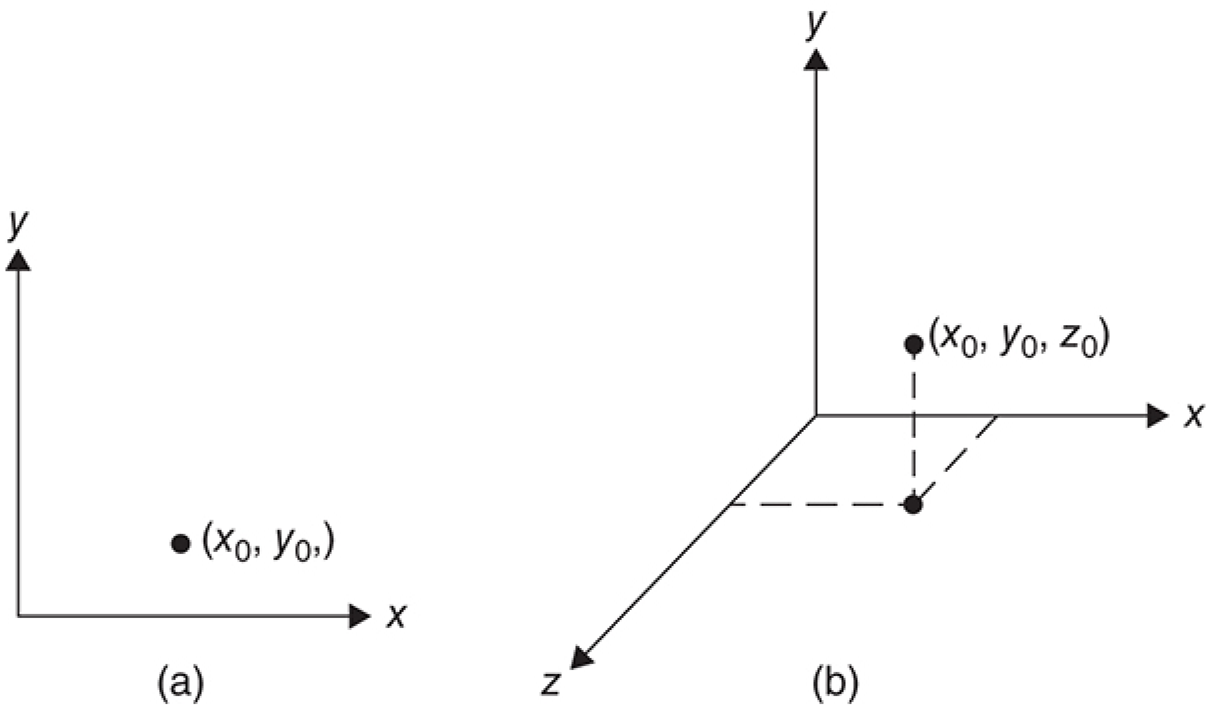


Figure 1.1 (a) Two-Dimensional and (b) Three-Dimensional Euclidean Spaces.

Vector spaces were first defined by the Italian mathematician Giuseppe Peano in 1888.

Example 1.1 *The Euclidean space R^n is a vector space whose elements are of the form $(x_1, x_2, \ldots, x_n)$, $n \geq 1$. For every pair of elements in R^n their sum is in R^n, and so is the product of a scalar and any elements that is in R^n. It is easy to verify that properties (1) through (8) in Definition 1.1 are satisfied. The zero element is $(0, 0, \ldots, 0)$.*

Example 1.2 *The set of all polynomials in x of of degree n or less of the form $\sum^n_{i=0} a_i x^i$, where the a_i's are scalars, is a vector space: the sum of any two such polynomials is a polynomial of the same form, and so is the product of a scalar with a polynomial. For the zero element, $a_i = 0$ for $\forall i$.*

Example 1.3 *The set of all positive functions defined on the closed interval $[-2, 2]$ is not a vector space since multiplying any such function by a negative scalar produces a function that is not in that set.*

Definition 1.2 (Vector Subspace) *Let V be a vector space over R, and let W be a subset of V. Then W is said to be a vector subspace of V if it satisfies the following conditions:*

1. The sum of any two elements in *W* is an element of *W*.
2. The product of any element in *W* by any scalar is an element in *W*.
3. The zero element of *V* is also an element of *W*.

It follows that for *W* to be a vector subspace of *V*, it must itself be a vector space. A vector subspace may consist of one element only, namely the zero element.

The set of all continuous functions defined on the closed interval $[a, b]$ is a vector subspace of all functions defined on the same interval. Also, the set of all points on the straight line $2x - 5y = 0$ is a vector subspace of R^2. However, the set of all points on any straight line in R^2 not going through the origin (0, 0) is not a vector subspace.

Example 1.4 *Let V_1, V_2, and V_3 be the sets of vectors having the forms* **x**, **y**, *and* **z**, *respectively:*

$$\mathbf{x} = \begin{bmatrix} \alpha \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} 0 \\ 0 \\ \beta \end{bmatrix}, \quad \text{and} \quad \mathbf{z} = \begin{bmatrix} \gamma \\ 0 \\ \delta \end{bmatrix} \quad \text{for real } \alpha, \ \beta, \ \gamma, \text{ and } \delta.$$

Then V_1, V_2, and V_3 each define a vector space, and they are all subspaces of R^3. Furthermore, V_1 and V_2 are each a subspace of V_3.

1.2 Base of a Vector Space

Suppose that every element in a vector space V can be expressed as a linear combination of a number of elements in V. The set consisting of such elements is said to *span* or *generate* the vector space V and is therefore called a *spanning set* for V.

Definition 1.3 *Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ be elements in a vector space V. If there exist scalars $a_1, a_2, \ldots, a_n$, not all equal to zero, such that $\sum_{i=1}^{n} a_i \boldsymbol{u}_i = \boldsymbol{0}$, then $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ are said to be linearly dependent. If, however, $\sum_{i=1}^{n} a_i \boldsymbol{u}_i = \boldsymbol{0}$ is true only if all the a_i's are zero, then $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ are said to be linearly independent.*

Note 1:

If $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ are linearly independent, then none of them can be zero. To see this, if, for example, $\boldsymbol{u}_1 = \boldsymbol{0}$, then $a_1\boldsymbol{u}_1 + 0\boldsymbol{u}_2 + \cdots + 0\boldsymbol{u}_n = \boldsymbol{0}$ for any $a_1 \neq 0$, which implies that $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ are linearly dependent, a contradiction. It follows that any set of elements of V that contains the zero element $\boldsymbol{0}$ must be linearly dependent. Furthermore, if $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ are linearly dependent, then at least one of them can be expressed as a linear combination of the remaining elements.

Example 1.5 *Consider the vectors*

$$\boldsymbol{x}_1 = \begin{bmatrix} 3 \\ -6 \\ 9 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 5 \\ -5 \end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix} 2 \\ 1 \\ 1 \end{bmatrix}, \boldsymbol{x}_4 = \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix}, \text{ and } \boldsymbol{x}_5 = \begin{bmatrix} 2 \\ -3 \\ 3 \end{bmatrix}.$$

It is clear that

$$2\boldsymbol{x}_1 + \boldsymbol{x}_4 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} -6 \\ 12 \\ -18 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \boldsymbol{0}, \quad (1.1)$$

that is,

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_4 = \boldsymbol{0} \quad \text{for} \quad \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

which is not zero. Therefore, $\boldsymbol{x}_1$ and $\boldsymbol{x}_4$ are linearly dependent. So also are $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$ because

$$2\boldsymbol{x}_1 + 3\boldsymbol{x}_2 - 3\boldsymbol{x}_3 = \begin{bmatrix} 6 \\ -12 \\ 18 \end{bmatrix} + \begin{bmatrix} 0 \\ 15 \\ -15 \end{bmatrix} + \begin{bmatrix} -6 \\ -3 \\ -3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \boldsymbol{0}. \quad (1.2)$$

In contrast to (1.1) and (1.2), consider

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 = \begin{bmatrix} 3a_1 \\ -6a_1 \\ 9a_1 \end{bmatrix} + \begin{bmatrix} 0 \\ 5a_2 \\ -5a_2 \end{bmatrix} = \begin{bmatrix} 3a_1 \\ -6a_1 + 5a_2 \\ 9a_1 - 5a_2 \end{bmatrix}. \qquad (1.3)$$

There are no values a_1 and a_2 which make (1.3) a zero vector other than $a_1 = 0 = a_2$. Therefore $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent.

Definition 1.4 *If the elements of a spanning set for a vector space V are linearly independent, then the set is said to be a basis for V. The number of elements in this basis is called the dimension of V and is denoted by dim(V).*

Note 2:

The reference in Definition 1.4 was to *a* basis and not the basis because for any vector space there are many bases. All bases of a space *V* have the same number of elements which is dim(*V*).

Example 1.6 *The vectors,*

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \text{ and } \quad \boldsymbol{x}_3 = \begin{bmatrix} 3 \\ 0 \\ 0 \end{bmatrix}$$

are all in R^3. Any two of them form a basis for the vector space whose typical vector is (α, β, 0)′ for α and β real. The dimension of the space is 2. (The space in this case is, of course, a subspace of R^3.)

Note 3:

If $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ form a basis for *V*, and if $\boldsymbol{u}$ is a given element in *V*, then there exists a unique set of scalars, $a_1, a_2, \ldots, a_n$, such that $\boldsymbol{u} = \sum_{i=1}^{n} a_i \boldsymbol{u}_i$. (see Exercise 1.4).

1.3 Linear Transformations

Linear transformations concerning two vector spaces are certain functions that map one vector space, U, into another vector space, V. More specifically, we have the following definition:

Definition 1.5 *Let U and V be two vector spaces over R. Suppose that T is a function defined on U whose values belong to V, that is, T maps U into V. Then, T is said to be a linear transformation on U into V if*

$$T(a_1\boldsymbol{u}_1 + a_2\boldsymbol{u}_2) = a_1T(\boldsymbol{u}_1) + a_2T(\boldsymbol{u}_2),$$

for all $\boldsymbol{u}_1$, $\boldsymbol{u}_2$ in U and any scalars a_1, a_2.

For example, let T be a function from R^3 into R^3 such that $T(x_1, x_2, x_3) = (x_1 - x_2, x_1 + x_3, x_3)$. It can be verified that T is a linear transformation.

Example 1.7 *In genetics, the three possible genotypes concerning a single locus on a chromosome at which there are only two alleles, G and g, are GG, Gg, and gg. Denoting these by g_1, g_2, and g_3, respectively, gene effects relative to these genotypes can be defined (e.g., Anderson and Kempthorne, 1954) in terms of a mean μ, a measure of gene substitution α, and a measure of dominance δ, such that*

$$\begin{aligned}
\mu &= \frac{1}{4}g_1 + \frac{1}{2}g_2 + \frac{1}{4}g_3 \\
\alpha &= \frac{1}{4}g_1 \qquad\qquad - \frac{1}{4}g_3 \\
\delta &= -\frac{1}{4}g_1 + \frac{1}{2}g_2 - \frac{1}{4}g_3
\end{aligned}$$

These equations represent a linear transformation of the vector $(g_1, g_2, g_3)'$ to $(\mu, \alpha, \delta)'$. In Chapter 2, it will be seen that such a transformation is determined by an array of numbers consisting of the coefficients of the g_i's in the above equations. This array is called a *matrix*.

1.3.1 The Range and Null Spaces of a Linear Transformation

Let T be a linear transformation that maps U into V, where U and V are two given vector spaces. The image of U under T, also called the *range* of T and is denoted by $\mathfrak{R}(T)$, is the set of all elements in V of the form $T(\boldsymbol{u})$ for $\boldsymbol{u}$ in U. The *null space*, or the *kernel*, of T consists of all elements $\boldsymbol{u}$ in U such that $T(\boldsymbol{u}) = \boldsymbol{0}_v$ where $\boldsymbol{0}_v$ is the zero element in V. This space is denoted by $\aleph(T)$. It is easy to show that $\mathfrak{R}(T)$ is a vector subspace of V and $\aleph(T)$ is a vector subspace of U (see Exercise 1.7). For example, let T be a linear transformation from R^3 into R^2 such that $T(x_1, x_2, x_3) = (x_1 - x_3, x_2 - x_1)$, then $\aleph(T)$ consists of all points in R^3 such that $x_1 - x_3 = 0$ and $x_2 - x_1 = 0$, or equivalently, $x_1 = x_2 = x_3$. These equations represent a straight line in R^3 passing through the origin.

Theorem 1.1 *Let T be a linear transformation from the vector space U into the vector space V. Let n = dim(U). Then, n = p + q, where p = dim[$\aleph(T)$] and q = dim[$\Re(T)$].*

Proof. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$ be a basis for $\aleph(T)$ and $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_q$ be a basis for $\Re(T)$. Furthermore, let $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ be elements in U such that $T(\boldsymbol{w}_i) = \boldsymbol{v}_i$ for $i = 1, 2, \ldots, q$. Then,

1. $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$; $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ are linearly independent.
2. The elements in (1) span U.

To show (1), suppose that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$; $\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ are not linearly independent, then there exist scalars $\alpha_1, \alpha_2, \ldots, \alpha_p$; $\beta_1, \beta_2, \ldots, \beta_q$ such that

$$\sum_{i=1}^{p} \alpha_i \boldsymbol{u}_i + \sum_{i=1}^{q} \beta_i \boldsymbol{w}_i = \boldsymbol{0}_u, \tag{1.4}$$

where $\boldsymbol{0}_u$ is the zero element in U. Mapping both sides of (1.4) under T, we get

$$\sum_{i=1}^{p} \alpha_i T(\boldsymbol{u}_i) + \sum_{i=1}^{q} \beta_i T(\boldsymbol{w}_i) = \boldsymbol{0}_v,$$

where $\boldsymbol{0}_v$ is the zero elements in V. Since the $\boldsymbol{u}_i$'s belong to the null space, then

$$\sum_{i=1}^{q} \beta_i T(\boldsymbol{w}_i) = \sum_{i=1}^{q} \beta_i \boldsymbol{v}_i = \boldsymbol{0}_v.$$

But, the $\boldsymbol{v}_i$'s are linearly independent, therefore, $\beta_i = 0$ for $i = 1, 2, \ldots, q$. From (1.4) it can be concluded that $\alpha_i = 0$ for $i = 1, 2, \ldots, p$ since the $\boldsymbol{u}_i$'s are linearly independent. It follows that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p; \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ are linearly independent.

To show (2), let $\boldsymbol{u}$ be any element in U, and let $T(\boldsymbol{u}) = \boldsymbol{v}$. There exist scalars $a_1, a_2, \ldots, a_q$ such that $\boldsymbol{v} = \sum_{i=1}^{q} a_i \boldsymbol{v}_i$. Hence,

$$\begin{aligned} T(\boldsymbol{u}) &= \sum_{i=1}^{q} a_i T(\boldsymbol{w}_i) \\ &= T\left(\sum_{i=1}^{q} a_i \boldsymbol{w}_i\right). \end{aligned}$$

It follows that

$$T\left(\boldsymbol{u} - \sum_{i=1}^{q} a_i \boldsymbol{w}_i\right) = \boldsymbol{0}_V.$$

This indicates that $\boldsymbol{u} - \sum_{i=1}^{q} a_i \boldsymbol{w}_i$ is an element in $\aleph(T)$. We can therefore write

$$\boldsymbol{u} - \sum_{i=1}^{q} a_i \boldsymbol{w}_i = \sum_{i=1}^{p} b_i \boldsymbol{u}_i, \tag{1.5}$$

for some scalars $b_1, b_2, \ldots, b_p$. From (1.5) it follows that $\boldsymbol{u}$ can be written as a linear combination of $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p; \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$, which proves (2).

From (1) and (2) we conclude that $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p; \boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_q$ form a basis for U. Hence, $n = p + q$.

Reference

Anderson, V. L. and Kempthorne, O. (1954). A model for the study of quantitative inheritance. *Genetics*, 39, 883–898.

Exercises

1. For $$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}, \mathbf{x}_2 = \begin{bmatrix} -1 \\ 3 \\ 2 \end{bmatrix}, \mathbf{x}_3 = \begin{bmatrix} -13 \\ -1 \\ 2 \end{bmatrix}, \quad \text{and } \mathbf{x}_4 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

   show the following:

   a. $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ are linearly dependent, and find a linear relationship among them.

   b. $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_4$ are linearly independent, and find the linear combination of them that equals $(a, b, c)'$.

2. Let U and V be two vector spaces over R. The *Cartesian product* $U \times V$ is defined as the set of all ordered pairs $(\boldsymbol{u}, \boldsymbol{v})$, where $\boldsymbol{u}$ and $\boldsymbol{v}$ are elements in U and V, respectively. The sum of two elements, $(\boldsymbol{u}_1, \boldsymbol{v}_1)$ and $(\boldsymbol{u}_2, \boldsymbol{v}_2)$ in $U \times V$ is defined as $(\boldsymbol{u}_1 + \boldsymbol{u}_2, \boldsymbol{v}_1 + \boldsymbol{v}_2)$, and if α is a scalar, then $\alpha(\boldsymbol{u}, \boldsymbol{v})$ is defined as $(\alpha\boldsymbol{u}, \alpha\boldsymbol{v})$, where $(\boldsymbol{u}, \boldsymbol{v})$ is an element in $U \times V$. Show that $U \times V$ is a vector space.

3. Let $\boldsymbol{V}$ be a vector space spanned by the vectors,

   $$\boldsymbol{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 3 \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} 0 \\ 4 \\ 6 \end{bmatrix}, \text{ and } \boldsymbol{v}_4 = \begin{bmatrix} 1 \\ 3 \\ 6 \end{bmatrix}.$$

   a. Show that $\boldsymbol{v}_1$ and $\boldsymbol{v}_2$ are linearly independent.

   b. Show that $\boldsymbol{V}$ has dimension 2.

4. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ be a basis for a vector space $\boldsymbol{U}$. Show that if $\boldsymbol{u}$ is any element in $\boldsymbol{U}$, then there exists a unique set of scalars, $a_1, a_2, \ldots, a_n$, such that $\boldsymbol{u} = \sum_{i=1}^{n} a_i\boldsymbol{u}_i$, which proves the assertion in Note 3.

5. Let $\boldsymbol{U}$ be a vector subspace of $\boldsymbol{V}$, $\boldsymbol{U} \neq \boldsymbol{V}$. Show that $\dim(\boldsymbol{U}) < \dim(\boldsymbol{V})$.

6. Let $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m$ be elements in a vector space $\boldsymbol{U}$. The collection of all linear combinations of the form $\sum_{i=1}^{m} a_i\boldsymbol{u}_i$, where $a_1, a_2, \ldots, a_m$ are scalars, is called the linear span of $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m$ and is denoted by $L(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m)$. Show that $L(\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_m)$ is a vector subspace of $\boldsymbol{U}$.

7. Let $\boldsymbol{U}$ and $\boldsymbol{V}$ be vector spaces and let T be a linear transformation from $\boldsymbol{U}$ into $\boldsymbol{V}$.

   1. Show that $\aleph(T)$, the null space of T, is a vector subspace of $\boldsymbol{U}$.

   2. Show that $\Re(T)$, the range of T, is a vector subspace of $\boldsymbol{V}$.

8. Consider a vector subspace of R^4 consisting of all $\boldsymbol{x} = (x_1, x_2, x_3, x_4)'$ such that $x_1 + 3x_2 = 0$ and $2x_3 - 7x_4 = 0$. What is the dimension of this vector subspace?

9. Suppose that T is a linear transformation from R^3 onto R (the image of R^3 under T is all of R) given by $T(x_1, x_2, x_3) = 3x_1 - 4x_2 + 9x_3$. What is the dimension of its null space?

10. Let T be a linear transformation from the vector space U into the vector space V. Show that T is one-to-one if and only if whenever $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are linearly independent in U, then $T(\boldsymbol{u}_1), T(\boldsymbol{u}_2), \ldots, T(\boldsymbol{u}_n)$ are linearly independent in V.

11. Let the functions x, e^x be defined on the closed interval $[0, 1]$. Show that these functions are linearly independent.

2
Matrix Notation and Terminology

We have seen in the Introduction that a matrix is defined in terms of a system of *m* linear equations in *n* variables of the form

$$\sum_{j=1}^{n} a_{ij}x_j = y_i, \quad i = 1, 2, \dots, m, \tag{2.1}$$

where the a_{ij}'s are given coefficients. These equations can be represented by a single equation as shown in (2.2)

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{bmatrix}. \tag{2.2}$$

This representation is in fact a linear transformation from the *n*-dimensional Euclidean space R^n to the *m*-dimensional Euclidean space R^m, that is, $R^n \to R^m$. The array of a_{ij} coefficients in (2.2) is called a *matrix* with *m* rows and *n* columns.

Thus a matrix is a rectangular or square array of numbers arranged in rows and columns. The rows are of equal length, as are the columns. The a_{ij} coefficient denotes the element in the ith row and jth column of a matrix $\boldsymbol{A}$. We then have the representation

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1j} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2j} & \cdots & a_{2n} \\ \vdots & & & \vdots & & \vdots & \\ a_{i1} & a_{i2} & a_{i3} & \cdots & a_{ij} & \cdots & a_{in} \\ \vdots & & & \vdots & & \vdots & \\ a_{m1} & a_{m2} & a_{m3} & \cdots & a_{mj} & \cdots & a_{mn} \end{bmatrix}. \tag{2.3}$$

The three dots indicate, in the first row, for example, that the elements a_{11}, a_{12}, and a_{13} continue in sequence up to a_{1j} and then up to a_{1n}; likewise in the first column the elements a_{11}, a_{21}, continue in sequence to a_{i1} and up to a_{m1}. The use of three dots to represent omitted values of a long sequence in this manner is standard and will be used extensively. This form of writing a matrix clearly specifies its elements, and also its size, that is, the number of rows and columns. An alternative and briefer form is

$$\mathbf{A} = \{a_{ij}\}, \quad \text{for} \quad i = 1, 2, \dots m, \quad \text{and} \quad j = 1, 2, \dots, n, \tag{2.4}$$

the curly brackets indicating that a_{ij} is a typical element, the limits of *i* and *j* being *m* and *n*, respectively.

The element a_{ij} is called the (i, j)th *element*, the first subscript referring to the row the element is in and the second to the column. Thus a_{23} is the element in the second row and third column. The size of the matrix, that is, the number of rows and columns, is referred to as its *order* or as its *dimensions*. (Sometimes the word *order* is also used for other characteristics of a matrix, but in this book it always refers to the number of rows and columns.) Thus **A** with m rows and n columns has order $m \times n$ (read as "m by n") and, to emphasize its order, the matrix can be written $\mathbf{A}_{m \times n}$. A simple example of a matrix of order 2 × 3 is

$$\mathbf{A}_{2\times3} = \begin{bmatrix} 4 & 0 & -3 \\ -7 & 2.73 & 1 \end{bmatrix}.$$

Notice that zero is legitimate as an element, that the elements need not all have the same sign, and that integers and decimal numbers can both be elements of the same matrix. The element in the first row and first column, a_{11}, the element in the upper left-hand corner of a matrix is called the *leading element* of the matrix.

When $r = c$, the number of rows equals the number of columns, **A** is square and is referred to as a *square matrix* and, provided there is no ambiguity, is described as being of order r. In this case the elements $a_{11}, a_{22}, a_{33}, \ldots, a_{rr}$ are referred to as the *diagonal elements* or *diagonal* of the matrix. For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 7 & 0 & 7 \\ 5 & 2 & 9 & 1 \\ 0 & 3 & 4 & 6 \\ 8 & 0 & 5 & 7 \end{bmatrix}$$

is square of order 4 and its diagonal elements are 1, 2, 4, and 7. The elements of a square matrix that lie in a line parallel to and just below the diagonal are sometimes referred to as being on the *subdiagonal elements*; in the example they are 5, 3, 5. Elements of a square matrix other than the diagonal elements are called *off-diagonal* or *nondiagonal* elements.

A square matrix having zero for all its nondiagonal elements is described as a *diagonal matrix*; for example,

$$\mathbf{A} = \begin{bmatrix} 3 & 0 & 0 \\ 0 & -17 & 0 \\ 0 & 0 & 99 \end{bmatrix}$$

is a diagonal matrix. When $a_1, a_2, \ldots, a_n$ (some of which may be zero) are the diagonal elements of a diagonal matrix, useful notations for that matrix are

$$\begin{aligned} D\{a_1, a_2, \ldots, a_n\} &= \operatorname{diag}\{a_1, a_2, \ldots, a_n\} \\ &= D\{a_i\} = \operatorname{diag}\{a_i\} \quad \text{for} \quad i = 1, 2, \ldots, n. \end{aligned}$$

A square matrix with all elements above (or below) the diagonal being zero is called a *triangular matrix*. For example,

$$\mathbf{B} = \begin{bmatrix} 1 & 5 & 13 \\ 0 & -2 & 9 \\ 0 & 0 & 7 \end{bmatrix} \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 0 & 0 \\ 8 & 3 & 0 \\ 1 & -1 & 2 \end{bmatrix}$$

are triangular matrices. **B** is an *upper triangular matrix* and **C** is a *lower triangular matrix*.

Example 2.1 *Suppose an experienced taxicab driver has ascertained that when in Town 1 there is a probability of .2 that the next fare will be within Town 1 and a probability of .8 that it will be to Town 2. But when in Town 2 the probabilities are .4 of going to Town 1 and .6 of staying in Town 2. These probabilities can be assembled in a matrix*

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} = \{p_{ij}\} \quad \text{for} \quad i, j = 1, 2. \tag{2.5}$$

Then p_{ij} is the probability when in Town i of the next fare being to Town j.

The matrix in (2.5) is an example of a transition probability matrix. It is an array of probabilities of making a transition from what is called state i to state j. (The states in the example are the two towns.) Thus p_{ij} is a transition probability, and hence the name: *transition probability matrix*.

In (2.5), the sum of the elements in each row (the *row sum*) is 1. This is so because whenever the taxi is in Town i it must, with its next fare, either stay in that town or go to the other one. This is a feature of all transition probability matrices. In general, for

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1j} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2j} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{i1} & p_{i2} & \cdots & p_{ij} & \cdots & p_{im} \\ \vdots & \vdots & & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mj} & \cdots & p_{mm} \end{bmatrix} = \{p_{ij}\} \quad \text{for} \quad i, j = 1, 2, \ldots, m, \tag{2.6}$$

corresponding to m different states, the sum of the probabilities of going from state i to anyone of the m states (including staying in state i) must be 1. Hence,

$$p_{i1} + p_{i2} + \cdots + p_{ij} + \cdots p_{im} = 1,$$

that is,

$$\sum_{j=1}^{m} p_{ij} = 1 \text{ for all } i = 1, 2, \ldots, m. \tag{2.7}$$

Situations like the taxicab example form a class of probability models known as *Markov chains*. They arise in a variety of ways: for example, in considering the probabilities of changes in the prime interest rate from week to week, or the probabilities of certain types of mating in genetics, or in studying the yearly mobility of labor where p_{ij} can be the probability of moving during a calendar year from category *i* in the labor force to category *j*. Assembling probabilities of this nature into a matrix then enables matrix algebra to be used as a tool for answering questions of interest; for example, the probabilities of moving from category *i* to category *j* in 2 years are given by $\mathbf{P}^2$, in 3 years by $\mathbf{P}^3$, and so on.

2.1 Plotting of a Matrix

There is an interesting method to provide a graphical representation of a matrix. Suppose that $\boldsymbol{M}$ is a matrix is of order $m \times n$. The elements in each column of $\boldsymbol{M}$ are plotted against the digits 1, 2, …, m which label the rows of $\boldsymbol{M}$. Thus if m_{ij} is the (i, j)th element of $\boldsymbol{M}$, then for each j, m_{1j}, m_{2j}, …, m_{mj} are plotted against 1, 2, …, m, respectively. In other words, each matrix column is plotted as a separate broken line and the resulting n of such lines form a graphical representation of the matrix $\boldsymbol{M}$. This is demonstrated in Figure 2.1 for the matrix,

$$\boldsymbol{M} = \begin{bmatrix} 15 & 2 & 4 & 14 \\ 5 & 10 & 11 & 7 \\ 8 & 8 & 6 & 13 \\ 5 & 14 & 13 & 1 \end{bmatrix}. \tag{2.8}$$

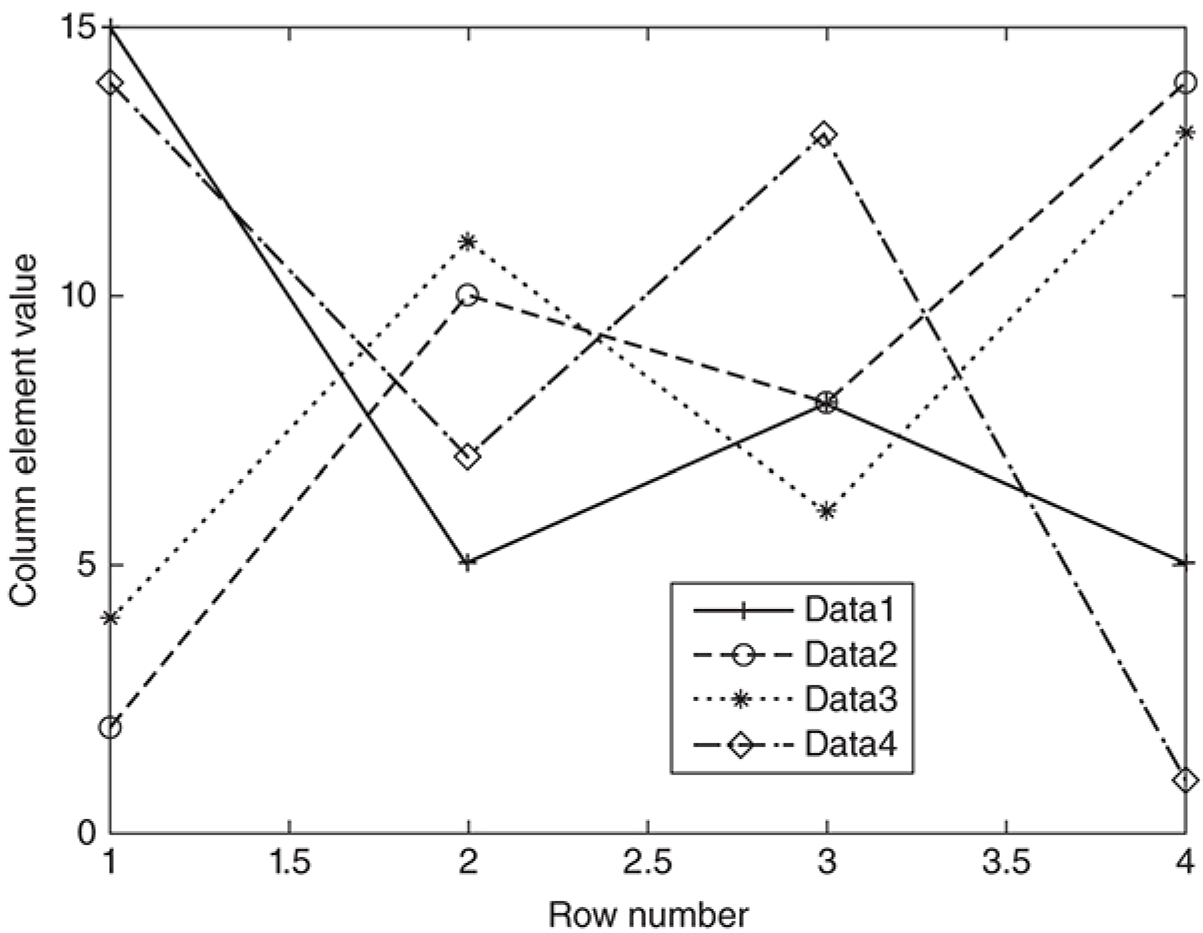


Figure 2.1 Plot of Matrix $\boldsymbol{M}$.

Plot of the 4 × 4 identity matrix,

$$\boldsymbol{I}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.9}$$

is shown in Figure 2.2. More detailed discussions concerning plotting matrices will be given in Chapter 16.

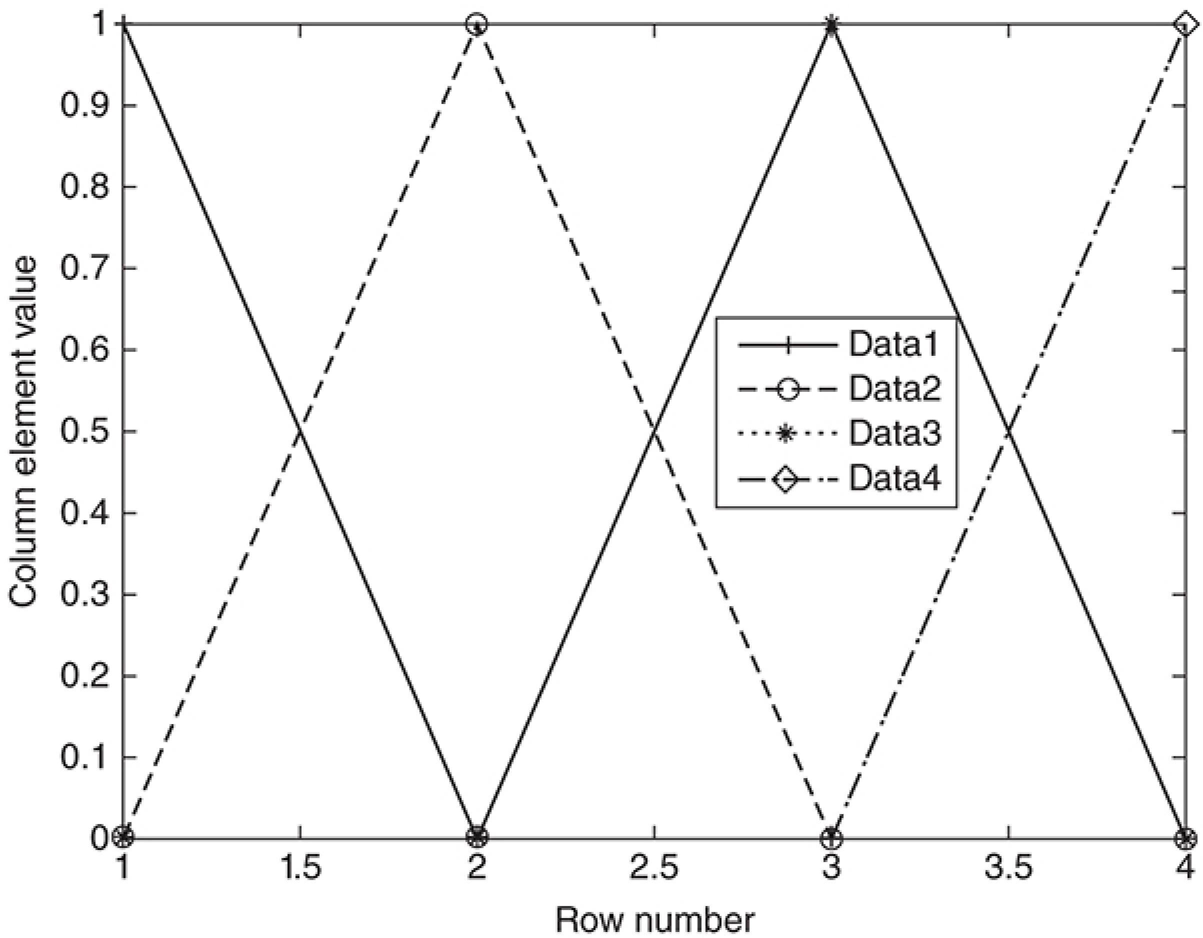


Figure 2.2 Plot of a 4 × 4 Identity Matrix.

2.2 Vectors and Scalars

A matrix consisting of only a single column is called a *column vector*; for example,

$$\mathbf{x} = \begin{bmatrix} 3 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

is a column vector of order 4. Likewise a matrix that is just a single row is a *row vector*:

$$\mathbf{y}' = [4 \quad 6 \quad -7]$$

is a row vector of order 3. A single number such as 2, 6.4 or –4 is called a *scalar*; the elements of a matrix are (usually) scalars. Sometimes it is convenient to think of a scalar as a matrix of order 1 × 1.

2.3 General Notation

A well-recognized notation is that of denoting matrices by uppercase letters and their elements by the lowercase counterparts with appropriate subscripts. Vectors are denoted by lowercase letters, often from the end of the alphabet, using the prime superscript to distinguish a row vector from a column vector. Thus **x** is a column vector and **x′** is a row vector. The lowercase Greek lambda, λ, is often used for a scalar.

Throughout this book the notation for displaying a matrix is

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

enclosing the array of elements in square brackets. Among the variety of forms that can be found in the literature are

$$\begin{pmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{pmatrix}, \quad \begin{Bmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{Bmatrix}, \quad \text{and} \quad \begin{Vmatrix} 1 & 4 & 6 \\ 0 & 2 & 3 \end{Vmatrix}.$$

Single vertical lines are seldom used, since they are usually reserved for determinants (to be defined later). Another useful notation that has already been described is

$$\mathbf{A} = \{a_{ij}\} \quad \text{for } i = 1, 2, \ldots, r \text{ and } j = 1, 2, \ldots, c.$$

The curly brackets indicate that a_{ij} is a typical term of the matrix **A** for all pairs of values of *i* and *j* from unity up to the limits shown, in this case *r* and *c*; that is, **A** is a matrix of *r* rows and *c* columns. This is by no means a universal notation and several variants of it can be found in the literature. Furthermore, there is nothing sacrosanct about the repeated use in this book of the letter **A** for a matrix. Any letter may be used.

Exercises

1. For the elements of the matrix

$$\mathbf{A} = \begin{bmatrix} -1 & 17 & 9 & -2 & 3 \\ 3 & 13 & 10 & 2 & 6 \\ 11 & -9 & 0 & -3 & 2 \\ -6 & -8 & 1 & 4 & 5 \end{bmatrix} = \{a_{ij}\}$$

for $i = 1, 2, \ldots, 4$ and $j = 1, 2, \ldots, 5$, show that

a. $a_{1\cdot} = 26$;

b. $a_{\cdot 3} = 20$;

c. $\sum_{i=1}^{3} a_{i2} = 21$;

d. $\sum_{\substack{i=1 \\ i \neq 2}}^{4} a_{i5} = 10$;

e. $\sum_{i=3}^{4} a_{i\cdot}^2 = \sum_{i=3}^{4} (a_{i\cdot})^2 = 17$;

f. $a_{\cdot\cdot} = 57$;

g. $\sum_{i=1}^{4} \frac{\sum_{j=1}^{5} a_{ij}^2}{a_{i\cdot}} = 203\frac{275}{442}$;

h. $\sum_{\substack{i=1 \\ i \neq 2}}^{4} \sum_{\substack{j=2 \\ j \neq 4}}^{5} a_{ij} = 20$;

i. $\prod_{j=1}^{3} a_{4j} = \prod_{i=1}^{4} a_{i4}$;

j. $\sum_{i=1}^{4} a_{i1} a_{i4} = -49$;

k. $\sum_{i=1}^{4} \sum_{j=1}^{5} a_{\cdot\cdot} = 1140$;

l. $\prod_{i=1}^{3} a_{i4} = 12;$

m. $\prod_{j=1}^{5} a_{2j}a_{3j} = 0;$

n. $\prod_{i=1}^{4} 2^{a_{i4}} = 2;$

o. $\sum_{i=1}^{4}\sum_{j=1}^{5} (-1)^{j} a_{ij} = -29;$

p. for $i = 2, \sum_{\substack{j=1 \\ j\neq 3}}^{5} (a_{ij} - a_{i+2,j})^2 = 527;$

q. $\prod_{j=1}^{5} 2^{(-1)^{j} a_{2j}} = 0.0625.$

2. Write down the following matrices:

$$\begin{array}{llllll} \mathbf{A} = \{a_{ij}\} & \text{for} & a_{ij} = i + j & \text{for} & i = 1, 2, 3 & \text{and} \quad j = 1, 2. \\ \mathbf{B} = \{b_{kt}\} & \text{for} & b_{kt} = k^{t-1} & \text{for} & k = 1, \ldots, 4 & \text{and} \quad t = 1, 3. \\ \mathbf{C} = \{c_{rs}\} & \text{for} & c_{rs} = 3r + 2(s-1) & \text{for} & r = 1, \ldots, 4 & \end{array}$$

and $s = 1, \ldots, 5$.

3. Prove the following identities:

a.

$$\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2 = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2.$$

b.

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}\right)^2 = \sum_{i=1}^{m} a_{i.}^2.$$

c.

$$\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ij}a_{hk} = a_{i.}a_{h.}, \quad \textit{for} \quad i \neq h.$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} 4a_{ij} = 4a_{..}.$$

d.

See Chapter 16 before doing Exercises 2.4 and 2.5.

4. Plot the following matrix:

$$\boldsymbol{A} = \begin{bmatrix} 13 & 4 & 6 & 16 \\ 7 & 8 & 11 & 9 \\ 9 & 10 & 8 & 15 \\ 6 & 12 & 15 & 2 \end{bmatrix}.$$

5. Plot the upper-triangular matrix,

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 & 7 \\ 0 & 7 & 3 \\ 0 & 0 & 9 \end{bmatrix}.$$

6. Construct a 3 × 3 matrix $\boldsymbol{A}$ whose (i, j)th element is $a_{ij} = (i + j)^2$. Conclude that the a_{ij}'s are symmetric with respect to the diagonal without doing any numerical computation.

7. Write down the matrices:

   a. $\boldsymbol{D}_1 = D\{3^{i-2}\}$ for $i = 1, 2, 3, 4$.

   b. $\boldsymbol{D}_2 = D\{i + 3^{i-2}\}$ for $i = 1, 2, 3, 4$.

8. If the elements a_{ij} of a matrix $\boldsymbol{A}$ of order $n \times n$ are symmetric with respect to its diagonal, how many different elements can $\boldsymbol{A}$ have above its diagonal?

9. Show that if the elements a_{ij} of a matrix $\boldsymbol{A}$ of order $n \times n$ are such that $a_{ij} = -a_{ji}$ for all $i, j(1, 2, \ldots, n)$, then $\boldsymbol{A}$ can have in general $n(n - 1)/2$ different elements above its diagonal, and the same identical elements (except for the negative sign) below the diagonal.

10. Suppose that

$$\begin{bmatrix} 2x & x - 2y \\ x - 3z & 3y + w \end{bmatrix} = \begin{bmatrix} 8 & 5 \\ 7 & 2 \end{bmatrix}.$$

Find the values of x, y, z, and w.

3
Determinants

In this chapter we discuss the calculation of a scalar value known as the *determinant* of a (square) matrix. Knowledge of this calculation is necessary for our discussion of matrix inversion and it is also useful in succeeding chapters for certain aspects of matrix theory proper that are used in applied problems. Moreover, since readers will undoubtedly encounter determinants elsewhere in their readings on matrices, it is appropriate to have an introduction to them here.

The literature of determinants is extensive and forms part of many texts on matrices. The presentation in this book is relatively brief, deals with elementary methods of evaluation and discusses selected topics arising therefrom. Extensive use is made of small numerical examples rather than rigorous mathematical proof. Several portions of the chapter can well be omitted at a first reading.

We begin with some general descriptions and proceed from them to the calculating of a determinant from its associated matrix. In practice, the method is useful mainly for deriving determinants of matrices of small order, but it is more informative about the structure of a determinant than is a formal definition. Nevertheless, a formal definition, which perforce is lengthy and quite difficult to follow, is given in Section 3.2.

3.1 Expansion by Minors

A determinant is a polynomial of the elements of a square matrix. It is a scalar. It is the sum of certain products of the elements of the matrix from which it is derived, each product being multiplied by +1 or −l according to certain rules.

Determinants are defined only for square matrices—the determinant of a nonsquare matrix is undefined, and therefore does not exist. The determinant of a (square) matrix of order $n \times n$ is referred to as an *n-order* determinant and the customary notation for the determinant of the matrix **A** is $|\mathbf{A}|$, it being assumed that **A** is square. Some texts use the notation $\|\mathbf{A}\|$ or det **A**, but $|\mathbf{A}|$ is common and is used throughout this book. Obtaining the value of $|\mathbf{A}|$ by adding the appropriate products of the elements of A (with the correct + 1 or −1 factor included in the product) is variously referred to as *evaluating* the determinant, *expanding* the determinant or *reducing* the determinant. The procedure for evaluating a determinant is first illustrated by a series of simple numerical examples.

3.1.1 First- and Second-Order Determinants

To begin, we have the elementary result that the determinant of a 1 × 1 matrix is the value of its sole element.

The value of a second-order determinant is scarcely more difficult. For example, the determinant of

$$\mathbf{A} = \begin{bmatrix} 7 & 3 \\ 4 & 6 \end{bmatrix} \quad \text{is written} \quad |\mathbf{A}| = \begin{vmatrix} 7 & 3 \\ 4 & 6 \end{vmatrix}$$

and is calculated as

$$|\mathbf{A}| = 7(6) - 3(4) = 30.$$

This illustrates the general result for expanding a second-order determinant: from the product of the diagonal terms subtract the product of the off-diagonal terms. In other words, the determinant of a 2 × 2 matrix consists of the product (multiplied by +1) of the diagonal terms plus the product (multiplied by –1) of the off-diagonal terms. Hence, in general

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} + (-1)a_{12}a_{21} = a_{11}a_{22} - a_{12}a_{21}.$$

Further examples are

$$\begin{vmatrix} 6 & 8 \\ 17 & 21 \end{vmatrix} = 6(21) - 8(17) = -10$$

and

$$\begin{vmatrix} 9 & -3.81 \\ 7 & -1.05 \end{vmatrix} = 9(-1.05) - (-3.81)(7) = 17.22.$$

The evaluation of a second-order determinant is patently simple.

3.1.2 Third-Order Determinants

A third-order determinant can be expanded as a linear function of three second-order determinants derived from it. Their coefficients are elements of a row (or column) of the main determinant, each product being multiplied by + 1 or − 1. For example, the expansion of

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix}$$

based on the elements of the first row, 1, 2, and 3, is

$$\begin{aligned} |\mathbf{A}| &= 1(+1)\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 2(-1)\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix} + 3(+1)\begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix} \\ &= 1(50-48) - 2(40-42) + 3(32-35) = -3. \end{aligned}$$

The determinant that multiplies each element of the chosen row (in this case the first row) is the determinant derived from $|\mathbf{A}|$ by crossing out the row and column containing the element concerned. For example, the first element, 1, is multiplied by the determinant $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$ which is obtained from $|\mathbf{A}|$ through crossing out the first row and column; and the element 2 is multiplied (apart from the factor of –1) by the determinant derived from $|\mathbf{A}|$ by deleting the row and column containing that element—namely, the first row and second column—so leaving $\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$. Determinants obtained in this way are called *minors* of $|\mathbf{A}|$, that is to say, $\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$ is the minor of the element 1 in $|\mathbf{A}|$, and $\begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$ is the minor of the element 2.

The (+1) and (−1) factors in the expansion are decided on according to the following rule: if $\mathbf{A}$ is written in the form $\mathbf{A} = \{a_{ij}\}$, the product of a_{ij} and its minor in the expansion of the determinant $|\mathbf{A}|$ is multiplied by $(-1)^{i+j}$. Hence because the element 1 in the example is the element a_{11}, its product with its minor is multiplied by $(-1)^{1+1} = +1$; and for the element 2, which is a_{12}, its product with its minor is multiplied by $(-1)^{1+2} = -1$.

Let us denote the minor of the element a_{11} by $|\mathbf{M}_{11}|$, where $\mathbf{M}_{11}$ is a submatrix of $\mathbf{A}$ obtained by deleting the first row and first column. Then in the preceding example

$|\mathbf{M}_{11}| = \begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix}$. Similarly, if $|\mathbf{M}_{12}|$ is the minor of a_{12}, then $|\mathbf{M}_{12}| = \begin{vmatrix} 4 & 6 \\ 7 & 10 \end{vmatrix}$; and if $|\mathbf{M}_{13}|$ is the minor of a_{13}, then $|\mathbf{M}_{13}| = \begin{vmatrix} 4 & 5 \\ 7 & 8 \end{vmatrix}$. With this notation, the preceding expansion of $|\mathbf{A}|$ is

$$|\mathbf{A}| = a_{12}(-1)^{1+1}|\mathbf{M}_{11}| + a_{12}(-1)^{1+2}|\mathbf{M}_{12}| + a_{13}(-1)^{1+3}|\mathbf{M}_{13}|.$$

This method of expanding a determinant is known as *expansion by the elements of a row* (*or column*) or sometimes as *expansion by minors*. It has been illustrated using elements of the first row but can also be applied to the elements of any row (or column). For example, the expansion of $|\mathbf{A}|$ just considered, using elements of the second row gives

$$\begin{aligned} |\mathbf{A}| &= 4(-1)\begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} + 5(+1)\begin{vmatrix} 1 & 3 \\ 7 & 10 \end{vmatrix} + 6(-1)\begin{vmatrix} 1 & 2 \\ 7 & 8 \end{vmatrix} \\ &= -4(-4) + 5(-11) - 6(-6) = -3 \end{aligned}$$

as before, and using elements of the first column the expansion also gives the same result:

$$\begin{aligned} |\mathbf{A}| &= 1(+1)\begin{vmatrix} 5 & 6 \\ 8 & 10 \end{vmatrix} + 4(-1)\begin{vmatrix} 2 & 3 \\ 8 & 10 \end{vmatrix} + 7(+1)\begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} \\ &= 1(2) - 4(-4) + 7(-3) = -3. \end{aligned}$$

The minors in these expansions are derived in exactly the same manner as in the first, by crossing out rows and columns of $|\mathbf{A}|$, and so are the (+1) and (− 1) factors; for example, the minor of the element 4 is $|\mathbf{A}|$ with second row and first column deleted, and since 4 is a_{21} its product with its minor is multiplied by $(-1)^{2+1} = -1$. Other terms are obtained in a similar manner.

The foregoing example illustrates the expansion of the general third-order determinant

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}.$$

Expanding this by elements of the first row gives

$$\begin{aligned} |\mathbf{A}| &= a_{11}(+1)\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + a_{12}(-1)\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}(+1)\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{33} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} + a_{12}a_{23}a_{31} \\ &\qquad + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

The reader should verify that expansion by the elements of any other row or column leads to the same result. No matter by what row or column the expansion is made, the value of the determinant is the same. Note that once a row or column is decided on and the sign calculated for the product of the first element therein with its minor, the signs for the succeeding products alternate from plus to minus and minus to plus.

The diagonal of a square matrix has already been defined. The *secondary diagonal* (as it is sometimes called) is the line of elements in a square matrix going from the lower left corner to the upper right, the "other" diagonal, so to speak. Thus the determinant of a matrix of order 2 is the product of elements on the diagonal minus that of the elements on the secondary diagonal. This extends to the determinant of a 3×3 matrix. From the end of the preceding paragraph, the expression for $|\mathbf{A}|$ can be rewritten with its three positive terms followed by its three negative terms:

$$\begin{aligned}|\mathbf{A}| = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21} \\ - \left(a_{31}a_{22}a_{13} + a_{11}a_{32}a_{23} + a_{12}a_{21}a_{33}\right).\end{aligned}$$

3.1.3 *n*-Order Determinants

The expansion of an n-order determinant by this method is an extension of the expansion of a third-order determinant as just given. Thus the determinant of the $n \times n$ matrix $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, 2, \ldots, n$, is obtained as follows. Consider the elements of any one row (or column): multiply each element, a_{ij}, of this row (or column) by its minor, $|\mathbf{M}_{ij}|$, the determinant derived from $|\mathbf{A}|$ by crossing out the row and column containing a_{ij}; multiply the product by $(-1)^{i+j}$; add the signed products and their sum is the determinant $|\mathbf{A}|$; that is, when expanding by elements of a row

$$|\mathbf{A}| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}|\mathbf{M}_{ij}| \quad \text{for any } i, \tag{3.1}$$

and when expanding by elements of a column

$$|\mathbf{A}| = \sum_{i=1}^{n} a_{ij}(-1)^{i+j}|\mathbf{M}_{ij}| \quad \text{for any } j. \tag{3.2}$$

This expansion is used recurrently when n is large, that is, each $|\mathbf{M}_{ij}|$ is expanded by the same procedure. Thus a fourth-order determinant is first expanded as four signed products each involving a third-order minor, and each of these is expanded as a sum of three signed products involving a second-order determinant. Consequently, a fourth-order determinant ultimately involves $4 \times 3 \times 2 = 24$ products of its elements, each product containing four elements. This leads us to the general statement that the determinant of a square matrix of order n is a sum *of* $(n!)$[1] signed products, each product involving n elements of the matrix. The determinant is referred to as an *n-order* determinant. Utilizing methods given by Aitken (1948), it can be shown that each product has one and only one element from each row and column and that all such products are included and none occurs more than once.

This method of evaluation requires lengthy computing for determinants of order exceeding 3 to 4, say, because for order n there are $n!$ terms to be calculated, and even for n as small as $n = 10$ this means 3,628,800 terms! Fortunately, easier methods exist, but because the method already discussed forms the basis of these easier methods, it has been considered first. Furthermore, it is useful in presenting various properties of determinants.

3.2 Formal Definition

As has been indicated, the determinant of an n-order square matrix is the sum of $n!$ signed products of the elements of the matrix, each product containing one and only one element from every row and column of the matrix. Evaluation through expansion by elements of a row or column, as represented in equations formulas (3.1) and (3.2), yields the requisite products with their correct signs when the minors are successively expanded at each stage by the same procedure. We have already referred to this as the method of expansion by minors.

Example 3.1 *In the expansion of*

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

the j_1, j_2, j_3 array in the product $a_{12}a_{23}a_{31}$ is $j_1 = 2, j_2 = 3$, and $j_3 = 1$. Now n_1 is the number of j's in this array that are less than j_1 but follows it; that is, that are less than 2 but follow it. There is only one, namely $j_3 = 1$; therefore $n_1 = 1$. Likewise $n_2 = 1$ because $j_2 = 3$ is followed by only one j less than 3, and finally $n_3 = 0$. Thus

$$n_1 + n_2 + n_3 = 1 + 1 + 0 = 2$$

and the sign of the product $a_{12}a_{23}a_{31}$ in $|\mathbf{A}|$ is therefore $(-1)^2 = +1$. That the sign is also positive in the expansion of $|\mathbf{A}|$ by minors is easily shown, for in expanding by elements of the first row the term involving a_{12} is

$$-a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} = -a_{12}(a_{21}a_{33} - a_{23}a_{31}),$$

which includes the term $+a_{12}a_{23}a_{31}$.

We shall now indicate the generality of the result implicit in the above example, that the formal definition of $|\mathbf{A}|$ agrees with the procedure of expansion by minors. The definition states that each product of n elements in $|\mathbf{A}|$ contains one element from every row and column. Suppose we set out to form one such product by selecting elements for it one at a time from each of the rows of $\mathbf{A}$. Starting with the first row there are n possible choices of an element as the first element of the product. Having made a choice there are then only $n - 1$ possible choices in the second row because the column containing the element chosen from the first row must be excluded from the choices in subsequent rows (in order to have one and only one element from each column as well as from each row). Similarly, the column containing the element chosen from the second row has to be excluded from the possible choices in the third and subsequent rows. Thus there are $(n - 2)$ possible choices in the third row, $(n - 3)$ in the fourth row, and so on. Hence, based on the definition, the total number of products in $|\mathbf{A}|$ is

$$n(n-1)(n-2)\cdots(3)(2)(1) = (n!).$$

On the other hand, expansion of $|\mathbf{A}|$ by minors initially gives $|\mathbf{A}|$ as a sum of n elements each multiplied (apart from sign) by its minor, a determinant of order $n - 1$. Each minor can be expanded in similar fashion as the sum of $(n - 1)$ elements multiplied (apart from sign) by minors that are determinants of order $n - 2$. In this way we see that the complete determinant consists of $n(n - 1)(n - 2)\cdots(3)(2)(1) = (n!)$ signed products each containing n elements. And at each stage of the expansion, the method of deleting a row and a column from a determinant to obtain the minor of an element ensures that in each product the n elements consist of one from every row and column of $\mathbf{A}$. Hence (apart from sign) both the definition and the procedure of expansion by minors lead to the same set of products.

We now show that the sign of each product is the same in both cases. The sign of any product in a determinant expanded by minors is the product of the signs applied to each minor involved in the derivation of the product. Consider the product as

$$a_{1j_1}a_{2j_2}a_{3j_3}\cdots a_{nj_n}.$$

Supposing the first expansion of $|\mathbf{A}|$ is by elements of the first row, the sign attached to the minor of a_{1j_1} is $(-1)^{1+j_1}$. If the expansion of this minor, is by elements of the second row of $|\mathbf{A}|$ which is now the first row of the minor, the minor therein of a_{2j_2} will have attached to it the sign $(-1)^{1+j_2-1}$ if j_1 is less than j_2 and $(-1)^{1+j_2}$ if j_1 exceeds j_2. Likewise the sign attached to the minor of a_{3j_3} in the minor of a_{2j_2} in the minor of a_{1j_1} in the expansion of $|\mathbf{A}|$ will be $(-1)^{1+j_2-2}$ if both j_1 and j_2 are less than j_3. It will be $(-1)^{1+j_2-1}$ if either of j_1 or j_2 is less than j_3 and the other exceeds it, and it will be $(-1)^{1+j_2}$ if both j_1 and j_2 exceed j_3. Hence the sign will be $(-1)^{1+j_3-m_3}$ where m_3 is the number of j's less than j_3 which precede it in the array $j_1j_2j_3\cdots j_n$. In general the sign of the ith minor involved will be $(-1)^{1+j_i-m_i}$ where m_i is the number of j's less than j_i which precede it in the array. Therefore the combined sign is

$$\prod_{i=1}^{n}(-1)^{1+j_i-m_i} = (-1)^q \quad \text{for} \quad q = \sum_{i=1}^{n}(1 + j_i - m_i).$$

Now, since the j_i's are the first n integers in some order, the number of them that are less than some particular j_i is $j_i - 1$. Therefore, by the definitions of n_i and m_i their sum is $j_i - 1$, that is, $n_i + m_i = j_i - 1$, which gives

$$q = \sum_{i=1}^{n}(1 + j_i - m_i) = \sum_{i-1}^{n}(2 + n_i) = 2n + \sum_{i=1}^{n} n_i.$$

Therefore $(-1)^q = (-1)^p$ for $p = \sum_{i=1}^{n} n_i$, so that when the determinant is expanded by minors the sign of a product is $(-1)^p$ as specified by the definition. Hence the definition and the method of expansion by minors are equivalent.

3.3 Basic Properties

Several basic properties of determinants are useful for circumventing both the tedious expansion by elements of a row (or column) described in Section 3.1 and the formal definition. Each is stated in the form of a theorem, followed by an example and proof.

3.3.1 Determinant of a Transpose

Theorem 3.1 *The determinant of the transpose of a matrix equals the determinant of the matrix itself:* $|\mathbf{A}'| = |\mathbf{A}|$.

Example 3.2 *On using expansion by the first row,*

$$|\mathbf{A}| = \begin{vmatrix} 1 & -1 & 0 \\ 2 & 1 & 2 \\ 4 & 4 & 9 \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 4 & 9 \end{vmatrix} + \begin{vmatrix} 2 & 2 \\ 4 & 9 \end{vmatrix} = 1 + 10 = 11$$

and

$$\begin{aligned} |\mathbf{A}'| = \begin{vmatrix} 1 & 2 & 4 \\ -1 & 1 & 4 \\ 0 & 2 & 9 \end{vmatrix} &= \begin{vmatrix} 1 & 4 \\ 2 & 9 \end{vmatrix} - 2\begin{vmatrix} -1 & 4 \\ 0 & 9 \end{vmatrix} + 4\begin{vmatrix} -1 & 1 \\ 0 & 2 \end{vmatrix} \\ &= 1 + 18 - 8 = 11. \end{aligned}$$

Proof. The formal definition of Section 3.2 and its equivalence to expansion by elements of a row provide the means for proving equality of that expansion with expansion by elements of a column. Therefore expansion of $|\mathbf{A}|$ by elements of its *i*th row is the same as expansion of $|\mathbf{A}'|$ by elements of its *i*th column, except that in $|\mathbf{A}|$ all minors will be of matrices that are the transpose of those in the corresponding minors in $|\mathbf{A}'|$. This will be true right down to the minors of the 2×2 matrices in each case. But these are equal, for example,

$$\begin{vmatrix} a & b \\ x & y \end{vmatrix} = ay - bx = \begin{vmatrix} a & x \\ b & y \end{vmatrix}.$$

Therefore, $|\mathbf{A}| = |\mathbf{A}'|$ ■

Corollary 3.1 *All properties of* $|\mathbf{A}|$ *in terms of rows can be stated equivalently in terms of columns—because expanding* $|\mathbf{A}'|$ *by rows is identical to expanding* $|\mathbf{A}|$ *by columns.*

3.3.2 Two Rows the Same

Theorem 3.2 *If two rows of* $\mathbf{A}$ *are the same,* $|\mathbf{A}| = 0$.

Example 3.3

$$\begin{vmatrix} 1 & 4 & 3 \\ 7 & 5 & 2 \\ 7 & 5 & 2 \end{vmatrix} = \begin{vmatrix} 5 & 2 \\ 5 & 2 \end{vmatrix} - 4\begin{vmatrix} 7 & 2 \\ 7 & 2 \end{vmatrix} + 3\begin{vmatrix} 7 & 5 \\ 7 & 5 \end{vmatrix} = 0,$$

because

$$\begin{vmatrix} a & b \\ a & b \end{vmatrix} = ab - ab = 0. \tag{3.3}$$

Proof. If **A** has two rows, the same, expand |**A**| by minors so that the 2 × 2 minors in the last step of the expansion are from two equal rows. Then (3.3) shows that all those minors are zero, and so |**A**| = 0.

3.3.3 Cofactors

The signed minor $(-1)^{i+j}|\mathbf{M}_{ij}|$ used in formulas (3.1) and (3.2) is called a cofactor:

$$c_{ij} = (-1)^{i+j}|\mathbf{M}_{ij}|, \tag{3.4}$$

where $\mathbf{M}_{ij}$ is **A** with its *i*th row and *j*th column deleted.

Two properties of cofactors are important. The first is that the sum of products of the elements of a row with their own cofactors is the determinant:

$$\sum_j a_{ij}c_{ij} = \sum_j a_{ij}(-1)^{i+j}|\mathbf{M}_{ij}| = |\mathbf{A}| \quad \text{for each } i. \tag{3.5}$$

This is just formula (3.1). Second, the sum of products of elements of a row with cofactors of some other row is zero:

$$\sum_j a_{ij}c_{kj} = 0 \quad \text{for} \quad i \neq k. \tag{3.6}$$

This is so because, for $i \neq k$,

$$\sum_j a_{ij}c_{kj} = \sum_j a_{ij}(-1)^{k+j}|\mathbf{M}_{kj}| = 0$$

with the last equality being true because it represents, by comparison with (3.5), expansion of |**A**| with its *k*th row replaced by its *i*th row and so is a determinant having two rows the same and hence is zero. Formulas (3.5) and (3.6) can also be restated in terms of columns of **A**.

Example 3.4 *In the determinant*

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 10 \end{vmatrix}$$

the cofactors of the elements of the first row are as follows:

$$\text{that of the 1 is}\quad (-1)^{1+1}\begin{vmatrix}5 & 6\\ 8 & 10\end{vmatrix} = 50-48 = 2;$$

$$\text{that of the 2 is}\quad (-1)^{1+2}\begin{vmatrix}4 & 6\\ 7 & 10\end{vmatrix} = -40+42 = 2;$$

$$\text{that of the 3 is}\quad (-1)^{1+3}\begin{vmatrix}4 & 5\\ 7 & 8\end{vmatrix} = 32-35 = -3.$$

Multiplying these elements of the first row by their cofactors gives $|\mathbf{A}|$:

$$|\mathbf{A}| = 1(2)+2(2)+3(-3) = -3.$$

But multiplying the elements of another row, the second say, by these cofactors gives zero:

$$4(2)+5(2)+6(-3) = 0.$$

The determinantal form of this last expression is

$$4(-1)^{1+1}\begin{vmatrix}5 & 6\\ 8 & 10\end{vmatrix} + 5(-1)^{1+2}\begin{vmatrix}4 & 6\\ 7 & 10\end{vmatrix} + 6(-1)^{1+3}\begin{vmatrix}4 & 5\\ 7 & 8\end{vmatrix} = \begin{vmatrix}4 & 5 & 6\\ 4 & 5 & 6\\ 7 & 8 & 10\end{vmatrix},$$

which is clearly zero because two rows are the same.

3.3.4 Adding Multiples of a Row (Column) to a Row (Column)

Theorem 3.3 *Adding to one row (column) of a determinant any multiple of another row (column) does not affect the value of the determinant.*

Example 3.5

$$\begin{aligned}|\mathbf{A}| &= \begin{vmatrix}1 & 3 & 2\\ 8 & 17 & 21\\ 2 & 7 & 1\end{vmatrix}\\ &= 1\,(17-147) - 3\,(8-42) + 2\,(56-34) = 16.\end{aligned}$$

And adding four times row 1 to row 2 does not affect the value of $|\mathbf{A}|$:

$$\begin{aligned}|\mathbf{A}| &= \begin{vmatrix}1 & 3 & 2\\ 8+4 & 17+12 & 21+8\\ 2 & 7 & 1\end{vmatrix} = \begin{vmatrix}1 & 3 & 2\\ 12 & 29 & 29\\ 2 & 7 & 1\end{vmatrix}\\ &= 1\,(29-203) - 3\,(12-58) + 2\,(84-58)\\ &= -174+138+52 = 16.\end{aligned} \tag{3.7}$$

Proof. Suppose $\mathbf{B}$ of order $n \times n$ has $[b_{11}\ b_{12} \cdots b_{1n}]$ and $[b_{21}\ b_{22} \cdots b_{2n}]$ as its first two rows. Let $\mathbf{A}$ be $\mathbf{B}$ with λ times its second row added to its first. We show that $|\mathbf{A}| = |\mathbf{B}|$.

With $|\mathbf{M}_{ij}|$ being the minor of b_{ij} in $|\mathbf{B}|$, expansion of $|\mathbf{A}|$ by elements of its first row, using formula (3.1), gives

$$|\mathbf{A}| = \sum_{j=1}^{n}(b_{1j} + \lambda b_{2j})(-1)^{1+j}|\mathbf{M}_{1j}|$$
$$= \sum_{j=1}^{n} b_{1j}(-1)^{1+j}|\mathbf{M}_{1j}| + \lambda \sum_{j=1}^{n} b_{2j}(-1)^{1+j}|\mathbf{M}_{1j}|.$$

The first sum here is $|\mathbf{B}|$, directly from (3.1). The second sum is an example of (3.6) and hence is zero. Therefore $|\mathbf{A}| = |\mathbf{B}|$

3.3.5 Products

Theorem 3.4 $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$ *when **A** and **B** are square and of the same order.*

Example 3.6 *With*

$$\mathbf{AB} = \begin{bmatrix} 1 & 2 \\ 3 & 8 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 6 & 6 \end{bmatrix} = \begin{bmatrix} 16 & 15 \\ 60 & 57 \end{bmatrix}, |\mathbf{AB}| = 912 - 900 = 12,$$

and

$$|\mathbf{A}||\mathbf{B}| = \begin{vmatrix} 1 & 2 \\ 3 & 8 \end{vmatrix} \begin{vmatrix} 4 & 3 \\ 6 & 6 \end{vmatrix} = 2(6) = 12.$$

Before proving the theorem, we first show that any determinant can be reduced to the determinant of a triangular matrix, and then consider two useful lemmas.

Reduction to Triangular Form

The determinant of a lower triangular matrix is the product of its diagonal elements. For example,

$$\begin{vmatrix} 6 & 0 & 0 \\ 3 & -1 & 0 \\ 7 & 3 & -5 \end{vmatrix} = 6(-1)(-5) = 30.$$

Verification is self-evident—through expansion by elements of successive rows.

Any determinant can be evaluated as the determinant of a lower triangular matrix. This is done by adding multiples of rows to other rows to reduce the determinant to triangular form. For example, consider

$$|\mathbf{P}| = \begin{vmatrix} 3 & 8 & 7 \\ 1 & 2 & 4 \\ -1 & 3 & 2 \end{vmatrix}.$$

We use row operations to reduce all elements above the diagonal to zero. Start first with the 2 of the (3, 3) element to reduce elements above it to zero; that is, to row 1 add (−7/2) times row 3 and to row 2 add (− 2) times row 3. This gives

$$|\mathbf{P}| = \begin{vmatrix} 6\frac{1}{2} & -2\frac{1}{2} & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix}.$$

Now use the (2, 2) element to reduce the elements above *it* to zero, by subtracting $-2\frac{1}{2}/(-4)$ times row 2 from row 1. Then

$$|\mathbf{P}| = \begin{vmatrix} 6\frac{1}{2} - 2\frac{1}{2}(3)/4 & 0 & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix} = \begin{vmatrix} 37/8 & 0 & 0 \\ 3 & -4 & 0 \\ -1 & 3 & 2 \end{vmatrix},$$

from which $|\mathbf{P}| = (37/8)(-4)(2) = -37$, as can be verified by any other form of expansion.

The method used in the example extends quite naturally to a determinant of any order n. Starting with the nth diagonal element, the elements above it are reduced to zero, then so are those above the $(n-1)$th,$(n-2)$th ,…,3rd, and 2nd diagonal elements. This yields a lower triangular matrix. (Similar calculations starting with the first diagonal element and using it to reduce all elements below it to zero, followed by doing the same to those below the 2nd, 3rd,…,$(n-1)$th, and nth diagonal elements yields an upper triangular matrix.) Then the determinant is the product of the resultant diagonal elements.

Two Useful Lemmas

Lemma 3.1 *For square matrices* **P**, **Q**, *and* ***X*** *of the same order n*

$$\begin{vmatrix} \mathbf{P} & \mathbf{0} \\ \mathbf{X} & \mathbf{Q} \end{vmatrix} = |\mathbf{P}||\mathbf{Q}|. \tag{3.8}$$

Proof. By row operations on the first n rows of the left-hand side of (3.8) reduce **P** to lower triangular form. Then expand the left-hand side by elements of successive rows for the first n rows. The result is $|\mathbf{P}||\mathbf{Q}|$, and so (3.8) is upheld.

Example 3.7 *Using* **P** *of the preceding example and*

$$\mathbf{Q} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix},$$

we have

$$\begin{vmatrix} 3 & 8 & 7 & 0 & 0 & 0 \\ 1 & 2 & 4 & 0 & 0 & 0 \\ -1 & 3 & 2 & 0 & 0 & 0 \\ x & y & z & 1 & 2 & 3 \\ p & q & r & 4 & 5 & 4 \\ a & b & c & 3 & 2 & 1 \end{vmatrix} = \begin{vmatrix} 37/8 & 0 & 0 & 0 & 0 & 0 \\ 3 & -4 & 0 & 0 & 0 & 0 \\ -1 & 3 & 2 & 0 & 0 & 0 \\ x & y & z & 1 & 2 & 3 \\ p & q & r & 4 & 5 & 4 \\ a & b & c & 3 & 2 & 1 \end{vmatrix} = |\mathbf{P}|\,|\mathbf{Q}|$$

$$= -37 \begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 4 \\ 3 & 2 & 1 \end{vmatrix}.$$

Lemma 3.2 *For* **R** *and* **S** *square and of the same order n*

$$\begin{vmatrix} \mathbf{0} & \mathbf{R} \\ -\mathbf{I} & \mathbf{S} \end{vmatrix} = |\mathbf{R}|. \tag{3.9}$$

Proof. Expand the left-hand side of (3.9) by elements of the successive columns through the $-\mathbf{I}$ to get

$$\begin{vmatrix} \mathbf{0} & \mathbf{R} \\ -\mathbf{I} & \mathbf{S} \end{vmatrix} = \left[(-1)^{n+1+1}(-1)\right]^n |\mathbf{R}| = (-1)^{n(n+3)}|\mathbf{R}| = |\mathbf{R}|.$$

Determinant of a Product

Consider the determinant of the matrix product

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix},$$

namely

$$\left| \begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} \right| = \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix}. \tag{3.10}$$

Note that using

$$\begin{bmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}$$

in premultiplication in (3.10) merely adds multiples of rows of

$$\left| \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{bmatrix} \right|$$

to other rows and so

$$\text{L. H. S. of (3.10)} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{B}|, \text{ by (3.8).}$$

But

$$\text{R. H. S. of (3.10)} = \begin{vmatrix} \mathbf{0} & \mathbf{AB} \\ -\mathbf{I} & \mathbf{B} \end{vmatrix} = |\mathbf{AB}|, \text{ by (3.9).}$$

Therefore $|\mathbf{A}||\mathbf{B}| = |\mathbf{AB}|$ or

$$|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|. \tag{3.11}$$

Note that for all three determinants in (3.11) to exist, **A** and **B** must both be square, of the same order. This is also implied by (3.10). The result (3.11) does not hold when **A** or **B** or both are rectangular; even if $|\mathbf{AB}|$ exists in the form $|\mathbf{A}_{\mathbf{r} \times \mathbf{c}}\mathbf{B}_{\mathbf{c} \times \mathbf{r}}|$, $|\mathbf{A}|$ and $|\mathbf{B}|$ do not and (3.11) does not apply.

There are several useful corollaries to this result.

Corollary 3.2

1. $|\mathbf{AB}| = |\mathbf{BA}|$*(because* $|\mathbf{A}||\mathbf{B}| = |\mathbf{B}||\mathbf{A}|$).
2. $|\mathbf{A}^2| = |\mathbf{A}|^2$ *(each equals* $|\mathbf{A}||\mathbf{A}|$); $|\mathbf{A}^k| = |\mathbf{A}|^k$ *is the extension.*
3. *For orthogonal* $\mathbf{A}$, $|\mathbf{A}| = \pm 1$ *(because* $\mathbf{AA}' = \mathbf{I}$ *implies* $|\mathbf{A}|^2 = 1$).
4. *For idempotent* $\mathbf{A}$, $|\mathbf{A}| = 0$ *or 1 (because* $\mathbf{A}^2 = \mathbf{A}$ *implies* $|\mathbf{A}|^2 = |\mathbf{A}|$).

3.4 Elementary Row Operations

In Section 3.3.4 we introduced the operation of adding a multiple of a row (column) of a determinant to another row (column). It did not alter the value of the determinant. That same operation on a matrix can be represented by a matrix product, for example,

$$\begin{bmatrix} 1 & 3 & 2 \\ 12 & 29 & 29 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} \tag{3.12}$$

is the matrix product representation of adding 4 times row 1 to row 2, as used in (3.7). Applying the product rule of Section 3.3.5 gives

$$\begin{vmatrix} 1 & 3 & 2 \\ 12 & 29 & 29 \\ 2 & 7 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{vmatrix}$$

because

$$\begin{vmatrix} 1 & 0 & 0 \\ 4 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1. \tag{3.13}$$

The matrix involved in (3.12) and (3.13)) is known as an *elementary operator matrix*. For the operation of adding λ times one row of a matrix to another, it is always an identity matrix with λ in an off-diagonal element; and its determinant is always unity. We denote such a matrix by $\mathbf{P}_{ij}(\lambda)$. Then $\mathbf{P}_{ij}(\lambda)\mathbf{A}$ is the operation on **A** of adding to its ith row λ times its jth row; and

$$\left|\mathbf{P}_{ij}(\lambda)\right| = 1; \quad \text{for example in (3.13)}, \quad \left|\mathbf{P}_{21}(4)\right| = 1. \tag{3.14}$$

Two other elementary operations are (i) interchanging two rows of a matrix, and (ii) multiplying a tow by a scalar. These are represented by matrices

$$\mathbf{E}_{ij} = \mathbf{I} \text{ with } i\text{th and } j\text{th rows interchanged}$$

and

$$\mathbf{R}_{ii}(\lambda) = \mathbf{I} \text{ with } i\text{th diagonal element replaced by } \lambda.$$

Then $\mathbf{E}_{ij}\mathbf{A}$ is **A** with its ith and jth rows interchanged, and $\mathbf{R}_{ii}(\lambda)\mathbf{A}$ is **A** with its ith row multiplied by λ

Example 3.8

$$\mathbf{E}_{12}\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 8 & 17 & 21 \\ 1 & 3 & 2 \\ 2 & 7 & 1 \end{bmatrix}$$

and

$$\mathbf{R}_{33}(5)\mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 2 & 7 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 2 \\ 8 & 17 & 21 \\ 10 & 35 & 5 \end{bmatrix}.$$

All three of these elementary operator matrices have simple determinants:

$$|\mathbf{P}_{ij}(\lambda)| = 1, \quad |\mathbf{R}_{ii}(\lambda)| = \lambda \quad \text{and} \quad |\mathbf{E}_{ij}| = -1. \tag{3.15}$$

Special cases of these elementary operator matrices used in combination with the product rule $|\boldsymbol{AB}| = |\boldsymbol{A}||\boldsymbol{B}|$ *provide useful techniques for simplifying the expansion of a determinant.*

3.4.1 Factorization

Theorem 3.5 *When* λ *(a nonzero scalar) is a factor of a row (column) of* $|\mathbf{A}|$ *then it is also a factor of* $|\mathbf{A}|$:

$$|\mathbf{A}| = \lambda |\,\mathbf{A} \text{ with } \lambda \text{ factored out of a row (column)}|. \tag{3.16}$$

Example 3.9

$$\begin{vmatrix} 4 & 6 \\ 1 & 7 \end{vmatrix} = 2\begin{vmatrix} 2 & 3 \\ 1 & 7 \end{vmatrix}, \qquad \textit{that is}, 4(7) - 1(6) = 22 = 2[2(7) - 1(3)].$$

Proof. When λ is a factor of every element in the ith row of $\mathbf{A}$,

$$\mathbf{A} = \mathbf{R}_{ii}(\lambda)[\mathbf{A} \text{ with } \lambda \text{ factored out of its } i\text{th row}]. \tag{3.17}$$

In the determinant of (3.17), use (3.11) and (3.15) to get (3.16).

Corollary 3.3 *If* $\boldsymbol{A}$ *is an* $n \times n$ *matrix and* λ *is a scalar, the determinant of the matrix* $\lambda\boldsymbol{A}$ *is* $\lambda^n|\boldsymbol{A}|$*; that is,* $|\lambda\boldsymbol{A}| = \lambda^n|\boldsymbol{A}|$*. (In this case* λ *is a factor of each of the n rows of* $\lambda\boldsymbol{A}$*, so that after factoring* λ *from each row the determinant* $|\boldsymbol{A}|$ *remains.)*

Example 3.10

$$\begin{bmatrix} 3 & 0 & 27 \\ -9 & 3 & 0 \\ 15 & 6 & -3 \end{bmatrix} = 3^3 \begin{bmatrix} 1 & 0 & 9 \\ -3 & 1 & 0 \\ 5 & 2 & -1 \end{bmatrix} = -2700.$$

Corollary 3.4 *If one row of a determinant is a multiple of another row, the determinant is zero. (Factoring out the multiple reduces the determinant to having two rows the same. Hence the determinant is zero.)*

Example 3.11

$$\begin{bmatrix} -3 & 6 & 12 \\ 2 & -4 & -8 \\ 7 & 5 & 9 \end{bmatrix} = -(1.5)\begin{bmatrix} 2 & -4 & -8 \\ 2 & -4 & -8 \\ 7 & 5 & 9 \end{bmatrix} = 0.$$

3.4.2 A Row (Column) of Zeros

Theorem 3.6 *When a determinant has zero for every element of a row (or column), the determinant is zero.*

Example 3.12

$$\begin{vmatrix} 0 & 0 & 0 \\ 3 & 6 & 5 \\ 2 & 9 & 7 \end{vmatrix} = 0\begin{vmatrix} 6 & 5 \\ 9 & 7 \end{vmatrix} - 0\begin{vmatrix} 3 & 5 \\ 2 & 7 \end{vmatrix} + 0\begin{vmatrix} 3 & 6 \\ 2 & 9 \end{vmatrix} = 0.$$

Proof. This is just (3.16) with $\lambda = 0$.

3.4.3 Interchanging Rows (Columns)

Theorem 3.7 *Interchanging two rows (columns) of a determinant changes its sign.*

Example 3.13 *Subtracting twice column 2 from column 1,*

$$\begin{bmatrix} 6 & 3 & 0 \\ -1 & 4 & 7 \\ 2 & 5 & 9 \end{bmatrix} = \begin{bmatrix} 0 & 3 & 0 \\ -9 & 4 & 7 \\ -8 & 5 & 9 \end{bmatrix} = -3(-81+56) = 75,$$

whereas

$$\begin{bmatrix} 2 & 5 & 9 \\ -1 & 4 & 7 \\ 6 & 3 & 0 \end{bmatrix} = \begin{bmatrix} -8 & 5 & 9 \\ -9 & 4 & 7 \\ 0 & 3 & 0 \end{bmatrix} = -3(-56+81) = -75.$$

Proof. Using (3.11) and (3.15)

$$|\mathbf{E}_{ij}\mathbf{A}| = |\mathbf{E}_{ij}||\mathbf{A}| = -|\mathbf{A}|.$$

3.4.4 Adding a Row to a Multiple of a Row

In Section 3.3.4 we added a multiple of a row to a row. The reader is cautioned against doing the reverse: adding a row to a multiple of a row is not the same thing. It leads to a different result. For example, instead of adding λ times row 2 to row 1 of |**A**|, and *not* altering |**A**|, the operation of adding row 2 to λ times row 1 gives λ|**A**|, for example,

$$\begin{vmatrix} 1 & \lambda \\ 0 & 1 \end{vmatrix}\begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2+5\lambda & 3+9\lambda \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = 3;$$

the operation which gives $\lambda|\mathbf{A}|$ is

$$\begin{vmatrix} \lambda & 1 \\ 0 & 1 \end{vmatrix} \begin{vmatrix} 2 & 3 \\ 5 & 9 \end{vmatrix} = \begin{vmatrix} 2\lambda + 5 & 3\lambda + 9 \\ 5 & 9 \end{vmatrix} = 3\lambda.$$

3.5 Examples

The foregoing properties can be applied in endless variation in expanding determinants. Efficiency in perceiving a procedure that leads to a minimal amount of effort in any particular case is largely a matter of practice, and beyond describing the possible steps available there is little more that can be said. The underlying method might be summarized as follows. By adding multiples of one row to other rows of the determinant results in reducing a column to having only one nonzero element. Expansion by elements of that column then involves only one minor, which is a determinant of order one less than the original determinant. Successive applications of this method reduce the determinant to one of order 2×2 whose expansion is obvious. If at any stage these reductions result in elements of a row containing a common factor, this can be factored out, and if they result in a row of zeros, or in two rows being identical, the determinant is zero.

Example 3.14

$$|\mathbf{A}| = \begin{vmatrix} 1 & 4 & 9 \\ -4 & 7 & 25 \\ 7 & 5 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 4 & 9 \\ -4+4(1) & 7+4(4) & 25+4(9) \\ 7-7(1) & 5-7(4) & 2-7(9) \end{vmatrix}$$

$$= \begin{vmatrix} 1 & 4 & 9 \\ 0 & 23 & 61 \\ 0 & -23 & -61 \end{vmatrix} = 0.$$

The 1 in the leading element of ***A*** *makes it easy to reduce the other elements of the first column to zero, whereupon adding the last two rows gives* $|\mathbf{A}| = 0$.

Example 3.15

$$|\mathbf{B}| = \begin{vmatrix} x & y & y & y \\ y & x & y & y \\ y & y & x & y \\ y & y & y & x \end{vmatrix}.$$

Observe that every column of ***B*** *sums to* $x + 3y$. *Therefore add every row* of ***B*** *to its first row and factor out* $x + 3y$:

$$|\mathbf{B}| = (x+3y)\begin{vmatrix} 1 & 1 & 1 & 1 \\ y & x & y & y \\ y & y & x & y \\ y & y & y & x \end{vmatrix}.$$

Now use the leading element of ***B*** *to reduce the remainder of the first row to zeros. This is achieved by subtracting the first column from each of the other columns:*

$$|\mathbf{B}| = (x+3y)\begin{vmatrix} 1 & 0 & 0 & 0 \\ y & x-y & 0 & 0 \\ y & 0 & x-y & 0 \\ y & 0 & 0 & x-y \end{vmatrix} = (x+3y)(x-y)^3,$$

the last step being obtained by simple expansion by minors. It involves the more general result that the determinant of a diagonal matrix is the product of its diagonal elements:

$$\left|\mathbf{D}\{d_i\}\right| = \begin{vmatrix} d_1 & 0 & 0 & & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ 0 & 0 & & & \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & \cdots & d_n \end{vmatrix} = \prod_{i=1}^{n} d_i = d_1 d_2 \cdots d_n.$$

Example 3.16

$$|\mathbf{C}| = \begin{vmatrix} 6 & 8 & 1 & 4 & 2 \\ 18 & 27 & 3 & 13 & 5 \\ -11 & -17 & -2 & -9 & 2 \\ 7 & 10 & 1 & 4 & 7 \\ 4 & 3 & 13 & 8 & 5 \end{vmatrix} = \begin{vmatrix} 6 & 8 & 1 & 4 & 2 \\ 0 & 3 & 0 & 1 & -1 \\ 1 & -1 & 0 & -1 & 6 \\ 1 & 2 & 0 & 0 & 5 \\ -74 & -101 & 0 & -44 & -21 \end{vmatrix}.$$

The occurrence of a 1 as the (1, 3) element of **C** *prompts using it to reduce all other elements in column 3 to zero. Doing so, expansion by elements of the third column gives*

$$|\mathbf{C}| = (-1)^{1+3}\begin{vmatrix} 0 & 3 & 1 & -1 \\ 1 & -1 & -1 & 6 \\ 1 & 2 & 0 & 5 \\ -74 & -101 & -44 & -21 \end{vmatrix}$$

$$= \begin{vmatrix} 0 & 0 & 0 & -1 \\ 1 & 17 & 5 & 6 \\ 1 & 17 & 5 & 5 \\ -74 & -164 & -65 & -21 \end{vmatrix} = 0.$$

The (− 1) of the (1, 4) element has been used here to reduce other elements of the first row to zero; then expansion by elements of the first row gives a value of zero because the only nonzero element is associated with a 3 × 3 minor having two rows the same.

*Alternatively, in the original form of |**C**|, we can observe that row 2 plus row 3 equals row 4. Hence |**C**| = 0. This is obviously easier—but this observation can be used only when it presents itself.*

Example 3.17

$$\begin{vmatrix} y & 7 & 7 & 7 \\ 7 & y & 7 & 7 \\ 7 & 7 & y & 7 \\ 7 & 7 & 7 & y \end{vmatrix} = 0 \text{ for } y = 7 \text{ or} -21$$

because for $y = 7$ all rows are the same, and for $y = -21$ the sum of all rows is a row of zeros. This determinant is a special case of Example 3.15.

Example 3.18

$$\begin{vmatrix} x & x & x \\ 4 & 3 & -9 \\ -3 & -2 & 10 \end{vmatrix} = 0 \quad \text{for any } x,$$

because the first row, after factoring out x, equals the sum of the other two rows.

3.6 Diagonal Expansion

A matrix can always be expressed as the sum of two matrices one of which is a diagonal matrix, that is, as $(\mathbf{A} + \mathbf{D})$ where $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, 2, \ldots, n$, and $\mathbf{D}$ is a diagonal matrix of order $n \times n$. The determinant of such a matrix can then be obtained as a polynomial of the elements of $\mathbf{D}$.

The minor of an element of a square matrix of order $n \times n$ is necessarily a determinant of order $n - 1$. But minors are not all of order $n - 1$. Deleting any r rows and any r columns from a square matrix of order $n \times n$ leaves a submatrix of order $(n - r) \times (n - r)$; the determinant of this submatrix is a *minor of order* $n - r$, or an $(n - r)$*-order minor*. It is useful to introduce an abbreviated notation for minors of the determinant of $\mathbf{A}$,

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{vmatrix}.$$

They will be denoted by just their diagonal elements; for example, $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$ is written as $\begin{vmatrix} a_{11} & a_{22} \end{vmatrix}$ and in similar fashion $\begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix}$ is written as $\begin{vmatrix} a_{12} & a_{23} \end{vmatrix}$. Combined with the notation $\mathbf{A} = \{a_{ij}\}$ no confusion can arise. For example, $\begin{vmatrix} a_{21} & a_{32} \end{vmatrix}$ denotes the 2×2 minor having a_{21} and a_{32} as diagonal elements, and from $|\mathbf{A}|$ we see that the elements in the same rows and columns as these are a_{22} and a_{31}, so that

$$\begin{vmatrix} a_{21} & a_{32} \end{vmatrix} = \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}.$$

Similarly,

$$\begin{vmatrix} a_{21} & a_{33} & a_{44} \end{vmatrix} = \begin{vmatrix} a_{21} & a_{23} & a_{24} \\ a_{31} & a_{33} & a_{34} \\ a_{41} & a_{43} & a_{44} \end{vmatrix}.$$

We will now consider the determinant $|\mathbf{A} + \mathbf{D}|$, initially for a 2×2 case, denoting the diagonal elements of $\mathbf{D}$ by x_1 and x_2; for example,

$$|\mathbf{A} + \mathbf{D}| = \begin{vmatrix} a_{11} + x_1 & a_{12} \\ a_{21} & a_{22} + x_2 \end{vmatrix}.$$

By direct expansion

$$|\mathbf{A} + \mathbf{D}| = (a_{11} + x_1)(a_{22} + x_2) - a_{12}a_{21}.$$

Written as a function of x_1 and x_2 this is

$$x_1x_2 + x_1a_{22} + x_2a_{11} + \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

In similar fashion it can be shown that

$$\begin{aligned}
&\begin{vmatrix} a_{11}+x_1 & a_{12} & a_{13} \\ a_{21} & a_{22}+x_2 & a_{23} \\ a_{31} & a_{32} & a_{33}+x_3 \end{vmatrix} \\
&\quad = x_1x_2x_3 + x_1x_2a_{33} + x_1x_3a_{22} + x_2x_3a_{11} \\
&\quad + x_1\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} + x_2\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} + x_3\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}
\end{aligned}$$

which, using the abbreviated notation, can be written as

$$\begin{aligned}
&x_1x_2x_3 + x_1x_2a_{33} + x_1x_3a_{22} + x_2x_3a_{11} \\
&\qquad + x_1\begin{vmatrix} a_{22} & a_{33} \end{vmatrix} + x_2\begin{vmatrix} a_{11} & a_{33} \end{vmatrix} + x_3\begin{vmatrix} a_{11} & a_{22} \end{vmatrix} \\
&\qquad + \begin{vmatrix} a_{11} & a_{22} & a_{33} \end{vmatrix}.
\end{aligned} \tag{3.18}$$

Considered as a polynomial in the x's we see that the coefficient of the product of all the x's is unity; the coefficients of the second-degree terms in the x's are the diagonal elements of **A**; the coefficients of the first-degree terms in the x's are the second-order minors of |**A**| having diagonals that are part (or all) of the diagonal of |**A**|; and the term independent of the x's is |**A**| itself. The minors of |**A**| in these coefficients, namely those whose diagonals are coincident with the diagonal of |**A**|, are called the *principal minors* of |**A**|.

This method of expansion is useful on many occasions because the determinantal form |**A** + **D**| occurs quite often, and when |**A**| is such that many of its principal minors are zero the expansion of |**A** + **D**| by this method is greatly simplified.

Another way to introduce some of the above items is given by the following definition:

Definition 3.1 *If* **A** *is a square matrix of order* $n \times n$*, and if rows* $i_1, i_2, \ldots, i_r$ *and columns* $i_1, i_2, \ldots, i_r$ $(r < n)$ *are deleted from* **A**, *then the resulting submatrix is called a principal submatrix of* **A** *(that is, this submatrix is obtained by deleting r rows and the same r columns from* **A***). The determinant of a principal submatrix is called a principal minor.*

If the deleted rows and columns are the last r rows and the last r columns, respectively, then the resulting submatrix is called a leading principal submatrix of **A**. *Its determinant is called a leading principal minor.*

Example 3.19 *If*

$$|\mathbf{X}| = \begin{vmatrix} 7 & 2 & 2 \\ 2 & 8 & 2 \\ 2 & 2 & 9 \end{vmatrix},$$

we have

$$|\mathbf{X}| = |\mathbf{A}+\mathbf{D}| = \left| \begin{bmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{bmatrix} + \begin{bmatrix} 5 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 7 \end{bmatrix} \right|.$$

Every element of **A** is a 2 so that |**A**| and all its 2 × 2 minors are zero. Consequently |**X**| evaluated by (3.18) consists of only the first four terms:

$$|\mathbf{X}| = 5(6)7 + 5(6)2 + 5(7)2 + 6(7)2 = 424.$$

Evaluating a determinant in this manner is also useful when all elements of the diagonal matrix **D** are the same, that is, when the x_i's are equal. The expansion (3.7) then becomes

$$x^3 + x^2(a_{11} + a_{22} + a_{33}) + x(|a_{11} \quad a_{22}| + |a_{11}a_{33}| + |a_{22}a_{33}|) + |\mathbf{A}|,$$

which is generally written as

$$x^3 + x^2\mathrm{tr}_1(\mathbf{A}) + x\mathrm{tr}_2(\mathbf{A}) + |\mathbf{A}|$$

where $\mathrm{tr}_1(\mathbf{A})$ is the trace of **A** (sum of diagonal elements) and tr_2(**A**) is the sum of the principal minors of order 2 of |**A**|. This method of expansion is known as expansion by diagonal elements or simply as diagonal expansion.

The general diagonal expansion of a determinant of order n,

$$|\mathbf{A}+\mathbf{D}| = \begin{vmatrix} a_{11}+x_1 & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22}+x_2 & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33}+x_3 & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn}+x_n \end{vmatrix},$$

consists of the sum of all possible products of the x_i taken r at a time for r = n, n − 1, …, 2, 1, 0, each product being multiplied by its complementary principal minor of order n − r in |**A**|. By complementary principal minor in |**A**| is meant the principal minor having diagonal elements other than those associated in |**A** + **D**| with the x's of the particular product concerned; for example, the complementary principal minor associated with $x_1x_3x_6$ is $|a_{22}a_{44}a_{55}a_{77}a_{88}\cdots a_{nn}|$. When the x's are all equal, the expression becomes $\sum_{i=0}^{n} x^{n-i}\mathrm{tr}_i(\mathbf{A})$ where $\mathrm{tr}_i(\mathbf{A})$ is the sum of the principal minors of order i of |**A**| and, by definition, tr_0(**A**) = 1. Note in passing that tr_n(**A**) = |**A**|.

Example 3.20 Diagonal expansion gives

$$
|\mathbf{A}+\mathbf{D}| = \begin{vmatrix} a+b & a & a & a \\ a & a+b & a & a \\ a & a & a+b & a \\ a & a & a & a+b \end{vmatrix}
$$

$$
= b^4 + b^3 \mathrm{tr}_1(\mathbf{A}) + b^2 \mathrm{tr}_2(\mathbf{A}) + b\,\mathrm{tr}_3(\mathbf{A}) + |\mathbf{A}|
$$

where $\mathbf{A}$ *is the* 4×4 *matrix whose every element is* a*. Thus* $|\mathbf{A}|$ *and all minors of order 2 or more are zero. Hence*

$$
|\mathbf{A}+\mathbf{D}| = b^4 + b^3(4a) = (4a+b)b^3.
$$

This is an extension of the example in Section 3.4.

3.7 The Laplace Expansion

In the expansion of

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{vmatrix}$$

the minor of a_{11} is $|a_{22}a_{33}a_{44}|$. An extension of this, easily verified, is that the coefficient of $|a_{11} \quad a_{22}|$ is $|a_{33} \quad a_{44}|$; namely, the coefficient in $|\mathbf{A}|$ of

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21} \text{ is } \begin{vmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{vmatrix} = a_{33}a_{44} - a_{34}a_{43}.$$

Likewise the coefficient of $|a_{11} \quad a_{24}|$ is $|a_{32} \quad a_{43}|$: that is, the coefficient of

$$|a_{11} \quad a_{24}| = \begin{vmatrix} a_{11} & a_{14} \\ a_{21} & a_{24} \end{vmatrix} = a_{11}a_{24} - a_{21}a_{14}$$

in the expansion of $|\mathbf{A}|$ is

$$|a_{32} \quad a_{43}| = \begin{vmatrix} a_{32} & a_{33} \\ a_{42} & a_{43} \end{vmatrix} = a_{32}a_{43} - a_{33}a_{42}.$$

Each determinant just described as the coefficient of a particular minor of $|\mathbf{A}|$ is the complementary minor in $|\mathbf{A}|$ of that particular minor: it is the determinant obtained from $|\mathbf{A}|$ by deleting from it all the rows and columns containing the particular minor. This is simply an extension of the procedure for finding the coefficient of an individual element in $|\mathbf{A}|$ as derived in the expansion by elements of a row or column discussed earlier. In that case, the particular minor is a single element and its coefficient in $|\mathbf{A}|$ is $|\mathbf{A}|$ amended by deletion of the row and column containing the element concerned. A sign factor is also involved, namely $(-1)^{i+j}$ for the coefficient of a_{ij} in $|\mathbf{A}|$. In the extension to coefficients of minors, the sign factor is minus one raised to the power of the sum of the subscripts of the diagonal elements of the chosen minor: for example, the sign factor for the coefficient of $|a_{32} \quad a_{43}|$ is $(-1)^{3+2+4+3} = -1$, as just given. The complementary minor multiplied by this sign factor can be appropriately referred to as the coefficient of the particular minor concerned. Furthermore, just as the expansion of a determinant is the sum of products of elements of a row (or column) with their coefficients, so also is the sum of products of all minors of order m that can be derived from any set of m rows, each multiplied by its coefficient as just defined. This generalization of the method of expanding a determinant by elements of a row to expanding it by minors of a set of rows was first established by Laplace and so bears his name. Aitken (1948) and Ferrar(1941) are two books where proof of the procedure is given; we shall be satisfied here with a general statement of the method and an example illustrating its use.

The Laplace expansion of a determinant |**A**| of order *n* can be obtained as follows. (i) Consider any *m* rows of |**A**|. They contain $n!/[m!(n-m)!]$ minors of order *m* (see footnote, Section 3.1.) (ii) Multiply each of these minors, **M** say, by its complementary minor and by a sign factor, where (a) the complementary minor of **M** is the $n-m$ order minor derived from |**A**| by deleting the *m* rows and columns containing **M**, and (b) the sign factor is $(-1)^{\mu}$ where μ is the sum of the subscripts of the diagonal elements of **M,** **A** being defined as $\mathbf{A} = \{a_{ij}\}, i, j = 1, 2, \ldots, n$. (iii) The sum of all such products is |**A**|.

Example 3.21 *Interchanging the second and fourth rows of*

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 3 & 0 & 0 \\ 1 & 0 & 4 & 2 & 3 \\ 2 & 0 & 1 & 4 & 5 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & 2 & 1 & 2 & 3 \end{vmatrix} \quad \text{gives } |\mathbf{A}| = -\begin{vmatrix} 1 & 2 & 3 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 2 & 0 & 1 & 4 & 5 \\ 1 & 0 & 4 & 2 & 3 \\ 0 & 2 & 1 & 2 & 3 \end{vmatrix}.$$

In this form we will expand |**A**| *using the Laplace expansion based on the first two rows, (m = 2). There are ten minors of order 2 in these two rows; seven of them are zero because they involve a column of zeros. Hence* |**A**| *can be expanded as the sum of three products involving the three 2 × 2 nonzero minors in the first two rows, namely as* $-|\mathbf{A}|$ =

$$(-1)^{1+1+2+2}\begin{vmatrix} 1 & 2 \\ -1 & 2 \end{vmatrix}\begin{vmatrix} 1 & 4 & 5 \\ 4 & 2 & 3 \\ 1 & 2 & 3 \end{vmatrix} + (-1)^{1+1+2+3}\begin{vmatrix} 1 & 3 \\ -1 & -1 \end{vmatrix}\begin{vmatrix} 0 & 4 & 5 \\ 0 & 2 & 3 \\ 2 & 2 & 3 \end{vmatrix}$$
$$+ (-1)^{1+2+2+3}\begin{vmatrix} 2 & 3 \\ 2 & -1 \end{vmatrix}\begin{vmatrix} 2 & 4 & 5 \\ 1 & 2 & 3 \\ 0 & 2 & 3 \end{vmatrix}.$$

The sign factors in these terms have been derived by envisaging **A** *as* $\{a_{ij}\}$. *Consequently the first 2 × 2 minor,* $\begin{vmatrix} 1 & 2 \\ -1 & 2 \end{vmatrix}$, *is* $\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$, *leading to* $(-1)^{1+1+2+2}$ *as its sign factor; likewise for the other terms. Simplification of the whole expression gives*

$$-|\mathbf{A}| = 4\begin{vmatrix} 1 & 4 & 5 \\ 3 & 0 & 0 \\ 1 & 2 & 3 \end{vmatrix} - 2(2)\begin{vmatrix} 4 & 5 \\ 2 & 3 \end{vmatrix} + (-8)\begin{vmatrix} 2 & 4 & 5 \\ 1 & 0 & 0 \\ 0 & 2 & 3 \end{vmatrix} = -24 - 8 + 16$$

and hence |**A**| = *16. It will be found that expansion by a more direct method leads to the same result.*

Numerous other methods of expanding determinants are based on extensions of the Laplace expansion, using it recurrently to expand a determinant not only by minors and their complementary minors but also to expand these minors themselves. Many of these expansions are identified by the names of their originators, for example, Cauchy, Binet–

Cauchy, and Jacoby. A good account of some of them is to be found in Aitken (1948) and Ferrar (1941)

3.8 Sums and Differences of Determinants

The sum of the determinants of each of two (or more) matrices generally does not equal the determinant of the sum. The simplest demonstration of this is

$$
\begin{aligned}
|\mathbf{A}| + |\mathbf{B}| &= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} \\
&= a_{11}a_{22} - a_{12}a_{21} + b_{11}b_{22} - b_{12}b_{21} \neq |\mathbf{A} + \mathbf{B}| \,.
\end{aligned}
$$

The same applies to the difference, $|\mathbf{A}| - |\mathbf{B}| \neq |\mathbf{A} - \mathbf{B}|$.

We may note in passing that both $|\mathbf{A}| + |\mathbf{B}|$ and $|\mathbf{A}| - |\mathbf{B}|$ have meaning even when **A** and **B** are square matrices of different orders, because the value of a determinant is a scalar. This contrasts with $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$, which have meaning only when the matrices are conformable for addition (have the same order).

Another point of interest is that although $|\mathbf{A}| + |\mathbf{B}|$ does not generally equal $|\mathbf{A} + \mathbf{B}|$ the latter can be written as the sum of certain other determinants. For example,

$$
\begin{aligned}
|\mathbf{A} + \mathbf{B}| &= \left| \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \right| = \begin{vmatrix} a_{11} + b_{11} & a_{12} + b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ a_{21} + b_{21} & a_{22} + b_{22} \end{vmatrix} \\
&= \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} \\ b_{21} & b_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} .
\end{aligned}
$$

In general if **A** and **B** are $n \times n$, $|\mathbf{A} + \mathbf{B}|$ can be expanded as the sum of 2^n n-order determinants.

Example 3.22 *The reader should verify that*

$$
\begin{vmatrix} a + b & a & a \\ a & a + b & a \\ a & a & a + b \end{vmatrix} = (3a + b)b^2,
$$

by expanding it as $|b\mathbf{I} + a\mathbf{J}|$.

3.9 A Graphical Representation of a 3 × 3 Determinant

A 3 × 3 determinant can be depicted using a parallelpiped in a three-dimensional Euclidean space. To demonstrate this, the following definition is needed:

Definition 3.2 *Let* $\boldsymbol{u} = (u_1, u_2.u_3)'$, $\boldsymbol{v} = (v_1, v_2, v_3)'$, $\boldsymbol{w} = (w_1, w_2, w_3)'$ *be three vectors in a three-dimensional Euclidean space. Then,*

Dot Product *The* dot product *of two vectors such as* $\boldsymbol{u}$ *and* $\boldsymbol{v}$, *denoted by* $\boldsymbol{u}.\boldsymbol{v}$, *is a scalar quantity equal to*

$$\boldsymbol{u}.\boldsymbol{v} = \| \boldsymbol{u} \| \| \boldsymbol{v} \| \cos(\theta), \tag{3.19}$$

where θ is the angle between the directions of $\boldsymbol{u}$ *and* $\boldsymbol{v}$, *and* $\| \boldsymbol{u} \|$ *and* $\| \boldsymbol{v} \|$ *are the Euclidean norms of* $\boldsymbol{u}$ *and* $\boldsymbol{v}$.

The dot product is also known as the scalar product. It can be shown that $\boldsymbol{u}.\boldsymbol{v}$ *is also equal to*

$$\boldsymbol{u}.\boldsymbol{v} = u_1 v_1 + u_2 v_2 + u_3 v_3. \tag{3.20}$$

Vector Product *The* vector product *of two vectors such as* $\boldsymbol{v}$ *and* $\boldsymbol{w}$, *denoted by* $\boldsymbol{v} \times \boldsymbol{w}$, *is a vector perpendicular to both* $\boldsymbol{v}$ *and* $\boldsymbol{w}$ *with the magnitude,*

$$\| \boldsymbol{v} \times \boldsymbol{w} \| = \| \boldsymbol{v} \| \| \boldsymbol{w} \| \sin(\phi), \tag{3.21}$$

where φ is the angle between the directions of $\boldsymbol{v}$ *and* $\boldsymbol{w}$. *The vector* $\boldsymbol{v} \times \boldsymbol{w}$ *is directed so that a rotation about it through the angle φ of not more than 180° carries* $\boldsymbol{v}$ *into* $\boldsymbol{w}$. *It is known that* $\boldsymbol{v} \times \boldsymbol{w}$ *can be expressed as*

$$\boldsymbol{v} \times \boldsymbol{w} = (v_2 w_3 - v_3 w_2)\boldsymbol{e}_1 + (v_3 w_1 - v_1 w_3)\boldsymbol{e}_2 + (v_1 w_2 - v_2 w_1)\boldsymbol{e}_3, \tag{3.22}$$

where e_1, e_2, *and* e_3 *are unit vectors in the positive directions of the coordinate axes, x, y, z, respectively [see, e.g., Rutherford (1957, p. 7)].*

Triple Scalar Product *The* triple scalar product *of the three vectors* $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$ *is a scalar quantity equal to* $\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w})$, *that is, the dot product of the two vectors,* $\boldsymbol{u}$ *and* $\boldsymbol{v} \times \boldsymbol{w}$. *Hence, it is equal to*

$$\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w}) = \| \boldsymbol{u} \| \| \boldsymbol{v} \| \| \boldsymbol{w} \| \sin(\phi)\cos(\eta), \tag{3.23}$$

where η is the angle between $\boldsymbol{u}$ *and* $\boldsymbol{v} \times \boldsymbol{w}$.

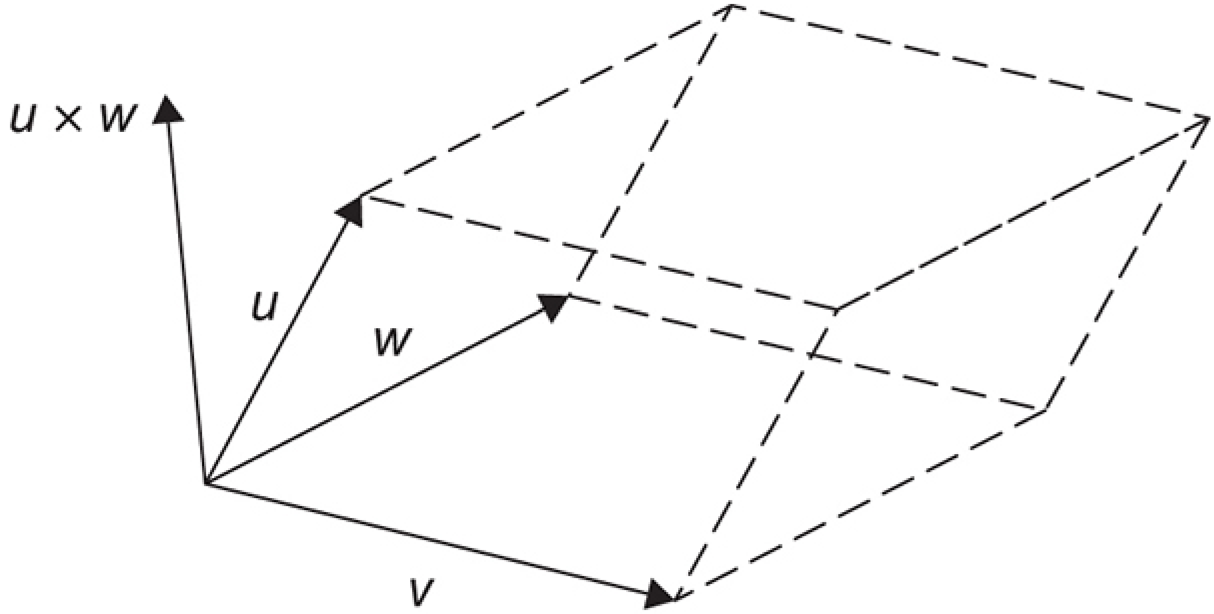
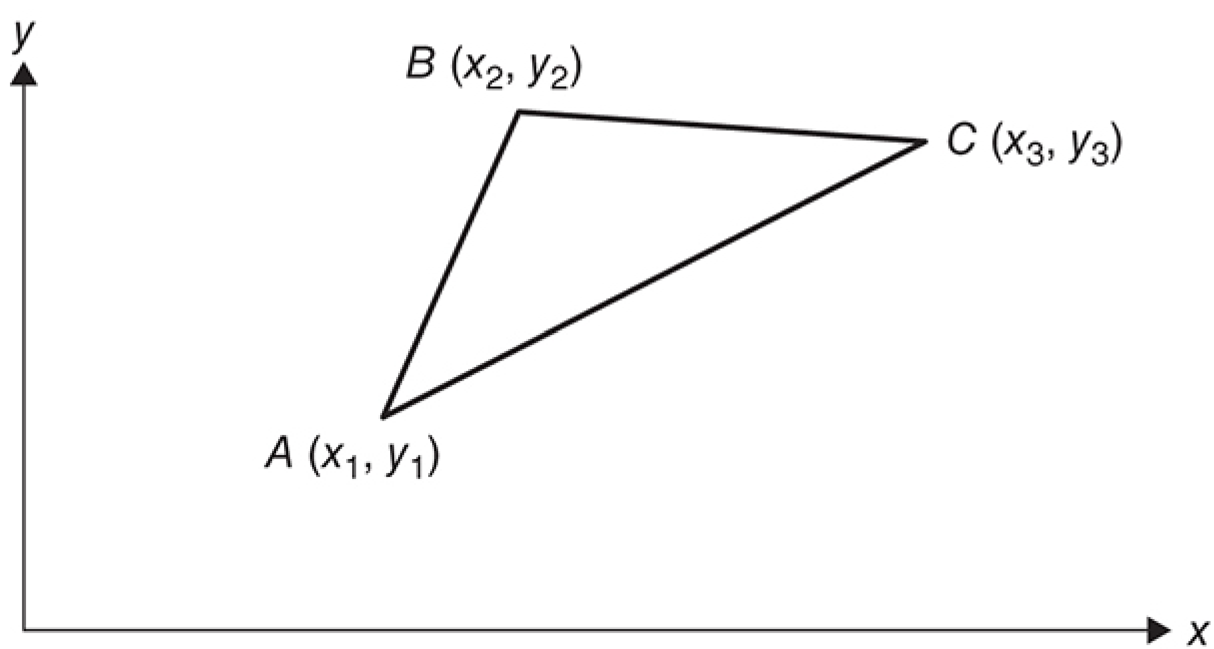


Figure 3.1 Parallelpiped Constructed from the Vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$.

It can be seen from (3.23) that the value of the triple scalar product $\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w})$ is equal to the volume of the parallelpiped constructed from the vectors $\boldsymbol{u}$, $\boldsymbol{v}$, $\boldsymbol{w}$. This is true since $\| \boldsymbol{v} \| \| \boldsymbol{w} \| \ \sin(\phi)$ is the area of the base defined by $\boldsymbol{v}$ and $\boldsymbol{w}$, and $\| \boldsymbol{u} \| \ \cos(\eta)$ is the perpendicular height of the parallelpiped (see Figure 3.1). From (3.20) and (3.22), it follows that the triple product $\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w})$ can be expressed as the value of the following determinant;

$$\boldsymbol{u}.(\boldsymbol{v} \times \boldsymbol{w}) = \begin{vmatrix} u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \\ w_1 & w_2 & w_3 \end{vmatrix}. \tag{3.24}$$

Thus we have a graphical depiction of the determinant in (3.24) as the volume of the parallelpiped in Figure 3.1 which is constructed from the vectors that make up the rows of the determinant.



Exercises

1. Show that both

   a. $\begin{vmatrix} 1 & 5 & -5 \\ 3 & 2 & -5 \\ 6 & -2 & -5 \end{vmatrix}$ and $\begin{vmatrix} -3 & 2 & -6 \\ -3 & 5 & -7 \\ -2 & 3 & -4 \end{vmatrix}$ equal -5;

   b. $\begin{vmatrix} 2 & 6 & 5 \\ -2 & 7 & -5 \\ 2 & -7 & 9 \end{vmatrix}$ and $\begin{vmatrix} 2 & -1 & 9 \\ -1 & 7 & 2 \\ 3 & -21 & 2 \end{vmatrix}$ equal 104.

2.

   a. Show that the determinant of $\mathbf{x1}'$ is zero.

   b. If $\mathbf{A1} = \mathbf{0}$, why does $\mathbf{A}$ have a zero determinant?

3. For

$$\mathbf{A} = \begin{bmatrix} 0 & -a & b & -c \\ a & 0 & -d & e \\ -b & d & 0 & -f \\ c & -e & f & 0 \end{bmatrix}$$

   a. Calculate $|\mathbf{A}|$.

   b. Verify that $|\mathbf{I} + \mathbf{A}| = 1 + (a^2 + b^2 + c^2 + d^2 + e^2 + f^2) + |\mathbf{A}|$ by direct expansion

4. Without expanding the determinants, suggest values of x that satisfy the following equations:

   a. $\begin{vmatrix} x & x & x \\ 2 & -1 & 0 \\ 7 & 4 & 5 \end{vmatrix} = 0;$

   b. $\begin{vmatrix} 1 & x & x^2 \\ 1 & 2 & 4 \\ 1 & -1 & 1 \end{vmatrix} = 0;$

5. Drop perpendiculars from A, B, C in the accompanying figure and by using the resulting trapezoids show that the area of the triangle ABC can be expressed as the absolute value of $\frac{1}{2}|\mathbf{M}|$ where

$$\mathbf{M} = \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}.$$

6. The equation of a line in two co-ordinates can be written as $1 + ax + by = 0$. Show that the equation of a line passing through (x_1, y_1) and (x_2, y_2) is

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

7. Denote the equations

$$\left.\begin{aligned} x_1 + 2x_2 + 3x_3 &= 26 \\ 3x_1 + 7x_2 + 10x_3 &= 87 \\ 2x_1 + 11x_2 + 7x_3 &= 73 \end{aligned}\right\} \quad \text{by} \quad \mathbf{Ax} = \mathbf{b}.$$

   a. Solve the equation using successive elimination.

   b. Replace each column of **A**, in turn, by **b**. In replacing column j, call the resulting matrix $\mathbf{A}_j$. Verify that the solutions in (a) are

$$x_j = |\mathbf{A}_j|/|\mathbf{A}|.$$

   This is known as Cramer's rule for solving linear equations.

   c. What condition must $|\mathbf{A}|$ satisfy for Cramer's rule to be workable?

8. Use the result in Section 3.3.4 to show that

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ a^2 & b^2 & c^2 \end{vmatrix} = (b-a)(c-a)(c-b).$$

The above determinant is known as *Vandermonde determinant* of order 3.

9. As in Exercise 8, use the result in Section 3.3.4 to show that the Vandermonde determinant is equal to the determinant,

$$\begin{vmatrix} 1 & 1 & 1 \\ a & b & c \\ bc & ac & ab \end{vmatrix}$$

10. Let $A = (a_{ij})$ be a square matrix of order $n \times n$. If $\sum^n_j a_{ij} = 1$ for $i = 1, 2, \ldots, n$, show that $|A - I| = 0$.

11. Expand the determinant on the left using its first row and show that

$$\begin{vmatrix} a_1 + \lambda\alpha_1 & b_1 + \lambda\beta_1 & c_1 + \lambda\gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} = \begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix} + \lambda \begin{vmatrix} \alpha_1 & \beta_1 & \gamma_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$$

12. Show that a determinant is unaltered in value when to any row , or column, is added a constant multiple of any other row, or column.

    (Hint: In Exercise 11, replace the second matrix on the right by a matrix whose second and third rows are identical to those in the matrix on the left, and its first row is equal to either the second row or the third row of the matrix on the left.)

13. Let $\boldsymbol{K}$ be a square matrix of order 12×12 such that $\boldsymbol{K}^5 = 3\boldsymbol{K}$. What is the numerical value of $|\boldsymbol{K}|$? Is there more than one possibility? If so, give all possible values.

14. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of order $m \times n$ and $n \times m$ ($n \geq m$), respectively. Show that $|\boldsymbol{I}_m - \boldsymbol{AB}| = |\boldsymbol{I}_n - \boldsymbol{BA}|$.

15. Let $\boldsymbol{a}$ be a column vector of n elements. Show that

$$|\lambda \boldsymbol{I}_n + \boldsymbol{1}_n \boldsymbol{a}'| = \lambda^{n-1}(\lambda + \boldsymbol{a}'\boldsymbol{1}_n).$$

(Hint: Use the result in Exercise 14.)

16. Show that $|a\boldsymbol{I}_n + b\boldsymbol{J}_n| = a^{n-1}(a + nb)$, where $\boldsymbol{J}_n$ is a matrix of ones of order $n \times n$.

Notes

[1]$n!$, read as "n factorial" is the product of all integers 1 through n; e.g., $4! = 1(2)(3)(4) = 24$.

References

Aitken, A. C. (1948). *Determinants and Matrices*, 5th ed. Oliver and Boyd, Edinburgh, UK.

Ferrar, W. L. (1941). *Algebra, A Text-Book of Determinants, Matrices and Algebraic Forms*. Oxford University Press, Oxford.

Rutherford, D. E. (1957). *Vector Methods*, 9th ed. Oliver and Boyd, Edinburgh, UK.

4
Matrix Operations

This chapter describes simple operations on matrices arising from their being rectangular arrays of numbers, and also the arithmetic of matrix addition, subtraction, and multiplication. Particular types of sums and products of, namely, the direct sum and direct product of matrices are defined. The trace, rank, and inverse of a matrix are also defined.

4.1 The Transpose of a Matrix

Consider the following two matrices,

$$\mathbf{A} = \begin{bmatrix} 18 & 17 & 11 \\ 19 & 13 & 6 \\ 6 & 14 & 9 \\ 9 & 11 & 4 \end{bmatrix},$$

and

$$\mathbf{B} = \begin{bmatrix} 18 & 19 & 6 & 9 \\ 17 & 13 & 14 & 11 \\ 11 & 6 & 9 & 4 \end{bmatrix}.$$

Although the elements of $\mathbf{B}$ are the same as those of $\mathbf{A}$, the matrices are clearly different; for example, they are not of the same order. $\mathbf{A}$ has order 4×3 and $\mathbf{B}$ is 3×4. The matrices are nevertheless related to each other, through the rows of one being the columns of the other. Whenever matrices are related in this fashion each is said to be the *transpose* of the other; for example, $\mathbf{B}$ is said to be the transpose of $\mathbf{A}$, and $\mathbf{A}$ is the transpose of $\mathbf{B}$.

A formal description is that the transpose of a matrix $\mathbf{A}$ is the matrix whose columns are the rows of $\mathbf{A}$, with order retained, from first to last. The transpose is written as $\mathbf{A}'$, although the notations $\mathbf{A}^t$ and $\mathbf{A}^T$ are also seen in the literature. We will use $\mathbf{A}'$.

An obvious consequence of the transpose operation is that the rows of $\mathbf{A}'$ are the same as the columns of $\mathbf{A}$, and if $\mathbf{A}$ is $r \times c$, the order of $\mathbf{A}'$ is $c \times r$. If a_{ij} is the term in the *i*th row and the *j*th column of $\mathbf{A}$, it is also the term in the *j*th row and *i*th column of $\mathbf{A}'$. Therefore, on defining a'_{ji} as the element in the *j*th row and *i*th column of $\mathbf{A}'$, we have $a'_{ji} = a_{ij}$. Equivalently, on interchanging *i* and *j*, we have

$$\mathbf{A}' = \{a'_{ij}\} \quad \text{for} \quad i = 1, \ldots, c \quad \text{and} \quad j = 1, \ldots, r, \quad \text{and} \quad a'_{ij} = a_{ji}.$$

A minor difficulty in notation arises when, for example, $\mathbf{A}_{r\times c}$ has *r* rows and *c* columns; but combining the notation $\mathbf{A}'$ for the transpose of $\mathbf{A}$ with the $r \times c$ subscript notation for the order of $\mathbf{A}$, to get $\mathbf{A}'_{r\times c}$, is ambiguous. $\mathbf{A}'_{r\times c}$ could mean either that $\mathbf{A}$ of order $r \times c$ has been transposed or that the transpose of $\mathbf{A}$ has order $r \times c$ (and so $\mathbf{A}$ has order $c \times r$). For clarity, one of the equivalent forms $(\mathbf{A}_{r\times c})'$ or $(\mathbf{A}')_{c\times r}$ must be used whenever it is necessary to have subscript notation for the order of a transposed matrix. Fortunately, the need for this clumsy notation seldom arises. The simple $\mathbf{A}'$ suffices in most contexts.

Two important consequences of the transpose operation are worth noting.

4.1.1 A Reflexive Operation

The transpose operation is reflexive: the transpose of a transposed matrix is the matrix itself, that is, $(\mathbf{A}')' = \mathbf{A}$. This is so because transposing $\mathbf{A}'$ yields a matrix whose rows are the columns of $\mathbf{A}'$, and these are the rows of $\mathbf{A}$. Hence $(\mathbf{A}')' = \mathbf{A}$. More formally,

$$(\mathbf{A}')' = \{a'_{ij}\}' = \{a_{ji}\}' = \{a'_{ji}\} = \{a_{ij}\} = \mathbf{A}.$$

4.1.2 Vectors

The transpose of a row vector is a column vector and vice versa. For example, the transpose of

$$\mathbf{x} = \begin{bmatrix} 1 \\ 6 \\ 4 \end{bmatrix} \qquad \text{is} \quad \mathbf{x}' = \begin{bmatrix} 1 & 6 & 4 \end{bmatrix}.$$

4.2 Partitioned Matrices

4.2.1 Example

Consider the matrix

$$\mathbf{B} = \begin{bmatrix} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ 4 & 3 & 6 & 1 & 2 & 1 \\ 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{bmatrix}.$$

Suppose we draw dashed lines between certain rows and columns as in

$$\mathbf{B} = \left[\begin{array}{cccc|cc} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ 4 & 3 & 6 & 1 & 2 & 1 \\ - & - & - & - & - & - \\ 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{array}\right]. \tag{4.1}$$

Each of the arrays of numbers in the four sections of $\mathbf{B}$ engendered by the dashed lines is a matrix:

$$\mathbf{B}_{11} = \begin{bmatrix} 1 & 6 & 8 & 9 \\ 2 & 4 & 1 & 6 \\ 4 & 3 & 6 & 1 \end{bmatrix}, \qquad \mathbf{B}_{12} = \begin{bmatrix} 3 & 8 \\ 1 & 1 \\ 2 & 1 \end{bmatrix}$$
$$\mathbf{B}_{21} = \begin{bmatrix} 9 & 1 & 4 & 6 \\ 6 & 8 & 1 & 4 \end{bmatrix}, \qquad \mathbf{B}_{22} = \begin{bmatrix} 8 & 7 \\ 3 & 2 \end{bmatrix}. \tag{4.2}$$

Using the matrices in (4.2), the matrix $\mathbf{B}$ of (4.1) can now be written as a matrix of matrices:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}. \tag{4.3}$$

This specification of $\mathbf{B}$ is called a *partitioning* of $\mathbf{B}$, and the matrices $\mathbf{B}_{11}, \mathbf{B}_{12}, \mathbf{B}_{21}, \mathbf{B}_{22}$ are said to be *submatrices* of $\mathbf{B}$; further, $\mathbf{B}$ of (4.3) is called a *partitioned* matrix.

Note that $\mathbf{B}_{11}$ and $\mathbf{B}_{21}$ have the same number of columns, as do $\mathbf{B}_{12}$ and $\mathbf{B}_{22}$. Likewise $\mathbf{B}_{11}$ and $\mathbf{B}_{12}$ have the same number of rows, as do $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$. This is the usual method of partitioning, as expressed in the general case for an $r \times c$ matrix:

$$\mathbf{A}_{r\times c} = \begin{bmatrix} \mathbf{K}_{p\times q} & \mathbf{L}_{p\times(c-q)} \\ \mathbf{M}_{(r-p)\times q} & \mathbf{N}_{(r-p)\times(c-q)} \end{bmatrix}$$

where $\mathbf{K}, \mathbf{L}, \mathbf{M}$, and $\mathbf{N}$ are the submatrices with their orders shown as subscripts.

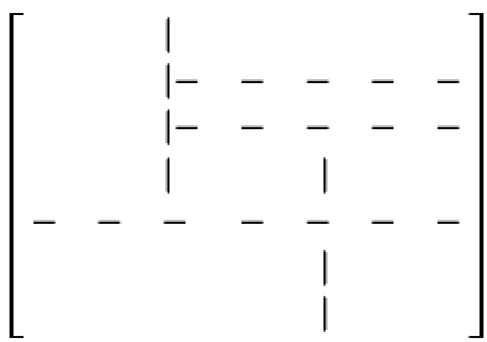
Partitioning is not restricted to dividing a matrix into just four submatrices; it can be divided into numerous rows and columns of matrices. Thus if

$$\mathbf{B}_{01} = \begin{bmatrix} 1 & 6 & 8 & 9 \\ 2 & 4 & 1 & 6 \end{bmatrix}, \qquad \mathbf{B}_{02} = \begin{bmatrix} 3 & 8 \\ 1 & 1 \end{bmatrix},$$
$$\mathbf{B}_{03} = \begin{bmatrix} 4 & 3 & 6 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{B}_{04} = \begin{bmatrix} 2 & 1 \end{bmatrix}$$

with $\mathbf{B}_{21}$ and $\mathbf{B}_{22}$ as in (4.2), then $\mathbf{B}$ can be written in partitioned form as

$$\mathbf{B} = \left[\begin{array}{cccc|cc} 1 & 6 & 8 & 9 & 3 & 8 \\ 2 & 4 & 1 & 6 & 1 & 1 \\ \hdashline 4 & 3 & 6 & 1 & 2 & 1 \\ \hdashline 9 & 1 & 4 & 6 & 8 & 7 \\ 6 & 8 & 1 & 4 & 3 & 2 \end{array}\right] = \begin{bmatrix} \mathbf{B}_{01} & \mathbf{B}_{02} \\ \mathbf{B}_{03} & \mathbf{B}_{04} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}. \tag{4.4}$$

It goes without saying that each such line must always go the full length (or breadth) of the matrix. Partitioning in any staggered manner such as



is not allowed.

4.2.2 General Specification

The example illustrates the simplicity of the operation of partitioning a matrix and, as is evident from (4.1) and (4.4), there is no single way in which any matrix can be partitioned. In general, a matrix $\mathbf{A}$ of order $p \times q$ can be partitioned into r rows and c columns of submatrices as

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} & \cdots & \mathbf{A}_{1c} \\ \mathbf{A}_{21} & \mathbf{A}_{22} & \cdots & \mathbf{A}_{2c} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_{r1} & \mathbf{A}_{r2} & \cdots & \mathbf{A}_{rc} \end{bmatrix}$$

where $\mathbf{A}_{ij}$ is the submatrix in the ith row and jth column of submatrices. If the ith row of submatrices has p_i rows of elements and the jth column of submatrices has q_j columns, then $\mathbf{A}_{ij}$ has order $p_i \times q_j$, where

$$\sum_{i=1}^{r} p_i = p \quad \text{and} \quad \sum_{j=1}^{c} q_j = q.$$

4.2.3 Transposing a Partitioned Matrix

The transpose of a partitioned matrix is the transposed matrix of transposed submatrices:

$$[\mathbf{X}\ \mathbf{Y}]' = \begin{bmatrix} \mathbf{X}' \\ \mathbf{Y}' \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} \mathbf{A} & \mathbf{B} & \mathbf{C} \\ \mathbf{D} & \mathbf{E} & \mathbf{F} \end{bmatrix}' = \begin{bmatrix} \mathbf{A}' & \mathbf{D}' \\ \mathbf{B}' & \mathbf{E}' \\ \mathbf{C}' & \mathbf{F}' \end{bmatrix}.$$

The reader should use the example in (4.1) to verify these results and to be assured that in general the transpose of $[\mathbf{X}\ \mathbf{Y}]$ is neither $\begin{bmatrix} \mathbf{X} \\ \mathbf{Y} \end{bmatrix}$ nor $[\mathbf{X}'\ \mathbf{Y}']$. For example,

$$\left[\begin{array}{c|cc} 2 & 8 & 9 \\ 3 & 7 & 4 \end{array}\right]' = \begin{bmatrix} 2 & & 3 \\ - & - & - \\ 8 & & 7 \\ 9 & & 4 \end{bmatrix}.$$

4.2.4 Partitioning Into Vectors

Suppose $\mathbf{a}_j$ is the jth column of $\mathbf{A}_{r\times c}$. Then

$$\mathbf{A} = [\mathbf{a}_1 \quad \mathbf{a}_2 \cdots \mathbf{a}_j \cdots \mathbf{a}_c] \tag{4.5}$$

is partitioned into its c columns. Similarly,

$$\mathbf{A} = \begin{bmatrix} \alpha_1' \\ \alpha_2' \\ \vdots \\ \alpha_i' \\ \vdots \\ \alpha_r' \end{bmatrix} \tag{4.6}$$

is partitioned into its r rows α'_i for $i = 1, \ldots, r$.

4.3 The Trace of a Matrix

The sum of the diagonal elements of a square matrix is called the *trace* of the matrix, written $tr(\mathbf{A})$; that is, for $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, \ldots, n$

For example,

$$tr(\mathbf{A}) = a_{11} + a_{22} + \cdots + a_{nn} = \sum_{i=1}^{n} a_{ii}.$$

$$tr\begin{bmatrix} 1 & 7 & 6 \\ 8 & 3 & 9 \\ 4 & -2 & -8 \end{bmatrix} = 1 + 3 - 8 = -4.$$

When $\mathbf{A}$ is not square, the trace is not defined; that is, it does not exist.

The trace of a transposed matrix is the same as the trace of the matrix itself:

$$tr(\mathbf{A}') = tr(\mathbf{A}).$$

For example,

$$tr\begin{bmatrix} 1 & 8 & 4 \\ 7 & 3 & -2 \\ 6 & 9 & -8 \end{bmatrix} = tr\begin{bmatrix} 1 & 7 & 6 \\ 8 & 3 & 9 \\ 4 & -2 & -8 \end{bmatrix} = 1 + 3 - 8 = -4.$$

Also, by treating a scalar as a 1×1 matrix we have

$$\text{tr(scalar)} = \text{scalar}; \quad \text{for example,} \quad \text{tr}(13) = 13.$$

4.4 Addition

We introduce the operation of addition by means of an example:

The numbers of lunches served one Saturday at a country club are shown in Table 4.1.

Table 4.1 Number of Lunches

	Beef	**Fish**	**Omelet**
Member	98	24	42
Guest	39	15	22

Let us write the data of the table as a 2 × 3 matrix:

$$\mathbf{A} = \begin{bmatrix} 98 & 24 & 42 \\ 39 & 15 & 22 \end{bmatrix}.$$

Then, with the same frame of reference, the data for the Sunday lunches might be

$$\mathbf{B} = \begin{bmatrix} 55 & 19 & 44 \\ 43 & 53 & 38 \end{bmatrix}.$$

Hence over the weekend the total number of members served a beef lunch is the sum of the elements in the first row and first column of each matrix, 98 + 55 = 153; and the total number of guests served a fish lunch is 15 + 53 = 68. In this way the matrix of all such sums is

$$\begin{bmatrix} 98+55 & 24+19 & 42+44 \\ 39+43 & 15+53 & 22+38 \end{bmatrix} = \begin{bmatrix} 153 & 43 & 86 \\ 82 & 68 & 60 \end{bmatrix},$$

which represents the numbers of lunches served over the weekend. This is the matrix sum $\mathbf{A}$ plus $\mathbf{B}$; it is the matrix formed by adding the matrices $\mathbf{A}$ and $\mathbf{B}$, element by element. Hence, if we write $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$ the matrix representing the sum of $\mathbf{A}$ and $\mathbf{B}$ is

$$\mathbf{A} + \mathbf{B} = \{a_{ij} + b_{ij}\} \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c;$$

that is, the sum of two matrices is the matrix of sums, element by element.

It is evident from this definition that matrix addition can take place only when the matrices involved are of the same order; that is, two matrices can be added only if they have the same number of rows and the same number of columns, in which case they are said to be *conformable for addition*.

Having defined addition, we can now consider the transpose of a sum and the trace of a sum:

i. The transpose of a sum of matrices is the sum of the transposed matrices, that is,

$$(\mathbf{A}+\mathbf{B})' = \mathbf{A}' + \mathbf{B}'.$$

ii. The trace of a sum of matrices is the sum of the traces, that is,

$$tr(\mathbf{A}+\mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B})$$

provided **A** and **B** are square, of the same order.

It is left to the reader to create examples illustrating these results. Formal proofs are as follows.

i. If

$$\mathbf{A}+\mathbf{B} = \mathbf{C} = \{c_{ij}\} = \{a_{ij} + b_{ij}\},$$

then

$$(\mathbf{A}+\mathbf{B})' = \mathbf{C}' = \{c'_{ij}\} = \{c_{ji}\} = \{a_{ji} + b_{ji}\} = \{a_{ji}\} + \{b_{ji}\} = \mathbf{A}' + \mathbf{B}'.$$

ii. $$tr(\mathbf{A}+\mathbf{B}) = \sum_{i=1}^{n}(a_{ii} + b_{ii}) = \sum_{i=1}^{n} a_{ii} + \sum_{i=1}^{n} b_{ii} = tr(\mathbf{A}) + tr(\mathbf{B})$$.

The difference between two matrices can be defined in a similar manner. For example, if $\boldsymbol{A} = \{a_{ij}\}$ and $\boldsymbol{B} = \{b_{ij}\}$ are two matrices of the same order, then $\boldsymbol{A} - \boldsymbol{B} = \{a_{ij} - b_{ij}\}$, that is, it is the matrix of differences, element by element.

4.5 Scalar Multiplication

We have just described matrix addition. A simple use of it shows that

$$\mathbf{A} + \mathbf{A} = \{a_{ij}\} + \{a_{ij}\} = \{2a_{ij}\} = 2\mathbf{A}.$$

Extending this to the case where λ is a positive integer, we have

$$\lambda\mathbf{A} = \mathbf{A} + \mathbf{A} + \mathbf{A} + \cdots + \mathbf{A},$$

there being λ **A**'s in the sum on the right. Carrying out these matrix additions gives

$$\lambda\mathbf{A} = \{\lambda a_{ij}\} \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c.$$

This result, extended to λ being any scalar, is the definition of *scalar multiplication of a matrix*. Thus the matrix **A** multiplied by the scalar λ is the matrix **A** with every element multiplied by λ. For example

$$3\begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 6 & -21 \\ 9 & 15 \end{bmatrix} \qquad \text{and} \quad \lambda\begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} 2\lambda & -7\lambda \\ 3\lambda & 5\lambda \end{bmatrix}.$$

Although $\lambda\mathbf{A}$ is the customary notation, rather than $\mathbf{A}\lambda$, the latter is more convenient when λ is a fraction; for example, $\mathbf{A}/17$ rather than $(1/17)\mathbf{A}$ or $\mathbf{A}(1/17)$. Thus

$$\begin{bmatrix} 2 & -7 \\ 3 & 5 \end{bmatrix}/17 = \begin{bmatrix} 2/17 & -7/17 \\ 3/17 & 5/17 \end{bmatrix}.$$

4.6 Equality and the Null Matrix

Two matrices are said to be equal when they are identical element by element. Thus $\mathbf{A} = \mathbf{B}$ when $\{a_{ij}\} = \{b_{ij}\}$, meaning that $a_{ij} = b_{ij}$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$. If

$$\mathbf{A} = \begin{bmatrix} 2 & 6 & -4 \\ 3 & 0 & 1 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 2 & 6 & -4 \\ 3 & 0 & 1 \end{bmatrix}, \quad \text{and} \quad \mathbf{C} = \begin{bmatrix} 2 & 6 & -4 \\ 2 & 0 & 1 \end{bmatrix},$$

$\mathbf{A}$ is equal to $\mathbf{B}$ but not equal to $\mathbf{C}$. It is also apparent that equality of two matrices has no meaning unless they are of the same order.

Combining the ideas of subtraction and equality leads to the definition of zero in matrix algebra. For when $\mathbf{A} = \mathbf{B}$, then $a_{ij} = b_{ij}$ for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, c$, and so

$$\mathbf{A} - \mathbf{B} = \{a_{ij} - b_{ij}\} = \{0\} = \mathbf{0}.$$

The matrix on the right is a matrix of zeros, that is, every element is zero. Such a matrix is called a *null matrix*; it is the zero of matrix algebra and is sometimes referred to as a *zero matrix*. It is, of course, not a unique zero because corresponding to every matrix there is a null matrix of the same order. For example, null matrices of order 2×4 and 3×3 are, respectively,

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

4.7 Multiplication

The method of multiplying matrices is developed by first considering the product of two vectors and then the product of a matrix and a vector. Each of these products is introduced by means of an illustration and is then given a formal definition.

4.7.1 The Inner Product of Two Vectors

Illustration. Consider buying supplies of experimental rats, mice, and rabbits for laboratory courses in the chemistry, biochemistry, nutrition, and physiology departments of a university. Suppose the price per animal of rats, mice, and rabbits in the home town is \$3, \$1, and \$10, and that the chemistry department needs 50, 100, and 30 animals, respectively. The total cost to the chemistry department is, very simply,

$$3(50) + 1(100) + 10(30) = 550 \tag{4.7}$$

dollars. Suppose the prices are written as a row vector $\mathbf{a}'$ and the numbers of animals needed as a column vector $\mathbf{x}$:

$$\mathbf{a}' = \begin{bmatrix} 3 & 1 & 10 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix}.$$

Then the total cost of the animals needed, 550, is the sum of products of the elements of $\mathbf{a}'$ each multiplied by the corresponding element of $\mathbf{x}$. This is the definition of the product $\mathbf{a}'\mathbf{x}$. It is written as

$$\mathbf{a}'\mathbf{x} = \begin{bmatrix} 3 & 1 & 10 \end{bmatrix} \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = 3(50) + 1(100) + 10(30) = 550. \tag{4.8}$$

This example illustrates the general procedure for obtaining $\mathbf{a}'\mathbf{x}$: multiply each element of the row vector $\mathbf{a}'$ by the corresponding element of the column vector $\mathbf{x}$ and add the products. The sum is $\mathbf{a}'\mathbf{x}$. Thus if

$$\mathbf{a}' = [a_1 \quad a_2 \cdots a_n] \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

their product $\mathbf{a}'\mathbf{x}$ is defined as

$$\mathbf{a}'\mathbf{x} = a_1x_1 + a_2x_2 + \cdots + a_nx_n = \sum_{i=1}^{n} a_ix_i.$$

It is called the *inner product* of the vectors $\mathbf{a}$ and $\mathbf{x}$. It exists only when $\mathbf{a}$ and $\mathbf{x}$ have the same order; when they are not of the same order the product $\mathbf{a}'\mathbf{x}$ is undefined.

4.7.2 A Matrix–Vector Product

Illustration. For the preceding illustration, suppose the animal prices in a neighboring town were \$2, \$2, and \$8, respectively. Let us represent them by the row vector [2 2 8]. Then purchasing the chemistry department's requirements in the neighboring town would cost

$$\begin{bmatrix} 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = 2(50) + 2(100) + 8(30) = 540 \tag{4.9}$$

dollars, calculated just like (4.8).

Now put the two sets of prices as the rows of a matrix,

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}.$$

Then the products (4.8) and (4.9) can be represented simultaneously as a single product of the matrix $\mathbf{A}$ and the vector $\mathbf{x}$:

$$\mathbf{Ax} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 50 \\ 100 \\ 30 \end{bmatrix} = \begin{bmatrix} 3(50) + 1(100) + 10(30) \\ 2(50) + 2(100) + 8(30) \end{bmatrix} = \begin{bmatrix} 550 \\ 540 \end{bmatrix}. \tag{4.10}$$

The result is a vector, its elements being the inner products (4.8) and (4.9). In terms of the rows of $\mathbf{A}$, this means that the elements of the vector $\mathbf{Ax}$ are derived in exactly the same way as the product $\mathbf{a}'\mathbf{x}$ developed earlier, using the successive rows of $\mathbf{A}$ as the vector $\mathbf{a}'$. The result is the product $\mathbf{Ax}$; that is, $\mathbf{Ax}$ is obtained by repetitions of the product $\mathbf{a}'\mathbf{x}$ using the rows of $\mathbf{A}$ successively for $\mathbf{a}'$ and writing the results as a column vector. Hence, on using the notation of (4.6), with α'_1 and α'_2 being the rows of $\mathbf{A}$, we see that (4.10) is

$$\mathbf{Ax} = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \end{bmatrix} \mathbf{x} = \begin{bmatrix} \alpha'_1\mathbf{x} \\ \alpha'_2\mathbf{x} \end{bmatrix}.$$

This generalizes at once to $\mathbf{A}$ having r rows:

$$\mathbf{Ax} = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \\ \vdots \\ \alpha'_t \\ \vdots \\ \alpha'_r \end{bmatrix} \mathbf{x} = \begin{bmatrix} \alpha'_1\mathbf{x} \\ \alpha'_2\mathbf{x} \\ \vdots \\ \alpha'_t\mathbf{x} \\ \vdots \\ \alpha'_r\mathbf{x} \end{bmatrix}. \tag{4.11}$$

Thus **Ax** is a column vector, with its ith element being the inner product of the ith row of **A** with the column vector **x**. Providing neither **A** nor **x** is a scalar, it is clear from this definition and from the example that **Ax** is defined only when the number of elements in each row of **A** (i.e., number of columns) is the same as the number of elements in the column vector **x**, and when this occurs **Ax** is a column vector having the same number of elements as there are rows in **A**. Therefore, when **A** has r rows and c columns and **x** is of order c, **Ax** is a column vector of order r; its ith element is $\Sigma^{c}_{k=1} a_{ik}x_k$ for $i = 1, 2, \ldots, r$. More formally, when

$$\mathbf{A} = \{a_{ij}\} \quad \text{and} \quad \mathbf{x} = \{x_j\} \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c,$$

then

$$\mathbf{Ax} = \left\{ \sum_{j=1}^{c} a_{ij}x_j \right\} \quad \text{for} \quad i = 1, 2, \ldots, r.$$

For example

$$\begin{bmatrix} 4 & 2 & 1 & 3 \\ 2 & 0 & -4 & 7 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 4(1) + 2(0) + & 1(-1) + 3(3) \\ 2(1) + 0(0) + & -4(-1) + 7(3) \end{bmatrix} = \begin{bmatrix} 12 \\ 27 \end{bmatrix}.$$

Each element of **Ax** is obtained by moving along a row of **A** and down the column **x**, multiplying each pair of corresponding elements and adding the products. This is always the procedure for calculating the product **Ax**. It is also, as we shall see, the basis for calculating the product **AB** of two matrices.

Typical of many uses of the matrix–vector product is one that occurs in the study of inbreeding, where what is known as the generation matrix is used to relate the frequencies of mating types in one generation to those in another. Kempthorne(1957, p. 108), for example, gives the result earlier stated by Fisher (1949) that if after one generation of full-sib mating $\mathbf{f}^{(1)}$ represents the vector of frequencies of the seven distinct types of mating possible in this situation, then $\mathbf{f}^{(1)} = \mathbf{Af}^{(0)}$, where $\mathbf{f}^{(0)}$ is the vector of initial frequencies and where **A** is the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & \frac{2}{16} & 0 & \frac{1}{4} & 0 & \frac{1}{16} & 0 \\ 0 & \frac{4}{16} & 1 & \frac{1}{4} & \frac{1}{2} & \frac{3}{16} & \frac{4}{16} \\ 0 & \frac{2}{16} & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{8}{16} & 0 & \frac{2}{4} & 0 & \frac{4}{16} & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{2}{16} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{6}{16} & \frac{8}{16} \\ 0 & 0 & 0 & 0 & 0 & 0 & \frac{4}{16} \end{bmatrix}.$$

This simply means that if $f^{(1)}_i$ is the relative frequency of the ith type of mating after a generation of full-sib mating (that is, the ith element of the vector $\mathbf{f}^{(1)}$) and if $f^{(0)}_i$ is the corresponding initial frequency [ith element of $\mathbf{f}^{(0)}$], then, for example,

$$f_1^{(1)} = f_1^{(0)} + \tfrac{2}{16} f_2^{(0)} + \tfrac{1}{4} f_4^{(0)} + \tfrac{1}{16} f_6^{(0)};$$

and as another example

$$f_4^{(1)} = \tfrac{8}{16} f_2^{(0)} + \tfrac{2}{4} f_4^{(0)} + \tfrac{4}{16} f_6^{(0)}.$$

These and five similar equations are implied in the vector equation $\mathbf{f}^{(1)} = \mathbf{A}\mathbf{f}^{(0)}$. The matrix $\mathbf{A}$ which represents the relationships between the two sets of frequencies is known in this context as the generation matrix.

4.7.3 A Product of Two Matrices

Multiplying two matrices can be explained as a simple repetitive extension of multiplying a matrix by a vector.

Illustration. Continuing the illustration of buying laboratory animals, suppose the biochemistry department needed 60 rats, 80 mice, and 40 rabbits. The costs of these, obtained in the manner of (4.10), are shown in the product

$$\begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 60 \\ 80 \\ 40 \end{bmatrix} = \begin{bmatrix} 660 \\ 600 \end{bmatrix}. \tag{4.12}$$

Similarly, if the nutrition department needed 90, 30, and 20 animals, and the physiology department needed 30, 20, and 10, respectively, their costs would be

$$\begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 90 \\ 30 \\ 20 \end{bmatrix} = \begin{bmatrix} 500 \\ 400 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 30 \\ 20 \\ 10 \end{bmatrix} = \begin{bmatrix} 210 \\ 180 \end{bmatrix}. \tag{4.13}$$

(The reader should verify the validity of these products.) Now, by writing the four column vectors of departmental requirements alongside one another as a matrix,

$$\mathbf{B} = \begin{bmatrix} 50 & 60 & 90 & 30 \\ 100 & 80 & 30 & 20 \\ 30 & 40 & 20 & 10 \end{bmatrix},$$

the products in (4.10), (4.12), and (2.13) can be represented as the single matrix product

$$\mathbf{AB} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \begin{bmatrix} 50 & 60 & 90 & 30 \\ 100 & 80 & 30 & 20 \\ 30 & 40 & 20 & 10 \end{bmatrix} = \begin{bmatrix} 550 & 660 & 500 & 210 \\ 540 & 600 & 400 & 180 \end{bmatrix}. \tag{4.14}$$

This example illustrates how the product **AB** is simply a case of obtaining the product of **A** with each column of **B** and setting the products alongside one another. Thus for

$$\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \mathbf{b}_3 \quad \mathbf{b}_4]$$

partitioned into columns in the manner of (4.5),

$$\mathbf{AB} = [\mathbf{Ab}_1 \quad \mathbf{Ab}_2 \quad \mathbf{Ab}_3 \quad \mathbf{Ab}_4].$$

Then, with **A** partitioned into its rows

$$\mathbf{A} = \begin{bmatrix} \alpha'_1 \\ \alpha'_2 \end{bmatrix} \quad \text{we have} \quad \mathbf{AB} = \begin{bmatrix} \alpha'_1\mathbf{b}_1 & \alpha'_1\mathbf{b}_2 & \alpha'_1\mathbf{b}_3 & \alpha'_1\mathbf{b}_4 \\ \alpha'_2\mathbf{b}_1 & \alpha'_2\mathbf{b}_2 & \alpha'_2\mathbf{b}_3 & \alpha'_2\mathbf{b}_4 \end{bmatrix}.$$

Hence, the element of **AB** in its *i*th row and *j*th column is the inner product $\alpha'_i\mathbf{b}_j$ of the *i*th row of **A** and the *j*th column of **B**. And this is true in general:

$$\begin{aligned} \mathbf{A}_{r\times c}\mathbf{B}_{c\times s} &= \{\alpha'_i\mathbf{b}_j\} \\ &= \{a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ic}b_{cj}\} \\ &= \left\{\sum_{k=1}^{c} a_{ik}b_{kj}\right\}, \quad \text{for} \quad i = 1, \ldots, r \quad \text{and} \quad j = 1, \ldots, s. \end{aligned} \tag{4.15}$$

The (ij)th element of **AB** can therefore be obtained by thinking of moving from element to element along the *i*th row of **A** and simultaneously down the *j*th column of **B**, summing the products of corresponding elements. The resulting sum is the (ij)th element of **AB**. Schematically the operation can be represented as follows: to obtain **AB** ,

$$\begin{bmatrix} \xrightarrow{i\text{th row}} \end{bmatrix}_{r\times c} \begin{bmatrix} \Big\downarrow \begin{matrix} j\text{th} \\ \text{column} \end{matrix} \end{bmatrix}_{c\times s} = \begin{bmatrix} \left\{ \begin{matrix} (ij)\text{th} \\ \text{element} \end{matrix} \right\} \end{bmatrix}_{r\times s}$$

for $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$. The arrows indicate moving along the *i*th row and simultaneously down the *j*th column, summing the products of corresponding elements to get the (ij)th element of the product.

Once again, this is a matrix operation defined only if a certain condition is met: the *i*th row of **A** (and hence all rows) must have the same number of elements as does the *j*th column of **B** (and hence all columns). Since the number of elements in a row of a matrix is the number of columns in the matrix (and the number of elements in a column is the number of rows), this means that there must be exactly as many columns in **A** as there are rows in **B**. Thus the matrix product **AB** is defined only if the number of columns of **A** equals the number of rows of **B**. Note also, particularly in the numerical examples, that the product **AB** has the same number of rows as **A** and the same number of columns as **B**. This is true in general.

The important consequences of the definition of matrix multiplication are therefore as follows. The product **AB** of two matrices **A** and **B** is defined and therefore exists only if the number of columns in **A** equals the number of rows in **B**; the matrices are then said to be *conformable for multiplication for the product* **AB**, and **AB** has the same number of rows as **A** and the same number of columns as **B**. And the (ij)th element of **AB** is the inner product of the ith row of **A** and jth column of **B**.

Example 4.1 *For*

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ -1 & 4 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 0 & 6 & 1 & 5 \\ 1 & 1 & 0 & 7 \\ 3 & 4 & 4 & 3 \end{bmatrix}$$

the element in the first row and first column of the product **AB** *is the inner product of the first row of* **A** *and the first column of* **B** *and is*

$$1(0) + 0(1) + 2(3) = 6;$$

that in the first row and second column is

$$1(6) + 0(1) + 2(4) = 14;$$

and the element of **AB** *in the second row and third column is*

$$-1(1) + 4(0) + 3(4) = 11.$$

In this way **AB** *is obtained as*

$$\mathbf{AB} = \begin{bmatrix} 6 & 14 & 9 & 11 \\ 13 & 10 & 11 & 32 \end{bmatrix}.$$

The reader should verify this result.

4.7.4 Existence of Matrix Products

Using subscript notation the product, $\boldsymbol{P}$, of two matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, of orders $r \times c$ and $c \times s$, respectively, can be written as

$$\mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \mathbf{P}_{r\times s},$$

a form which provides both for checking the conformability of **A** and **B** and for ascertaining the order of their product. Repeated use of this also simplifies determining the order of a matrix derived by multiplying several matrices together. Adjacent subscripts (which must be equal for conformability) “cancel out,” leaving the first and last subscripts as the order of the product. For example, the product

$$\mathbf{A}_{r\times c}\mathbf{B}_{c\times s}\mathbf{C}_{s\times t}\mathbf{D}_{t\times u}$$

is a matrix of order $r \times u$.

This notation also demonstrates what is by now readily apparent from the definition of matrix multiplication, namely that the product **BA** does not necessarily exist, even if **AB** does. For **BA** can be written as $\mathbf{B}_{c\times s}\mathbf{A}_{r\times c}$, which we see at once is a legitimate product only if $s = r$. Otherwise $\mathbf{BA}$ is not defined. There are therefore three situations regarding the product of two matrices $\mathbf{A}$ and $\mathbf{B}$. If $\mathbf{A}$ is of order $r \times c$:

i. $\mathbf{AB}$ exists only if $\mathbf{B}$ has c rows.

ii. $\mathbf{BA}$ exists only if $\mathbf{B}$ has r columns.

iii. $\mathbf{AB}$ and $\mathbf{BA}$ both exist only if $\mathbf{B}$ is $c \times r$.

A corollary to (iii) is that $\mathbf{A}^2$ exists only when $\mathbf{A}$ is square. Another corollary is that both $\mathbf{AB}$ and $\mathbf{BA}$ always exist and are of the same order when $\mathbf{A}$ and $\mathbf{B}$ are square and of the same order. But as shall be shown subsequently, the two products are not necessarily equal. Their inequality will be discussed when considering the commutative law of multiplication, but meanwhile we simply state that they are not in general equal.

As a means of distinction, $\mathbf{AB}$ is described as $\mathbf{A}$ *postmultiplied* by $\mathbf{B}$, or as $\mathbf{A}$ *multiplied on the right* by $\mathbf{B}$; and $\mathbf{BA}$ is either $\mathbf{A}$ *premultiplied* by $\mathbf{B}$, or $\mathbf{A}$ *multiplied on the left* by $\mathbf{B}$
.

4.7.5 Products With Vectors

Both the inner product of two vectors and the product of a matrix postmultiplied by a column vector are special cases of the general matrix product $\mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \mathbf{P}_{r\times s}$. Thus for the inner product of two vectors, $r = 1$ and $s = 1$, which means that $\mathbf{A}_{r\times c}$ becomes $\mathbf{A}_{1\times c}$, a row vector $(\mathbf{a}')_{1\times c}$, say; and $\mathbf{B}_{c\times s}$ becomes $\mathbf{B}_{c\times 1}$, a column vector $\mathbf{b}_{c\times 1}$. Thus we have the inner product

$$(\mathbf{a}')_{1\times c}\mathbf{b}_{c\times 1} = \mathbf{p}_{1\times 1},$$

a scalar. And the product in reverse order, where $\mathbf{a}$ and $\mathbf{b}$ can now be of different orders, is

$$\mathbf{b}_{c\times 1}(\mathbf{a}')_{1\times r} = \mathbf{P}_{c\times r},$$

a matrix: it is called the *outer product of* $\mathbf{b}$ *and* $\mathbf{a}'$. For

$$\mathbf{A}_{r\times c}\mathbf{b}_{c\times 1} = \mathbf{p}_{r\times 1},$$

the product being a column vector; similarly, a row vector postmultiplied by a matrix is a row vector:

$$(\mathbf{a}')_{1\times c}\mathbf{B}_{c\times r} = \mathbf{p}'_{1\times r}.$$

In words, these four results (for comformable products) are as follows:

i. A row vector postmultiplied by a column vector is a scalar.

ii. A column vector postmultiplied by a row vector is a matrix.

iii. A matrix postmultiplied by a column vector is a column vector.

iv. A row vector postmultiplied by a matrix is a row vector.

Example 4.2

Given

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix}, \quad \mathbf{a}' = \begin{bmatrix} 1 & 5 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 1 \end{bmatrix},$$

then

$$\text{(i) } \mathbf{a}'\mathbf{b} = \begin{bmatrix} 1 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = 8.$$

$$\text{(ii) } \mathbf{b}\mathbf{a}' = \begin{bmatrix} 3 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 5 \end{bmatrix} = \begin{bmatrix} 3 & 15 \\ 1 & 5 \end{bmatrix}.$$

$$\text{(iii) } \mathbf{A}\mathbf{b} = \begin{bmatrix} 1 & 2 \\ 7 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 25 \end{bmatrix}.$$

$$\text{(iv) } \mathbf{a}'\mathbf{B} = \begin{bmatrix} 1 & 5 \end{bmatrix} \begin{bmatrix} 6 & 3 & 1 \\ -1 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 13 & 26 \end{bmatrix}.$$

Illustration (Markov Chain). In the taxicab illustration of Example 2.1, suppose the driver lives in Town 1 and starts work from there each day. Denote this by a vector $\mathbf{x}_0' = [1 \quad 0]$. In general $\mathbf{x}'$ is called a *state probability vector* (or simply *state vector*). The subscript 0 represents, the beginning of the day, and the elements 1 and 0 are probabilities of starting in Town 1 and in Town 2, respectively. This being so, the probabilities of being in Town 1 or Town 2 after the morning's first fare are, using the transition probability matrix **P** of equation (2.5).

$$\mathbf{x}_1' = \mathbf{x}_0'\mathbf{P} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.2 & 0.8 \end{bmatrix}. \tag{4.16}$$

And after the second fare the probabilities are

$$\mathbf{x}_2' = \mathbf{x}_1'\mathbf{P} = \begin{bmatrix} 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 0.36 & 0.64 \end{bmatrix}. \tag{4.17}$$

Illustration (Linear Programming). An order for 1600 boxes of coronas, 500 boxes of half coronas, and 2000 boxes of cigarillos comes to a cigar-making company. The daily production and operating costs of the company's two small plants are shown in the following table. The company president needs to decide how many days each plant should operate to fill the order.

	Daily Production (100s of boxes)		
	Plant 1	**Plant 2**	**Order Received**
Product			
Corona	8	2	16
Half corona	1	1	5
Cigarillo	2	7	20
Daily operating cost	$1000	$2000	

The decision will be made on the criterion of filling the order with minimum cost. Let Plant 1 operate for x_1 days and Plant 2 for x_2 days to fill the order. Then to minimize cost we need to

$$\text{Minimize } f = 1000x_1 + 2000x_2,$$

and to fill the order x_1 and x_2 must satisfy

$$\begin{aligned} 8x_1 + 2x_2 &\geq 16 \\ x_1 + x_2 &\geq 5 \\ 2x_1 + 7x_2 &\geq 20, \end{aligned}$$

and because plants cannot operate for negative days we must also have

$$x_1 \geq 0 \quad \text{and} \quad x_2 \geq 0.$$

This is an example of what is called a linear programming problem. It can be stated generally as

$$\text{Minimize } f = \mathbf{c}'\mathbf{x}, \quad \text{subject to} \quad \mathbf{Ax} \geq \mathbf{r} \quad \text{and} \quad \mathbf{x} \geq 0 \tag{4.18}$$

where the inequality sign is used in vector and matrix statements in exactly the same manner as the equality sign: $\mathbf{u} \geq \mathbf{v}$ means each element of $\mathbf{u}$ is equal to or greater than the corresponding element of $\mathbf{v}$, that is, $u_i \geq v_i$ for all i. In our example (4.18) has

$$\mathbf{c} = \begin{bmatrix} 1000 \\ 2000 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 8 & 2 \\ 1 & 1 \\ 2 & 7 \end{bmatrix}, \quad \text{and} \quad \mathbf{r} = \begin{bmatrix} 16 \\ 5 \\ 20 \end{bmatrix}.$$

Linear programming has its own vast literature with many problems involving more than just two variables (as does our example), solution of which can be quite difficult. Nevertheless, being able to state problems succinctly in matrix terminology, as in (4.18), provides a basis for being able to work efficiently with characteristics of a problem that lead to its solution. For example, it is known that (4.18) is equivalent to the problem

$$\text{Maximize } g = \mathbf{r}'\mathbf{z}, \quad \text{subject to} \quad \mathbf{A}'\mathbf{z} \leq \mathbf{c} \quad \text{and} \quad \mathbf{z} \geq \mathbf{0}. \tag{4.19}$$

Furthermore, if $\mathbf{x}_0$ and $\mathbf{z}_0$ are solutions of (4.18) and (4.19), respectively, then the minimum f in (4.18) and the maximum g in (4.19) are equal: $f_{\min} = \mathbf{c}'\mathbf{x}_0 = g_{\max} = \mathbf{r}'\mathbf{z}_0$. Derivation of such results and development of solutions $\mathbf{x}_0$ and $\mathbf{z}_0$ are beyond the scope of this book, but the ability to state complicated problems and relationships in terms of simple matrix products is to be noted.

4.7.6 Products With Scalars

To the extent that a scalar can be considered a 1 × 1 matrix, certain cases of the scalar multiplication of a matrix in Section 4.4 are included in the general matrix product $\mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \mathbf{P}_{r\times s}$, to give $a_{1\times 1}\mathbf{b}'_{1\times c} = a\mathbf{b}'$ and $(\mathbf{a}')_{r\times 1}b_{1\times 1} = b\mathbf{a}'$.

4.7.7 Products With Null Matrices

For any matrix $\mathbf{A}_{r\times s}$, pre- or post-multiplication by a null matrix of appropriate order results in a null matrix. Thus if $\mathbf{0}_{c\times r}$ is a null matrix of order $c \times r$,

$$\mathbf{0}_{c\times r}\mathbf{A}_{r\times s} = \mathbf{0}_{c\times s} \quad \text{and} \quad \mathbf{A}_{r\times s}\mathbf{0}_{s\times p} = \mathbf{0}_{r\times p}.$$

For example,

$$\begin{bmatrix} 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}.$$

In writing this equality as **0A** = **0**, it is important to notice that the two **0**'s are not of the same order.

4.7.8 Products With Diagonal Matrices

A diagonal matrix is defined as a square matrix having all off-diagonal elements zero. Multiplication by a diagonal matrix is particularly easy: premultiplication of a matrix **A** by a diagonal matrix **D** gives a matrix whose rows are those of **A** multiplied by the respective diagonal elements of **D**. For example, in the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix}$$

of (4.10), rows represent prices of rats, mice, and rabbits bought locally or in a neighboring town. Suppose prices increase 5% locally and 20% in the neighboring town. Then with

$$\mathbf{D} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.20 \end{bmatrix}$$

the matrix of new prices is

$$\mathbf{DA} = \begin{bmatrix} 1.05 & 0 \\ 0 & 1.20 \end{bmatrix} \begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} = \begin{bmatrix} 3.15 & 1.05 & 10.50 \\ 2.40 & 2.40 & 9.60 \end{bmatrix}.$$

4.7.9 Identity Matrices

A diagonal matrix having all diagonal elements equal to unity is called an *identity matrix*, or sometimes a *unit matrix*. It is usually denoted by the letter $\mathbf{I}$, with a subscript for its order when necessary for clarity; for example,

$$\mathbf{I}_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{I}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

When $\mathbf{A}$ is of order $p \times q$

$$\mathbf{I}_p \mathbf{A}_{p\times q} = \mathbf{A}_{p\times q} \mathbf{I}_q = \mathbf{A}_{p\times q};$$

that is, multiplication of a matrix $\mathbf{A}$ by an (conformable) identity matrix does not alter $\mathbf{A}$. For example,

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 2 & -4 \\ 9 & 7 & 2 \end{bmatrix}.$$

4.7.10 The Transpose of a Product

The transpose of a product matrix is the product of the transposed matrices taken in reverse sequence, that is, $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Example 4.3

$$\mathbf{AB} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 0 & 2 & 4 \\ 3 & 0 & 7 \end{bmatrix} = \begin{bmatrix} -2 & 1 & -6 \\ 11 & 0 & 19 \end{bmatrix}$$

$$\mathbf{B}'\mathbf{A}' = \begin{bmatrix} 1 & 0 & 3 \\ 1 & 2 & 0 \\ 1 & 4 & 7 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ -1 & 3 \end{bmatrix} = \begin{bmatrix} -2 & 11 \\ 1 & 0 \\ -6 & 9 \end{bmatrix} = (\mathbf{AB})'.$$

Consideration of order and conformability for multiplication confirms this result. If $\mathbf{A}$ is $r \times s$ and $\mathbf{B}$ is $s \times t$, the product $\mathbf{P} = \mathbf{AB}$ is $r \times t$; that is, $\mathbf{A}_{r\times s}\mathbf{B}_{s\times t} = \mathbf{P}_{r\times t}$. But $\mathbf{A}'$ is $s \times r$ and $\mathbf{B}'$ is $t \times s$ and the only product to be derived from these is $(\mathbf{B}')_{t\times s}(\mathbf{A}')_{s\times r} = \mathbf{Q}_{t\times r}$ say. That $\mathbf{Q} = \mathbf{B}'\mathbf{A}'$ is indeed the transpose of $\mathbf{P} = \mathbf{AB}$ is apparent from the definition of multiplication: the (ij)th term of $\mathbf{Q}$ is the inner product of the ith row of $\mathbf{B}'$ and the jth column of $\mathbf{A}'$, which in turn is the inner product of the ith column of $\mathbf{B}$ and the jth row of $\mathbf{A}$, and this by the definition of multiplication is the (ji)th term of $\mathbf{P}$. Hence, $\mathbf{Q} = \mathbf{P}'$, or $\mathbf{B}'\mathbf{A}' = (\mathbf{AB})'$. More formally,

$$\mathbf{AB} = \mathbf{P} = \{p_{ij}\} = \left\{ \sum_{k=1}^{s} a_{ik} b_{kj} \right\},$$

This result for the transpose of the product of two matrices extends directly to the product of more than two. For example, $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$ and $(\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'$. Proof is left as an exercise for the reader.

4.7.11 The Trace of a Product

The trace of a matrix is defined in Section 4.3. Thus if $\mathbf{A} = \{a_{ij}\}$ for $i, j = 1, 2, \ldots, n$, the trace of $\mathbf{A}$ is $tr(\mathbf{A}) = \Sigma_{i=1}^{n} a_{ii}$, the sum of the diagonal elements. We now show that for the product $\mathbf{AB}$, $tr(\mathbf{AB}) = tr(\mathbf{BA})$ and hence $tr(\mathbf{ABC}) = tr(\mathbf{BCA}) = tr(\mathbf{CAB})$. Note that $tr(\mathbf{AB})$ exists only if $\mathbf{AB}$ is square, which occurs only when $\mathbf{A}$ is $r \times c$ and $\mathbf{B}$ is $c \times r$. Then if $\mathbf{AB} = \mathbf{P} = \{p_{ij}\}$ and $\mathbf{BA} = \mathbf{T} = \{t_{ij}\}$,

$$tr(\mathbf{AB}) = \sum_{i=1}^{r} p_{ii} = \sum_{i=1}^{r}\left(\sum_{j=1}^{c} a_{ij}b_{ji}\right) = \sum_{i=1}^{r}\left(\sum_{j=1}^{c} b_{ji}a_{ij}\right)$$
$$= \sum_{j=1}^{c}\left(\sum_{i=1}^{r} b_{ji}a_{ij}\right) = \sum_{j=1}^{c} t_{jj} = \mathrm{tr}(\mathbf{BA}).$$

Extension to products of three or more matrices is obvious.

Notice that the intermediate result

$$tr(\mathbf{AB}) = \sum_{i=1}^{r}\sum_{j=1}^{c} a_{ij}b_{ji}$$

can also, using $\mathbf{B}' = \{b'_{ij} = b_{ji}\}$, be expressed as

$$tr(\mathbf{AB}) = \sum_{i=1}^{r}\sum_{j=1}^{c} a_{ij}b'_{ij},$$

which is the sum of products of each element of $\mathbf{A}$ multiplied by the corresponding element of $\mathbf{B}'$. And if $\mathbf{B} = \mathbf{A}'$, we have

$$tr(\mathbf{AA}') = tr(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^{r}\sum_{j=1}^{c} a_{ij}^2,$$

that the trace of $\mathbf{AA}'$ (and of $\mathbf{A}'\mathbf{A}$) is the sum of squares of elements of $\mathbf{A}$. It is left as an exercise for the reader to verify these results numerically.

4.7.12 Powers of a Matrix

Since $\mathbf{A}_{r\times c}\mathbf{A}_{r\times c}$ exists only if $r = c$, that is, only if $\mathbf{A}$ is square, we see that $\mathbf{A}^2$ exists only when $\mathbf{A}$ is square; and then $\mathbf{A}^k$ exists for all positive integers k. And, in keeping with scalar arithmetic where $x^0 = 1$, we take $\mathbf{A}^0 = \mathbf{I}$ for $\mathbf{A}$ square.

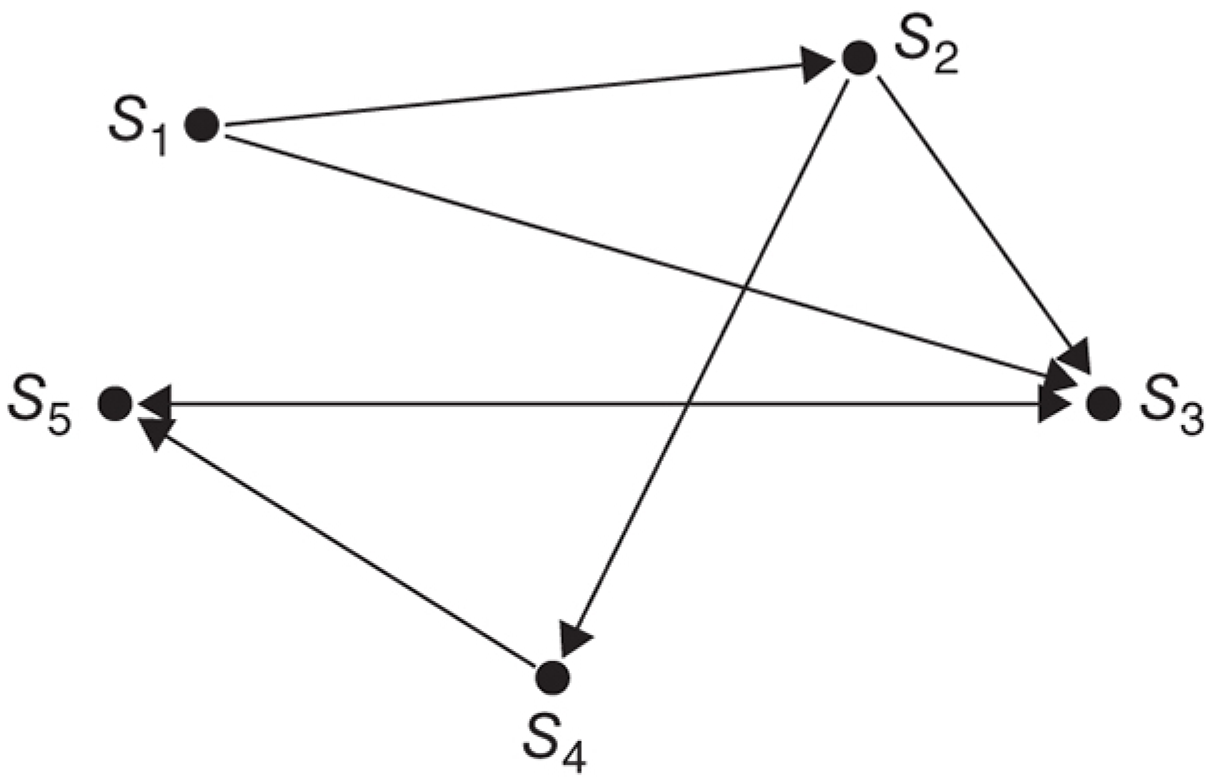
For example, in the taxicab illustration, it is clear from (4.16) that (4.17) is

$$\mathbf{x}_2' = \mathbf{x}_1'\mathbf{P} = \mathbf{x}_0'\mathbf{P}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix}^2$$

$$= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 0.36 & 0.64 \\ 0.32 & 0.68 \end{bmatrix} = \begin{bmatrix} 0.36 & 0.64 \end{bmatrix}.$$

Similarly, $\mathbf{x}_3' = \mathbf{x}_0'\mathbf{P}^3$, and in general $\mathbf{x}_n' = \mathbf{x}_0'\mathbf{P}^n$.

Illustration (Graph Theory). Suppose in a communications network of five stations messages can be sent only in the directions of the arrows of the accompanying diagram.



This chart of the possible message routes can also be represented by a matrix $\mathbf{T} = \{t_{ij}\}$ say, where $t_{ij} = 0$ except $t_{ij} = 1$ if a message can be sent from S_i to S_j. Hence,

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

In the rth power of $\mathbf{T}$, say $\mathbf{T}^r = \{t_{ij}^{(r)}\}$, the element $t^{(r)}_{ij}$ is then the number of ways of getting a message from station i to station j in exactly r steps. Thus,

$$\mathbf{T}^2 = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{T}^3 = \begin{bmatrix} 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

show that messages can be transmitted in two ways from S_2 to S_5 in two steps and in two ways from S_2 to S_3 in three steps. In this manner the pathways through a network can be counted simply by looking at powers of **T**. More than that, if there is no direct path from S_i to S_j, that is, $t_{ij} = 0$, the powers of **T** can be used to ascertain if there are any indirect paths. For example, $t_{25} = 0$ but $t^{(2)}{}_{25} = 2$, showing that the path from S_2 to S_5 cannot be traversed directly but there are two indirect routes of two steps each. Thus only if $\Sigma^{\infty}{}_{r=0} t_{ij}{}^{(r)} = 0$ is there no path at all from S_i to S_j.

Applications of matrices like **T** arise in many varied circumstances. For example, instead of the S's being stations in a communications network, they could be people in a social or business group, whereupon each arrow of the figure could represent the dominance of one person by another. Then $\mathbf{T}^2$ represents the two-stage dominance existing within the group of people; and so on. Or the figure could represent the spreading of a rumor between people or groups of people; or it could be the winning of sports contests; and so on. In this way, graph theory, supported by matrix representations and attendant matrix algebra, can provide insight to a host of real-life problems.

4.7.13 Partitioned Matrices

When matrices **A** and **B** are partitioned so that their submatrices are appropriately conformable for multiplication, the product **AB** can be expressed in partitioned form having submatrices that are functions of the submatrices of **A** and **B**. For example, if

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{21} \end{bmatrix},$$

then

$$\mathbf{AB} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} \\ \mathbf{B}_{21} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} \end{bmatrix}.$$

This means that the partitioning of **A** along its columns must be the same as that of **B** along its rows. Then $\mathbf{A}_{11}$ (and $\mathbf{A}_{21}$) has the same number of columns as $\mathbf{B}_{11}$ has rows, and $\mathbf{A}_{12}$ (and $\mathbf{A}_{22}$) has the same number of columns as $\mathbf{B}_{21}$ has rows. We see at once that when two matrices are appropriately partitioned the submatrices of their product are obtained by treating the submatrices of each of them as elements in a normal matrix product, and the individual elements of the product are derived in the usual way from the products of the submatrices.

In general, if **A** is $p \times q$ and is partitioned as

$$\mathbf{A}_{p\times q} = \{\mathbf{A}_{ij}(p_i \times q_j)\}, \quad \text{for} \quad i = 1, 2, \ldots, r \quad \text{and} \quad j = 1, 2, \ldots, c$$

with $\Sigma_{i=1}^{r} p_i = p$ and $\Sigma_{j=1}^{c} q_j = q$, where $p_i \times q_j$ is the order of the submatrix $\mathbf{A}_{ij}$, and likewise if

$$\mathbf{B}_{q\times s} = \{\mathbf{B}_{jk}(q_j \times s_k)\}, \quad \text{for} \quad j = 1, 2, \ldots, c \quad \text{and} \quad k = 1, 2, \ldots, d$$

with $\Sigma_{j=1}^{c} q_j = q$ and $\Sigma^{d}_{k=1} s_k = s$, then

$$(\mathbf{AB})_{p\times s} = \left\{\sum_{j=1}^{c} \mathbf{A}_{ij}\mathbf{B}_{jk}(p_i \times s_k)\right\},$$

for $i = 1, 2, \ldots, r$ and $k = 1, 2, \ldots, d$.

Illustration. Feller (1968, p. 439) gives an example of a matrix of transition probabilities that can be partitioned as

$$\mathbf{P} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{U}_1 & \mathbf{V}_1 & \mathbf{T} \end{bmatrix},$$

where **A** and **B** are also transition probability matrices. It is readily shown that for $\mathbf{U}_n = \mathbf{U}_1\mathbf{A}^{n-1} + \mathbf{T}\mathbf{U}_{n-1}$ and $\mathbf{V}_n = \mathbf{V}_1\mathbf{B}^{n-1} + \mathbf{T}\mathbf{V}_{n-1}$,

$$\mathbf{P}^n = \begin{bmatrix} \mathbf{A}^n & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B}^n & \mathbf{0} \\ \mathbf{U}_n & \mathbf{V}_n & \mathbf{T}^n \end{bmatrix}.$$

4.7.14 Hadamard Products

The definition of a product that has been presented at some length is the one most generally used. But because matrices are arrays of numbers, they provide opportunity for defining products in several different ways.

The Hadamard product of matrices $\mathbf{A} = \{a_{ij}\}$ and $\mathbf{B} = \{b_{ij}\}$ is defined only when **A** and **B** have the same order. It is the matrix of the element-by-element products of corresponding elements in **A** and **B**:

$$\mathbf{A} \cdot \mathbf{B} = \{a_{ij}b_{ij}\}.$$

Thus the (ij)th element of the Hadamard product $\mathbf{A} \cdot \mathbf{B}$ is the product of the (ij)th elements of **A** and **B**:

$$\begin{bmatrix} 2 & 0 & -1 \\ 4 & 1 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 3 & 7 \\ 2 & 0 & 6 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -7 \\ 8 & 0 & 18 \end{bmatrix}.$$

For example, suppose that in the taxicab illustration that the average fare between the two towns is \$8 and within Town 1 it is \$6 and within Town 2 it is \$5. Then, corresponding to the transition probability matrix, the reward matrix is

$$\mathbf{R} = \begin{bmatrix} 6 & 8 \\ 8 & 5 \end{bmatrix}.$$

Therefore, since the expected reward for a fare from Town i to Town j is $p_{ij}r_{ij}$, the matrix of expected rewards is the Hadamard product of **P** and **R**:

$$\mathbf{P} \cdot \mathbf{R} = \begin{bmatrix} .2 & .8 \\ .4 & .6 \end{bmatrix} \cdot \begin{bmatrix} 6 & 8 \\ 8 & 5 \end{bmatrix} = \begin{bmatrix} 1.2 & 6.4 \\ 3.2 & 3.0 \end{bmatrix}. \tag{4.20}$$

4.8 The Laws of Algebra

We now give formal consideration to the associative, commutative, and distributive laws of algebra as they relate to the addition and multiplication of matrices.

4.8.1 Associative Laws

The addition of matrices is associative provided the matrices are conformable for addition. For if $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ have the same order,

$$(\mathbf{A}+\mathbf{B})+\mathbf{C}=\{a_{ij}+b_{ij}\}+\{c_{ij}\}=\{a_{ij}+b_{ij}+c_{ij}\}=\mathbf{A}+\mathbf{B}+\mathbf{C}.$$

Also,

$$\{a_{ij}+b_{ij}+c_{ij}\}=\{a_{ij}\}+\{b_{ij}+c_{ij}\}=\mathbf{A}+(\mathbf{B}+\mathbf{C}),$$

so proving the associative law of addition.

In general, the laws of algebra that hold for matrices do so because matrix results follow directly from corresponding scalar results for their elements—as illustrated here. Further proofs in this section are therefore omitted.

The associative law is also true for multiplication, provided the matrices are conformable for multiplication. For if $\mathbf{A}$ is $p \times q$, $\mathbf{B}$ is $q \times r$, and $\mathbf{C}$ is $r \times s$, then $(\mathbf{AB})\mathbf{C}=\mathbf{A}(\mathbf{BC})=\mathbf{ABC}$.

4.8.2 The Distributive Law

The distributive law holds true. For example,

$$\mathbf{A}(\mathbf{B}+\mathbf{C})=\mathbf{AB}+\mathbf{AC}$$

provided both $\mathbf{B}$ and $\mathbf{C}$ are conformable for addition (necessarily of the same order) and $\mathbf{A}$ and $\mathbf{B}$ are conformable for multiplication (and hence $\mathbf{A}$ and $\mathbf{C}$ also).

4.8.3 Commutative Laws

Addition of matrices is commutative (provided the matrices are conformable for addition). If $\mathbf{A}$ and $\mathbf{B}$ are of the same order

$$\mathbf{A}+\mathbf{B}=\{a_{ij}+b_{ij}\}=\{b_{ij}+a_{ij}\}=\mathbf{B}+\mathbf{A}.$$

Multiplication of matrices is not in general commutative. As seen earlier, there are two possible products that can be derived from matrices $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{AB}$ and $\mathbf{BA}$, and if $\mathbf{A}$ is of order $r \times c$ both products exist only if $\mathbf{B}$ is of order $c \times r$. $\mathbf{AB}$ is then square, of order $r \times r$, and $\mathbf{BA}$ is also square, of order $c \times c$. Possible equality of $\mathbf{AB}$ and $\mathbf{BA}$ can therefore be considered only where $r = c$, in which case $\mathbf{A}$ and $\mathbf{B}$ are both square and have the same order $r \times r$. The products are then

$$\mathbf{AB} = \left\{ \sum_{k=1}^{r} a_{ik} b_{kj} \right\} \quad \text{and} \quad \mathbf{BA} = \left\{ \sum_{k=1}^{r} b_{ik} a_{kj} \right\} \quad \text{for} \quad i, j = 1, \ldots, r.$$

It can be seen that the (ij)th elements of these products do not necessarily have even a single term in common in their sums of products, let alone are they equal. Therefore, even when $\mathbf{AB}$ and $\mathbf{BA}$ both exist and are of the same order, they are not in general equal; for example,

$$\begin{aligned} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} &= \begin{bmatrix} 2 & -3 \\ 4 & -7 \end{bmatrix} \\ &\neq \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} -3 & -4 \\ -2 & -2 \end{bmatrix}. \end{aligned}$$

But in certain cases $\mathbf{AB}$ and $\mathbf{BA}$ *are* equal; for example,

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 8 \\ 12 & 18 \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 3 & 3 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}.$$

Two special cases of matrix multiplication being commutative are $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$ and $\mathbf{0A} = \mathbf{A0} = \mathbf{0}$ for $\mathbf{A}$ being square. When $\mathbf{A}$ is rectangular the former holds, with identity matrices of different orders, that is, $\mathbf{I}_r \mathbf{A}_{r \times c} = \mathbf{A}_{r \times c} \mathbf{I}_c = \mathbf{A}_{r \times c}$; the latter can be expressed more generally as two separate statements: $\mathbf{0}_{p \times r} \mathbf{A}_{r \times c} = \mathbf{0}_{p \times c}$ and $\mathbf{A}_{r \times c} \mathbf{0}_{c \times s} = \mathbf{0}_{r \times s}$.

4.9 Contrasts With Scalar Algebra

The definition of matrix multiplication leads to results in matrix algebra that have no counterpart in scalar algebra. In fact, some matrix results contradict their scalar analogues. We give some examples.

First, although in scalar algebra $ax + bx$ can be factored either as $x(a + b)$ or as $(a + b)x$, this duality is not generally possible with matrices:

$$\mathbf{AX} + \mathbf{BX} = (\mathbf{A} + \mathbf{B})\mathbf{X} \quad \text{and} \quad \mathbf{XA} + \mathbf{XB} = \mathbf{X}(\mathbf{A} + \mathbf{B}),$$

but

$$\mathbf{XP} + \mathbf{QX} \text{ generally does not have } \mathbf{X} \text{ as a factor.}$$

Another example of factoring is that, similar to $xy - x = x(y - 1)$ in scalar algebra, $\mathbf{XY} - \mathbf{X} = \mathbf{X}(\mathbf{Y} - \mathbf{I})$ in matrix algebra, but with the second term inside the parentheses being $\mathbf{I}$ and not the scalar 1. (It cannot be 1 because 1 is not conformable for subtraction from $\mathbf{Y}$.) Furthermore, the fact that it is $\mathbf{I}$ emphasizes, because of conformability for subtraction from $\mathbf{Y}$, that $\mathbf{Y}$ must be square. Of course this can also be gleaned directly from $\mathbf{XY} - \mathbf{X}$ itself: if $\mathbf{X}$ is $r \times c$, then $\mathbf{Y}$ must have c rows in order for $\mathbf{XY}$ to be defined, and $\mathbf{Y}$ must have c columns in order for $\mathbf{XY}$ and $\mathbf{X}$ to be conformable for subtraction. This simple example illustrates the need for constantly keeping conformability in mind, especially when matrix symbols do not have their orders attached.

Another consequence of matrix multiplication is that even when $\mathbf{AB}$ and $\mathbf{BA}$ both exist, they may not be equal. Thus for

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} \tag{4.21}$$

we have

$$\mathbf{AB} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \neq \mathbf{BA} = \begin{bmatrix} 2 & 2 \\ -2 & -2 \end{bmatrix}.$$

Notice here that even though $\mathbf{AB} = \mathbf{0}$, neither $\mathbf{A}$ nor $\mathbf{B}$ is $\mathbf{0}$. This illustrates an extremely important feature of matrices: the equation $\mathbf{AB} = \mathbf{0}$ does *not* always lead to the conclusion that $\mathbf{A}$ or $\mathbf{B}$ is $\mathbf{0}$, as would be the case with scalars. A further illustration arises from observing that for $\mathbf{A}$ and $\mathbf{B}$ of (4.21), we have $\mathbf{BA} = 2\mathbf{B}$. This can be rewritten as $\mathbf{BA} - 2\mathbf{B} = \mathbf{0}$, equivalent to $\mathbf{B}(\mathbf{A} - 2\mathbf{I}) = \mathbf{0}$. But from this we cannot conclude either that $\mathbf{A} - 2\mathbf{I}$ is $\mathbf{0}$ or that $\mathbf{B} = \mathbf{0}$. Similarly, not even the equation $\mathbf{X}^2 = \mathbf{0}$ means that $\mathbf{X} = \mathbf{0}$; for example, with

$$\mathbf{X} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix} \quad \text{we have} \quad \mathbf{X}^2 = \mathbf{0}.$$

Likewise $\mathbf{Y}^2 = \mathbf{I}$ implies neither $\mathbf{Y} = \mathbf{I}$ nor $\mathbf{Y} = -\mathbf{I}$; for example,

$$\mathbf{Y} = \begin{bmatrix} 1 & 0 \\ 4 & -1 \end{bmatrix}, \quad \text{but} \quad \mathbf{Y}^2 = \mathbf{I} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Similarly, we can have $\mathbf{M}^2 = \mathbf{M}$ with both $\mathbf{M} \neq \mathbf{I}$ and $\mathbf{M} \neq \mathbf{0}$; for example,

$$\mathbf{M} = \begin{bmatrix} 3 & -2 \\ 3 & -2 \end{bmatrix} = \mathbf{M}^2.$$

Such a matrix $\mathbf{M}$ with $\mathbf{M} = \mathbf{M}^2$ is said to be *idempotent*.

4.10 Direct Sum of Matrices

Direct sums and direct products are matrix operations defined in terms of partitioned matrices. They are discussed in this and the next section.

The direct sum of two matrices $\mathbf{A}$ and $\mathbf{B}$ is defined as

$$\mathbf{A} \oplus \mathbf{B} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \tag{4.22}$$

and extends very simply to more than two matrices:

$$\mathbf{A} \oplus \mathbf{B} \oplus \mathbf{C} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{C} \end{bmatrix}$$

and

$$\bigoplus_{i=1}^{k} \mathbf{A}_i = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & & \vdots \\ \vdots & & & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{A}_k \end{bmatrix} = \text{diag}\{\mathbf{A}_i\} \text{ for } i = 1, \ldots, k.$$

The definition (4.22) and its extensions apply whether or not the submatrices are of the same order; and all null matrices are of appropriate order. For example,

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \oplus \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & 8 & 9 \end{bmatrix}. \tag{4.23}$$

Transposing a direct sum gives the direct sum of the transposes. It is clear from (4.22) that $\mathbf{A} \oplus (-\mathbf{A}) \neq \mathbf{0}$ unless $\mathbf{A}$ is null. Also,

$$(\mathbf{A} \oplus \mathbf{B}) + (\mathbf{C} \oplus \mathbf{D}) = (\mathbf{A} + \mathbf{C}) \oplus (\mathbf{B} + \mathbf{D})$$

only if the necessary conditions of conformability for addition are met. Similarly,

$$(\mathbf{A} \oplus \mathbf{B})(\mathbf{C} \oplus \mathbf{D}) = \mathbf{AC} \oplus \mathbf{BD}$$

provided that conformability for multiplication is satisfied. The direct sum $\mathbf{A} \oplus \mathbf{B}$ is square only if $\mathbf{A}$ is $p \times q$ and $\mathbf{B}$ is $q \times p$. The determinant of $\mathbf{A} \oplus \mathbf{B}$ equals $|\mathbf{A}||\mathbf{B}|$ if both $\mathbf{A}$ and $\mathbf{B}$ are square, but otherwise it is zero or nonexistent.

4.11 Direct Product of Matrices

The direct product of two matrices $\mathbf{A}_{p\times q}$ and $\mathbf{B}_{m\times n}$ is defined as

$$\mathbf{A}_{p\times q} \otimes \mathbf{B}_{m\times n} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1q}\mathbf{B} \\ \vdots & & \vdots \\ a_{p1}\mathbf{B} & \cdots & a_{pq}\mathbf{B} \end{bmatrix} \tag{4.24}$$

and is sometimes called the Kronecker product or Zehfuss product [(see Henderson et al. (1981)]. Clearly, (4.24) is partitioned into as many submatrices as there are elements of $\mathbf{A}$, each submatrix being $\mathbf{B}$ multiplied by an element of $\mathbf{A}$. Therefore the elements of the direct product consist of all possible products of an element of $\mathbf{A}$ multiplied by an element of $\mathbf{B}$. It has order $pm \times qn$. For example,

$$\begin{bmatrix} 1 & 2 & 3 \end{bmatrix} \otimes \begin{bmatrix} 6 & 7 \\ 8 & 9 \end{bmatrix} = \begin{bmatrix} 6 & 7 & 12 & 14 & 18 & 21 \\ 8 & 9 & 16 & 18 & 24 & 27 \end{bmatrix}.$$

The transpose of a direct product is the direct product of the transposes—as is evident from transposing (4.24).

Direct products have many useful and interesting properties, some of which are as follows.

i. In contrast to $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$, we have $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$.

ii. For $\mathbf{x}$ and $\mathbf{y}$ being vectors: $\mathbf{x}' \otimes \mathbf{y} = \mathbf{y}\mathbf{x}' = \mathbf{y} \otimes \mathbf{x}'$.

iii. For λ being a scalar: $\lambda \otimes \mathbf{A} = \lambda\mathbf{A} = \mathbf{A} \otimes \lambda = \mathbf{A}\lambda$.

iv. For partitioned matrices, although

$$\begin{aligned} [\mathbf{A}_1 \quad \mathbf{A}_2] \otimes \mathbf{B} &= [\mathbf{A}_1 \otimes \mathbf{B} \quad \mathbf{A}_2 \otimes \mathbf{B}], \\ \mathbf{A} \otimes [\mathbf{B}_1 \quad \mathbf{B}_2] &\neq [\mathbf{A} \otimes \mathbf{B}_1 \quad \mathbf{A} \otimes \mathbf{B}_2]. \end{aligned}$$

v. Provided conformability requirements for regular matrix multiplication are satisfied, $(\mathbf{A} \otimes \mathbf{B})(\mathbf{X} \otimes \mathbf{Y}) = \mathbf{AX} \otimes \mathbf{BY}$.

vi. For $\mathbf{D}_k$ being a diagonal matrix of order $k \times k$ with elements d_i: $\mathbf{D}_k \otimes \mathbf{A} = d_1\mathbf{A} \oplus d_2\mathbf{A} \oplus \cdots \oplus d_k\mathbf{A}$.

vii. The trace obeys product rules: $tr(\mathbf{A} \otimes \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$.

viii. Provided $\mathbf{A}$ and $\mathbf{B}$ are square, $|\mathbf{A}_{p\times p} \otimes \mathbf{B}_{m\times m}| = |\mathbf{A}|^m |\mathbf{B}|^p$.

These results are readily illustrated with simple numerical examples, and most of them are not difficult to prove.

Sometimes $\mathbf{A} \otimes \mathbf{B}$ is referred to as the *right direct product* to distinguish it from $\mathbf{B} \otimes \mathbf{A}$ which is then called the *left direct product;* and on rare occasions the right-hand side of

(4.24) will be found defined as $\mathbf{B} \otimes \mathbf{A}$. Whatever names are used, it is apparent from (4.24) that for $\mathbf{A}_{p\times q} = \{a_{ij}\}$ and $\mathbf{B}_{m\times n} = \{b_{rs}\}$ the elements of both $\mathbf{A} \otimes \mathbf{B}$ and $\mathbf{B} \otimes \mathbf{A}$ consist of all possible products $a_{ij}b_{rs}$. In fact, $\mathbf{B} \otimes \mathbf{A}$ is simply $\mathbf{A} \otimes \mathbf{B}$ with the rows and columns each in a different order.

Illustration. A problem in genetics that uses direct products is concerned with *n* generations of random mating starting with the progeny obtained from crossing two autotetraploid plants which both have genotype AAaa. Normally the original plants would produce gametes AA, Aa, and aa in the proportion 1: 4: 1. But suppose the proportion is u: $1 - 2u$: u where, for example, u might take the value $(1 - \alpha)/6$, for α being a measure of "diploidization" of the plants: $\alpha = 0$ is the case of autotetraploids with chromosome segregation and $\alpha = 1$ is the diploid case with all gametes being Aa. The question now is, what are the genotypic frequencies in the population after *n* generations of random mating? Let $\mathbf{u}_i$ be the vector of gametic frequencies and $\mathbf{f}_i$ the vector of genotype frequencies in the ith generation of random mating, where $\mathbf{u}_0$ is the vector of gametic frequencies in the initial plants. Then

$$\mathbf{u}_0 = \begin{bmatrix} u \\ 1 - 2u \\ u \end{bmatrix}$$

and $\mathbf{f}_{i+1} = \mathbf{u}_i \otimes \mathbf{u}_i$ for $i = 0, 1, 2, \ldots, n$. Furthermore, the relationship between $\mathbf{u}_i$ and $\mathbf{f}_i$ at any generation is $\mathbf{u}_i = \mathbf{B}\mathbf{f}_i$ where

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

Thus

$$\begin{aligned} \mathbf{f}_i &= \mathbf{u}_{i-1} \otimes \mathbf{u}_{i-1} = \mathbf{B}\mathbf{f}_{i-1} \otimes \mathbf{B}\mathbf{f}_{i-1} = (\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-1} \otimes \mathbf{f}_{i-1}) \\ &= (\mathbf{B} \otimes \mathbf{B})[(\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2}) \otimes (\mathbf{B} \otimes \mathbf{B})(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2})] \\ &= (\mathbf{B} \otimes \mathbf{B})[(\mathbf{B} \otimes \mathbf{B}) \otimes (\mathbf{B} \otimes \mathbf{B})][(\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2}) \otimes (\mathbf{f}_{i-2} \otimes \mathbf{f}_{i-2})]. \end{aligned}$$

It is easily seen (and can be verified by induction) that

$$\mathbf{f}_i = \otimes^2\mathbf{B}(\otimes^4\mathbf{B})(\otimes^8\mathbf{B}) \cdots (\otimes^{2^{i-1}}\mathbf{B})(\otimes^{2^i}\mathbf{u}_0)$$

where $\otimes^n\mathbf{B}$ means the direct product of $n\mathbf{B}$'s.

4.12 The Inverse of a Matrix

Let $\boldsymbol{A} = (a_{ij})$ be a matrix of order $n \times n$ whose determinant is not equal to zero. This matrix is said to be *nonsingular*. The *inverse* of $\boldsymbol{A}$, denoted by $\boldsymbol{A}^{-1}$, is an $n \times n$ matrix that satisfies the condition $\boldsymbol{A}\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$. Such a matrix is unique.

The inverse of $\boldsymbol{A}$ can be computed as follows: let c_{ij} denote the cofactor of a_{ij} (see Section 3.3.3). Define the matrix $\boldsymbol{C}$ as $\boldsymbol{C} = (c_{ij})$. The inverse of $\boldsymbol{A}$ is given by

$$\boldsymbol{A}^{-1} = \frac{\boldsymbol{C}'}{|\boldsymbol{A}|},$$

where $\boldsymbol{C}'$ is the transpose of $\boldsymbol{C}$. The matrix $\boldsymbol{C}'$ is called the *adjugate*, or *adjoint*, matrix of $\boldsymbol{A}$, and is denoted by $adj\boldsymbol{A}$. For example, if $\boldsymbol{A}$ is the matrix

$$\boldsymbol{A} = \begin{bmatrix} 2 & 5 \\ 3 & 9 \end{bmatrix},$$

then

$$adj\boldsymbol{A} = \begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix},$$

and

$$\boldsymbol{A}^{-1} = \frac{1}{3}\begin{bmatrix} 9 & -5 \\ -3 & 2 \end{bmatrix},$$

since the determinant of $\boldsymbol{A}$ is equal to 3.

The following are some properties associated with the inverse operation [see, e.g., Harville (1997, Chapter 8)]:

(a) $(\boldsymbol{A}\boldsymbol{B})^{-1} = \boldsymbol{B}^{-1}\boldsymbol{A}^{-1}$.

(b) $(\boldsymbol{A}')^{-1} = (\boldsymbol{A}^{-1})'$.

(c) $|\boldsymbol{A}^{-1}| = \frac{1}{|\boldsymbol{A}|}$, if $\boldsymbol{A}$ is nonsingular.

(d) $(\boldsymbol{A}^{-1})^{-1} = \boldsymbol{A}$.

(e) $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$.

(f) $(\boldsymbol{A} \oplus \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \oplus \boldsymbol{B}^{-1}$.

(g) If $\boldsymbol{A}$ is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{ij} is of order $n_i \times n_j$ ($i, j = 1, 2$), then

$$|A| = \begin{cases} |A_{11}||A_{22} - A_{21}A_{11}^{-1}A_{12}|, & \text{if } A_{11} \text{ is nonsingular} \\ |A_{22}||A_{11} - A_{12}A_{22}^{-1}A_{21}|, & \text{if } A_{22} \text{ is nonsingular} \end{cases}$$

(h) If A is partitioned as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{ij} is of order $n_i \times n_j$ ($i, j = 1, 2$), then the inverse of A is given by

$$A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where

$$\begin{aligned} B_{11} &= (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \\ B_{12} &= -B_{11}A_{12}A_{22}^{-1}, \\ B_{21} &= -A_{22}^{-1}A_{21}B_{11}, \\ B_{22} &= A_{22}^{-1} + A_{22}^{-1}A_{21}B_{11}A_{12}A_{22}^{-1}, \end{aligned}$$

provided that the inverses of the matrices that appear in the above four B_{ij} expressions do exist.

The matrix

$$S = A_{22} - A_{21}A_{11}^{-1}A_{12} \tag{4.25}$$

occurring in Property (g) is called the *Schur complement* of A_{11} in

$$Q = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

when A_{11} is nonsingular. For singular A_{11}, the matrix $A_{22} - A_{21}A_{11}^{-}A_{12}$ is said to be the *generalized Schur complement* relative to A_{11}^{-} (see Chapter 8). Extensive properties of these complements are to be found in the literature, but this book is not the place for their development. Typical of results, for example, concern inverses of the form

$$(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} = A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1}. \tag{4.26}$$

Readers interested in these and other extensions are referred to Marsaglia and Styan(1974a,b), to Henderson and Searle (1981) for results on inverses, and to Ouellette

(1981) for general discussion of uses in statistics. All four papers have extensive references.

A further example of the algebra of inverses is a matrix analogue of the scalar result that for scalar $x \neq 1$

$$1 + x + x^2 + \cdots + x^{n-1} = (x^n - 1)/(x - 1).$$

A matrix counterpart is that, provided $(\mathbf{X} - \mathbf{I})^{-1}$ exists

$$\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1} = (\mathbf{X}^n - \mathbf{I})(\mathbf{X} - \mathbf{I})^{-1}. \tag{4.27}$$

This is established by noting that the product

$$\begin{aligned}(\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1})(\mathbf{X} - \mathbf{I}) \\ = \mathbf{X} + \mathbf{X}^2 + \mathbf{X}^3 + \cdots + \mathbf{X}^{n-1} + \mathbf{X}^n - (\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \cdots + \mathbf{X}^{n-1}) \\ = \mathbf{X}^n - \mathbf{I},\end{aligned} \tag{4.28}$$

and therefore if $(\mathbf{X} - \mathbf{I})^{-1}$ exists, postmultiplication of both sides of (4.28) by $(\mathbf{X} - \mathbf{1})^{-1}$ yields (4.27). Similarly, it can also be shown that

$$\mathbf{I} + \mathbf{X} + \mathbf{X}^2 + \ldots + \mathbf{X}^{n-1} = (\mathbf{X} - \mathbf{I})^{-1}(\mathbf{X}^n - \mathbf{I}).$$

4.13 Rank of a Matrix—Some Preliminary Results

This section discusses an important characteristic of a matrix, namely its rank. Before defining what the rank is, the following theorems are needed:

Theorem 4.1 *A set of linearly independent (LIN) vectors of n elements, each cannot contain more than n such vectors.*

Proof 4.1 Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ be n LIN vectors of order $n \times 1$. Let $\mathbf{u}_{n+1}$ be any other non-null vector of order $n \times 1$. We show that it and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ are linearly dependent.

Since $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n]$ has LIN columns, $|\mathbf{U}| \neq \mathbf{0}$ (why?) and $\mathbf{U}^{-1}$ exists. Let $\mathbf{q} = -\mathbf{U}^{-1}\mathbf{u}_{n+1} \neq \mathbf{0}$, because $\mathbf{u}_{n+1} \neq \mathbf{0}$; that is, not all elements of $\mathbf{q}$ are zero. Then $\mathbf{Uq} + \mathbf{u}_{n+1} = \mathbf{0}$, which can be rewritten as

$$q_1\mathbf{u}_1 + q_2\mathbf{u}_2 + \cdots + q_n\mathbf{u}_n + \mathbf{u}_{n+1} = \mathbf{0} \tag{4.29}$$

with not all the q's being zero, This means $\mathbf{u}_1, \mathbf{u}_2, \ldots,, \mathbf{u}_{n+1}$ are linearly dependent; that is, with $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ being LIN it is impossible to put another vector $\mathbf{u}_{n+1}$ with them and have all $n + 1$ vectors being LIN. ■

Corollary 4.1 *When p vectors of order n × 1 are LIN then p ≤ n.*

It is important to note that this theorem is *not* stating that there is only a single set of n vectors of order $n \times 1$ that are LIN. What it is saying is that if we do have a set of n LIN vectors of order $n \times 1$, then there is no larger set of LIN vectors of order $n \times 1$; that is, there are no sets of $n + 1, n + 2, n + 3, \cdots$ vectors of order $n \times 1$ that are LIN. Although there are many sets of n LIN vectors of order $n \times 1$, an infinite number of them in fact, for each of them it is impossible to put another vector (or vectors) with them and have the set, which then contains more than n vectors, still be LIN.

Example 4.4

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 3 \\ 5 \end{bmatrix}$$

are a set of two LIN vectors of order 2 × 1. Put any other vector of order 2 × 1 with these two vectors and the set will be linearly dependent, that is,

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 7 \\ 8 \end{bmatrix} \quad \text{are linearly dependent;}$$

$$\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 3 \\ 5 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} a \\ b \end{bmatrix} \quad \text{are linearly dependent;}$$

This latter result is true of course because $\begin{bmatrix} a \\ b \end{bmatrix}$ is a linear combination of $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$. But the theorem is not saying that $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 3 \\ 5 \end{bmatrix}$ are the only set of two LIN vectors of order 2 × 1. For example, $\begin{bmatrix} 5 \\ 7 \end{bmatrix}$ and $\begin{bmatrix} 9 \\ 13 \end{bmatrix}$ also form such a set; and any other vector of order 2 × 1 put with them forms a linearly dependent set of three vectors. This is so because that third vector is a linear combination of these two; for example,

$$\begin{bmatrix} a \\ b \end{bmatrix} = \tfrac{1}{2}(13a - 9b)\begin{bmatrix} 5 \\ 7 \end{bmatrix} + \tfrac{1}{2}(-7a + 5b)\begin{bmatrix} 9 \\ 13 \end{bmatrix}.$$

No matter what the values of *a* and *b* are, this expression holds true—that is, every second-order vector can be expressed as a linear combination of a set of two LIN vectors of order 2 × 1. In general, every nth-order vector can be expressed as a linear combination of any set of *n* independent vectors of order *n* × 1. The maximum number of non-null vectors in a set of independent vectors is therefore *n*, the order of the vectors. This is simply a restatement of the theorem.

4.14 The Number of LIN Rows and Columns in a Matrix

Explanation has been given of how a determinant is zero when any of its rows (or columns) are linear combinations of other rows (or columns). In other words, a determinant is zero when its rows (or columns) do not form a set of LIN vectors. Evidently, therefore, a determinant cannot have both its rows forming a dependent set and its columns an independent set, a statement which prompts the more general question of the relationship between the number of LIN rows of a matrix and the number of LIN columns. The relationship is simple.

Theorem 4.2 *The number of LIN rows in a matrix is the same as the number of LIN columns.*

Before proving this, notice that independence of rows (columns) is a property of rows that is unrelated to their sequence within a matrix. For example, because the rows of

$$\begin{bmatrix} 1 & 2 & 3 \\ 6 & 9 & 14 \\ 3 & 0 & 1 \end{bmatrix} \quad \text{are LIN, so are the rows of} \quad \begin{bmatrix} 6 & 9 & 14 \\ 3 & 0 & 1 \\ 1 & 2 & 3 \end{bmatrix}.$$

Therefore, insofar as general discussion of independence properties is concerned, there is no loss of generality for a matrix that has k LIN rows in assuming that they are the first k rows. This assumption is therefore often made when discussing properties and consequences of independence of rows (columns) of a matrix.

We now prove the theorem.

Proof Let $\mathbf{A}_{p \times q}$ have k LIN rows and m LIN columns. We show that $k = m$.

Assume that the first k rows $\mathbf{A}$ are LIN and similarly the first m columns. Then partition $\mathbf{A}$ as

$$\mathbf{A}_{p\times q} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X}_{k\times m} & \mathbf{Y}_{k\times(q-m)} \\ \mathbf{Z}_{(p-k)\times m} & \mathbf{W}_{(p-k)\times(q-m)} \end{bmatrix} \leftarrow k \,\textbf{LIN}\text{ rows.} \qquad (4.30)$$

$$\uparrow$$

$$m\,\text{LIN columns}$$

Thus the k rows of $\mathbf{A}$ through $\mathbf{X}$ and $\mathbf{Y}$ are LIN (as are the m columns of $\mathbf{A}$ through $\mathbf{X}$ and $\mathbf{Z}$). Since $\mathbf{A}$ has only k LIN rows the other rows of $\mathbf{A}$ (those through $\mathbf{Z}$ and $\mathbf{W}$) are linear combinations of the first k rows. In particular, the rows of $\mathbf{Z}$ are linear combinations of the rows of $\mathbf{X}$. Hence, these rows can be expressed as $\mathbf{Z} = \mathbf{TX}$ for some matrix $\mathbf{T}$. Now assume that the columns of $\mathbf{X}$ are linearly dependent, that is,

$$\mathbf{Xa} = \mathbf{0} \quad \text{for some vector} \quad \mathbf{a} \neq \mathbf{0}. \qquad (4.31)$$

Then $\mathbf{Za} = \mathbf{TXa} = \mathbf{0}$ and so

$$\begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix} \mathbf{a} = \mathbf{0} \quad \text{for that same} \quad \mathbf{a} \neq \mathbf{0}.$$

But this is a statement of the linear dependence of the columns of $\begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix}$, that is, of the first m columns of **A**. These columns, however, have been taken in (4.30) as being LIN. This is a contradiction, and so assumption (4.31), from which it is derived, is false; that is, the columns of **X** are not dependent. Hence they must be LIN.

Having shown the columns of **X** to be LIN, observe from (4.30) that there are m of them and they are of order $k \times 1$. Hence by the theorem in the preceding section, $m \leq k$. A similar argument based on the rows of **X** and **Y**, rather than the columns of **X** and **Y**, shows that $k \leq m$. Hence $m = k$.

It is important to notice that this theorem says nothing about which rows (columns) of a matrix are LIN—it is concerned solely with how many of them are LIN. This means, for example, that if there are two LIN rows in a matrix of order 5×4 then once two rows are ascertained as being LIN the other three rows can be expressed as linear combinations of those two. And there may be (there usually is) more than one set of two rows that are LIN. For example, in

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 0 & 1 \\ 1 & -1 & 3 & 2 \\ 3 & 0 & 6 & 5 \\ 2 & 1 & 3 & 3 \\ 6 & 0 & 12 & 10 \end{bmatrix}$$

rows 1 and 2 are LIN, and each of rows 3, 4, and 5 is a linear combination of those first two. Rows 1 and 3 are also LIN, and the remaining rows are linear combinations of them; but not all pairs of rows are LIN. For example, rows 3 and 5 are not linearly independent.

4.15 Determination of The Rank of a Matrix

The theorem in Section 4.14 shows that every matrix has the same number of linearly independent rows as it does linearly independent columns.

Definition 4.1

The *rank* of a matrix is the number of linearly independent rows (and columns) in the matrix.

Notation: The rank of **A** will be denoted equivalently by $r_{\mathbf{A}}$ or $r(\mathbf{A})$. Thus if $r_{\mathbf{A}} \equiv r(\mathbf{A}) = k$, then **A** has k LIN rows and k LIN columns. The symbol r is often used for rank, that is, $r_{\mathbf{A}} \equiv r$.

Notice again that no specific set of LIN rows (columns) is identified by knowing the rank of a matrix. Rank indicates only how many are LIN and not where they are located in the matrix. The following properties and consequences of rank are important.

i. $r_{\mathbf{A}}$ is a positive integer, except that $r_{\mathbf{0}}$ is defined as $r_{\mathbf{0}} = 0$.

ii. $r(\mathbf{A}_{p\times q}) \leq p$ and $\leq q$: the rank of a matrix equals or is less than the smaller of its number of rows or columns.

iii. $r(\mathbf{A}_{n\times n}) \leq n$: a square matrix has rank not exceeding its order.

iv. When $r_{\mathbf{A}} = r \neq 0$ there is at least one square submatrix of **A** having order $r \times r$ that is nonsingular. Equation (4.30) with $k = m = r$ is

$$\mathbf{A}_{p\times q} = \begin{bmatrix} \mathbf{X}_{r\times r} & \mathbf{Y}_{r\times(q-r)} \\ \mathbf{Z}_{(p-r)\times r} & \mathbf{W}_{(p-r)\times(q-r)} \end{bmatrix} \tag{4.32}$$

and $\mathbf{X}_{r\times r}$, the intersection of r linearly independent rows and r linearly columns, is nonsingular. All square submatrices of order greater than $r \times r$ are singular.

v. When $r(\mathbf{A}_{n\times n}) = n$ then by (iv) **A** is nonsingular, that is, $\mathbf{A}^{-1}$ exists. [In (4.32) $\mathbf{A} \equiv \mathbf{X}$.]

vi. When $r(\mathbf{A}_{n\times n}) < n$ then **A** is singular and $\mathbf{A}^{-1}$ does not exist.

vii. When $r(\mathbf{A}_{p\times q}) = p < q$, **A** is said to have *full row rank*, or to be of full row rank. Its rank equals its number of rows.

viii. When $r(\mathbf{A}_{p\times q}) = q < p$, **A** is said to have *full column rank*, or to be of full column rank. Its rank equals its number of columns.

ix. When $r(\mathbf{A}_{n\times n}) = n$, **A** is said to have *full rank*, or to be of full rank. Its rank equals its order, it is nonsingular, its inverse exists, and it is said to be *invertible*.

x. $r(\mathbf{A}) = r(\mathbf{A}')$.

xi. The rank of $\boldsymbol{A}$ does not change if it is pre-multiplied or post-multiplied by a nonsingular matrix. Thus, if $\boldsymbol{A}$ is an $m \times n$ matrix, and $\boldsymbol{B}$ and $\boldsymbol{C}$ are nonsingular matrices of orders $m \times m$ and $n \times n$, respectively, then $r(\boldsymbol{A}) = r(\boldsymbol{BA}) = r(\boldsymbol{AC})$.

xii. The ranks of $\boldsymbol{AA'}$ and $\boldsymbol{A'A}$ are equal and each is equal to $r(\boldsymbol{A})$.

xiii. For any matrices, $\boldsymbol{A}_1, \boldsymbol{A}_2, \cdots, \boldsymbol{A}_k$, having the same number of rows and columns,

$$r\left(\sum_{i=1}^{k} \boldsymbol{A}_i\right) \leq \sum_{i=1}^{k} r(\boldsymbol{A}_i).$$

xiv. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of orders $m \times n$ and $n \times q$, respectively, then

$$r(\boldsymbol{A}) + r(\boldsymbol{B}) - n \leq r(\boldsymbol{AB}) \leq \min\{r(\boldsymbol{A}), r(\boldsymbol{B})\}$$

This is known as *Sylvester's law*.

xv. $r(\boldsymbol{A} \otimes \boldsymbol{B}) = r(\boldsymbol{A})r(\boldsymbol{B})$.

xvi. $r(\boldsymbol{A} \oplus \boldsymbol{B}) = r(\boldsymbol{A}) + r(\boldsymbol{B})$.

Rank is one of the most useful and important characteristics of any matrix. It occurs again and again in this book and plays a vital role throughout all aspects of matrix algebra. For example, from items (vi) and (ix) we see at once that ascertaining whether $|\mathbf{A}|$ is zero or not for determining the existence of $\mathbf{A}^{-1}$ can be replaced by ascertaining whether $r_{\mathbf{A}} < n$or $r_{\mathbf{A}} = n$. And almost always it is far easier to work with rank than determinants.

4.16 Rank and Inverse Matrices

A square matrix has an inverse if and only if its rank equals its order. This and other equivalent statements are summarized in Table 4.2. In each half of the table any one of the statements implies all the others: the first and second statements are basically definitional and the last six are equivalences. Hence whenever assurance is needed for the existence of $\mathbf{A}^{-1}$, we need only to establish any one of the last six statements in the first column of the table. The easiest of these is usually that concerning rank: when $r_{\mathbf{A}} = n$ then $\mathbf{A}^{-1}$ exists, and when $r_{\mathbf{A}} < n$ then $\mathbf{A}^{-1}$ does not exist. The problem of ascertaining the existence of an inverse is therefore equivalent to ascertaining if the rank of a square matrix is less than its order.

More generally, there are often occasions when the rank of a matrix is needed exactly. Although locating $r_{\mathbf{A}}$ linearly independent rows in $\mathbf{A}$ may not always be easy, deriving $r_{\mathbf{A}}$ itself is conceptually not difficult.

Table 4.2 Equivalent Statements for the Existence of $\mathbf{A}^{-1}$ of Order $n \times n$

Inverse Existing	**Inverse Not Existing**
$\mathbf{A}^{-1}$ exists	$\mathbf{A}^{-1}$ does not exist
$\mathbf{A}$ is nonsingular	$\mathbf{A}$ is singular
$\lvert\mathbf{A}\rvert \neq 0$	$\lvert\mathbf{A}\rvert = 0$
$\mathbf{A}$ has full rank	$\mathbf{A}$ has less than full rank
$r_{\mathbf{A}} = n$	$r_{\mathbf{A}} < n$
$\mathbf{A}$ has n linearly independent rows	$\mathbf{A}$ has fewer than n linearly independent rows
$\mathbf{A}$ has n linearly independent columns	$\mathbf{A}$ has fewer than n linearly independent columns
$\mathbf{Ax} = \mathbf{0}$ has sole solution, $\mathbf{x} = \mathbf{0}$	$\mathbf{Ax} = \mathbf{0}$ has many solutions, $\mathbf{x} \neq \mathbf{0}$

4.17 Permutation Matrices

Proof of the theorem in Section 4.14 begins by assuming that all the linearly independent rows come first in a matrix, and the linearly independent columns likewise. But this is not always so. For example, in

$$\boldsymbol{M} = \begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 4 & 4 & 12 \\ 2 & 2 & 5 \end{bmatrix}$$

there are two linearly independent rows and two linearly independent columns—but not the first two. Now consider $\boldsymbol{E}_{24}\boldsymbol{M}$ for

$$\boldsymbol{E}_{24} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

$$\boldsymbol{E}_{24}\boldsymbol{M} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1 & 1 & 3 \\ 1 & 1 & 3 \\ 4 & 4 & 12 \\ 2 & 2 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 2 & 5 \\ 4 & 4 & 12 \\ 1 & 1 & 3 \end{bmatrix}. \tag{4.33}$$

Note that $\boldsymbol{E}_{24}$ is an identity matrix with its second and fourth rows interchanged, and $\boldsymbol{E}_{24}\boldsymbol{M}$ is $\boldsymbol{M}$ with those same two rows interchanged.

$\boldsymbol{E}_{24}$ of (4.33) exemplifies what shall be called an *elementary permutation matrix* in general, $\boldsymbol{E}_{rs}$ is an identity matrix with its rth and sth rows interchanged. And $\boldsymbol{E}_{rs}\boldsymbol{M}$ is $\boldsymbol{M}$ with those same two rows interchanged. (The order of $\boldsymbol{E}_{rs}$ is determined by the product it is used in.) By virtue of its definition, $\boldsymbol{E}_{rs}$, is always symmetric. It is also orthogonal (see Section 5.4). Thus $\boldsymbol{E}_{rs}\boldsymbol{E}'_{rs} = \boldsymbol{E}_{rs}\boldsymbol{E}_{rs} = \boldsymbol{I}$ the second equality being true because $\boldsymbol{E}_{rs}\boldsymbol{E}_{rs}$ is $\boldsymbol{E}_{rs}$ with its rth and sth rows interchanged and so is $\boldsymbol{I}$, that is, $\boldsymbol{E}_{rs}\boldsymbol{E}_{rs} = \boldsymbol{I}$.

In the same way that premultiplication of $\boldsymbol{M}$ by $\boldsymbol{E}_{rs}$ interchanges rows *r* and *s* of $\boldsymbol{M}$, so does postmultiplication interchange columns. Thus using (4.33),

$$\boldsymbol{E}_{24}\boldsymbol{M}\boldsymbol{E}_{23} = \begin{bmatrix} 1 & 1 & 3 \\ 2 & 2 & 5 \\ 4 & 4 & 12 \\ 1 & 1 & 3 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 3 & 1 \\ 2 & 5 & 2 \\ 4 & 12 & 4 \\ 1 & 3 & 1 \end{bmatrix} \tag{4.34}$$

is $\boldsymbol{E}_{24}\boldsymbol{M}$ with its second and third columns interchanged.

Note in (4.34) that $\mathbf{E}_{24}\mathbf{M}\mathbf{E}_{23}$ has its first two rows linearly independent and its first two columns linearly independent also. Hence by premultiplying **M** by one elementary permutation matrix and postmultiplying it by another we get a matrix to which the argument of the proof of the theorem applies. And so, because the theorem applies only to the number of linearly independent rows and linearly independent columns, whatever holds for (4.34) in this connection also holds for **M**.

This use of permutation matrices extends to cases involving more than just a single interchange of rows and/or columns. Consider

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 3 & 1 \\ 1 & 1 & 3 & 2 \\ 3 & 3 & 9 & 6 \\ 2 & 2 & 5 & 4 \\ 1 & 1 & 7 & 8 \end{bmatrix}. \tag{4.35}$$

In premultiplying **A** first by $\mathbf{E}_{34}$ (to interchange rows 3 and 4) and then by $\mathbf{E}_{25}$ (to interchange rows 2 and 5), define **P** as $\mathbf{P} = \mathbf{E}_{25}\mathbf{E}_{34}$. Then,

$$\mathbf{PA} = \mathbf{E}_{25}\mathbf{E}_{34}\mathbf{A} = \mathbf{E}_{25}\begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & 1 & 3 & 2 \\ 2 & 2 & 5 & 4 \\ 3 & 3 & 9 & 6 \\ 1 & 1 & 7 & 8 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 3 & 2 \\ 1 & 1 & 7 & 8 \\ 2 & 2 & 5 & 4 \\ 3 & 3 & 9 & 6 \\ 1 & 1 & 3 & 2 \end{bmatrix},$$

where

$$\mathbf{P} = \mathbf{E}_{25}\mathbf{E}_{34} = \mathbf{E}_{25}\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}.$$

Then, for $\mathbf{Q} = \mathbf{E}_{24}$,

$$\mathbf{PAQ} = \begin{bmatrix} 1 & 2 & 3 & 1 \\ 1 & 8 & 7 & 1 \\ 2 & 4 & 5 & 2 \\ 3 & 8 & 9 & 3 \\ 1 & 2 & 3 & 1 \end{bmatrix}$$

and this is a matrix with its first three rows linearly independent and its first three columns likewise.

In this case **P** is a product of two **E**-matrices. It is an example of a *permutation matrix* which, in general, is an identity matrix with its rows resequenced in some order. Because such a matrix is always a product of elementary permutation matrices (the **E**-matrices), **P** is not necessarily symmetric, but it is always orthogonal—because it is a product of orthogonal **E**-matrices (see Exercise 9). Furthermore, $\mathbf{P}^{-1} = \mathbf{P}'$ is also a permutation matrix, because although **P** is defined as an identity matrix with its rows resequenced, it is also an identity matrix with its columns resequenced. Therefore $\mathbf{P}'$ is an identity matrix with its rows resequenced and so $\mathbf{P}^{-1} = \mathbf{P}'$ is a permutation matrix, too.

One of the great uses of permutation matrices is that for situations like that of the theorem in Section 4.14, permutation matrices provide a mechanism for resequencing rows and columns in a matrix so that a matrix having k linearly independent rows can be resequenced into one having its first k rows and its first k columns linearly independent. Properties of the permutation matrices then allow many properties concerning linear independence of the resequenced matrices to also apply to the original matrix.

4.18 Full-Rank Factorization

An immediate consequence of the notion of rank is that a $p \times q$ matrix of rank $r \neq 0$ can be partitioned into a group of r independent rows and a group of $p - r$ rows that are linear combinations of the first group. This leads to a useful factorization.

4.18.1 Basic Development

Equation (4.32) shows the partitioning of $\mathbf{A}_{p \times q}$ of rank r when the first r rows and first r columns are assumed to be linearly independent, and where **X** is nonsingular. The first r rows in (4.32) are those of [**X Y**] and are linearly independent, and so the rows of [**Z W**] are linear combinations of those of [**X Y**]; hence for some matrix **F**,

$$[\mathbf{Z} \quad \mathbf{W}] = \mathbf{F}[\mathbf{X} \quad \mathbf{Y}]. \tag{4.36}$$

Similar reasoning applied to the columns of **A** gives, for some matrix **H**,

$$\begin{bmatrix} \mathbf{Y} \\ \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X} \\ \mathbf{Z} \end{bmatrix} \mathbf{H}. \tag{4.37}$$

But (4.36) has **Z** = **FX** and **W** = **FY**, and (4.37) has **Y** = **XH**. Hence **W** = **FY** = **FXH**, and so therefore

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{XH} \\ \mathbf{FX} & \mathbf{FXH} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{F} \end{bmatrix} \mathbf{X}[\mathbf{I} \quad \mathbf{H}]. \tag{4.38}$$

Furthermore, because X (4.32) is nonsingular, its inverse $\mathbf{X}^{-1}$ exists, so that **Z** = **FX** and **Y** = **XH** give $\mathbf{F} = \mathbf{ZX}^{-1}$ and $\mathbf{H} = \mathbf{X}^{-1}\mathbf{Y}$ and then $\mathbf{W} = \mathbf{FY} = \mathbf{ZX}^{-1}\mathbf{Y}$. Hence, (4.38) can be written as

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Z} & \mathbf{ZX}^{-1}\mathbf{Y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} \\ \mathbf{ZX}^{-1} \end{bmatrix} \mathbf{X}[\mathbf{I} \quad \mathbf{X}^{-1}\mathbf{Y}]. \tag{4.39}$$

Equation (4.38) can be further rewritten in two equivalent ways as

$$\mathbf{A} = \begin{bmatrix} \mathbf{X} \\ \mathbf{FX} \end{bmatrix} [\mathbf{I} \quad \mathbf{H}] = \begin{bmatrix} \mathbf{I} \\ \mathbf{F} \end{bmatrix} [\mathbf{X} \quad \mathbf{XH}], \tag{4.40}$$

each of which is of the form

$$\mathbf{A}_{p\times q} = \mathbf{K}_{p\times r}\mathbf{L}_{r\times q} \tag{4.41}$$

where **K** has full column rank $r = r_A$, and **L** has full row rank r. We call the matrix product in (4.41) the *full-rank factorization*, after Ben-Israel and Greville(1974, p. 22). It has also been called the *full-rank decomposition* by Marsaglia and Styan (1974a, p. 271). Whatever its name, it can always be done (see Section 4.18.2, which follows) and it has many uses in matrix algebra.

Example 4.5 *Equation (4.39) is illustrated by*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 1 & 5 \\ 2 & 5 & 1 & 14 \\ 4 & 9 & 3 & 24 \end{bmatrix} \tag{4.42}$$

$$= \begin{bmatrix} & \mathbf{I}_2 & \\ (4 \quad 9) & \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^{-1} & \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \left[\mathbf{I}_2 \quad \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 1 & 5 \\ 1 & 14 \end{pmatrix} \right]$$

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

and from (4.40) we have the two equivalent forms of (4.42):

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & 1 & 5 \\ 2 & 5 & 1 & 14 \end{bmatrix} \tag{4.43}$$

$$= \begin{bmatrix} 1 & 2 \\ 2 & 5 \\ 4 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 3 & -3 \\ 0 & 1 & -1 & 4 \end{bmatrix}.$$

A particular case of interest is when $r_{\mathbf{A}} = 1$, *whereupon* $\mathbf{A} = \mathbf{xy}'$; *that is,*

$$\begin{bmatrix} 2 & -2 & 4 \\ 3 & -3 & 6 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix} [1 \quad -1 \quad 2].$$

4.18.2 The General Case

Development of (4.36)–(4.40) rests upon the first r rows (and r columns) of $\mathbf{A}$ in (4.32) being linearly independent. But suppose this is not the case, as in (4.35). It is in just such a situation that permutation matrices play their part. Let $\mathbf{M} = \mathbf{PAQ}$ where $\mathbf{P}$ and $\mathbf{Q}$ are permutation matrices, as discussed in Section 4.17, and where $\mathbf{M}$ has the form (4.32). Then $\mathbf{M}$ can be expressed as $\mathbf{M} = \mathbf{KL}$ as in (4.41). Therefore, on using the orthogonality of $\mathbf{P}$ and $\mathbf{Q}$ (that is, $\mathbf{P}^{-1} = \mathbf{P}'$, as at the end of Section 4.17,

$$\mathbf{A} = \mathbf{P}^{-1}\mathbf{M}\mathbf{Q}^{-1} = \mathbf{P}'\mathbf{KLQ}' = (\mathbf{P}'\mathbf{K})(\mathbf{LQ}') \tag{4.44}$$

where $\mathbf{P}'\mathbf{K}$ and $\mathbf{LQ}'$ here play the same roles as $\mathbf{K}$ and $\mathbf{L}$ do in (4.41). Hence, (4.41) can be derived quite generally, for any matrix.

4.18.3 Matrices of Full Row (Column) Rank

When $\mathbf{A}_{p \times q}$ has full row rank, there will be no $\mathbf{Z}$ and $\mathbf{W}$ in (4.32); and $\mathbf{K}$ of (4.41) will be an identity matrix. Then (4.32) is $\mathbf{A} = [\mathbf{X}\ \mathbf{Y}]$, and if $\mathbf{X}$ is singular, postmultiplication of $\mathbf{A}$ by a permutation matrix $\mathbf{Q}$ can lead to the partitioning

$$\mathbf{AQ} = [\mathbf{M} \quad \mathbf{L}]$$

where $\mathbf{M}$ is nonsingular. We then have the following lemma.

Lemma 4.1 *A matrix of full row rank can always be written as a product of matrices one of which has the partitioned form [**I H**] for some matrix **H**.*

Proof. From the preceding equations $\mathbf{A} = [\mathbf{M} \quad \mathbf{L}]\mathbf{Q}^{-1} = \mathbf{M}[\mathbf{I} \quad \mathbf{M}^{-1}\mathbf{L}]\mathbf{Q}' = \mathbf{M}[\mathbf{I} \quad \mathbf{H}]\mathbf{Q}'$ for $\mathbf{H} = \mathbf{M}^{-1}\mathbf{L}$, and we have the desired result. ($\mathbf{M}^{-1}$ exists because $\mathbf{M}$ is nonsingular, and $\mathbf{Q}^{-1} = \mathbf{Q}'$ because $\mathbf{Q}$ is a permutation matrix.)

Exercises

1. For $\mathbf{A} = \begin{bmatrix} 3 & 6 \\ 2 & 1 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & -1 & -1 & 1 \end{bmatrix}$, $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}$ and $\mathbf{y} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$

show that

a. $\mathbf{AB} = \begin{bmatrix} 3 & -6 & 3 & 12 \\ 2 & -1 & 5 & 5 \end{bmatrix}$ and $\mathbf{A'B} = \begin{bmatrix} 3 & -2 & 7 & 8 \\ 6 & -1 & 17 & 13 \end{bmatrix}$;

b. $(\mathbf{A} + \mathbf{A}')\mathbf{B} = \begin{bmatrix} 6 & 8 \\ 8 & 2 \end{bmatrix} \mathbf{B} = \begin{bmatrix} 6 & -8 & 10 & 20 \\ 8 & -2 & 22 & 18 \end{bmatrix} = \mathbf{AB} + \mathbf{A'B}$;

c. $\mathbf{BB'} = \begin{bmatrix} 14 & -1 \\ -1 & 3 \end{bmatrix}$ and $\mathbf{B'B} = \begin{bmatrix} 1 & 0 & 3 & 2 \\ 0 & 1 & 1 & -1 \\ 3 & 1 & 10 & 5 \\ 2 & -1 & 5 & 5 \end{bmatrix}$;

d. $tr(BB') = tr(B'B) = 17$;

e. $\mathbf{Bx} = \begin{bmatrix} -1 \\ -2 \end{bmatrix}$, $\quad \mathbf{B'Bx} = \begin{bmatrix} -1 \\ 2 \\ -1 \\ -4 \end{bmatrix}$, $\quad \mathbf{x'B'Bx} = 5 = (\mathbf{Bx})'\mathbf{Bx}$;

f. $\mathbf{Ay} = \begin{bmatrix} -3 \\ 1 \end{bmatrix}$, $\quad \mathbf{A'Ay} = \begin{bmatrix} -7 \\ -17 \end{bmatrix}$, $\quad \mathbf{y'A'Ay} = 10 = (\mathbf{Ay})'\mathbf{Ay}$;

g. $\mathbf{A}^2 - 4\mathbf{A} - 9\mathbf{I} = \mathbf{0}$;

h. $\frac{1}{9}\mathbf{A} \begin{bmatrix} -1 & 6 \\ 2 & -3 \end{bmatrix} = \frac{1}{9} \begin{bmatrix} -1 & 6 \\ 2 & -3 \end{bmatrix} \mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

2. Confirm:

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

b. if $\mathbf{A} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$, then $\mathbf{A}^2 = \mathbf{A}$;

c. if $\mathbf{B} = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{3} & 1/\sqrt{2} \\ -2/\sqrt{6} & 1/\sqrt{3} & 0 \\ 1/\sqrt{6} & 1/\sqrt{3} & -1/\sqrt{2} \end{bmatrix}$, $\mathbf{BB'} = \mathbf{B'B} = \mathbf{I}_3$;

d. if $\mathbf{C} = \begin{bmatrix} 6 & -4 \\ 9 & -6 \end{bmatrix}$, then $\mathbf{C}^2$ is null;

e. $\frac{1}{9}\begin{bmatrix} 4 & -5 & -1 \\ 1 & 1 & 2 \\ 4 & 4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 0 & 4 & -1 \end{bmatrix}$ is an identity matrix.

3. For $\mathbf{X} = \begin{bmatrix} 1 & 2 & 3 \\ 0 & -1 & -2 \\ -1 & 0 & 7 \end{bmatrix}$ and $\mathbf{Y} = \begin{bmatrix} 6 & 0 & 0 \\ -3 & 4 & 0 \\ 0 & -5 & 2 \end{bmatrix}$

find $\mathbf{X}^2, \mathbf{Y}^2, \mathbf{XY}$, and $\mathbf{YX}$, and show that

$$(\mathbf{X}+\mathbf{Y})^2 = \mathbf{X}^2 + \mathbf{XY} + \mathbf{YX} + \mathbf{Y}^2 = \begin{bmatrix} 40 & 5 & 44 \\ -28 & 13 & -33 \\ -1 & -62 & 88 \end{bmatrix}.$$

4. Given $\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 0 & 2 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & -1 \\ 2 & 3 & 0 \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 6 & 5 & 7 \\ 2 & 2 & 4 \\ 3 & 3 & 6 \end{bmatrix}$,

show that $\mathbf{AX} = \mathbf{BX}$ even though $\mathbf{A} \neq \mathbf{B}$.

5. a. If, in general $\mathbf{A}$ is $r \times c$, what must be the order of $\mathbf{B}$ so that $\mathbf{A} + \mathbf{B}'$ exists? Why?

   b. Explain why $\mathbf{AVA}' = \mathbf{BVB}'$ implies that $\mathbf{A}$ and $\mathbf{B}$ have the same order and $\mathbf{V}$ is square.

6. Under what conditions do both $tr(\mathbf{ABC})$ and $tr(\mathbf{BAC})$ exist? When they exist, do you expect them to be equal? Why? Calculate $tr(\mathbf{ABC})$ and $tr(\mathbf{BAC})$ for

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

7. For $\mathbf{A}$ and $\mathbf{B}$ having the same order, explain why

$$\begin{aligned}(\mathbf{A}+\mathbf{B})(\mathbf{A}+\mathbf{B})' &= (\mathbf{A}+\mathbf{B})(\mathbf{A}'+\mathbf{B}') \\ &= \mathbf{AA}' + \mathbf{AB}' + \mathbf{BA}' + \mathbf{BB}'.\end{aligned}$$

Will these expressions also generally equal $(\mathbf{A}+\mathbf{B})'(\mathbf{A}+\mathbf{B})$? Why must $\mathbf{A}$ and $\mathbf{B}$ have the same order here? Verify these equalities for

$$\mathbf{A} = \begin{bmatrix} 2 & 3 & 7 \\ 6 & 1 & -2 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} -1 & 0 & 1 \\ 0 & 1 & 7 \end{bmatrix}.$$

8. For a matrix $\mathbf{A}$ from one of the preceding exercises, calculate $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ and verify that the trace of each is the sum of squares of elements of $\mathbf{A}$. Conclude that $tr(\mathbf{A}'\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$.

9. a. When does $(\mathbf{A}+\mathbf{B})(\mathbf{A}-\mathbf{B}) = \mathbf{A}^2 - \mathbf{B}^2$?

   b. When $\mathbf{A} = \mathbf{A}'$, prove that $tr(\mathbf{AB}) = tr(\mathbf{AB}')$.

10. A generation matrix given by Kempthorne (1957, p. 120) is $\mathbf{A} = \begin{bmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$. Given that $\mathbf{f}^{(i)} = \mathbf{A}\mathbf{f}^{(i-1)}$, show that

$$\mathbf{f}^{(2)} = \begin{bmatrix} 1 & \frac{3}{4} \\ 0 & \frac{1}{4} \end{bmatrix} \mathbf{f}^{(0)}, \quad \mathbf{f}^{(3)} = \begin{bmatrix} 1 & \frac{7}{8} \\ 0 & \frac{1}{8} \end{bmatrix} \mathbf{f}^{(0)}$$

and $\mathbf{f}^{(n)} = \mathbf{A}^n\mathbf{f}^{(0)} = \begin{bmatrix} 1 & 1-1/2^n \\ 0 & 1/2^n \end{bmatrix} \mathbf{f}^{(0)}$.

11. With $\mathbf{A} = \begin{bmatrix} 1 & 0 & 3 \\ 2 & -1 & 0 \\ 0 & 4 & 1 \end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix} 3 & 3 & 1 \\ -4 & 2 & -1 \\ -2 & 0 & 0 \end{bmatrix}$:

   a. Partition **A** and **B** as

   $$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix}$$

   where both $\mathbf{A}_{21}$, and $\mathbf{B}_{21}$ have order 1×2.

   b. Calculate **AB** both with and without the partitioning, to demonstrate the validity of multiplication of partitioned matrices.

   c. Calculate $\mathbf{AB}'$, showing that

   $$\mathbf{B}' = \begin{bmatrix} \mathbf{B}'_{11} & \mathbf{B}'_{21} \\ \mathbf{B}'_{12} & \mathbf{B}'_{22} \end{bmatrix}.$$

12. By considering the inner product of $\mathbf{y} - (\mathbf{x}'\mathbf{y}/\mathbf{x}'\mathbf{x})\mathbf{x}$ with itself, prove that

$$\left(\sum_{i=1}^{n} x_i y_i\right)^2 \leq \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2.$$

   This is the Cauchy–Schwarz inequality. Use it to prove that a product–moment correlation (apart from sign) can never exceed unity.

13. Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are two matrices of order $n \times n$. Show that the following equalities are equivalent:

   a. $\boldsymbol{AA'BB'} = \boldsymbol{AB'AB'}$.

   b. $tr(\boldsymbol{AA'BB'}) = tr(\boldsymbol{AB'AB'})$

   c. $\boldsymbol{A'B} = \boldsymbol{B'A}$.

14. a. Let $\boldsymbol{A}$ be a matrix of order $n \times n$. Show that the sum of all elements of $\boldsymbol{A}$ is equal to $\mathbf{1}'_n\boldsymbol{A}\mathbf{1}_n$, where $\mathbf{1}_n$ is a vector whose n elements are all equal to one.

b. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two matrices of order $n \times n$. Show that $\boldsymbol{AB}$ and $\boldsymbol{BA}$ have the same sum of diagonal elements.

(Hint: Use the vector u_i whose elements are all equal to zero except for the ith element which is equal to one, $i = 1, 2, \ldots, n$.)

15. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be two matrices of order $n \times n$ such that $\boldsymbol{A} = \boldsymbol{A}^2$ and $\boldsymbol{B} = \boldsymbol{B}^2$. Show that if $(\boldsymbol{A} - \boldsymbol{B})^2 = \boldsymbol{A} - \boldsymbol{B}$, then $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{B}$.

16. Let $\boldsymbol{A}$ be a square matrix of order $n \times n$ and $\mathbf{1}$ be a vector of ones of order $n \times 1$. Suppose that $\boldsymbol{x} \neq \mathbf{0}$ exists such that $\boldsymbol{Ax} = \mathbf{0}$ and $\boldsymbol{x}'\mathbf{1} = 0$. Prove that

   a. $\boldsymbol{A} + \lambda\mathbf{11}'$ is singular for any scalar λ,

   b. $\boldsymbol{A} + \boldsymbol{f}\mathbf{1}'$ is also singular for any vector $\boldsymbol{f}$.

17. Let $\boldsymbol{X}$ be a matrix of order $n \times p$ and rank p ($n \geq p$). What is the rank of $\boldsymbol{XX'X}$?

18. Let $\boldsymbol{X}$ be a matrix of order $n \times n$ that satisfies the equation

$$\boldsymbol{X}^2 + 2\boldsymbol{X} + \boldsymbol{I}_n = \mathbf{0}.$$

   a. Show that $\boldsymbol{X} + \boldsymbol{I}_n$ is singular.

   b. Show that $\boldsymbol{X}$ is nonsingular.

   c. Show that $\boldsymbol{X} + 2\boldsymbol{I}_n$ is nonsingular.

   d. Derive an expression for $\boldsymbol{X}^{-1}$.

19. Prove that $\boldsymbol{vv}' - \boldsymbol{v}'\boldsymbol{vI}$ is singular.

20. Show that if

   a. $$\mathbf{A} = \begin{bmatrix} 6 & 13 \\ 5 & 12 \end{bmatrix}, \text{ then } |\mathbf{A}| = 7, \quad \mathbf{A}^{-1} = \tfrac{1}{7}\begin{bmatrix} 12 & -13 \\ -5 & 6 \end{bmatrix}$$

   and $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$;

   b. $$\mathbf{B} = \begin{bmatrix} 3 & -4 \\ 7 & 14 \end{bmatrix}, \text{then } |\mathbf{B}| = 70, \quad \mathbf{B}^{-1} = \tfrac{1}{70}\begin{bmatrix} 14 & 4 \\ -7 & 3 \end{bmatrix}$$

   and $\mathbf{BB}^{-1} = \mathbf{B}^{-1} = \mathbf{B} = \mathbf{I}$

21. Demonstrate the reversal rule for the inverse of a product of two matrices, using $\mathbf{A}$ and $\mathbf{B}$ given in Exercise 20.

22. Show that if

   a. $$\mathbf{A} = \begin{bmatrix} -1 & 3 & 0 \\ 0 & 2 & 1 \\ 1 & 0 & 4 \end{bmatrix}, |\mathbf{A}| = -5, \quad \mathbf{A}^{-1} = \tfrac{1}{5}\begin{bmatrix} -8 & 12 & -3 \\ -1 & 4 & -1 \\ 2 & -3 & 2 \end{bmatrix}$$ and $\mathbf{A}^{-1}\mathbf{A} = \mathbf{AA}^{-1} =$ $\mathbf{I}$;

b. $\mathbf{B} = \begin{bmatrix} 10 & 6 & -1 \\ 6 & 5 & 4 \\ -1 & 4 & 17 \end{bmatrix}, |\mathbf{B}| = 25, \quad \mathbf{B}^{-1} = \begin{bmatrix} 2.76 & -4.24 & 1.16 \\ -4.24 & 6.76 & -1.84 \\ 1.16 & -1.84 & 0.56 \end{bmatrix}$ and $\mathbf{B}^{-1}\mathbf{B} = \mathbf{B}\mathbf{B}^{-1} = \mathbf{I}$;

c. $\mathbf{C} = \frac{1}{10}\begin{bmatrix} 0 & -6 & 8 \\ -10 & 0 & 0 \\ 0 & -8 & -6 \end{bmatrix}$, then $|\mathbf{C}| = 1, \quad \mathbf{C}^{-1} = \mathbf{C}'$ and $\mathbf{C}\mathbf{C}' = \mathbf{C}'\mathbf{C} = \mathbf{I}$.

23. For $\boldsymbol{H} = \boldsymbol{I} - 2\boldsymbol{w}\boldsymbol{w}'$ with $\boldsymbol{w}'\boldsymbol{w} = 1$, prove that $\boldsymbol{H} = \boldsymbol{H}^{-1}$.

References

Ben-Israel, A. and Greville, T. N. E. (1974). *Generalized Inverses: Theory and Applications*. John Wiley & Sons, New York.
Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, Vol. 1, 3rd ed. John Wiley & Sons, New York.
Fisher, R. A. (1949). *The Theory of Inbreeding*. Oliver and Boyd, Edingburgh.
Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.
Henderson, H. V. and Searle, S. R. (1981). On deriving the inverse of a sum of matrices. *SIAM Rev.*, 23, 53–60.
Henderson, H. V., Pukelsheim, F., and Searle, S. R. (1981). On the history of the Kronecker product. Technical Report BU-733-M, Biometrics Unit, Cornell University, Ithaca, NY.
Kempthorne, O. (1957). *An Introduction to Genetic Statistics*. John Wiley & Sons, New York.
Marsaglia, G. and Styan, G. P. H. (1974a). Equalities and inequalities for ranks of matrices. *Linear and Multilinear Algebra*, 2, 269–292.
Marsaglia, G. and Styan, G. P. H. (1974b). Rank conditions for generalized inverses of partitioned matrices. *Sankhyā*, 36, 437–442.
Ouellette, D. V. (1981). Schur complements and statisticsSchur complements and statistics. *Linear Algebra Appl.*, 36, 187–195.

5
Special Matrices

Numerous matrices with particular properties have attracted special names. Although their historical origins are in specific mathematical problems or applications, they were later found to have properties of broader interest. Of the vast number of such matrices, this chapter presents but a small selection of special matrices that are commonly used in statistics. Other matrices appear in subsequent chapters.

5.1 Symmetric Matrices

Definition 5.1 *A square matrix is defined as symmetric when it equals its transpose; that is,*

$$\boldsymbol{A} \text{ is symmetric when } \boldsymbol{A} = \boldsymbol{A}', \text{ with } a_{ij} = a_{ji} \tag{5.1}$$

for i, j = i, …, r for $\mathbf{A}_{r \times r}$.

Symmetric matrices have many useful special properties, a few of which (relating to matrix products) are noted here. Others are noted in later chapters.

5.1.1 Products of Symmetric Matrices

Products of symmetric matrices are not generally symmetric. If $\mathbf{A} = \mathbf{A}'$ and $\mathbf{B} = \mathbf{B}'$, then **A** and **B** are symmetric but the transpose of their product (when it exists) is

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}' = \mathbf{BA}.$$

Since **BA** is generally not the same as **AB**, this means **AB** is generally not symmetric.

Example 5.1 *With*

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 3 & 7 \\ 7 & 6 \end{bmatrix},$$
$$(\mathbf{AB})' = \begin{bmatrix} 17 & 19 \\ 27 & 32 \end{bmatrix}' = \begin{bmatrix} 17 & 27 \\ 19 & 32 \end{bmatrix} = \mathbf{BA} \neq \mathbf{AB}.$$

5.1.2 Properties of AA′ and A′A

Products of a matrix and its transpose always exist and are symmetric:

$$(\mathbf{AA}')' = (\mathbf{A}')'\mathbf{A}' = \mathbf{AA}' \quad \text{and} \quad (\mathbf{A}'\mathbf{A})' = \mathbf{A}'(\mathbf{A}')' = \mathbf{A}'\mathbf{A}. \tag{5.2}$$

Observe the method used for showing that a matrix is symmetric: transpose the matrix and show that the result equals the matrix itself.

Although both products $\mathbf{AA}'$ and $\mathbf{A}'\mathbf{A}$ are symmetric, they are not necessarily equal. In fact, only when **A** is square might they be equal, because it is only then that both products have the same order. Thus for $\mathbf{A}_{r \times c}$, the product $\mathbf{AA}' = \mathbf{A}_{r \times c}(\mathbf{A}')_{c \times r}$ has order $r \times r$, and $\mathbf{A}'\mathbf{A} = (\mathbf{A}')_{c \times r}\mathbf{A}_{r \times c}$ has order $c \times c$.

Example 5.2 *For*

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

$$\mathbf{AA}' = \begin{bmatrix} 6 & 2 \\ 2 & 10 \end{bmatrix}, \quad \textit{and} \quad \mathbf{A}'\mathbf{A} = \begin{bmatrix} 10 & 2 & 2 \\ 2 & 4 & -2 \\ 2 & -2 & 2 \end{bmatrix}.$$

Note that matrix multiplication ensures that elements of **AA**′ are inner products of rows of **A** with themselves and with each other:

$$\mathbf{AA}' = \{\text{inner product of } i\text{th and } k\text{th rows of } \mathbf{A}\} \text{ for } i,\, k = 1, \ldots, r. \tag{5.3}$$

In particular, the ith diagonal element of **AA**′ is the sum of squares of the elements of the ith row of **A**, namely $\sum_{j=1}^{c} a_{ij}^2$ for **A** of order $r \times c$. Confining attention to real matrices, we can use the property of real numbers that a sum of squares of them is positive (unless they are all zero), and so observe that for real matrices **A**, the matrix **AA**′ has diagonal elements that are positive (or zero). Similar results hold for **A**′**A** in terms of columns of **A**:

$$\mathbf{A}'\mathbf{A} = \{\text{inner product of } j\text{th } \textit{and } m\text{th } \textit{columns of } \mathbf{A}\} \text{ for } j,\, m = 1, \ldots, c. \tag{5.4}$$

The reader can verify these results for the example.

Recall that if a sum of squares of real numbers is zero, then each of the numbers is zero; that is, for real numbers $x_1, x_2, \ldots, x_n, \Sigma_{i=1}^{n} x_i^2 = 0$ implies $x_1 = 0 = x_2 = \cdots = x_n$. This is the basis for proving, for any real matrix **A**, that

$$\mathbf{A}'\mathbf{A} = \mathbf{0} \quad \text{implies} \quad \mathbf{A} = \mathbf{0} \tag{5.5}$$

and

$$tr(\mathbf{A}'\mathbf{A}) = \mathbf{0} \quad \text{implies} \quad \mathbf{A} = \mathbf{0}. \tag{5.6}$$

Result (5.5) is true because $\mathbf{A}'\mathbf{A} = \mathbf{0}$ implies that every diagonal element of **A**′**A** is zero; and for the jth such element this means, by (5.3), that the inner product of the jth column of **A** with itself is zero; that is, for **A** of order $r \times c$, that $\sum_{k=1}^{r} a_{kj}^2 = 0$ for $j = 1, \ldots, c$. Hence $a_{kj} = 0$ for $k = 1, \ldots, r$; and this is true for $j = 1, \ldots, c$. Therefore, $\mathbf{A} = \mathbf{0}$. Proof of (5.6) is similar:

$$tr(\mathbf{A}'\mathbf{A}) = \sum_{j=1}^{c} (j\text{th diagonal element of } \mathbf{A}'\mathbf{A}) = \sum_{j=1}^{c} \sum_{k=1}^{r} a_{kj}^2$$

and so $tr(\mathbf{A}'\mathbf{A}) = \mathbf{0}$ implies that every a_{kj} is zero; that is, every element of **A** is zero. Hence $\mathbf{A} = \mathbf{0}$.

Results (5.5) and (5.6) are seldom useful for the sake of some particular matrix **A**, but they are often helpful in developing other results in matrix algebra when **A** is a function of other matrices. For example, by means of (5.5) we can prove, for real matrices **P**, **Q**, and **X**, that

$$\mathbf{PXX}' = \mathbf{QXX}' \quad \text{implies} \quad \mathbf{PX} = \mathbf{QX}. \tag{5.7}$$

The proof consists of observing that

$$(\mathbf{PXX}' - \mathbf{QXX}')(\mathbf{P}' - \mathbf{Q}') \equiv (\mathbf{PX} - \mathbf{QX})(\mathbf{PX} - \mathbf{QX})'. \tag{5.8}$$

Hence if $\mathbf{PXX}' = \mathbf{QXX}'$, the left-hand side of (5.8) is null and so, therefore, is the right-hand side; hence by (5.5), $\mathbf{PX} - \mathbf{QX} = \mathbf{0}$; that is, $\mathbf{PX} = \mathbf{QX}$, and (5.7) is established.

5.1.3 Products of Vectors

The inner product of two vectors is a scalar and is therefore symmetric: $\mathbf{x}'\mathbf{y} = (\mathbf{x}'\mathbf{y})' = \mathbf{y}'\mathbf{x}$. In contrast, the outer product of two vectors (see Section 4.7.5) is not necessarily symmetric: $\mathbf{xy}' = (\mathbf{yx}')' \neq (\mathbf{xy}')'$. Indeed, such a product is not necessarily even square.

Example 5.3 *For* $\mathbf{x}' = [1 \quad 0 \quad 2]$, $\mathbf{y}' = [2 \quad 2 \quad 1]$, *and* $\mathbf{z}' = [3 \quad 2]$,

$$\mathbf{xy}' = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \begin{bmatrix} 2 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 2 & 1 \\ 0 & 0 & 0 \\ 4 & 4 & 2 \end{bmatrix} \tag{5.9}$$

is not symmetric. Neither is

$$\mathbf{yx}' = \begin{bmatrix} 2 \\ 2 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 4 \\ 2 & 0 & 4 \\ 1 & 0 & 2 \end{bmatrix}, \tag{5.10}$$

but its transpose does, of course, equal $\mathbf{xy}'$*; that is, (5.10) equals the transpose of (5.9); namely* $\mathbf{yx}' = (\mathbf{xy}')'$*. And* $\mathbf{xz}'$ *and* $\mathbf{zx}'$ *are rectangular:*

$$\mathbf{xz}' = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \begin{bmatrix} 3 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 2 \\ 0 & 0 \\ 6 & 4 \end{bmatrix}$$

and

$$\mathbf{zx}' = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 & 6 \\ 2 & 0 & 4 \end{bmatrix}.$$

5.1.4 Sums of Outer Products

Let $\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_c$ be the columns of $\boldsymbol{A}$, and $\boldsymbol{\beta}'_1, \boldsymbol{\beta}'_2, \cdots, \boldsymbol{\beta}'_c$ be the rows of $\boldsymbol{B}$, then the product $\boldsymbol{AB}$ expressed as,

$$\mathbf{AB} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 \cdots \mathbf{a}_c\end{bmatrix}\begin{bmatrix}\beta'_1 \\ \beta'_2 \\ \vdots \\ \beta'_c\end{bmatrix} = \sum_{j=1}^{c} \mathbf{a}_j \beta'_j, \tag{5.11}$$

is the sum of outer products of the columns of **A** with the corresponding rows in **B**.

Example 5.4

$$\begin{aligned}\mathbf{AB} &= \begin{bmatrix}1 & 4 \\ 2 & 5 \\ 3 & 6\end{bmatrix}\begin{bmatrix}7 & 8 \\ 9 & 10\end{bmatrix} = \begin{bmatrix}43 & 48 \\ 59 & 66 \\ 75 & 84\end{bmatrix} \\ &= \begin{bmatrix}1 \\ 2 \\ 3\end{bmatrix}\begin{bmatrix}7 & 8\end{bmatrix} + \begin{bmatrix}4 \\ 5 \\ 6\end{bmatrix}\begin{bmatrix}9 & 10\end{bmatrix} = \begin{bmatrix}7 & 8 \\ 14 & 16 \\ 21 & 24\end{bmatrix} + \begin{bmatrix}36 & 40 \\ 45 & 50 \\ 54 & 60\end{bmatrix}.\end{aligned}$$

A special case of (5.11) is when **B** = **A**′:

$$\mathbf{AA}' = \sum_{j=1}^{c} \mathbf{a}_j \mathbf{a}'_j. \tag{5.12}$$

AA′ *is thus the sum of outer products of each column of* **A** *with itself; that is,*

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

5.1.5 Elementary Vectors

A special case of (5.12) is

$$\mathbf{I}_n = \sum_{i=1}^{n} \mathbf{e}_i \mathbf{e}'_i$$

for $\mathbf{e}_i$ being the ith column of $\mathbf{I}_n$, namely a vector with unity for its ith element and zeros elsewhere. $\mathbf{e}_i$ is called an *elementary vector*. When necessary, its order n can be identified by denoting $\mathbf{e}_i$ as $\mathbf{e}^{(n)}{}_i$.

The outer product of one elementary vector with another is a null matrix except for one element of unity: for example, with

$$\mathbf{e}_1 = \begin{bmatrix}1 \\ 0 \\ 0\end{bmatrix} \quad \text{and} \quad \mathbf{e}_2 = \begin{bmatrix}0 \\ 1 \\ 0\end{bmatrix}, \quad \mathbf{E}_{12} = \mathbf{e}_1\mathbf{e}'_2 = \begin{bmatrix}0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0\end{bmatrix}.$$

In general, $\mathbf{E}_{ij} = \mathbf{e}_i\mathbf{e}'_j$ is null except for element (i, j) being unity; and, of course, $\mathbf{I} = \Sigma_i \mathbf{E}_{ii}$. These $\mathbf{E}_{ij}$-matrices are particularly useful in applications of calculus to matrix algebra (see Section 9.4.3)

The **e**-vectors are also useful for delineating individual rows and columns of a matrix. Thus $\mathbf{e}'_i\mathbf{A} = \alpha_i'$, the ith row of **A** and $\mathbf{Ae}_j = \mathbf{a}_j$, the jth column of **A**; that is, for

$$\mathbf{e}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{A} = \begin{bmatrix} 2 & 3 \\ 5 & 7 \end{bmatrix}, \quad \mathbf{e}'_2\mathbf{A} = [5 \quad 7] \quad \text{and} \quad \mathbf{Ae}_2 = \begin{bmatrix} 3 \\ 7 \end{bmatrix}.$$

5.1.6 Skew-Symmetric Matrices

A symmetric matrix **A** has the property **A** = **A**′; in contrast there are also matrices **B** having the property **B**′ = −**B**. Their diagonal elements are zero and each off-diagonal element is minus its symmetric partner; that is, $b_{ii} = 0$ and $b_{ij} = -b_{ji}$. An example is

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

Such matrices, having **B**′ = −**B**, are called *skew-symmetric*.

5.2 Matrices Having all Elements Equal

Vectors whose every element is unity are called *summing vectors* because they can be used to express a sum of numbers in matrix notation as an inner product.

Example 5.5 *The row vector* $\mathbf{1}' = [1 \quad 1 \quad 1 \quad 1]$ *is the summing vector of order* 1×4 *and for* $\mathbf{x}' = [3 \quad 6 \quad 8 \quad -2]$

$$\mathbf{1}'\mathbf{x} = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 3 \\ 6 \\ 8 \\ -2 \end{bmatrix} = 3 + 6 + 8 - 2$$

$$= 15 = \begin{bmatrix} 3 & 6 & 8 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \mathbf{x}'\mathbf{1}.$$

When necessary to avoid confusion, the order of a summing vector can be denoted in the usual way: $\mathbf{1}'_4 = [1 \quad 1 \quad 1 \quad 1]$. *For example,*

$$\mathbf{1}'_3\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ -5 & -3 \\ 4 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 1 \end{bmatrix} = \mathbf{1}'_2.$$

Were this equation in $\mathbf{X}$ *to be written as* $\mathbf{1}'\mathbf{X} = \mathbf{1}'$, *one might be tempted to think that both* $\mathbf{1}'$ *vectors had the same order. That they do not is made clear by denoting* $\mathbf{X}$ *of order* $r \times c$ *as* $\mathbf{X}_{r \times c}$ *so that* $\mathbf{1}'\mathbf{X}_{r \times c} = \mathbf{1}'$ *is* $\mathbf{1}'_r\mathbf{X}_{r \times c} = \mathbf{1}'_c$.

The inner product of a summing vector with itself is a scalar, the vector’s order:

$$\mathbf{1}'_n\mathbf{1}_n = n. \tag{5.13}$$

And outer products are matrices with all elements unity:

$$\mathbf{1}_3\mathbf{1}'_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \mathbf{J}_{3\times2}.$$

In general, $\mathbf{1}_r\mathbf{1}'_s$ has order $r \times s$ and is often denoted by the symbol $\mathbf{J}$ or $\mathbf{J}_{r \times s}$:

$$\mathbf{1}_r\mathbf{1}'_s = \mathbf{J}_{r\times s}, \text{ having all elements unity.} \tag{5.14}$$

Clearly, $\lambda\mathbf{J}_{r \times s}$ has all elements λ.

Products of $\mathbf{J}$’s with each other and with $\mathbf{1}$’s are, respectively, $\mathbf{J}$’s and $\mathbf{1}$’s (multiplied by scalars):

$$\mathbf{J}_{r\times s}\mathbf{J}_{s\times t} = s\mathbf{J}_{r\times t}, \qquad \mathbf{1}'_r\mathbf{J}_{r\times s} = r\mathbf{1}'_s \quad \text{and } \mathbf{J}_{r\times s}\mathbf{1}_s = s\mathbf{1}_r. \tag{5.15}$$

Particularly useful are square **J**'s and a variant thereof:

$$\mathbf{J}_n = \mathbf{1}_n\mathbf{1}'_n \quad \text{with} \quad \mathbf{J}_n^2 = n\mathbf{J}_n; \tag{5.16}$$

and

$$\bar{\mathbf{J}}_n = \frac{1}{n}\mathbf{J}_n \quad \text{with} \quad \bar{\mathbf{J}}_n^2 = \bar{\mathbf{J}}_n.$$

And for statistics

$$\mathbf{C}_n = \mathbf{I} - \bar{\mathbf{J}}_n = \mathbf{I} - \frac{1}{n}\mathbf{J}_n, \tag{5.17}$$

known as a *centering matrix*, is especially useful, as is now illustrated. First observe that

$$\mathbf{C} = \mathbf{C}' = \mathbf{C}^2, \qquad \mathbf{C1} = \mathbf{0} \quad \text{and} \quad \mathbf{CJ} = \mathbf{JC} = \mathbf{0}, \tag{5.18}$$

which the reader can easily verify.

Example 5.6 *The mean and sum of squares about the mean for data $x_1, x_2, \ldots, x_n$ can be expressed in terms of **1**-vectors and **J**-matrices. Define*

$$\mathbf{x}' = [x_1 \quad x_2 \quad \cdots \quad x_n].$$

Then the mean of the x's is

$$\bar{x} = (x_1 + x_2 + \cdots + x_n)/n = \sum_{i=1}^{n} x_i/n = \frac{1}{n}\mathbf{x}'\mathbf{1} = \frac{1}{n}\mathbf{1}'\mathbf{x},$$

the last equality arising from $\mathbf{x}'\mathbf{y} = \mathbf{y}'\mathbf{x}$. *Using* **C** *of* (5.11) and (5.18), *we get*

$$\mathbf{x}'\mathbf{C} = \mathbf{x}' - \mathbf{x}'\bar{\mathbf{J}} = \mathbf{x}' - \frac{1}{n}\mathbf{x}'\mathbf{11}' = \mathbf{x}' - \bar{\mathbf{x}}\mathbf{1}' = \{\mathbf{x}_i - \bar{\mathbf{x}}\}$$

*is the data vector with each observation expressed as a deviation from $\bar{x}$. (This is the origin of the name centering matrix for **C**.) Postmultiplying **x'C** by **x** then gives*

$$\mathbf{x}'\mathbf{Cx} = (\mathbf{x}' - \bar{\mathbf{x}}\mathbf{1}')\mathbf{x} = \mathbf{x}'\mathbf{x} - \bar{\mathbf{x}}(\mathbf{1}'\mathbf{x}) = \mathbf{x}'\mathbf{x} - n\bar{x}^2.$$

Hence, using a standard result in statistics, we get

$$\sum_{l-1}^{n}(x_l - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2 = \mathbf{x}'\mathbf{x} - n\bar{x}^2 = \mathbf{x}'\mathbf{Cx}. \tag{5.19}$$

*Thus for **x'** being a data vector, **x'1**/n is the mean, **x'C** is the vector of deviations from the mean, and **x'Cx** is the sum of squares about the mean.*

Expression (5.19) is a special case of the form $\mathbf{x}'\mathbf{A}\mathbf{x}$, known as a quadratic form (see Section 5.6), which can be used for sums of squares generally. Expressed in this manner, and with the aid of other matrix concepts (notably idempotency; see Section 5.3), sums of squares of normally distributed x's are known to be distributed as χ^2 under very simply stated (in matrix notation) conditions (see Section 10.5).

Example 5.7 *It was noted in Example 2.1 that the row sums of a transition probability matrix are always unity [see formula (2.6)]. For example, with*

$$\mathbf{P} = \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix}, \qquad \begin{bmatrix} 0.2 + 0.8 \\ 0.4 + 0.6 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}; \text{ i.e., } \begin{bmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

This last result is

$$\mathbf{P1} = \mathbf{1},$$

which is true generally for any transition probability matrix. Furthermore, because $\mathbf{P}^2\mathbf{1} = \mathbf{P}(\mathbf{P1}) = \mathbf{P1} = \mathbf{1}$ *this in turn extends to*

$$\mathbf{P}^n\mathbf{1} = \mathbf{1},$$

showing that row sums of any power of a transition probability matrix are also unity.

5.3 Idempotent Matrices

The matrix $\bar{\mathbf{J}}_n$ in (5.17) has the characteristic that its square equals itself. Many different matrices are of this nature; they are called idempotent matrices. Thus when $\mathbf{K}$ is such that $\mathbf{K}^2 = \mathbf{K}$, we say $\mathbf{K}$ is *idempotent* (from Latin, idem meaning "same," and potent "power"). All idempotent matrices are square (otherwise $\mathbf{K}^2$ does not exist); identity matrices and square null matrices are idempotent. When $\mathbf{K}$ is idempotent, all powers of $\mathbf{K}$ equal $\mathbf{K}$; that is, $\mathbf{K}^r = \mathbf{K}$ for r being a positive integer, and $(\mathbf{I} - \mathbf{K})$ is idempotent. Thus

$$\mathbf{K}^2 = \mathbf{K} \text{ implies } (\mathbf{I} - \mathbf{K})^2 = \mathbf{I} - \mathbf{K},$$

but $\mathbf{K} - \mathbf{I}$ is not idempotent. A product of two idempotent matrices is idempotent if the matrices commute in multiplication.

Idempotent matrices occur in many applications of matrix algebra and they play an especially important role in statistics.

Example 5.8 *The matrix* $\mathbf{I} - \bar{\mathbf{J}}$ *in (5.17) is idempotent; so is* $\mathbf{GA}$ *whenever* $\mathbf{G}$ *is such that* $\mathbf{AGA} = \mathbf{A}$ *[because then* $(\mathbf{GA})^2 = \mathbf{GAGA} = \mathbf{GA}$*]. A matrix* $\mathbf{G}$ *of this nature is called a generalized inverse of* $\mathbf{A}$*; its properties are considered in Chapter 8.*

A matrix $\mathbf{A}$ satisfying $\mathbf{A}^2 = \mathbf{0}$ is called *nilpotent*, and that for which $\mathbf{A}^2 = \mathbf{I}$ could be called *unipotent*.

Example 5.9

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 5 \\ 2 & 4 & 10 \\ -1 & -2 & -5 \end{bmatrix} \textit{ is nilpotent;} \quad \mathbf{B} = \begin{bmatrix} \mathbf{I} & \mathbf{X} \\ 0 & -\mathbf{I} \end{bmatrix} \textit{ is unipotent.}$$

Variations on these definitions are $\mathbf{A}^k = \mathbf{A}$, $\mathbf{A}^k = \mathbf{0}$, and $\mathbf{A}^k = \mathbf{I}$ for some positive integer $k > 2$. An example is the matrix

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix},$$

for which $\mathbf{B}^3 = \mathbf{I}$, but $\mathbf{B}^2 \neq \mathbf{I}$.

The following theorem is very useful in determining the rank of an idempotent matrix:

Theorem 5.1 *If* $\boldsymbol{K}$ *is an* $n \times n$ *idempotent matrix, then its rank* r *is equal to its trace, that is,* $r = r(\boldsymbol{K}) = tr(\boldsymbol{K})$.

Proof. Let $\boldsymbol{l}_1, \boldsymbol{l}_2, \ldots, \boldsymbol{l}_r$ be linearly independent vectors that span (form a basis for) the column space of $\boldsymbol{K}$. Let $\boldsymbol{L} = [\boldsymbol{l}_1 : \boldsymbol{l}_2 : \ldots : \boldsymbol{l}_r]$, then $\boldsymbol{L}$ is of order $n \times r$ and rank r. The ith column, k_i of $\boldsymbol{K}$ can then be expressed as a linear combination of the columns of $\boldsymbol{L}$ (i = 1, 2, …, n). We can therefore write $\boldsymbol{k}_i = \boldsymbol{L}\boldsymbol{m}_i$ where $\boldsymbol{m}_i$ is a vector of coefficients consisting of r elements (i = 1, 2, …, n). Let $\boldsymbol{M}$ be a matrix of order $r \times n$ whose columns are $\boldsymbol{m}_1, \boldsymbol{m}_2, \ldots, \boldsymbol{m}_n$. Thus $\boldsymbol{K}$ can be written as

$$\boldsymbol{K} = \boldsymbol{L}\boldsymbol{M}. \tag{5.20}$$

It follows that $r(\boldsymbol{K}) \leq r(\boldsymbol{M}) \leq r$, since the rank of $\boldsymbol{M}$ cannot exceed the number of its rows. But, $r(\boldsymbol{K}) = r$. We conclude that $r(\boldsymbol{M}) = r$.

Since $\boldsymbol{K}$ is idempotent, $\boldsymbol{K}^2 = \boldsymbol{K}$. We then have from (5.20),

$$\boldsymbol{L}\boldsymbol{M}\boldsymbol{L}\boldsymbol{M} = \boldsymbol{L}\boldsymbol{M}. \tag{5.21}$$

Furthermore, because $\boldsymbol{L}$ is of full-column rank, multiplying both sides of (5.21) on the left by $\boldsymbol{L}'$ and noting that $\boldsymbol{L}'\boldsymbol{L}$ is nonsingular by the fact that it is of order $r \times r$ of rank r, we get, after multiplying both of the resulting sides of (5.21) on the left by the inverse of $\boldsymbol{L}'\boldsymbol{L}$,

$$\boldsymbol{M}\boldsymbol{L}\boldsymbol{M} = \boldsymbol{M}. \tag{5.22}$$

Similarly, $\boldsymbol{M}$ being of full-row rank, the matrix $\boldsymbol{M}\boldsymbol{M}'$ is nonsingular. Multiplying the two sides of (5.22) on the right by $\boldsymbol{M}'$ and then multiplying the resulting equation on the right by the inverse of $\boldsymbol{M}\boldsymbol{M}'$, we get

$$\boldsymbol{M}\boldsymbol{L} = \boldsymbol{I}_r. \tag{5.23}$$

It follows from (5.20) and (5.23) that $r(\boldsymbol{K}) = r = tr(\boldsymbol{M}\boldsymbol{L}) = tr(\boldsymbol{L}\boldsymbol{M}) = tr(\boldsymbol{K})$. Thus, $r(\boldsymbol{K}) = tr(\boldsymbol{K})$.

Since the trace of $\boldsymbol{K}$, being the sum of its diagonal elements, is easy to compute, this theorem facilitates the finding of the rank of $\boldsymbol{K}$, which, in general, is more difficult to determine, when $\boldsymbol{K}$ is idempotent.

5.4 Orthogonal Matrices

Definition 5.2 *Another useful class of matrices is that for which* $\mathbf{A}$ *has the property* $\mathbf{AA}' = \mathbf{I} = \mathbf{A}'\mathbf{A}$. *Such matrices are called orthogonal. We lead up to them with the following definitions:*

The norm of a real vector $\mathbf{x}' = [x_1 \quad x_2 \quad \cdots \quad x_n]$ *is defined as*

$$\text{norm of } \mathbf{x} = \sqrt{\mathbf{x}'\mathbf{x}} = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}} \tag{5.24}$$

For example, the norm of $\mathbf{x}' = [1 \quad 2 \quad 2 \quad 4]$ is $(1+4+4+16)^{\frac{1}{2}} = 5$. (The square root is taken as positive.) A vector is said to be either *normal* or a *unit vector* when its norm is unity; that is, when $\mathbf{x}'\mathbf{x} = 1$. An example is $\mathbf{x}' = [.2 \quad .4 \quad .4 \quad .8]$. Any non-null vector $\mathbf{x}$ can be changed into a unit vector by multiplying it by the scalar $1/\sqrt{\mathbf{x}'\mathbf{x}}$; that is,

$$\mathbf{u} = \left(\frac{1}{\sqrt{\mathbf{x}'\mathbf{x}}}\right)\mathbf{x}$$

is the *normalized* form of $\mathbf{x}$ (because $\mathbf{u}'\mathbf{u} = 1$).

Non-null vectors $\mathbf{x}$ and $\mathbf{y}$ are described as being *orthogonal* when $\mathbf{x}'\mathbf{y} = 0$ (equivalent, of course, to $\mathbf{y}'\mathbf{x} = 0$); for example, $\mathbf{x}' = [1 \quad 2 \quad 2 \quad 4]$ and $\mathbf{y}' = [6 \quad 3 \quad -2 \quad -2]$ are orthogonal vectors because

$$\mathbf{x}'\mathbf{y} = \begin{bmatrix}1 & 2 & 2 & 4\end{bmatrix}\begin{bmatrix}6\\3\\-2\\-2\end{bmatrix} = 6+6-4-8 = 0.$$

Two vectors are defined as *orthonormal vectors* when they are orthogonal and normal. Thus $\mathbf{u}$ and $\mathbf{v}$ are orthonormal when $\mathbf{u}'\mathbf{u} = 1 = \mathbf{v}'\mathbf{v}$ and $\mathbf{u}'\mathbf{v} = 0$; for example, $\mathbf{u}' = \frac{1}{6}[1 \quad 1 \quad 3 \quad 3 \quad 4]$ and $\mathbf{v}' = [-0.1 \quad -0.9 \quad -0.1 \quad -0.1 \quad 0.4]$ are orthonormal vectors.

A group, or collection, of vectors all of the same order is called a *set* of vectors. A set of vectors $\mathbf{x}_i$ for $i = 1, 2, \ldots, n$ is said to be an orthonormal set of vectors when every vector in the set is normal, $\mathbf{x}'_i\mathbf{x}_i = 1$ for all i, and when every pair of different vectors in the set is orthogonal, $\mathbf{x}'_i\mathbf{x}_j = 0$ for $i \neq j = 1, 2, \ldots, n$. We can say that the vectors of an orthonormal set are all normal, and pairwise orthogonal.

A matrix $\mathbf{P}_{r \times c}$ whose rows constitute an orthonormal set of vectors is said to have orthonormal rows, whereupon $\mathbf{PP}' = \mathbf{I}_r$. But then $\mathbf{P}'\mathbf{P}$ is not necessarily an identity matrix $\mathbf{I}_c$, as the following example shows:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{PP}' = \mathbf{I}_2 \quad \text{but} \quad \mathbf{P}'\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \neq \mathbf{I}_2.$$

Conversely, when $\mathbf{P}_{r \times c}$ has orthonormal columns $\mathbf{P}'\mathbf{P} = \mathbf{I}_c$ but $\mathbf{PP}'$ may not be an identity matrix.

Square matrices having orthonormal rows are in a special class: their columns are also orthonormal (why?). The matrix $\boldsymbol{P}$, say, is then such that

$$\mathbf{PP}' = \mathbf{P}'\mathbf{P} = \boldsymbol{I}. \tag{5.25}$$

This equation defines $\mathbf{P}$ as being an *orthogonal matrix*; it implies that $\boldsymbol{P}$ is square and that $\boldsymbol{P}$ has orthonormal rows and orthonormal columns. It can also be seen that $\boldsymbol{P}$ is nonsingular since the absolute value of its determinant is equal to 1. Furthermore, the inverse of $\boldsymbol{P}$ is equal to $\boldsymbol{P}'$. These are characteristics of any orthogonal matrix $\boldsymbol{P}$. Actually, it can be shown that any two of the conditions (i) $\boldsymbol{P}$ is a square matrix, (ii) $\mathbf{P}'\mathbf{P} = \mathbf{I}$, and (iii) $\mathbf{PP}' = \mathbf{I}$ imply the third, that is, imply that $\boldsymbol{P}$ is orthogonal (see Exercise 5.20).

A simple property of orthogonal matrices is that products of them are orthogonal. (see Exercise 5.9).

Theorem 5.2 (The QR Decomposition) *Let $\boldsymbol{A}$ be an $n \times p$ matrix of rank p $(n \geq p)$. Then, $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{QR}$, where $\boldsymbol{Q}$ is an $n \times p$ matrix whose columns are orthonormal and $\boldsymbol{R}$ is an upper-triangular matrix of order $p \times p$.*

This theorem is based on applying the so-called *Gram-Schmidt* orthogonalization of the columns of $\boldsymbol{A}$. Details of the proof can be found in Harville(1997, Section 6.4).

Example 5.10

$$\mathbf{A} = \frac{1}{\sqrt{6}} \begin{bmatrix} \sqrt{2} & \sqrt{2} & \sqrt{2} \\ \sqrt{3} & -\sqrt{3} & 0 \\ 1 & 1 & -2 \end{bmatrix}$$

is an orthogonal matrix, as the reader may easily verify.

The boundless variety of matrices that are orthogonal includes many that are carefully proscribed, three of which are now briefly illustrated and described.

5.4.1 Special Cases

Helmert, Givens, and Householder matrices are all orthogonal matrices.

Helmert Matrices

The Helmert matrix of order 4×4 is

$$\mathbf{H}_4 = \begin{bmatrix} 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} & 1/\sqrt{4} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} & 0 \\ 1/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} & -3/\sqrt{12} \end{bmatrix}.$$

The Helmert matrix $\mathbf{H}_n$ of order $n \times n$ has $n^{-\frac{1}{2}}\mathbf{1}'_n$ for its first row, and each of its other $n - 1$ rows for $i = 1, \ldots, n - 1$, has the partitioned form

$$[\mathbf{1}'_i \,|\, -i \,|\, 0\mathbf{1}'_{n-i-1}] \,/\, \sqrt{\lambda_i} \quad \text{with } \lambda_i = i(i+1).$$

The notation $0\mathbf{1}'$ used here simply indicates a row vector with every element zero.

Givens Matrices

The orthogonal matrix

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

is a Givens matrix of order 2×2. It is the basis of Givens matrices of order higher than 2×2. Those of order 3×3 are

$$\mathbf{G}_{12} = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{G}_{13} = \begin{bmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{bmatrix}$$

$$\mathbf{G}_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\gamma & \sin\gamma \\ 0 & -\sin\gamma & \cos\gamma \end{bmatrix}. \tag{5.26}$$

The general form of $\mathbf{G}_{rs} = \mathbf{G}_{sr}$ of order $n \times n$ is an identity matrix except for four elements: $g_{rr} = g_{ss} = \cos\theta_{rs}$ and, for $r > s$, $-g_{rs} = g_{sr} = \sin\theta_{rs}$. All such matrices are orthogonal, as are products of any number of them.

A sometimes useful operation is that of triangularizing a square matrix; that is, if **A** is square, of premultiplying it by some matrix **G** so that **GA** is an upper triangular matrix. This can be done by deriving **G** as a product of Givens matrices. The operation of going from **A** to **GA** is called a *Givens transformation*. A lower triangular matrix can be obtained from **A** in similar fashion, by postmultiplying **A** by Givens matrices.

Householder Matrices

Other orthogonal matrices useful for triangularizing square matrices are the Householder matrices, the general form of which is

$$\mathbf{H} = \mathbf{I} - 2\mathbf{h}\mathbf{h}' \quad \text{for} \quad \mathbf{h}'\mathbf{h} = 1, \tag{5.27}$$

with **h**, obviously, being a non-null column vector. Then **H** is not only orthogonal but also symmetric; and for any non-null vector

$$\mathbf{x}' = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \quad \text{there exists} \quad \mathbf{h}' = \begin{bmatrix} h_1 & h_2 & \cdots & h_n \end{bmatrix}$$

such that

$$\mathbf{Hx} = \begin{bmatrix} \lambda \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad \text{for} \quad \begin{cases} \lambda = -(\text{sign of } x_1)\sqrt{\mathbf{x}'\mathbf{x}} \\ h_1 = \sqrt{\frac{1}{2}(1 - x_1/\lambda)} \\ h_i = -x_i/2h_1\lambda \quad \text{for} \quad i = 2, 3, \ldots, n. \end{cases}$$

Example 5.11 *For* $\mathbf{x}' = [1 \quad 2 \quad 3 \quad 5 \quad 5]$ *we find* $\mathbf{h}' = \frac{1}{12}[9 \quad 2 \quad 3 \quad 5 \quad 5]$ *and* $\mathbf{Hx} = [-8 \quad 0 \quad 0 \quad 0 \quad 0]'$.

Triangularization of a square matrix **A** proceeds by developing Householder matrices from successive columns of **A**.

5.5 Parameterization of Orthogonal Matrices

The n^2 elements of an orthogonal matrix $\boldsymbol{A}$ of order $n \times n$ are not independent. This follows from the fact that $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}_n$ which implies that the elements of $\boldsymbol{A}$ are subject to $n(n + 1)/2$ equality constraints. Hence, the number of independent elements is $n^2 - n(n + 1)/2 = n(n - 1)/2$. This means that $\boldsymbol{A}$ can be represented by $n(n - 1)/2$ independent parameters. Knowledge of such a representation is needed in order to generate orthogonal matrices that are used in several statistical applications. Khuri and Myers (1981) adopted this approach to construct response surface designs that are robust to non-normality of the distribution of the error in the response surface model. Another application is the generation of random orthogonal matrices to be used in simulation experiments [see Heiberger et al. (1983) and Anderson et al. (1987)].

Khuri and Good (1989) reviewed several methods to parameterize an orthogonal matrix. Two of these methods are described here.

i. **Exponential Representation** [Gantmacher (1959)]

   If $\boldsymbol{A}$ is an orthogonal matrix with determinant equal to 1, then it can be represented as

$$\boldsymbol{A} = \exp(\boldsymbol{T}), \tag{5.28}$$

   where $\boldsymbol{T}$ is a skew-symmetric matrix. The elements of $\boldsymbol{T}$ above its main diagonal can be used to parameterize $\boldsymbol{A}$. The exponential function in (5.28) is represented as the sum of the infinite series of matrices,

$$\exp(\boldsymbol{T}) = \sum_{i=0}^{\infty} \frac{1}{i!}\boldsymbol{T}^i, \tag{5.29}$$

   where $\boldsymbol{T}^0 = \boldsymbol{I}_n$.

ii. **Cayley's Representation** [Gantmacher (1959, p. 289)]

   If $\boldsymbol{A}$ is an orthogonal matrix of order $n \times n$ that does not have the eigenvalue -1, then it can be written in Cayley's form, namely,

$$\boldsymbol{A} = (\boldsymbol{I}_n - \boldsymbol{U})(\boldsymbol{I}_n + \boldsymbol{U})^{-1},$$

   where $\boldsymbol{U}$ is a skew-symmetric matrix of order $n \times n$.

5.6 Quadratic Forms

The illustration (5.19),

$$\sum_{i=1}^{n}(x_i - \bar{x}.)^2 = \mathbf{x}'\mathbf{C}\mathbf{x},$$

is the product of a row vector $\mathbf{x}'$, a matrix $\mathbf{C}$, and the column vector $\mathbf{x}$. It is called a *quadratic form*; its general form for any matrix $\mathbf{A}$ is $\mathbf{x}'\mathbf{A}\mathbf{x}$. Expressions of this form have many uses, particularly in the general theory of analysis of variance in statistics wherein, with appropriate choice of $\mathbf{A}$, any sum of squares can be represented as $\mathbf{x}'\mathbf{A}\mathbf{x}$.

Consider the example

$$\mathbf{x}'\mathbf{A}\mathbf{x} = [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}.$$

Straightforward multiplication gives

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 &+ 4x_2x_1 + 2x_3x_1 + 2x_1x_2 \\ &+ 7x_2^2 - 2x_3x_2 + 3x_1x_3 + 6x_2x_3 + 5x_3^2, \end{aligned} \tag{5.30}$$

which simplifies to $\mathbf{x}'\mathbf{A}\mathbf{x}$ =

$$\begin{aligned} &x_1^2 + x_1x_2(4+2) + x_1x_3(2+3) + 7x_2^2 + x_2x_3(-2+6) + 5x_3^2 \\ &= x_1^2 + 7x_2^2 + 5x_3^2 + 6x_1x_2 + 5x_1x_3 + 4x_2x_3. \end{aligned} \tag{5.31}$$

This is a quadratic function of the x's; hence the name quadratic form. Two characteristics of its development are noteworthy. First, in (5.30) we see that $\mathbf{x}'\mathbf{A}\mathbf{x}$ is the sum of products of all possible pairs of the x_i's, each multiplied by an element of $\mathbf{A}$; thus in (5.30) the second term, $4x_2x_1$, is x_2x_1 multiplied by the element of $\mathbf{A}$ in the second row and first column. Second, in simplifying (5.30) to (5.31) we see that the coefficient of x_1x_2, for example, is the sum of two elements in $\mathbf{A}$: the one in the first column and second row plus that in the second column and first row. These results are true generally.

If $\boldsymbol{x}$ is a vector of order $n \times 1$ with elements x_i for $i = 1, 2, \ldots, n$, and if $\mathbf{A}$ is a square matrix of order $n \times n$ with elements a_{ij} for $i, j = 1, 2, \ldots, n$, then

$$\begin{aligned} \mathbf{x}'\mathbf{A}\mathbf{x} &= \sum_i \sum_j x_i x_j a_{ij} \quad \text{[similar to (5.30)]} \\ &= \sum_i x_i^2 a_{ii} + \sum\sum_{i \neq j} x_i x_j a_{ij} \end{aligned}$$

and as in (5.31) this is

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + \sum_{j>i}\sum x_i x_j \left(a_{ij} + a_{ji}\right). \tag{5.32}$$

Thus $\mathbf{x}'\mathbf{A}\mathbf{x}$ is the sum of squares of the elements of $\mathbf{x}$, each square multiplied by the corresponding diagonal element of $\mathbf{A}$, plus the sum of products of the elements of $\mathbf{x}$, each product multiplied by the sum of the corresponding elements of $\mathbf{A}$; that is, the product of the ith and jth element of $\mathbf{x}$ is multiplied by $(a_{ij} + a_{ji})$.

Returning to the example, note that

$$\begin{aligned}
\mathbf{x}'\mathbf{A}\mathbf{x} &= x_1^2 + 7x_2^2 + 5x_3^2 + 6x_1x_2 + 5x_1x_3 + 4x_2x_3 \\
&= x_1^2 + 7x_2^2 + 5x_3^2 + x_1x_2\,(1+5) + x_1x_3\,(1+4) + x_2x_3\,(0+4) \\
&= \mathbf{x}'\begin{bmatrix} 1 & 1 & 1 \\ 5 & 7 & 0 \\ 4 & 4 & 5 \end{bmatrix}\mathbf{x}.
\end{aligned}$$

In this way we see that

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 2 & 3 \\ 4 & 7 & 6 \\ 2 & -2 & 5 \end{bmatrix}\mathbf{x} \text{ is the same as } \mathbf{x}'\mathbf{B}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 1 & 1 \\ 5 & 7 & 0 \\ 4 & 4 & 5 \end{bmatrix}\mathbf{x},$$

where $\mathbf{B}$ is different from $\mathbf{A}$. Note that the quadratic form is the same, even though the associated matrix is not the same. In fact, there is no unique matrix $\mathbf{A}$ for which any particular quadratic form can be expressed as $\mathbf{x}'\mathbf{A}\mathbf{x}$. Many matrices can be so used. Each one has the same diagonal elements, and in each of them the sum of each pair of symmetrically placed off-diagonal elements a_{ij} and a_{ji} must be the same; for example, (5.30) can also be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 2342 & -789 \\ -2336 & 7 & 1.37 \\ 794 & 2.63 & 5 \end{bmatrix}\mathbf{x}. \tag{5.33}$$

In particular, if we write

$$\mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 + 7x_2^2 + 5x_3^2 + x_1x_2\,(3+3) + x_1x_3\left(2\tfrac{1}{2} + 2\tfrac{1}{2}\right) + x_2x_3\,(2+2)$$

we see that it can be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\begin{bmatrix} 1 & 3 & 2\frac{1}{2} \\ 3 & 7 & 2 \\ 2\frac{1}{2} & 2 & 5 \end{bmatrix}\mathbf{x} \tag{5.34}$$

where **A** is now a symmetric matrix. As such it is unique; that is to say, for any particular quadratic form there is a unique symmetric matrix **A** for which the quadratic form can be expressed as $\mathbf{x}'\mathbf{A}\mathbf{x}$. It can be found in any particular case by rewriting the quadratic $\mathbf{x}'\mathbf{A}\mathbf{x}$ where **A** is not symmetric as $\mathbf{x}'[\frac{1}{2}(\mathbf{A}+\mathbf{A}')]\mathbf{x}$, because $\frac{1}{2}(\mathbf{A}+\mathbf{A}')$ is symmetric. For example, if **A** is the matrix used in (5.33), it is easily observed that $\frac{1}{2}(\mathbf{A}+\mathbf{A}')$ is the symmetric matrix used in (5.34).

Taking **A** as symmetric with $a_{ij} = a_{ji}$, we see from (5.32) that the quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ can be expressed as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_i x_i^2 a_{ii} + 2\sum_{j>i}\sum x_i x_j a_{ij}. \tag{5.35}$$

For example,

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= \mathbf{x}'\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33}\end{bmatrix}\mathbf{x} \\ &= a_{11}x_1^2 + a_{22}x_2^2 + a_{33}x_3^2 + 2(a_{12}x_1x_2 + a_{13}x_1x_3 + a_{23}x_2x_3).\end{aligned}$$

The importance of the symmetric **A** is that, when writing any particular quadratic function as $\mathbf{x}'\mathbf{A}\mathbf{x}$, there are many different matrices that can be used, but there is only one symmetric matrix—and it is unique for that particular function. Therefore, when dealing with quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$, we always take **A** as symmetric. This is convenient not only because the symmetric **A** is unique for any particular quadratic form, but also because symmetric matrices have many properties that are useful in studying quadratic forms, particularly those associated with analysis of variance. Hereafter, whenever we deal with a quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$, we assume $\mathbf{A} = \mathbf{A}'$.

A slightly more general (but not so useful) function is the second-degree function in two sets of variables **x** and **y**, say. For example,

$$\begin{aligned}\mathbf{x}'\mathbf{M}\mathbf{y} &= [x_1 \quad x_2]\begin{bmatrix} 2 & 4 & 3 \\ 7 & 6 & 5\end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \\ y_3\end{bmatrix} \\ &= 2x_1y_1 + 4x_1y_2 + 3x_1y_3 + 7x_2y_1 + 6x_2y_2 + 5x_2y_3.\end{aligned}$$

It is called a *bilinear form* and, as illustrated here, its matrix **M** does not have to be square as does the matrix in a quadratic form. Clearly, quadratic forms are special cases of bilinear forms—when **M** is square and $\mathbf{y} = \mathbf{x}$.

5.7 Positive Definite Matrices

All quadratic forms $\mathbf{x}'\mathbf{A}\mathbf{x}$ are zero for $\mathbf{x} = \mathbf{0}$. For some matrices $\mathbf{A}$ the corresponding quadratic form is zero *only* for $\mathbf{x} = \mathbf{0}$. For example,

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 2 & 2 & 1 \\ 2 & 5 & 1 \\ 1 & 1 & 2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= 2x_1^2 + 5x_2^2 + 2x_3^2 + 4x_1x_2 + 2x_1x_3 + 2x_2x_3 \\ &= (x_1 + 2x_2)^2 + (x_1 + x_3)^2 + (x_2 + x_3)^2,\end{aligned}$$

and, by the nature of this last expression, we see that (for elements of $\mathbf{x}$ being real numbers) it is positive unless all elements of $\mathbf{x}$ are zero, that is, $\mathbf{x} = \mathbf{0}$. Such a quadratic form is described as being positive definite. More formally,

$$\text{when } \mathbf{x}'\mathbf{A}\mathbf{x} > 0 \text{ for all } \mathbf{x} \text{ other than } \mathbf{x} = \mathbf{0},$$

then $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a *positive definite* quadratic form, and $\mathbf{A} = \mathbf{A}'$ is correspondingly a *positive definite* (p.d.) matrix.

There are also symmetric matrices $\mathbf{A}$ for which $\mathbf{x}'\mathbf{A}\mathbf{x}$ is zero for some non-null $\mathbf{x}$ as well as for $\mathbf{x} = \mathbf{0}$; for example,

$$\begin{aligned}\mathbf{x}'\mathbf{A}\mathbf{x} &= [x_1 \quad x_2 \quad x_3]\begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \\ &= 37x_1^2 + 13x_2^2 + 17x_3^2 - 4x_1x_2 - 48x_1x_3 - 6x_2x_3 \\ &= (x_1 - 2x_2)^2 + (6x_1 - 4x_3)^2 + (3x_2 - x_3)^2.\end{aligned}$$

This is zero for $\mathbf{x}' = [2 \quad 1 \quad 3]$, and for any scalar multiple thereof, as well as for $\mathbf{x} = \mathbf{0}$. This kind of quadratic form is called positive semidefinite and has for its formal definition:

$$\text{when } \mathbf{x}'\mathbf{A}\mathbf{x} \geq 0 \text{ for all } \mathbf{x} \textit{ and } \mathbf{x}'\mathbf{A}\mathbf{x} = 0 \text{ for some } \mathbf{x} \neq \mathbf{0}$$

then $\mathbf{x}'\mathbf{A}\mathbf{x}$ is a *positive semidefinite* quadratic form and hence $\mathbf{A} = \mathbf{A}'$ is a *positive semidefinite* (p.s.d.) matrix. The two classes of (forms and) matrices taken together, positive definite and positive semidefinite, are called *non-negative definite* (n.n.d.).

Example 5.12 *The sum of squares of (5.19),*

$$\sum_{t=1}^{n} (x_t - \bar{x})^2 = \mathbf{x}'\mathbf{C}\mathbf{x},$$

*is a positive semidefinite quadratic form because it is positive, except for being zero when all the x_i's are equal. Its matrix, $\mathbf{I} - \bar{\mathbf{J}}$, which is idempotent, is also p.s.d., as are all symmetric idempotent matrices (except **I**, which is the only p.d. idempotent matrix).*

Unfortunately there is no universal agreement on the definition of positive semidefinite. Most writers use it with the meaning defined here, but some use it in the sense of meaning non-negative definite. But on one convention there is universal agreement: p.d., p.s.d., and n.n.d. matrices are always taken as being symmetric. This is so because the definitions of these matrices are in terms of quadratic forms which can always be expressed utilizing symmetric matrices.

Theorem 5.3

a. For a matrix $\boldsymbol{X}$, the product $\boldsymbol{X}'\boldsymbol{X}$ is positive definite if $\boldsymbol{X}$ is of full column rank and is positive semidefinite otherwise.
b. For a matrix $\boldsymbol{Y}$, the product $\boldsymbol{Y}\boldsymbol{Y}'$ is positive definite if $\boldsymbol{Y}$ is of full row rank and is positive semidefinite otherwise.

Proof. See Exercise 5.23.

Exercises

1. Show that $\begin{bmatrix} 3 & 8 & 4 \\ 8 & 7 & -1 \\ 4 & -1 & 2 \end{bmatrix} + \begin{bmatrix} 1 & -1 & 3 \\ -1 & 2 & 4 \\ 3 & 4 & 6 \end{bmatrix}$ is symmetric.
2. If $\mathbf{x}$ and $\mathbf{y}$ are $n \times 1$ column vectors and $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices, which of the following expressions are undefined? Of those that are defined, which are bilinear forms, quadratic forms?
   a. $\mathbf{y} = \mathbf{Ax}$.
   b. $\mathbf{xy} = \mathbf{A'B}$.
   c. $\mathbf{x'Bx}$.
   d. $\mathbf{yBx}$.
   e. $\mathbf{y'B'Ax}$.
   f. $\mathbf{x'} = \mathbf{y'B'}$.
   g. $\mathbf{x'Ay}$.
   h. $\mathbf{y'A'By}$.
   i. $\mathbf{xy'} = \mathbf{B'}$.
3. If $\mathbf{A}$ is skew symmetric, prove that (a) $a_{ii} = 0$ and $a_{ij} = -a_{ji}$; and (b) $\mathbf{I} + \mathbf{A}$ is positive definite.
4. For a square matrix $\mathbf{A}$, prove that it is the sum of a symmetric and a skew-symmetric matrix.
5. If the product of two symmetric matrices is symmetric, prove that the matrices commute in multiplication.
6. Show that if $\mathbf{X'X} = \mathbf{X}$, then $\mathbf{X} = \mathbf{X'} = \mathbf{X}^2$
7. a. Show that the only real symmetric matrix whose square is null is the null matrix itself.
   b. Explain why $\mathbf{X'XGX'X} = \mathbf{X'X}$ implies $\mathbf{XGX'X} = \mathbf{X}$. **(Note:** The matrix $\boldsymbol{G}$ is a g-inverse of $\boldsymbol{X'X}$. Such an inverse will be defined in Chapter 8).
8. If $X_{r\times c} = [x_1\ x_2\ \dots\ x_c]$, prove that for symmetric $\mathbf{A}$ $tr[(\mathbf{AXX'})^2] = \Sigma_{j=1}^{c}\Sigma_{k=1}^{c}(\mathbf{x}_j'\ \mathbf{Ax}_k)^2$.
9. Prove that a product of orthogonal matrices is orthogonal.
10. a. If $\mathbf{A}$ is idempotent and symmetric, prove that it is positive semidefinite.

b. When **X** and **Y** are idempotent, prove that **XY** is, provided **X** and **Y** commute in multiplication.

c. Prove that $\mathbf{I} + \mathbf{KK}'$ is positive definite for real **K**.

11. Using $\mathbf{x}' = (1, 3, 5, 7, 9)$, derive or state the numerical value of **A**, **B**, and **C** such that

a. $1^2 + 3^2 + 5^2 + 7^2 + 9^2 = \mathbf{x}'\mathbf{Ax}$;

b. $(1 + 3 + 5 + 7 + 9)^2/5 = \mathbf{x}'\mathbf{Bx}$;

c. $(1-5)^2 + (3-5)^2 + (5-5)^2 + (7-5)^2 + (9-5)^2 = \mathbf{x}'\boldsymbol{C}\boldsymbol{x}$.

12. Explain why

a. $\mathbf{x}'\mathbf{Ax} = \mathbf{x}'\mathbf{A}'\mathbf{x}$, even when **A** is not symmetric;

b. $\mathbf{x}'\mathbf{Bx} = tr(\mathbf{x}'\mathbf{Bx})$;

c. $\mathbf{x}'\mathbf{Cx} = tr(\mathbf{Cxx}')$.

13. a. Suppose that the columns of the matrix $\boldsymbol{P}_{r\times c}$ are orthonormal, that is, $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_c$. Give an example to demonstrate that the rows of $\boldsymbol{P}$ are not necessarily orthonormal, that is, $\boldsymbol{PP}'$ is not necessarily an identity matrix.

b. Show that a square matrix having orthonormal rows must also have orthonormal columns.

14. Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are symmetric idempotent matrices.

a. If $\boldsymbol{AB} = \boldsymbol{B}$, show that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite.

b. Show that the reverse of (a) is true, that is, if $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite, then $\boldsymbol{AB} = \boldsymbol{B}$.

c. Show that $\boldsymbol{AB} = \boldsymbol{B}$ if and only if $(\boldsymbol{A} - \boldsymbol{B})^2 = \boldsymbol{A} - \boldsymbol{B}$.

15. Suppose that $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices of order $m \times n$. Show that

$$[tr(\boldsymbol{A}'\boldsymbol{B})]^2 \le tr(\boldsymbol{A}'\boldsymbol{A})tr(\boldsymbol{B}'\boldsymbol{B}).$$

This is known as the *Cauchy–Schwarz inequality* for matrices. Equality holds if and only if one of the matrices is a multiple of the other.

16. Suppose that $\boldsymbol{A}$ is a symmetric matrix of order $n \times n$ and let $\boldsymbol{B}$ be a skew-symmetric matrix of order $n \times n$. Show that $tr(\boldsymbol{AB}) = 0$.

17. Let $\boldsymbol{A}$ be a matrix of order $n \times n$. Suppose that $tr(\boldsymbol{PA}) = 0$ for every skew-symmetric matrix of order $n \times n$. Show that $\boldsymbol{A}$ is symmetric.

18. Suppose that the matrix $\boldsymbol{A}$ is positive definite and that the matrix $\boldsymbol{P}$ is nonsingular. Show that $\boldsymbol{P}'\boldsymbol{AP}$ is positive definite. What can be said about $\boldsymbol{P}'\boldsymbol{AP}$ if $\boldsymbol{A}$ is assumed to be positive semidefinite?

19. Let $\boldsymbol{V}$ be an $n \times n$ symmetric positive definite matrix such that $\boldsymbol{V} = \boldsymbol{P}'\boldsymbol{P}$, where $\boldsymbol{P}$ is a matrix of order $n \times n$, and let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix. Show that

    a. $\boldsymbol{P}$ is a nonsingular matrix.

    b. $\boldsymbol{PAP}'$ is idempotent if and only if $\boldsymbol{AV}$ is idempotent.

    c. If $\boldsymbol{AV}$ is idempotent, then $r(\boldsymbol{PAP}') = tr(\boldsymbol{AV})$.

20. Show that any two of the conditions,

    (i) $\boldsymbol{P}$ is a square matrix of order $n \times n$, (ii) $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_n$, and (iii) $\boldsymbol{PP}' = \boldsymbol{I}_n$ imply the third, that is, imply that $\boldsymbol{P}$ is orthogonal.

21. Consider the following two-way crossed classification model concerning two factors, denoted by A and B, having 4 and 3 levels, respectively, with two replications per treatment combination:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk},$$

where α_i represents the ith level of factor A, β_j represents the jth level of factor B, $(\alpha\beta)_{ij}$ is the corresponding interaction term, and ε_{ijk} is an experimental error term, $i = 1, 2, 3, 4$; $j = 1, 2, 3$; $k = 1, 2$. Consider the following sums of squares from the corresponding analysis of variance table:

$$SS_A = 6\sum_{i=1}^{4}(\bar{y}_{i..} - \bar{y}_{...})^2$$

$$SS_B = 8\sum_{j=1}^{3}(\bar{y}_{.j.} - \bar{y}_{...})^2$$

$$SS_{AB} = 2\sum_{i=1}^{4}\sum_{j=1}^{3}(\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...})^2,$$

where

$$\bar{y}_{i..} = \frac{1}{6}\sum_{j=1}^{3}\sum_{k=1}^{2} y_{ijk}$$

$$\bar{y}_{.j.} = \frac{1}{8}\sum_{i=1}^{4}\sum_{k=1}^{2} y_{ijk}$$

$$\bar{y}_{ij.} = \frac{1}{2}\sum_{k=1}^{2} y_{ijk}$$

$$\bar{y}_{...} = \frac{1}{24}\sum_{i=1}^{4}\sum_{j=1}^{3}\sum_{k=1}^{2} y_{ijk}.$$

Derive the matrices $\boldsymbol{P}_1$, $\boldsymbol{P}_2$, and $\boldsymbol{P}_3$ such that

$$SS_A = \boldsymbol{y}'\boldsymbol{P}_1\boldsymbol{y}$$
$$SS_B = \boldsymbol{y}'\boldsymbol{P}_2\boldsymbol{y}$$
$$SS_{AB} = \boldsymbol{y}'\boldsymbol{P}_3\boldsymbol{y},$$

where $\boldsymbol{y}$ is the column vector of all 24 values of y_{ijk} with k varying first from 1 to 2, followed by j varying from 1 to 3, and then i varying from 1 to 4.

22. Let $\boldsymbol{A}$ be an $n \times n$ positive definite matrix and let $\boldsymbol{d}$ be a vector of n elements. Consider the matrix,

$$\boldsymbol{B} = \boldsymbol{A} - \frac{1}{c}\boldsymbol{dd}',$$

where c is a scalar such that $c > \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d}$. Show that the inverse of $\boldsymbol{B}$ exists and is given by

$$\boldsymbol{B}^{-1} = \boldsymbol{A}^{-1} + \gamma\boldsymbol{A}^{-1}\boldsymbol{dd}'\boldsymbol{A}^{-1},$$

where $\gamma = (c - \boldsymbol{d}'\boldsymbol{A}^{-1}\boldsymbol{d})^{-1}$.

23. Show that if

    a. $\boldsymbol{X}$ is of full column rank, then $\boldsymbol{X}'\boldsymbol{X}$ is positive definite, otherwise $\boldsymbol{X}'\boldsymbol{X}$ is positive semidefinite.

    b. $\boldsymbol{Y}$ is of full row rank, then $\boldsymbol{YY}'$ is positive definite, otherwise $\boldsymbol{YY}'$ is positive semidefinite.

References

Andreson, T. W., Olkin, I., and Underhill, L. G. (1987). Generation of random orthogonal matrices. *SIAM J. Sci. Stat. Comput.*, 8, 625–629.

Gantmacher, F. R. (1959). *The Theory of Matrices*, Vols. I and II. Chelsea, New York.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.

Heiberger, R. M., Velleman, P. F., and Ypelaar, M. A. (1983). Generating test data with independently controllable features for multivariate general linear forms. *J. Am. Stat. Assoc.*, 78, 585–595.

Khuri, A. I. and Good, I. J. (1989). The parameterization of orthogonal matrices: A review mainly for statisticians. *South African Stat. J.*, 23, 231–250.

Khuri, A. I. and Myers, R. H. (1981). Design related robustness of tests in regression models. *Comm. Stat. Theory Methods*, 10, 223–235.

6
Eigenvalues and Eigenvectors

This chapter discusses a particular linear transformation which maps a vector into a scalar multiple of itself. More specifically, let $\boldsymbol{A}$ be a matrix of order $n \times n$ which serves as a linear transformation of a vector $\boldsymbol{u}$ such that

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u}, \tag{6.1}$$

where λ is a scalar. It is of interest to know the conditions under which $\boldsymbol{u}$ and λ can exist so that (6.1) is valid for $\boldsymbol{u} \neq 0$.

6.1 Derivation of Eigenvalues

Equation (6.1) can be written as

$$(\mathbf{A} - \lambda \boldsymbol{I}_n)\mathbf{u} = \mathbf{0}, \tag{6.2}$$

from which we know that if $\mathbf{A} - \lambda \boldsymbol{I}_n$ is nonsingular, the only solution is **u** = **0**. But if we were to have a non-null solution, then (6.2) indicates that the columns of $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ are linearly dependent and therefore the rank of $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ must be less than *n*, which implies that $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ is a singular matrix (see Section 4.15). This shows that the determinant of the matrix in (6.2) must be equal to zero, that is,

$$|\mathbf{A} - \lambda \boldsymbol{I}_n| = 0. \tag{6.3}$$

Hence, (6.3) is the condition for **u** and λ to exist such that (6.2) is true, that is, pick λ so that the determinant of $\mathbf{A} - \lambda \boldsymbol{I}_n$ is zero.

Equation (6.3) is called the *characteristic equation* of **A**. For **A** of order *n* × *n*, the characteristic equation is a polynomial equation in λ of degree *n*, with *n* roots to be denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$, some of which may be zero. These roots are called *latent roots, characteristic roots, or eigenvalues*. Corresponding to each root λ_i is a vector $\mathbf{u}_i$ satisfying (6.1):

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{for} \quad i = 1, \ldots, n, \tag{6.4}$$

and these vectors $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_n$ are correspondingly called *latent vectors, characteristic vectors*, or *eigenvectors*. Eigenvalues and eigenvectors will be used in this book. The prefix eigen originated from the German word "eigen" which means "characteristic," "particular," "special," or "unique to."

Example 6.1 *The matrix*

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

has the characteristic equation

$$\left| \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0; \quad \textit{that is,} \quad \begin{vmatrix} 1-\lambda & 4 \\ 9 & 1-\lambda \end{vmatrix} = 0. \tag{6.5}$$

Expanding the determinant in (6.5) gives

$$(1-\lambda)^2 - 36 = 0; \quad \textit{that is,} \quad \lambda = -5 \text{ or } 7.$$

Note that characteristic equations are, from (6.3), always of the form shown in (6.5): equated to zero is the determinant of ***A*** *amended by subtracting λ from each diagonal element.*

The derivation of eigenvectors corresponding to solutions of a characteristic equation is discussed in Section 6.3, but meanwhile it can be seen here that

$$\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5 \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7 \begin{bmatrix} 2 \\ 3 \end{bmatrix}, \tag{6.6}$$

these being examples of (6.1). Thus $\begin{bmatrix} 2 \\ -3 \end{bmatrix}$ *is an eigenvector corresponding to the eigenvalue –5, and* $\begin{bmatrix} 2 \\ 3 \end{bmatrix}$ *is a vector for the root 7.*

Example 6.2 *The characteristic equation for*

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix} \quad \text{is} \quad \begin{vmatrix} 2-\lambda & 2 & 0 \\ 2 & 1-\lambda & 1 \\ -7 & 2 & -3-\lambda \end{vmatrix} = 0.$$

An expansion of the determinant gives, after simple arithmetic,

$$\lambda^3 - 13\lambda + 12 = 0, \quad \textit{equivalent to} \quad (\lambda - 1)(\lambda^2 + \lambda - 12) = 0.$$

Solutions are $\lambda = 1, 3,$ *and* -4*, and these are the eigenvalues of* $\boldsymbol{A}$*.*

Note: If an eigenvalue λ of a matrix $\boldsymbol{A}$ is greater than one, then the action of $\boldsymbol{A}$ on a corresponding eigenvector $\boldsymbol{x}$ amounts to “stretching” it without changing its direction. However, if $0 < \lambda < 1$, then $\boldsymbol{A}$ acts on $\boldsymbol{x}$ by shrinking its length without a change in direction. In the event $\lambda < 0$, then $\boldsymbol{Ax}$ has an opposite direction to that of $\boldsymbol{x}$. These three cases are illustrated in Figure 6.1.

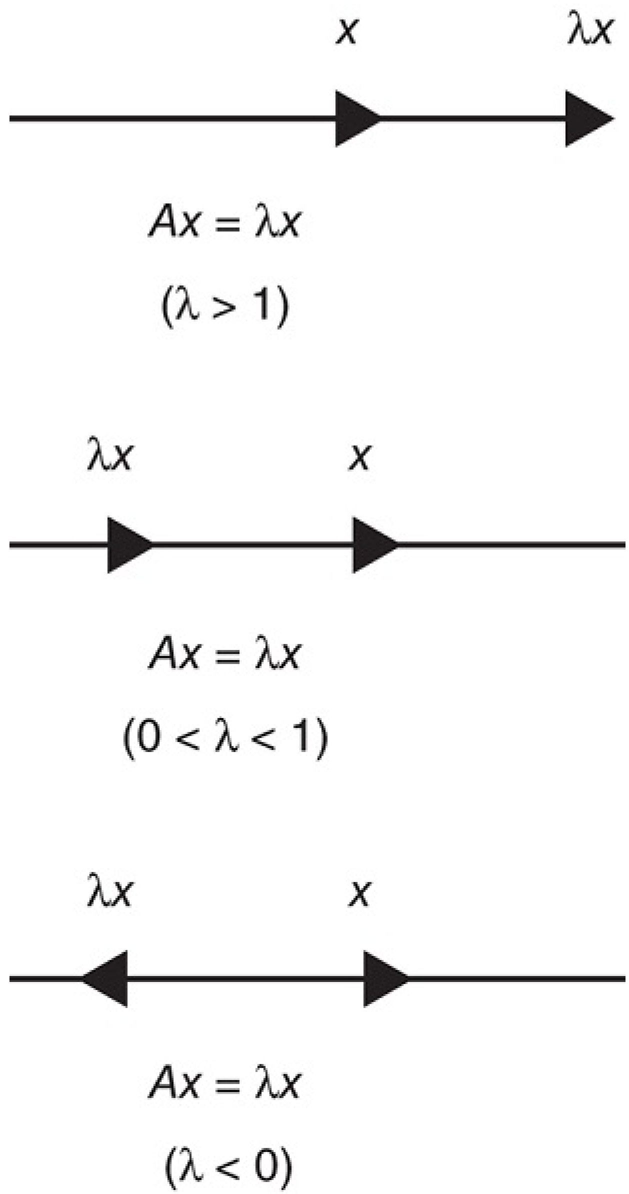


Figure 6.1 The Matrix ***A*** with Its Eigenvalue λ and Eigenvector ***x***.

6.1.1 Plotting Eigenvalues

Some of the eigenvalues of a matrix can be complex numbers with each having a real part and an imaginary part. Figure 6.2 shows a scatter plot of eigenvalues of a matrix of order 50 × 50 whose elements were randomly generated from a standard normal distribution with mean zero and standard deviation 1. Each point in the plot represents an eigenvalue with a real part (on the horizontal axis) and a possible imaginary part (on the vertical axis). If the latter value is equal to zero, then the eigenvalue is a real number. This plot provides a convenient way to graphically describe the distribution of the eigenvalues of a matrix. The plot was derived using the MATLAB computer package (see Chapter 16 for more details).

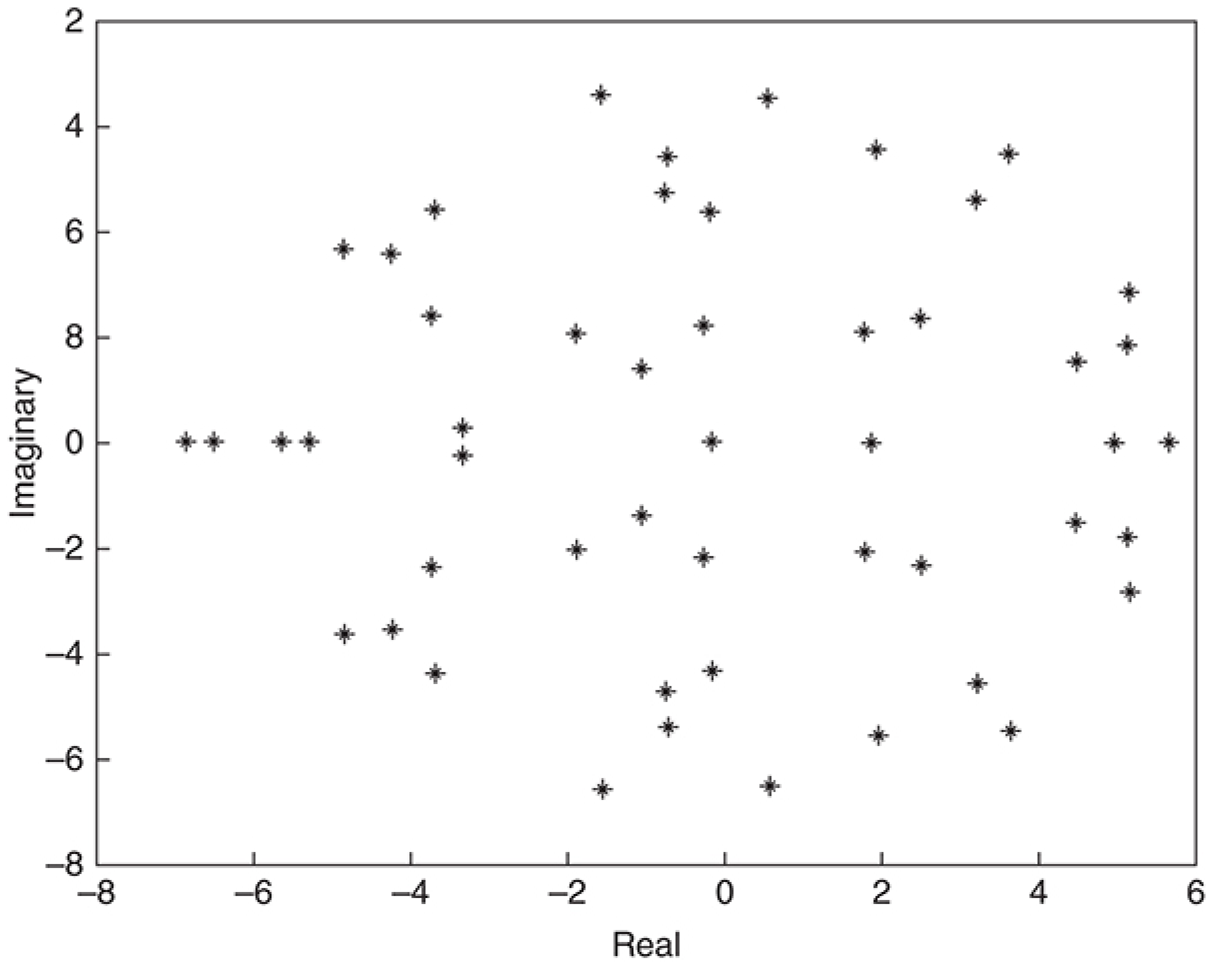


Figure 6.2 A Scatter Plot of Eigenvalues of a Matrix of Order 50 × 50.

6.2 Elementary Properties of Eigenvalues

Several properties of eigenvalues that stem directly from their definition in (6.1) and the resulting characteristic equation (6.3) command attention before considering eigenvectors.

6.2.1 Eigenvalues of Powers of a Matrix

The defining equation is $\mathbf{Au} = \lambda\mathbf{u}$ of (6.1). Premultiplying this by $\mathbf{A}$ and using (6.1) again gives

$$\mathbf{A}^2\mathbf{u} = \mathbf{A}\lambda\mathbf{u} = \lambda\mathbf{Au} = \lambda(\lambda\mathbf{u}) = \lambda^2\mathbf{u}.$$

Comparing this equation with (6.1), we see that it defines λ^2 as being an eigenvalue of $\mathbf{A}^2$. Similarly, λ^3 is an eigenvalue of $\mathbf{A}^3$: $\mathbf{A}^3\mathbf{u} = \mathbf{A}\lambda^2\mathbf{u} = \lambda^2\mathbf{Au} = \lambda^3\mathbf{u}$, and in general λ^k is an eigenvalue of $\mathbf{A}^k$:

$$\mathbf{A}^k\mathbf{u} = \lambda^k\mathbf{u}. \tag{6.7}$$

Furthermore, when $\mathbf{A}$ is nonsingular (6.1) gives $\mathbf{u} = \mathbf{A}^{-1}\lambda\mathbf{u} = \lambda\mathbf{A}^{-1}\mathbf{u}$ and so

$$\mathbf{A}^{-1}\mathbf{u} = \lambda^{-1}\mathbf{u}. \tag{6.8}$$

Hence, when λ is an eigenvalue of $\boldsymbol{A}$, then λ^k is an eigenvalue of $\mathbf{A}^k$ where k is positive if $\mathbf{A}$ is singular, and k is positive or negative if $\mathbf{A}$ is nonsingular. In particular, when $\mathbf{A}$ is nonsingular with eigenvalue λ, the inverse $\mathbf{A}^{-1}$ has $1/\lambda$ as an eigenvalue. (Recall that $k = 0$ gives $\mathbf{A}^k = \mathbf{A}^0 = \mathbf{I}$).

6.2.2 Eigenvalues of a Scalar-by-Matrix Product

Multiplying $\mathbf{Au} = \lambda\mathbf{u}$ of (6.1) by a scalar c gives

$$c\mathbf{Au} = c\lambda\mathbf{u}. \tag{6.9}$$

Hence, when λ is an eigenvalue of $\boldsymbol{A}$, we have $c\lambda$ being an eigenvalue of $c\mathbf{A}$.

Rewriting (6.9) as $\mathbf{A}(c\mathbf{u}) = \lambda(c\mathbf{u})$ shows that when $\mathbf{u}$ is an eigenvector of $\boldsymbol{A}$ so also is $c\,\boldsymbol{u}$, both vectors corresponding to the same eigenvalue λ.

When $\boldsymbol{A}$ has eigenvalue λ then $\boldsymbol{A} + c\mathbf{I}$ for scalar c has eigenvalue $\lambda + c$. This is so because

$$(\mathbf{A} + c\mathbf{I})\mathbf{u} = \mathbf{Au} + c\mathbf{u} = \lambda\mathbf{u} + c\mathbf{u} = (\lambda + c)\mathbf{u}.$$

Combining this with (6.8) shows that when $(\mathbf{A} + c\mathbf{I})^{-1}$ exists, it has $1/(\lambda + c)$ as an eigenvalue.

6.2.3 Eigenvalues of Polynomials

A consequence of the two preceding sections is that when **A** has an eigenvalue λ then a polynomial in ***A***, say $f(\boldsymbol{A})$, has an eigenvalue $f(\lambda)$. For example, consider the polynomial $f(\mathbf{A}) = \mathbf{A}^3 + 7\mathbf{A}^2 + \mathbf{A} + 5\mathbf{I}$. If λ and ***u*** are a corresponding eigenvalue and vector of ***A***,

$$\mathbf{A}^3\mathbf{u} = \lambda^3\mathbf{u}, \quad \mathbf{A}^2\mathbf{u} = \lambda^2\mathbf{u}, \quad \mathbf{A}\mathbf{u} = \lambda\mathbf{u}, \quad \text{and} \quad 5\mathbf{I}\mathbf{u} = 5\mathbf{u}.$$

Hence,

$$\begin{aligned}(\mathbf{A}^3 + 7\mathbf{A}^2 + \mathbf{A} + 5\mathbf{I})\mathbf{u} &= \mathbf{A}^3\mathbf{u} + 7\mathbf{A}^2\mathbf{u} + \mathbf{A}\mathbf{u} + 5\mathbf{I}\mathbf{u} \\ &= \lambda^3\mathbf{u} + 7\lambda^2\mathbf{u} + \lambda\mathbf{u} + 5\mathbf{u} \\ &= \left(\lambda^3 + 7\lambda^2 + \lambda + 5\right)\mathbf{u}.\end{aligned}$$

Thus $f(\lambda) = \lambda^3 + 7\lambda^2 + \lambda + 5$ is an eigenvalue. Extension to any polynomial $f(\mathbf{A})$ gives the result that $f(\mathbf{A})$ has eigenvalue $f(\lambda)$. An important special case is that $e^{\mathbf{A}} = \sum_{i=0}^{\infty} \mathbf{A}^i/i!$ has eigenvalue e^{λ}.

Example 6.3 (continuation of Example 6.1)

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix}$$

has eigenvalues –5 and 7. The eigenvalues of

$$\begin{aligned} f(\mathbf{A}) &= 2\mathbf{A}^2 + 2\mathbf{A} - 12\mathbf{I} \\ &= 2\begin{bmatrix} 37 & 8 \\ 18 & 37 \end{bmatrix} + 2\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} - 12\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 64 & 24 \\ 54 & 64 \end{bmatrix} \end{aligned}$$

are given by $\begin{vmatrix} 64-\lambda & 24 \\ 54 & 64-\lambda \end{vmatrix} = 0$*, that is,* $\lambda^2 - 128\lambda + 2800 = 0$*, so that* λ *is 100 or 28. And* $100 = 2(7)^2 + 2(7) - 12 = f(7)$*, and* $28 = f(-5)$.

6.2.4 The Sum and Product of Eigenvalues

The characteristic equation (6.3) of degree 3,

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} = 0, \tag{6.10}$$

reduces to

$$-\lambda^3 + (-\lambda)^2\, tr_1(\mathbf{A}) + (-\lambda)\, tr_2(\mathbf{A}) + |\mathbf{A}| = 0 \tag{6.11}$$

using diagonal expansion of determinants and the $\mathrm{tr}_i(\mathbf{A})$ notation defined in Section 3.6. For ***A*** of order $n \times n$ this takes the form

$$(-\lambda)^n + (-\lambda)^{n-1}\, tr_1\,(\mathbf{A}) + (-\lambda)^{n-2}\, tr_2(\mathbf{A}) \\ + \cdots + (-\lambda)\, tr_{n-1}\,(\mathbf{A}) + tr_n\,(\mathbf{A}) = 0. \tag{6.12}$$

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are roots of this equation, then it is equivalent to

$$(\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda) = 0,$$

which expands to

$$(-\lambda)^n + (-\lambda)^{n-1} \sum \lambda_l + (-\lambda)^{n-2} \sum_{l \neq j} \lambda_l \lambda_j + \cdots + \prod_{l=1}^{n} \lambda_i = 0. \tag{6.13}$$

Equating the coefficients of $(-\lambda)^{n-1}$ in (6.12) and (6.13), and also equating the final terms, gives

$$\sum_{l=1}^{n} \lambda_i = tr_1\,(\mathbf{A}) = tr\,(\mathbf{A}) \quad \text{and} \quad \prod_{i=1}^{n} \lambda_l = tr_n\,(\mathbf{A}) = |\mathbf{A}|\,. \tag{6.14}$$

Hence, the sum of the eigenvalues of a matrix equals its trace, and their product equals its determinant.

Example 6.4 (continuation of Example 6.2)

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix} \quad \text{has determinant} \quad \begin{vmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{vmatrix} = -12$$

*and its eigenvalues are 1, 3, and − 4. Their sum is 1 + 3 − 4 = 0 as is tr(**A**) = 2 + 1 − 3 = 0; and their product is* $1(3)(-4) = -12$, $= |\mathbf{A}|$.

6.3 Calculating Eigenvectors

Suppose λ_k is an eigenvalue of the $n \times n$ matrix $\mathbf{A}$. It is a solution of the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$. Calculating an eigenvector corresponding to λ_k requires finding a non-null $\mathbf{u}$ to satisfy $\mathbf{Au} = \lambda_k\mathbf{u}$, which is equivalent to solving

$$(\mathbf{A} - \lambda_k\mathbf{I})\mathbf{u} = \mathbf{0}. \tag{6.15}$$

We recall that $\mathbf{A} - \lambda_k\mathbf{I}$ is a singular matrix. Let ρ denote its rank which must be smaller than n. It follows that the orthogonal complement of the row space of $\boldsymbol{A} - \lambda_k\boldsymbol{I}$ is of dimension $n - \rho > 0$ and is therefore non-empty. This orthogonal complement forms a subspace of the n-dimensional Euclidean space. Any solution to (6.15) must belong to this subspace. Hence, (6.15) has at least one non-null solution $\boldsymbol{u}$. To find such a solution, ρ linearly independent equations are selected from (6.15), then $n - \rho$ elements of $\boldsymbol{u}$ are chosen arbitrarily and used to solve for the remaining ρ elements of $\boldsymbol{u}$. Thus, infinitely many non-null solutions to (6.15) can be found in this manner. These solutions form a subspace called the *eigenspace* corresponding to λ_k.

6.3.1 Simple Roots

Since λ_k is a solution to a polynomial equation it can be a solution more than once, in which case it is called a *multiple root*. We deal first with λ_k being a solution only once, in which case we call λ_k a *simple root*.

Whenever λ_k is a simple root, the rank ρ of $\mathbf{A} - \lambda_k\mathbf{I}$ is equal to $n - 1$ (see Section 7.1). Hence, $n - \rho = 1$. This implies that the eigenspace corresponding to λ_k is of dimension 1. Therefore, there exists only one linearly independent eigenvector corresponding to the simple root. This means that all values of $\mathbf{u}_k$ for a given λ_k are multiples of one another.

Example 6.5 (continuation of Example 6.2)

The eigenvalues of $\mathbf{A} = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix}$ *are $\lambda_1 = 1$, $\lambda_2 = 3$ and $\lambda_3 = -4$. For $\lambda_1 = 1$ we get,*

$$\mathbf{A} - \mathbf{I} = \begin{bmatrix} 1 & 2 & 0 \\ 2 & 0 & 1 \\ -7 & 2 & -4 \end{bmatrix}$$

Making the substitution in (6.15), we obtain

$$u_1 + 2u_2 = 0,$$
$$2u_1 + u_3 = 0,$$
$$-7u_1 + 2u_2 - 4u_3 = 0.$$

Here, $\rho = 2$ and $n - \rho = 1$. Thus, one of u_1, u_2, u_3 can be given an arbitrary value. For example, we choose $u_1 = -2$. Using the first two of the above three equations, which are linearly independent, and then solving for u_2 and u_3, we get $u_2 = 1$ and $u_3 = 4$. Hence, an eigenvector corresponding to $\lambda_1 = 1$ is $(-2, 1, 4)'$.

Similarly, with $\lambda_2 = 3$, we get the equations,

$$-u_1 + 2u_2 = 0$$
$$2u_1 - 2u_2 + u_3 = 0$$
$$-7u_1 + 2u_2 - 6u_3 = 0.$$

The first two equations are linearly independent. Assigning a value to u_1, for example $u_1 = -2$, and solving for u_2 and u_3, we get $u_2 = -1$, $u_3 = 2$. *The resulting eigenvector is* $(-2, -1, 2)$.

Finally, for $\lambda_3 = -4$ *and following similar steps as before, we obtain* $(1, -3, 13)'$ *as a corresponding eigenvector.*

6.3.2 Multiple Roots

We now deal with eigenvalues λ_k that are solutions of the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$ more than one time. Any λ_k for which this occurs is called a *multiple eigenvalue* and the number of times that it is a solution is called its *multiplicity*. In general, we formulate $\boldsymbol{A}$ of order $n \times n$ as having s distinctly different eigenvalues $\lambda_1, \ldots, \lambda_s$ with λ_k having multiplicity m_k for $k = 1, 2, \ldots, s$ and, of course, $\sum_{k=1}^{s} m_k = n$. Note two features of this formulation. If zero is an eigenvalue (as it can be), it is one of the λ_k'a; and simple eigenvalues are also included, their multiplicities each being 1.

As already noted, each eigenvalue λ_k having multiplicity $m_k = 1$ leads to $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1$ and hence, correspondingly just one linearly independent eigenvector. In contrast, for any multiple eigenvalue λ_k, the rank of $\mathbf{A} - \lambda_k\mathbf{I}$ must be ascertained for each value because, as shall be seen, this rank value plays an important role in subsequent developments.

Example 6.6 *The characteristic equation of*

$$\mathbf{A} = \begin{bmatrix} -1 & -2 & -2 \\ 1 & 2 & 1 \\ -1 & -1 & 0 \end{bmatrix} \quad \text{reduces to} \quad (\lambda - 1)\left(\lambda^2 - 1\right) = 0$$

so that the eigenvalues are $\lambda_1 = 1$ *with* $m_1 = 2$, *and* $\lambda_2 = -1$ *with* $m_2 = 1$. *For* $\lambda_1 = 1$,

$$\mathbf{A}-\lambda_1\mathbf{I}=\begin{bmatrix}-2 & -2 & -2\\ 1 & 1 & 1\\ -1 & -1 & -1\end{bmatrix} \quad \text{has rank } 1, \tag{6.16}$$

so that there are 3 − 1 = 2 *linearly independent eigenvectors of* **A** *corresponding to* λ_1 = 1. *Using* (6.16), *the elements* u_1, u_2, u_3 *satisfy the following three equations:*

$$\begin{aligned}-2u_1-2u_2-2u_3&=0\\ u_1+u_2+u_3&=0\\ -u_1-u_2-u_3&=.\end{aligned}$$

Only one equation is linearly independent. Selecting, for example, the second one, we have

$$u_1+u_2+u_3=0. \tag{6.17}$$

Two of the u'_i*s are linearly independent. We can therefore choose* u_2 and u_3 *to have arbitrary values and use* (*6.17*) *to solve for* u_1. *Since we need to have two linearly independent eigenvectors for the eigenvalue* λ_1, *we can assign two pairs of values for* u_2 *and* u_3 *so that, with the corresponding two* u_1 *values, we obtain two linearly independent eigenvectors. For example, choosing the two pairs* (1, −1), (1, 1) *for* u_2 *and* u_3, *the corresponding two values of* u_1 *(from using (*6.17*)) are 0, −2. This results in the two linearly independent eigenvectors (0, 1, −1)′, (− 2, 1, 1)′ that correspond to* λ_1 = 1.

For λ_2 = −1 *we get, from using (*6.15*)*

$$\mathbf{A}-\lambda_2\mathbf{I}=\begin{bmatrix}0 & -2 & -2\\ 1 & 3 & 1\\ -1 & -1 & 1,\end{bmatrix}$$

The rank of this matrix is 2. Hence, only one linearly independent eigenvector exists for λ_2. *Using a similar procedure as before, we find a corresponding eigenvector given by (2, −1, 1)′. All three eigenvectors thus obtained are linearly independent.*

Example 6.7 *The characteristic equation of*

$$\mathbf{A}=\begin{bmatrix}0 & 1 & 2\\ 2 & 3 & 0\\ 0 & 4 & 5\end{bmatrix} \quad \text{reduces to} \quad (\lambda-1)^2(\lambda-6)=0,$$

so that eigenvalues are λ_1 = 1 *with* m_1 = 2, *and* λ_2 = 6 *with* m_2 = 1. *Then,*

$$\mathbf{A}-\lambda_1\mathbf{I}=\begin{bmatrix}-1 & 1 & 2\\ 2 & 2 & 0\\ 0 & 4 & 4\end{bmatrix} \quad \text{and} \quad \mathbf{A}-\lambda_2\mathbf{I}=\begin{bmatrix}-6 & 1 & 2\\ 2 & -3 & 0\\ 0 & 4 & -1\end{bmatrix} \tag{6.18}$$

both have rank 2, so there is $3 - 2 = 1$ *linearly independent eigenvector corresponding to each of* λ_1 *and* λ_2*. Using the above two matrices, it can be verified that the following are corresponding eigenvectors:* $\mathbf{u}'_1 = (1, -1, 1)$ *and* $\mathbf{u}_2' = (3, 2, 8)$*. Note that in this example although* $n = 3$ *there are only* $2(< n)$ *linearly independent eigenvectors. We refer to this subsequently.*

6.4 The Similar Canonical Form

6.4.1 Derivation

Every eigenvalue λ_i has a corresponding eigenvector $\mathbf{u}_i$ for which

$$\mathbf{A}\mathbf{u}_i = \lambda_i \mathbf{u}_i \quad \text{for} \quad i = 1, 2, \ldots, n. \tag{6.19}$$

This is true for all *n* roots of the characteristic equation, which means that for λ_k of multiplicity m_k there are m_k equations like (6.19), each involving the same λ_k; if there are m_k linearly independent eigenvectors $\mathbf{u}_k$ corresponding to λ_k (as in Example 6.6), there will be one equation for each vector, whereas if there are not m_k linearly independent eigenvectors $\mathbf{u}_k$ (as in Example 6.7), there can still be m_k equations with some of the $\mathbf{u}_k$'s being repetitions of others. In all cases, the array of equations represented by 6.19 can be written as

$$\mathbf{A}\begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Then, on defining $\boldsymbol{U}$ as the matrix of *n* eigenvectors and $\boldsymbol{D}$ as the diagonal matrix of eigenvalues, namely,

$$\mathbf{U} = [\mathbf{u}_1 \quad \mathbf{u}_2 \quad \cdots \quad \mathbf{u}_n] \quad \text{and} \quad \mathbf{D} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \tag{6.20}$$

we have

$$\mathbf{AU} = \mathbf{UD}. \tag{6.21}$$

$\boldsymbol{D}$ is known as the *canonical form under similarity* or equivalently as the *similar canonical form*.

The repetition of $\mathbf{u}_k$'s is illustrated for Example 6.7 by

$$\mathbf{U} = \begin{bmatrix} 1 & 1 & 3 \\ -1 & -1 & 2 \\ 1 & 1 & 8 \end{bmatrix} \quad \text{with} \quad \mathbf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}. \tag{6.22}$$

Through permitting this repetition whenever necessary, the form (6.21) can always be made to exist. However, in cases involving repeated $\mathbf{u}_k$'s it is clear that $\boldsymbol{U}$ is singular. But for the cases when $\boldsymbol{U}$ is nonsingular, that is, when all *n* eigenvectors are linearly independent, (6.21) can be expressed as

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{D} = \mathrm{diag}\left\{\lambda_1, \lambda_2, \ldots, \lambda_n\right\}, \tag{6.23}$$

The product on the left is known as reduction to the similar canonical form; $\boldsymbol{D}$ on the right is the diagonal matrix of eigenvalues.

The existence of (6.23) depends on $\boldsymbol{U}$ being nonsingular. $\boldsymbol{AU} = \boldsymbol{UD}$ of (6.21) always exists, but rearranging it as $\mathbf{U}^{-1}\boldsymbol{AU} = \boldsymbol{D}$ of (6.23) requires $\boldsymbol{U}$ to be nonsingular. Within the context of eigenvalues and eigenvectors, the existence of nonsingular $\boldsymbol{U}$ is sometimes referred to as *$\boldsymbol{A}$being diagonable*, since $\boldsymbol{D}$ is diagonal. An important theorem provides us with conditions for ascertaining whether $\boldsymbol{U}$ is singular or not. We refer to it as the diagonability theorem.

Theorem 6.1 (Diagonability Theorem)

$\mathbf{A}_{n \times n}$, *having eigenvalues* λ_k *with multiplicity* m_k *for* $k = 1, 2, \ldots, s$ *and* $\sum_{k=1}^{s} m_k = n$, *has n eigenvectors that are linearly independent if and only if* $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ for all $k = 1, 2, \ldots, s$; *whereupon* $\boldsymbol{U}$ of (6.20) *is nonsingular and* $\mathbf{A}$ *is diagonable as* $\mathbf{U}^{-1}\boldsymbol{AU} = \boldsymbol{D}$ *of (6.23).*

Proof of this theorem is lengthy and is given in Section 7.1, together with prerequisite lemmas.

Existence of $\mathbf{U}^{-1}$ depends upon ascertaining if

$$r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k. \tag{6.24}$$

Each eigenvalue satisfying (6.24) is called a *regular* eigenvalue. When every eigenvalue is regular, $\mathbf{U}^{-1}$ exists and $\boldsymbol{A}$ is called a *regular matrix*. Whenever one or more eigenvalues are not regular, $\mathbf{U}^{-1}$ does not exist. Thus, a single violation of (6.24) ensures the nonexistence of $\mathbf{U}^{-1}$, whereupon $\mathbf{U}^{-1}\mathbf{AU}$ does not exist and $\boldsymbol{A}$ is said to be a *deficient* or *defective matrix*. Note, though, that there is no need to check (6.24) for simple eigenvalues (multiplicity $m_k = 1$) because, as already alluded to and proved in Section 7.1, (6.24) is always satisfied for such values.

Example 6.8 (continuation of Example 6.2)

Each eigenvalue has $m_k = 1$, *we know that (6.24) is satisfied, and so* $\mathbf{U}^{-1}$ *exists. Assembling the eigenvectors into a matrix* $\boldsymbol{U}$ *we get*

$$\mathbf{U} = \begin{bmatrix} -2 & -2 & 1 \\ 1 & -1 & -3 \\ 4 & 2 & 13 \end{bmatrix} \quad \text{and} \quad \mathbf{U}^{-1} = \frac{1}{70}\begin{bmatrix} -7 & 28 & 7 \\ -25 & -30 & -5 \\ 6 & -4 & 4 \end{bmatrix}. \tag{6.25}$$

With these values

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D} = \mathrm{diag}\{1, 3, -4\} \quad \text{and} \quad \mathbf{A} = \mathbf{UDU}^{-1}. \tag{6.26}$$

The reader should check these calculations.

Example 6.9 (continuation of Example 6.6)

Equation (6.16) has $\lambda_1 = 1$ *with* $m_1 = 2$*; and* $r(\mathbf{A} - \lambda_1\mathbf{I}) = 1 = 3 - 2 = n - m_1$ *satisfies (6.24). The other eigenvalue is* $\lambda_2 = -1$ *with* $m_2 = 1$ *which we know satisfies (6.24); that is, (6.24) is satisfied by all* λ_k*'s. Hence,* $\mathbf{U}^{-1}$ *exists:*

$$\mathbf{U} = \begin{bmatrix} 0 & -2 & 2 \\ 1 & 1 & -1 \\ -1 & 1 & 1 \end{bmatrix}, \quad \text{with} \quad \mathbf{U}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix},$$

and it is easily checked that

$$\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D} = \operatorname{diag}\{1, 1, -1\} \quad \text{and} \quad \mathbf{A} = \mathbf{UDU}^{-1}.$$

Example 6.10 (continuation of Example 6.7)

Here we have $\lambda_1 = 2$ *with* $m_1 = 2$ *and from (6.18),*

$$r(\mathbf{A} - \lambda_1\mathbf{I}) = 2 \neq n - m_1 = 3 - 2 = 1.$$

Hence (6.24) is not satisfied, $\mathbf{U}^{-1}$ *does not exist, and so* $\boldsymbol{A}$ *is not diagonable as* $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$*; but of course* $\boldsymbol{AU} = \boldsymbol{UD}$ *exists, with* $\boldsymbol{U}$ *and* $\boldsymbol{D}$ *of (6.22) wherein the singularity of* $\boldsymbol{U}$ *is clearly apparent.*

Since (6.24) is satisfied for every simple eigenvalue, the diagonability theorem is satisfied for any matrix having all n eigenvalues distinct. It is also satisfied for symmetric matrices, as discussed in Section 7.1. Further properties of the eigenvalues and vectors of symmetric matrices are given in Section 7.2.

6.4.2 Uses

Uses of the similar canonical form are many and varied. One of the most important applications is the ease provided by the diagonability of $\boldsymbol{A}$ for calculating powers of $\mathbf{A}$. This is because $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$ implies

$$\mathbf{A} = \mathbf{UDU}^{-1} \tag{6.27}$$

and hence $\mathbf{A}^2 = \mathbf{UDU}^{-1}\mathbf{UDU}^{-1} = \mathbf{UD}^2\mathbf{U}^{-1}$ and in general, for positive integers p,

$$\mathbf{A}^p = \mathbf{UD}^p\mathbf{U}^{-1}. \tag{6.28}$$

Similarly, for $\mathbf{A}$ nonsingular

$$\mathbf{A}^{-1} = \mathbf{UD}^{-1}\mathbf{U}^{-1}, \tag{6.29}$$

in which case (6.28) also holds for negative integers p. Because $\mathbf{D}$ is diagonal, (6.28) is easy to calculate: $\mathbf{D}^p$ is just $\mathbf{D}$ with its (diagonal) nonzero elements raised to the pth power.

Example 6.11 (continuation of Example 6.2)

From (6.25) and (6.26), equation (6.28) is

$$\mathbf{A}^p = \begin{bmatrix} 2 & 2 & 0 \\ 2 & 1 & 1 \\ -7 & 2 & -3 \end{bmatrix}^p \tag{6.30}$$

$$= \begin{bmatrix} -2 & -2 & 1 \\ 1 & -1 & -3 \\ 4 & 2 & 13 \end{bmatrix} \begin{bmatrix} 1^p & 0 & 0 \\ 0 & 3^p & 0 \\ 0 & 0 & (-4)^p \end{bmatrix} \frac{1}{70} \begin{bmatrix} -7 & 28 & 7 \\ -25 & -30 & -5 \\ 6 & -4 & 4 \end{bmatrix}.$$

The reader should verify this for $p = -1, 1,$ *and* 2.

Example 6.12 *Economists and others often have occasions to use systems of linear equations known as linear difference equations. An example of a system of first-order linear difference equations is*

$$\mathbf{x}_t = \mathbf{A}\mathbf{x}_{t-1} + \mathbf{d} \tag{6.31}$$

for $t = 1, 2, 3, \ldots,$ *and where* $\mathbf{A}$, $\mathbf{x}_0$, *and* $\mathbf{d}$ *are known. The problem is to calculate values of* $\mathbf{x}_t$. *From (6.31) we have*

$$\mathbf{x}_t = \mathbf{A}(\mathbf{A}\mathbf{x}_{t-2} + \mathbf{d}) + \mathbf{d} = \mathbf{A}^2\mathbf{x}_{t-2} + (\mathbf{A} + \mathbf{I})\,\mathbf{d} = \cdots$$
$$= \mathbf{A}^t\mathbf{x}_0 + (\mathbf{A}^{t-1} + \mathbf{A}^{t-2} + \cdots + \mathbf{A} + \mathbf{I})\mathbf{d}.$$

Then, providing that $\mathbf{A}^k \to \mathbf{0}$ *as* $k \to \infty$ *and that* $(\mathbf{I} - \mathbf{A})^{-1}$ *exists, use can be made of (4.27) to yield*

$$\mathbf{x}_t = \mathbf{A}^t\mathbf{x}_0 + (\mathbf{I} - \mathbf{A}^t)(\mathbf{I} - \mathbf{A})^{-1}\,\mathbf{d}. \tag{6.32}$$

The similar canonical form provides a method of calculating $\mathbf{A}^t$; *and if, as has already been assumed,* $\mathbf{A}^t \to \mathbf{0}$ as $t \to \infty$, *then we see from (6.32) that the corresponding limit for* $\mathbf{x}_t$ *is* $(\mathbf{I} - \mathbf{A})^{-1}\mathbf{d}$.

An extension of (6.32) as the solution to (6.31) is to adapt linear difference equations of higher than the first order to being first-order equations. For example, suppose we wish to solve the third-order difference equation

$$y_t = \alpha y_{t-1} + \beta y_{t-2} + \gamma y_{t-3} + \delta, \tag{6.33}$$

for $t = 3, 4, \ldots,$ *knowing* α, β, γ, *and* δ *and the initial values* y_0, y_1, *and* y_2. *Define* $z_t = y_{t-1}$ for $t = 1, 2, \ldots$ *and* $w_t = y_{t-2} = z_{t-1}$ for $t = 2, 3, \ldots$, *and observe that using these and (6.33)*

$$\begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix} = \begin{bmatrix} \alpha & \beta & \gamma \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ z_{t-1} \\ w_{t-1} \end{bmatrix} + \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix}. \tag{6.34}$$

Hence on defining

$$\mathbf{x}_t = \begin{bmatrix} y_t \\ z_t \\ w_t \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} \alpha & \beta & \gamma \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{d} = \begin{bmatrix} \delta \\ 0 \\ 0 \end{bmatrix},$$

(6.34) can be expressed in the form of (6.31) and so (6.32) yields its solution, which is the solution of (6.33).

6.5 Symmetric Matrices

Symmetric matrices have sufficient notable properties in regard to eigenvalues and eigenvectors as to warrant special attention. Furthermore, the widespread use of symmetric matrices in statistics, through their involvement in quadratic forms (see Section 5.6), make it worthwhile to discuss these properties in detail. Only real, symmetric matrices are considered.

6.5.1 Eigenvalues All Real

The eigenvalues of a matrix of order $n \times n$ are roots of a polynomial equation of degree n and so are not necessarily real numbers. Some may be complex numbers, occurring in pairs as a + ib and its complex conjugate $a - ib$, where $i = \sqrt{-1}$ and a and b are real. However, when **A** is symmetric (and has elements that are real numbers) then all its eigenvalues are real. We state this formally as a lemma.

Lemma 6.1 *The eigenvalues of every real symmetric matrix are real.*

Proof.[1] Suppose λ is a complex eigenvalue of the symmetric matrix $\mathbf{M}_{n \times n}$ with $\mathbf{u}$ being a corresponding eigenvector. If $\lambda = \alpha + i\beta$, define $\bar{\lambda} = \alpha - i\beta$, and for $\mathbf{u} = \{u_k\} = \mathbf{a} + i\mathbf{b}$ define $\bar{\mathbf{u}} = \{\bar{u}_k\} = \mathbf{a} - i\mathbf{b}$. Then by definition, $\mathbf{Mu} = \lambda\mathbf{u}$ so that $\bar{\mathbf{u}}'\mathbf{Mu} = \bar{\mathbf{u}}'\lambda\mathbf{u} = \lambda\bar{\mathbf{u}}'\mathbf{u}$. But $\mathbf{M}\bar{\mathbf{u}} = \bar{\lambda}\bar{\mathbf{u}}$ so that we also have $\bar{\mathbf{u}}'\mathbf{Mu} = (\mathbf{M}\bar{\mathbf{u}})'\,\mathbf{u} = (\bar{\lambda}\bar{\mathbf{u}})'\mathbf{u} = \bar{\lambda}\bar{\mathbf{u}}'\mathbf{u}$. Equating these two expressions for $\bar{\mathbf{u}}'\mathbf{Mu}$ gives $\lambda\bar{\mathbf{u}}'\mathbf{u} = \bar{\lambda}\bar{\mathbf{u}}'\mathbf{u}$ and since $\bar{\mathbf{u}}'\mathbf{u}$ is a sum of squares of real numbers, it is nonzero and so $\lambda = \bar{\lambda}$, that is, $a + ib = a - ib$ and hence $b = 0$. Thus every λ is real.

This lemma means that when dealing with symmetric matrices, all eigenvalues are real. Corresponding to each eigenvalue is an eigenvector **u**, say, that is real.

6.5.2 Symmetric Matrices Are Diagonable

For every eigenvalue of a symmetric matrix **A**, condition (6.24) is satisfied: $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ for every λ_k (Proof is given in Section 7.1). Therefore, $\mathbf{A} = \mathbf{A}'$ is diagonable. Hence, for any symmetric matrix **A**, we have $\mathbf{U}^{-1}$ of $\mathbf{AU} = \mathbf{UD}$ existing, and $\mathbf{A} = \mathbf{UDU}^{-1}$ and $\mathbf{A}^p = \mathbf{UD}^p\mathbf{U}^{-1}$. Moreover, as is now developed, **U** is orthogonal.

6.5.3 Eigenvectors Are Orthogonal

Symmetric matrices have eigenvectors that are orthogonal to one another. We establish this in two cases: (1) eigenvectors corresponding to different eigenvalues, and (2) eigenvectors corresponding to a multiple eigenvalue.

1. Different Eigenvalues

Consider two different eigenvalues $\lambda_1 \neq \lambda_2$ with $\mathbf{u}_1$, $\mathbf{u}_2$ being corresponding eigenvectors. Then with $\mathbf{A} = \mathbf{A}'$ and $\mathbf{Au}_k = \lambda_k\mathbf{u}_k$ we have

$$\lambda_1\mathbf{u}_2'\mathbf{u}_1 = \mathbf{u}_2'\lambda_1\mathbf{u}_1 = \mathbf{u}_2'\mathbf{A}\mathbf{u}_1 = \mathbf{u}_1'\mathbf{A}'\mathbf{u}_2 = \mathbf{u}_1'\mathbf{A}\mathbf{u}_2 = \mathbf{u}_1'\lambda_2\mathbf{u}_2 = \lambda_2\mathbf{u}_1'\mathbf{u}_2 = \lambda_2\mathbf{u}_2'\mathbf{u}_1;$$

that is, $\lambda_1\mathbf{u}'_2\mathbf{u}_1 = \lambda_2\mathbf{u}'_2\mathbf{u}_1$. But $\lambda_1 \neq \lambda_2$. Therefore $\mathbf{u}'_2\mathbf{u}_1 = 0$; that is, $\mathbf{u}_1$ and $\mathbf{u}_2$ are orthogonal. Hence, eigenvectors corresponding to different eigenvalues are orthogonal.

Before discussing case (2), the following lemma is needed:

Lemma 6.2 *Let* $\boldsymbol{B}$ *be an* $n \times n$ *matrix. Solutions to* $\mathbf{Bx} = \mathbf{0}$ *can always be found that are orthogonal to one another.*

Proof. Suppose that $\boldsymbol{B}$ has rank r. Then, $\mathbf{Bx} = \mathbf{0}$ has $n - r$ linearly independent solutions. Suppose $\mathbf{x}_1$ is a solution. Consider the equations in $\mathbf{x}$:

$$\begin{bmatrix} \mathbf{B} \\ \mathbf{x}_1' \end{bmatrix} \mathbf{x} = \mathbf{0}. \tag{6.35}$$

Because $\mathbf{x}_1$ is a solution to $\mathbf{Bx} = \mathbf{0}$, $\boldsymbol{x}'_1$ is orthogonal to the rows of $\boldsymbol{B}$ and hence linearly independent of these rows. Therefore, (6.35) has $n - r - 1$ linearly independent solutions. Suppose $\mathbf{x}_2$ is such a solution. Then clearly $\mathbf{x}_2$ satisfies $\mathbf{Bx} = \mathbf{0}$ and is also orthogonal to $\mathbf{x}_1$. Similarly, if $n - r > 2$ a third solution can be obtained by solving

$$\begin{bmatrix} \mathbf{B} \\ \mathbf{x}_1' \\ \mathbf{x}_2' \end{bmatrix} \mathbf{x} = \mathbf{0}$$

which has $n - r - 2$ linearly independent solutions for $\mathbf{x}$. This process can be continued until $n - r$ solutions have been obtained, all orthogonal to one another.

2. Multiple Eigenvalues

If $\mathbf{A} = \mathbf{A}'$ has λ_k as an eigenvalue with multiplicity m_k, then $\mathbf{A} - \lambda_k\mathbf{I}$ has rank $n - m_k$ and is singular; and there are m_k eigenvectors corresponding to λ_k which are linearly independent solutions to $(\mathbf{A} - \lambda_k\mathbf{I})\mathbf{u} = \mathbf{0}$ that are also orthogonal to one another by Lemma 6.2. Hence, if $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$, there exist m_k linearly independent orthogonal eigenvectors of $\mathbf{A}$ corresponding to λ_k.

Theorem 6.2 (The Spectral Decomposition Theorem)

Let $\boldsymbol{A}$ *be a symmetric matrix of order* $n \times n$*. There exists an orthogonal matrix* $\boldsymbol{P}$ *such that*

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}', \tag{6.36}$$

where $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_n)$ *is a diagonal matrix whose diagonal elements are the eigenvalues of* $\boldsymbol{A}$*, and the columns of* $\boldsymbol{P}$ *are eigenvectors of* $\boldsymbol{A}$ *that correspond to* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *in the following manner: If* $\boldsymbol{P}$ *is partitioned as*

$$\boldsymbol{P} = [\boldsymbol{p}_1\ \boldsymbol{p}_2\ \ldots\ \boldsymbol{p}_n],$$

*where $\mathbf{p}_i$ is the ith column of **P**, then $\mathbf{p}_i$ is an eigenvector of **A** corresponding to λ_i* (i = 1, 2, …, n). *The matrix **A** can then be expressed as*

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{p}_i \boldsymbol{p}_i'. \tag{6.37}$$

Proof. We have established that **A** = **A**′ is diagonable, that eigenvectors corresponding to different eigenvalues are orthogonal, and that m_k linearly independent eigenvectors corresponding to any eigenvalue λ_k of multiplicity m_k can be obtained such that they are orthogonal. Thus the m_k eigenvectors corresponding to λ_k are orthogonal, not only to one another, but also to the m_t eigenvectors corresponding to each other eigenvalue λ_t; and this is true for every eigenvalue λ_k. Hence eigenvectors for a symmetric matrix can always be found such that they are *all* orthogonal to one another. On normalizing each vector (see Section 5.4) by changing ***u*** to $(1/\sqrt{\mathbf{u}'\mathbf{u}})\mathbf{u}$ and arraying the normalized vectors in a matrix ***P***, we then have ***P*** as an orthogonal matrix and hence

$$\mathbf{P}'\mathbf{A}\mathbf{P} = \boldsymbol{\Lambda} \quad \text{with} \quad \mathbf{P}\mathbf{P}' = \mathbf{I}.$$

It follows that

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'. \tag{6.38}$$

The proof of the next theorem can be found in Harville (1997, Theorem 14.5.16, p. 231).

Theorem 6.3 (The Cholesky Decomposition)

*Let **A** be a symmetric matrix of order n* × *n.*

a. *If **A** is positive definite, then there exists a unique upper triangular matrix **T** with positive diagonal elements such that*

$$\boldsymbol{A} = \boldsymbol{T}'\boldsymbol{T}.$$

b. *If **A** is non-negative definite with rank equal to r, then there exists a unique upper triangular matrix **U*** with *r positive diagonal elements and with n* − *r zero rows such that*

$$\boldsymbol{A} = \boldsymbol{U}'\boldsymbol{U}.$$

Example 6.13

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

has a characteristic equation which reduces to $(\lambda + 1)^2(\lambda - 5) = 0$. *Hence,* $\lambda_1 = 5$ *with* $m_1 = 1$, *and* $\lambda_2 = -1$ *with* $m_2 = 2$. *For* $\lambda_1 = 5$, *we get* $\boldsymbol{u}_1 = (1, 1, 1)'$ *as an eigenvector. For*

$\lambda_2 = -1$ *we find* $\boldsymbol{u}_2 = (-2, 1, 1)'$ *as an eigenvector that is orthogonal to* $\boldsymbol{u}_1$. *Since* $m_2 = 2$ *there should be another eigenvector that is orthogonal to* $\boldsymbol{u}_2$ *by Case 2 in Section 6.5.3. Such an eigenvector can be obtained by finding a vector* $\boldsymbol{v}_2$ *such that* $\boldsymbol{v}_2'\boldsymbol{u}_2 = 0$, *that is,* $v_{21}(-2) + v_{22}(1) + v_{23}(1) = 0$, where the v_{2i}'s *are the elements of* $\boldsymbol{v}_2$. *We find that* $\boldsymbol{u}_3 = (0, -1, 1)'$ *satisfies this condition and is orthogonal to* $\boldsymbol{u}_1$. *Arraying the normalized forms of these vectors as a matrix gives*

$$\boldsymbol{P} = \frac{1}{\sqrt{6}}\begin{bmatrix} \sqrt{2} & -2 & 0 \\ \sqrt{2} & 1 & -\sqrt{3} \\ \sqrt{2} & 1 & \sqrt{3} \end{bmatrix}.$$

The reader should verify that $\boldsymbol{P}'\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda} = \text{diag}\,(5, -1, -1)$ *and* $\mathbf{P}\mathbf{P}' = \mathbf{I}$.

6.5.4 Rank Equals Number of Nonzero Eigenvalues for a Symmetric Matrix

Define $z_{\mathbf{A}}$ as the number of zero eigenvalues of a symmetric matrix $\boldsymbol{A}$ of order $n \times n$; then $n - z_{\mathbf{A}}$ is the number of nonzero eigenvalues. Since, for $\boldsymbol{A}$ being symmetric, $\mathbf{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$ for nonsingular (orthogonal) $\boldsymbol{P}$, the ranks of $\boldsymbol{A}$ and $\boldsymbol{\Lambda}$ are equal, $r_{\mathbf{A}} = r_{\boldsymbol{\Lambda}}$. But, the only nonzero elements in the diagonal matrix $\boldsymbol{\Lambda}$ are the nonzero eigenvalues, and so its rank is the number of such eigenvalues, $n - z_{\mathbf{A}}$. Hence

$$r_{\mathbf{A}} = n - z_{\mathbf{A}}, \quad \text{for} \quad \mathbf{A} = \mathbf{A}'; \tag{6.39}$$

that is, for symmetric matrices rank equals the number of nonzero eigenvalues.

This result is true not only for all symmetric matrices (because they are diagonable), but also for all diagonable matrices. Other than diagonability there is nothing inherent in the development of (6.39) that uses the symmetry of $\boldsymbol{A}$. Nevertheless, (6.39) is of importance because it applies to all symmetric matrices.

6.6 Eigenvalues of orthogonal and Idempotent Matrices

Theorem 6.4 *If $\boldsymbol{A}$* is an *$n \times n$ orthogonal matrix and λ is an eigenvalue of $\boldsymbol{A}$, then* λ = ±1.

Proof. We have that $\boldsymbol{Ax} = \lambda\boldsymbol{x}$, where $\boldsymbol{x}$ is an eigenvector corresponding to λ. Since $\boldsymbol{A}'\boldsymbol{A} = \boldsymbol{I}$, multiplying both sides on the left by $\boldsymbol{x}'$ and on the right by $\boldsymbol{x}$ we get $\boldsymbol{x}'\boldsymbol{A}'\boldsymbol{Ax} = \boldsymbol{x}'\boldsymbol{x}$. This implies that $\lambda^2\boldsymbol{x}'\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{x}$. We conclude that $\lambda^2 = 1$, that is, λ = ±1 since $\boldsymbol{x} \neq \boldsymbol{0}$.

Theorem 6.5 *If $\boldsymbol{B}$ is an $n \times n$ idempotent matrix, then its eigenvalues are equal to 0 or 1. Furthermore, if the rank of $\boldsymbol{B}$ is r, then $\boldsymbol{B}$* has *r eigenvalues all equal to 1.*

Proof. Let λ be an eigenvalue of $\boldsymbol{B}$ with an eigenvector $\boldsymbol{x}$. Then, $\boldsymbol{B}^2\boldsymbol{x} = \boldsymbol{Bx} = \lambda\boldsymbol{x}$. But, $\boldsymbol{B}^2\boldsymbol{x} = \boldsymbol{BBx} = \boldsymbol{B}\lambda\boldsymbol{x} = \lambda\boldsymbol{Bx} = \lambda^2\boldsymbol{x}$. We conclude that $\lambda\boldsymbol{x} = \lambda^2\boldsymbol{x}$. Hence, $\lambda = \lambda^2$, that is, $\lambda = 0$or 1 since $\boldsymbol{x} \neq \boldsymbol{0}$. Furthermore, since $\boldsymbol{B}$ is idempotent, then by Theorem 5.1, $r(\boldsymbol{B}) = tr(\boldsymbol{B})$. But, $r = r(\boldsymbol{B})$ and $tr(\boldsymbol{B})$ is the sum of the nonzero eigenvalues, which are equal to 1, as the remaining ones are equal to zero, we conclude that $\boldsymbol{B}$ has *r* nonzero eigenvalues that are equal to 1.

6.6.1 Eigenvalues of Symmetric Positive Definite and Positive Semidefinite Matrices

Theorem 6.6 *The symmetric matrix $\boldsymbol{A}$ is positive definite if and only if its eigenvalues are positive.*

Proof. Let λ and $\boldsymbol{x}$ be an eigenvalue and corresponding eigenvector of $\boldsymbol{A}$. If $\boldsymbol{A}$ is positive definite, then $\boldsymbol{x}'\boldsymbol{Ax} = \lambda\boldsymbol{x}'\boldsymbol{x}$. Since $\boldsymbol{x}'\boldsymbol{Ax} > 0$, then λ > 0. Vice versa, if the eigenvalues of $\boldsymbol{A}$ are positive, then by the spectral decomposition theorem (Theorem 6.2), $\boldsymbol{x}'\boldsymbol{Ax} = \boldsymbol{x}'\boldsymbol{P\Lambda P}'\boldsymbol{x}$, where $\boldsymbol{P}$ is an orthogonal matrix and $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\boldsymbol{A}$. Since the diagonal elements of $\boldsymbol{\Lambda}$ are all positive, then $\boldsymbol{x}'\boldsymbol{Ax} > 0$ for all nonzero $\boldsymbol{x}$.

The proof of the following theorem is left as an exercise.

Theorem 6.7 *Let $\boldsymbol{A}$ be a symmetric matrix. Then, $\boldsymbol{A}$ is positive semidefinite if and only if its eigenvalues are non-negative with at least one equal to zero.*

The proof of the next theorem can be found in Banerjee and Roy (2014, Theorem 13.18).

Theorem 6.8 *Let $\boldsymbol{A}$ be a symmetric matrix. Then, $\boldsymbol{A}$ is non-negative definite if and only if all its principal minors (see Section 3.6) are non-negative.*

Example 6.14 *The characteristic equation for the positive semidefinite matrix of Section 5.7*

$$\mathbf{A}=\begin{bmatrix} 37 & -2 & -24 \\ -2 & 13 & -3 \\ -24 & -3 & 17 \end{bmatrix} \quad \text{is} \quad \lambda(\lambda-14)(\lambda-53)=0, \tag{6.40}$$

so that the eigenvalues are 0, 14, and 53.

A further feature of non-negative definite matrices (see Section 5.7) is that each is the square of some other matrix of real elements. For if $\mathbf{A}$ *is non-negative definite we have, by the Spectral Decomposition Theorem,* $\mathbf{A}=\mathbf{PDP}'$ *where*

$$\mathbf{D}=\operatorname{diag}\{\lambda_i\}=\mathbf{D}^{\frac{1}{2}}\mathbf{D}^{\frac{1}{2}} \quad \text{for} \quad \mathbf{D}^{\frac{1}{2}}=\operatorname{diag}\{\sqrt{\lambda_i}\}$$

with $\sqrt{\lambda_i}$ *being the positive square root. Hence,*

$$\mathbf{A}=\mathbf{PDP}'=\mathbf{PD}^{\frac{1}{2}}\mathbf{P}'\mathbf{PD}^{\frac{1}{2}}\mathbf{P}'=\mathbf{H}^2 \quad \text{for} \quad \mathbf{H}=\mathbf{PD}^{\frac{1}{2}}\mathbf{P}'. \tag{6.41}$$

For positive definite matrices no λ_i *is zero and so* $\mathbf{D},\mathbf{D}^{\frac{1}{2}}$*, and* $\mathbf{H}$ *are all of full rank giving* $\mathbf{A}^{-1}=(\mathbf{H}^{-1})^2$.

Also, if $\mathbf{P}'\mathbf{AP}=\mathbf{D}=\begin{bmatrix} \Delta^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$ *where* $\Delta=\operatorname{diag}\{\sqrt{\lambda_i}\}$ *for just the* $r_{\mathbf{A}}$ *nonzero eigenvalues, then with*

$$\mathbf{S}=\begin{bmatrix} \Delta^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n-r_A} \end{bmatrix} \text{ and } \mathbf{T}=\mathbf{PS}, \text{ we have } \mathbf{T}'\mathbf{AT}=\begin{bmatrix} \mathbf{I}_{r_A} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{6.42}$$

where $\mathbf{T}$ is nonsingular.

Example 6.15 (continued from Example 6.14)

Eigenvectors of $\mathbf{A}$ *in (6.40) corresponding to the eigenvalues 14, 53, and 0 are*

$$\begin{bmatrix} 2 \\ -13 \\ 3 \end{bmatrix}, \quad \begin{bmatrix} -3 \\ 0 \\ 2 \end{bmatrix}, \quad \text{and} \quad \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \tag{6.43}$$

respectively. Hence $\mathbf{H}$ *of (6.41) is*

$$\mathbf{H}=\begin{bmatrix} 2 & -3 & 2 \\ -13 & 0 & 1 \\ 3 & 2 & 3 \end{bmatrix}\begin{bmatrix} 1/\sqrt{182} & 0 & 0 \\ 0 & 1/\sqrt{13} & 0 \\ 0 & 0 & 1/\sqrt{14} \end{bmatrix}$$

$$\times\begin{bmatrix} \sqrt{14} & 0 & 0 \\ 0 & \sqrt{53} & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} 1/\sqrt{182} & 0 & 0 \\ 0 & 1/\sqrt{13} & 0 \\ 0 & 0 & 1/\sqrt{14} \end{bmatrix}\begin{bmatrix} 2 & -13 & 3 \\ -3 & 0 & 2 \\ 2 & 1 & 3 \end{bmatrix},$$

with the second and fourth matrices in this product representing the normalization of the eigenvectors in (6.43). After simplification, $\mathbf{H}$ *can be written as*

$$\mathbf{H} = \frac{1}{13\sqrt{14}} \begin{bmatrix} 4+9a & -26 & 6-6a \\ -26 & 169 & -39 \\ 6-6a & -39 & 9+4a \end{bmatrix} \quad \text{with} \quad a = \sqrt{14(53)}.$$

The reader can verify that $\mathbf{H}^2 = \mathbf{A}$*, as in (6.41). With*

$$\mathbf{S} = \begin{bmatrix} 1/\sqrt{14} & 0 & 0 \\ 0 & 1/\sqrt{53} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and}$$

$$\mathbf{T} = \mathbf{PS} = \begin{bmatrix} 2 & -3 & 2 \\ -13 & 0 & 1 \\ 3 & 2 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{14(182)} & 0 & 0 \\ 0 & 1/\sqrt{53(13)} & 0 \\ 0 & 0 & \sqrt{1/14} \end{bmatrix},$$

(6.42) can also be verified.

Theorem 6.9 *Let* $\boldsymbol{A}$ be an $n \times n$ *symmetric matrix of rank* r*. Then* $\boldsymbol{A}$ *can be written as*

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{L}', \tag{6.44}$$

where $\boldsymbol{L}$ *is* $n \times r$ *of rank* r*, that is, of* $\boldsymbol{L}$ *is of full column rank.*

Proof. By the spectral decomposition theorem (Theorem 6.2), $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}'$, where $\boldsymbol{\Lambda}$ is a diagonal matrix of eigenvalues of $\boldsymbol{A}$ and $\boldsymbol{P}$ is an orthogonal matrix of corresponding eigenvectors of $\boldsymbol{A}$. The matrix $\boldsymbol{\Lambda}$ can be written as

$$\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{D}_r, \mathbf{0}), \tag{6.45}$$

where $\boldsymbol{D}_r$ is a diagonal matrix of order $r \times r$ whose diagonal elements are the nonzero eigenvalues of $\boldsymbol{A}$, and the first r columns of $\boldsymbol{P}$ are the corresponding eigenvectors. we then have

$$\boldsymbol{A} = \boldsymbol{P}\begin{bmatrix} \boldsymbol{D}_r^{1/2} \\ \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{D}_r^{1/2} & \mathbf{0}' \end{bmatrix} \boldsymbol{P}' = \boldsymbol{L}\boldsymbol{L}', \tag{6.46}$$

where

$$\boldsymbol{L} = \boldsymbol{P}\begin{bmatrix} \boldsymbol{D}_r^{1/2} \\ \mathbf{0} \end{bmatrix}.$$

Note that $\boldsymbol{L}$ is of order $n \times r$ and rank r. Its elements are not necessarily real since some of the the diagonal elements of $\boldsymbol{D}_r$ may be negative.

Corollary 6.1 *If* $\boldsymbol{A}$ *is a symmetric positive semidefinite matrix of order* $n \times n$ *and rank* r*, then it can be written as* $\boldsymbol{A} = \boldsymbol{K}\boldsymbol{K}'$ *where* $\boldsymbol{K}$ *is a real matrix of order* $n \times r$ *and rank* r*.*

Proof. This result follows from Theorem 6.7 since $\boldsymbol{A}$, being positive semidefinite, its nonzero eigenvalues are positive. Hence, the diagonal elements of the diagonal matrix $\boldsymbol{D}_r$ are all positive which makes the matrix $\boldsymbol{L}$ a real matrix. In this case, $\boldsymbol{K} = \boldsymbol{L}$.

Corollary 6.2 *If $\boldsymbol{A}$ is a symmetric positive definite matrix of order $n \times n$, then $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{M}'$, where $\boldsymbol{M}$ is a nonsingular matrix.*

Proof. This follows directly from Theorem 6.6 since the diagonal elements of the diagonal matrix $\boldsymbol{\Lambda}$ are positive. Hence, $\boldsymbol{A}$ can be written as $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{M}'$, where $\boldsymbol{M} = \boldsymbol{P}\boldsymbol{\Lambda}^{1/2}$, which is a nonsingular matrix.

6.7 Eigenvalues of Direct Products and Direct Sums of Matrices

Theorem 6.10 *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be matrices of orders* $m \times m$ *and* $n \times n$*, respectively. Let* λ *and* $\boldsymbol{x}$ *be an eigenvalue and a corresponding eigenvector of* $\boldsymbol{A}$*. Likewise, let* μ *and* $\boldsymbol{y}$ *denote an eigenvalue and a corresponding eigenvector of* $\boldsymbol{B}$*. Then,* $\lambda\mu$ *is an eigenvalue of* $\boldsymbol{A} \otimes \boldsymbol{B}$ *with a corresponding eigenvector* $\boldsymbol{x} \otimes \boldsymbol{y}$.

Proof. We have that $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ and $\boldsymbol{B}\boldsymbol{y} = \mu\boldsymbol{y}$. Hence,

$$(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{x} \otimes \boldsymbol{y}) = (\boldsymbol{A}\boldsymbol{x}) \otimes (\boldsymbol{B}\boldsymbol{y}) = (\lambda\boldsymbol{x}) \otimes (\mu\boldsymbol{y}) = \lambda\mu(\boldsymbol{x} \otimes \boldsymbol{y}).$$

This shows that $\lambda\mu$ is an eigenvalue of $\boldsymbol{A} \otimes \boldsymbol{B}$ with an eigenvector $\boldsymbol{x} \otimes \boldsymbol{y}$.

Based on this theorem it can be seen that if $\lambda_1, \lambda_2, \ldots, \lambda_m$ are eigenvalues of $\boldsymbol{A}$, and $\mu_1, \mu_2, \ldots, \mu_n$ are eigenvalues of $\boldsymbol{B}$, then $\lambda_i\mu_j$ ($i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$) are eigenvalues of $\boldsymbol{A} \otimes \boldsymbol{B}$.

Corollary 6.3 The rank of $\boldsymbol{A} \otimes \boldsymbol{B}$ is the product of the ranks of $\boldsymbol{A}$ and $\boldsymbol{B}$.

Proof. From Section 4.15 (item xii), we can write $r(\boldsymbol{A} \otimes \boldsymbol{B}) = r\{(\boldsymbol{A} \otimes \boldsymbol{B})'(\boldsymbol{A} \otimes \boldsymbol{B})\} = r\{(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})\}$. Clearly, $(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})$ is symmetric. Thus, by Section 6.5.4, its rank is equal to the number of nonzero eigenvalues of $(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})$, which are real scalars. If τ_i is the ith eigenvalue of $\boldsymbol{A}'\boldsymbol{A}$ and κ_j is the jth eigenvalue of $\boldsymbol{B}'\boldsymbol{B}$, then by Theorem 6.7 $\tau_i\kappa_j$ is an eigenvalue of $(\boldsymbol{A}'\boldsymbol{A}) \otimes (\boldsymbol{B}'\boldsymbol{B})$, which is nonzero if and only if both τ_i and κ_j are nonzero. It follows that the number of nonzero eigenvalues of $(\boldsymbol{A}'\boldsymbol{A}) \times (\boldsymbol{B}'\boldsymbol{B})$ is the product of the nonzero eigenvalues of $\boldsymbol{A}'\boldsymbol{A}$, which is equal to the rank of $\boldsymbol{A}'\boldsymbol{A}$, or equivalently, the rank of $\boldsymbol{A}$ by Section 4.15 (item xii), and the number of nonzero eigenvalues of $\boldsymbol{B}'\boldsymbol{B}$, which is the rank of $\boldsymbol{B}'\boldsymbol{B}$, or the rank of $\boldsymbol{B}$. We then conclude that

$$r(\boldsymbol{A} \otimes \boldsymbol{B}) = r(\boldsymbol{A})r(\boldsymbol{B}).$$

Corollary 6.3 confirms the result given in Section 4.15 (item xv).

Theorem 6.11 *Let* $\boldsymbol{A}$ *and* $\boldsymbol{B}$ *be matrices of orders* $m \times m$ *and* $n \times n$*, respectively. Let* λ *be an eigenvalue of* $\boldsymbol{A}$ *with a corresponding eigenvector* $\boldsymbol{x}$*, and let* μ *be an eigenvalues of* $\boldsymbol{B}$ *with a corresponding eigenvector* $\boldsymbol{y}$*. Then,* $\lambda \oplus \mu$ *is an eigenvalue of* $\boldsymbol{A} \oplus \boldsymbol{B}$ with a corresponding eigenvector $\boldsymbol{x} \oplus \boldsymbol{y}$.

Proof.

$$\begin{aligned}(\boldsymbol{A} \oplus \boldsymbol{B})(\boldsymbol{x} \oplus \boldsymbol{y}) &= (\boldsymbol{A}\boldsymbol{x}) \oplus (\boldsymbol{B}\boldsymbol{y}) \\ &= (\lambda\boldsymbol{x}) \oplus (\mu\boldsymbol{y}) \\ &= (\boldsymbol{x} \oplus \boldsymbol{y})(\lambda \oplus \mu).\end{aligned}$$

We conclude that $\lambda \oplus \mu$ is eigenvalue of $\boldsymbol{A} \oplus \boldsymbol{B}$ with a corresponding eigenvector $\boldsymbol{x} \oplus \boldsymbol{y}$. Thus, if $\lambda_1, \lambda_2, \ldots, \lambda_m$ are eigenvectors of $\boldsymbol{A}$, and $\mu_1, \mu_2, \ldots, \mu_n$ are eigenvectors of $\boldsymbol{B}$, then $\lambda_1, \lambda_2, \ldots, \lambda_m; \mu_1, \mu_2, \ldots, \mu_n$ are eigenvalues of $\boldsymbol{A} \oplus \boldsymbol{B}$.

Based on this theorem it is easy to see that $tr(\boldsymbol{A} \oplus \boldsymbol{B}) = tr(\boldsymbol{A}) + tr(\boldsymbol{B})$.

6.8 Nonzero Eigenvalues of *AB* and *BA*

The following theorem relates the nonzero eigenvalues of $\boldsymbol{AB}$ to those of $\boldsymbol{BA}$. It has useful applications in statistics as will be seen later:

Theorem 6.12 Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of orders $m \times n$ and $n \times m$, respectively ($n \geq m$). Let $\lambda \neq 0$. Then,

a. $\left|\boldsymbol{BA} - \lambda \boldsymbol{I}_n\right| = (-\lambda)^{n-m} \left|\boldsymbol{AB} - \lambda \boldsymbol{I}_m\right|$.
b. *The nonzero eigenvalues of* $\boldsymbol{BA}$ *are the same as those of* $\boldsymbol{AB}$. *Furthermore, the multiplicity of a nonzero eigenvalue of* $\boldsymbol{BA}$ *is the same as when regarded as an eigenvalue of* $\boldsymbol{AB}$.
c. *If* $m = n$, *then all the eigenvalues of* $\boldsymbol{BA}$ (*not just the nonzero ones*) *are the same as those of* $\boldsymbol{AB}$.

Proof.

a. Consider the determinant of the matrix $\boldsymbol{L}$ partitioned as

$$|\boldsymbol{L}| = \begin{vmatrix} \boldsymbol{I}_m & (1/\lambda)\boldsymbol{A} \\ \boldsymbol{B} & \boldsymbol{I}_n \end{vmatrix}, \quad \lambda \neq 0. \tag{6.47}$$

By applying property (g) in Section 4.12, the determinant of $\boldsymbol{L}$ can be written as

$$|\boldsymbol{L}| = \left|\boldsymbol{I}_n - \boldsymbol{B}(1/\lambda)\boldsymbol{A}\right|, \text{ since here } \boldsymbol{A}_{11} = \boldsymbol{I}_m, \quad \lambda \neq 0,$$
$$|\boldsymbol{L}| = \left|\boldsymbol{I}_m - (1/\lambda)\boldsymbol{AB}\right|, \text{ since here } \boldsymbol{A}_{22} = \boldsymbol{I}_n, \quad \lambda \neq 0.$$

We conclude that

$$\left|\boldsymbol{I}_n - \boldsymbol{B}(1/\lambda)\boldsymbol{A}\right| = \left|\boldsymbol{I}_m - (1/\lambda)\boldsymbol{AB}\right|, \quad \lambda \neq 0.$$

Hence,

$$\lambda^{-n} \left|\lambda \boldsymbol{I}_n - \boldsymbol{BA}\right| = \lambda^{-m} \left|\lambda \boldsymbol{I}_m - \boldsymbol{AB}\right|, \quad \lambda \neq 0.$$

This can be written as

$$\left|\boldsymbol{BA} - \lambda \boldsymbol{I}_n\right| = (-\lambda)^{n-m} \left|\boldsymbol{AB} - \lambda \boldsymbol{I}_m\right|, \quad \lambda \neq 0. \tag{6.48}$$

b. It can be seen from (6.48) that the nonzero eigenvalues of $\boldsymbol{BA}$ are the same as those of $\boldsymbol{AB}$. Furthermore, each nonzero eigenvalue of $\boldsymbol{BA}$ has the same multiplicity as when regarded as a nonzero eigenvalue of $\boldsymbol{AB}$.
c. This is obvious from part (a).

Exercises

1. Find the eigenvalues and eigenvectors of the following symmetric matrices. In each case combine the eigenvectors into an orthogonal matrix $\boldsymbol{P}$ and verify that $\mathbf{P'AP} = \mathbf{\Lambda}$ where $\mathbf{\Lambda}$ is the diagonal matrix of the eigenvalues.

$$\mathbf{B} = \begin{bmatrix} 4 & -2 & 0 \\ -2 & 3 & -2 \\ 0 & -2 & 2 \end{bmatrix} \mathbf{C} = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & -4 \\ 1 & -4 & 1 \end{bmatrix}$$

$$\mathbf{E} = \begin{bmatrix} -1 & -2 & 1 \\ -2 & 2 & -2 \\ 1 & -2 & -1 \end{bmatrix} \mathbf{F} = \begin{bmatrix} -9 & 2 & 6 \\ 2 & -9 & 6 \\ 6 & 6 & 7 \end{bmatrix}$$

2. Show that if $\mathbf{B} = \mathbf{A}^2 + \mathbf{A}$, and if λ is an eigenvalue of $\mathbf{A}$ then $\lambda^2 + \lambda$ is an eigenvalue of $\mathbf{B}$.

3. Show that the eigenvectors of

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \quad \text{are of the form} \quad \begin{bmatrix} -b \\ a - \lambda_1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} -b \\ a - \lambda_2 \end{bmatrix}$$

   where λ_1 and λ_2 are the eigenvalues. Verify that $2a = \lambda_1 + \lambda_2$ when $a = d$.

4. Factorize the characteristic equation of

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

   using the eigenvalue $\lambda = a - b$.

5.
   a. Find the eigenvalues of $\mathbf{I} - \boldsymbol{xx'}$ for $\boldsymbol{x}$ being a vector of order $n \times 1$.

   b. Find orthogonal eigenvectors of $\mathbf{I} - \boldsymbol{xx'}$.

   c. Let $\mathbf{\Lambda}$ be the diagonal matrix of eigenvalues of $\boldsymbol{I} - \boldsymbol{xx'}$ and let $\boldsymbol{P}$ be the corresponding orthogonal matrix of eigenvectors. Find these matrices and show that $\boldsymbol{I} - \boldsymbol{xx'} = \boldsymbol{P\Lambda P'}$.

6. Using the characteristic equation, show for λ being an eigenvalue of $\boldsymbol{A}$ that $1/(1 + \lambda)$ is an eigenvalue of $(\mathbf{I} + \boldsymbol{A})^{-1}$.

7. When eigenvalues of $\mathbf{A}$ are positive, prove that those of $\mathbf{A} + \mathbf{A}^{-1}$ are equal or greater than 2.

8. For $\mathbf{A}$ being symmetric of order $n \times n$, with eigenvalues λ_i, prove that

$$\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^2 = \sum_{k=1}^{n} \lambda_k^2.$$

9. Show that if λ is an eigenvalue of an orthogonal matrix, then so is 1/λ.
10. Prove that the eigenvalues of $\boldsymbol{BA}$ are the same as those of $\boldsymbol{ABA}$ when $\boldsymbol{A}$ is idempotent.
11. Suppose $\boldsymbol{T}$ and $\boldsymbol{K}$ commute in multiplication. For $\boldsymbol{x}$ being an eigenvector of $\boldsymbol{T}$, show that $\boldsymbol{Kx}$ is also.
12. The eigenvalues of **A** with respect to **V** are defined as the solutions for λ to $|\mathbf{A} - \lambda\mathbf{V}| = 0$; and then $\mathbf{At} = \lambda\mathbf{Vt}$ defines **t** as an eigenvector of **A** with respect to **V**. If **V** is positive definite, symmetric,
    a. Show that λ is an eigenvalue of $\mathbf{V}^{-\frac{1}{2}}\mathbf{AV}^{-\frac{1}{2}}$.
    b. Find **t** in terms of an eigenvector of $\mathbf{V}^{-\frac{1}{2}}\mathbf{AV}^{-\frac{1}{2}}$.
13. For any idempotent matrix **A**, prove that
    a. $\mathbf{A}^k$ has the same eigenvalues as **A**,
    b. $\mathbf{A}^k$ has rank $r_{\mathbf{A}}$

       and
    c. $r(\mathbf{I} - \mathbf{A}) = n - r_{\mathbf{A}}$ for **A** of order $n \times n$.
14. Let $\boldsymbol{Q}$, $\boldsymbol{R}$, and $\boldsymbol{Y}$ be matrices of order $n \times n$ such that $\boldsymbol{Q} = \boldsymbol{RR}'$. Prove that the eigenvalues of $\boldsymbol{YQ}$ are also eigenvalues of $\boldsymbol{R}'\boldsymbol{YR}$.
15. Find the eigenvalues of $a\mathbf{I} + b\mathbf{J}$ where $\boldsymbol{J}$ has order $r \times r$ and every element is equal to one. Find also the corresponding eigenvectors.
16. Let $\boldsymbol{A}$ be an $n \times n$ matrix and $\boldsymbol{C}$ be a nonsingular matrix of the same order. Show that $\boldsymbol{A}$, $\boldsymbol{C}^{-1}\boldsymbol{AC}$, and $\boldsymbol{CAC}^{-1}$ have the same set of eigenvalues.
17. It might seem "obvious" that all eigenvalues of a matrix being zero implies the matrix is null.
    a. Prove this, for real, symmetric matrices.
    b. Create a numerical example (e.g., a 2×2 matrix) illustrating that the statement is not true for nonsymmetric matrices. Note that your example also illustrates that Section 6.5.4 is not valid for nonsymmetric matrices.
18. Let $\boldsymbol{A}$ be a symmetric $n \times n$ matrix.
    a. Show that

$$\lambda_{min} \le \frac{\boldsymbol{x}'\boldsymbol{Ax}}{\boldsymbol{x}'\boldsymbol{x}} \le \lambda_{max}, \quad \text{for any } \boldsymbol{x} \neq \mathbf{0},$$

       where λ_{min} is the smallest eigenvalue of $\boldsymbol{A}$ and λ_{max} is the largest eigenvalue.

b. For what values of $\boldsymbol{x}$ can the lower and upper bound in part (a) be achieved?

Note: The ratio $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x}/\boldsymbol{x}'\boldsymbol{x}$ is called *Rayleigh's quotient*.

19. If $\boldsymbol{A}$ is a symmetric matrix of order $n \times n$, λ_{min} and λ_{max} are the same as in Exercise 6.18, then

$$\lambda_{min} \leq \frac{1}{n} \sum_{ij} a_{ij} \leq \lambda_{max},$$

where a_{ij} is the (ij)th element of $\boldsymbol{A}$.

20. Let $\boldsymbol{A}$ be a symmetric matrix of order $n \times n$. Show that $\boldsymbol{A}$ can be written in the form:

$$\boldsymbol{A} = \sum_{i=1}^{n} \lambda_i \boldsymbol{A}_i,$$

where $\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n$ are idempotent matrices of rank 1, $\boldsymbol{A}_i\boldsymbol{A}_j = \boldsymbol{0}$, $i \neq j$, and λ_i is the ith eigenvalue of $\boldsymbol{A}$, $i = 1, 2, \ldots, n$.

21. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $n \times n$ symmetric matrices with $\boldsymbol{A}$ non-negative definite.Show that

$$e_{min}(\boldsymbol{B})tr(\boldsymbol{A}) \leq tr(\boldsymbol{AB}) \leq e_{max}(\boldsymbol{B})tr(\boldsymbol{A}),$$

where e_{min} and e_{max} are the smallest and largest eigenvalues of $\boldsymbol{A}$, respectively.

22. Let $\boldsymbol{A} = (a_{ij})$ be a symmetric matrix of order $n \times n$. Show that

$$e_{min} \leq a_{ii} \leq e_{max}, \quad i = 1, 2, \ldots, n.$$

23. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be symmetric of the same order and $\boldsymbol{B}$ is positive definite. Prove that solutions for θ to $|\boldsymbol{A} - \theta\boldsymbol{B}| = 0$ are real.

24. Let $\boldsymbol{A}$ be positive semidefinite of order $n \times n$. Show that

$$tr(\boldsymbol{A}^2) \leq e_{max}(\boldsymbol{A})tr(\boldsymbol{A}).$$

25. Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be matrices of orders $n \times n$ and $m \times m$, respectively. The *Kronecker Sum* of $\boldsymbol{A}$ and $\boldsymbol{B}$, denoted by $\boldsymbol{A} \uplus \boldsymbol{B}$, is defined as

$$\boldsymbol{A} \uplus \boldsymbol{B} = \boldsymbol{A} \otimes \boldsymbol{I}_m + \boldsymbol{I}_n \otimes \boldsymbol{B}.$$

Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be eigenvectors corresponding to the eigenvalues λ and μ of $\boldsymbol{A}$ and $\boldsymbol{B}$, respectively. Show that $\lambda + \mu$ is an eigenvalue of $\boldsymbol{A} \uplus \boldsymbol{B}$ with an eigenvector $\boldsymbol{x} \otimes \boldsymbol{y}$.

26. Find the eigenvalues and corresponding eigenvectors of $\boldsymbol{A} \uplus \boldsymbol{B}$, where

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

27. If the full-rank factorization of $\boldsymbol{A}$ having order $n \times n$ and rank r is $\boldsymbol{BC}$, show that the characteristic equation of $\boldsymbol{A}$ is $(-\lambda)^{n-r}|\boldsymbol{CB} - \lambda \boldsymbol{I}_r|$.

28. Let $\boldsymbol{A}$ be matrix of order $n \times n$. Show that $\boldsymbol{A}$ and its transpose, $\boldsymbol{A}'$, have the same set of eigenvalues.

29. Prove that the eigenvalues of a skew-symmetric matrix (defined in Section 5.1.6) are zero or imaginary.

Notes

[1]Thanks go to G. P. H. Styan for the brevity of this proof.

References

Banerjee, S. and Roy, A. (2014). *Linear Algebra and Matrix Analysis for Statistics*. CRC Press, Boca Raton, FL.

Harville, D. A. (1997). *Matrix Algebra From a Statistician's Perspective*. Springer, New York.

7
Diagonalization of Matrices

Eigenvalues and eigenvectors are a foundation for extending both applications and theory far beyond the horizons of this book. A few indications of this are given here. First, with two prerequisite lemmas, comes proof of the diagonability theorem, together with proof that all symmetric matrices are regular, that is, diagonable. Second are some results for the simultaneous diagonalization of symmetric matrices; third is the Cayley–Hamilton theorem and finally, the very useful and important singular value decomposition of a matrix.

Notation: Most of the following notation comes from Chapter 6, but is summarized here for convenience.

$\mathbf{A}$ is square, of order $n \times n$ and rank $r_{\mathbf{A}} = r$.

s = number of distinct eigenvalues.

m_k = multiplicity of the eigenvalue λ_k, with $\sum_{k=1}^{s} m_k = n$.

$\mathbf{z_A}$ is the number of zero eigenvalues of $\mathbf{A}$.

$\boldsymbol{D}$ = diagonal matrix of all n eigenvalues.

$\boldsymbol{U}$ = a matrix of eigenvectors used in $\boldsymbol{AU} = \boldsymbol{UD}$.

$\boldsymbol{B} = \boldsymbol{A} - \lambda\boldsymbol{I}$ for λ_k an eigenvalue of $\boldsymbol{A}$.

7.1 Proving the Diagonability Theorem

7.1.1 The Number of Nonzero Eigenvalues Never Exceeds Rank

It is shown in Section 6.5.4, formula (6.39), that for symmetric matrices rank equals the number of nonzero eigenvalues, $r_{\mathbf{A}} = n - z_{\mathbf{A}}$; and we remarked there that this is so for any diagonable matrix **A**. This is part of a more general result given in the following lemma applicable to any square matrix, whether diagonable or not.

Lemma 7.1 *For* $\mathbf{A}_{n \times n}$ *of rank* $r_{\mathbf{A}}$, *with* $z_{\mathbf{A}}$ *zero eigenvalues,*

$$r_{\mathbf{A}} \geq n - z_{\mathbf{A}}; \tag{7.1}$$

that is, the number of nonzero eigenvalues never exceeds rank.

Proof Let $r_{\mathbf{A}} = r$. Then, when the full-rank factorization of **A** is **XY**, its characteristic equation can be expressed (Exercise 27 in Chapter 6) as

$$(-\lambda)^{n-r}\,|\mathbf{YX} - \lambda\mathbf{I}| = 0. \tag{7.2}$$

Therefore $\lambda = 0$ is certainly a root $n - r$ times. But to the extent that **YX** has zero eigenvalues, $\lambda = 0$ can be an eigenvalue of **A** more than $n - r$ times. Therefore $z_{\mathbf{A}} \geq n - r_{\mathbf{A}}$. ∎

Example 7.1

$$\mathbf{A} = \begin{bmatrix} -1 & 1 & 2 \\ 2 & 2 & 0 \\ 0 & 4 & 4 \end{bmatrix} \tag{7.3}$$

has rank 2 (row 3 = twice row 1 + row 2); and its characteristic equation reduces to $\lambda^2(\lambda - 5) = 0$. *Hence* $r_{\mathbf{A}} = 2 > n - z_{\mathbf{A}} = 3 - 2 = 1$, *so illustrating the inequality in (7.1).*

7.1.2 A Lower Bound on $r(\mathbf{A} - \lambda_k\mathbf{I})$

A matrix **A** is diagonable only if $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ for every eigenvalue λ_k of multiplicity m_k. Example 6.10 at the end of Section 6.4 illustrates this condition not being satisfied. For cases like this, the following lemma shows that $r(\mathbf{A} - \lambda_k\mathbf{I}) \geq n - m_k$, that is, when $r(\mathbf{A} - \lambda_k\mathbf{I})$ does not equal $n - m_k$, it always exceeds $n - m_k$.

Lemma 7.2 *When* λ_k *is an eigenvalue of* $\mathbf{A}_{n \times n}$ *with multiplicity* m_k, *then*

$$r(\mathbf{A} - \lambda_k\mathbf{I}) \geq n - m_k. \tag{7.4}$$

Proof. Define $\mathbf{B} = \mathbf{A} - \lambda_k\mathbf{I}$ and $p_k = n - r_{\mathbf{B}}$. From Lemma 7.1

$$z_{\mathbf{B}} \geq p_k. \tag{7.5}$$

For θ being an eigenvalue of **B**, the corresponding eigenvalue of $\mathbf{A} = \mathbf{B} + \lambda_k\mathbf{I}$ is $\theta + \lambda_k$. Since by (7.5) we have θ = 0 not fewer than p_k times, $0 + \lambda_k = \lambda_k$ is an eigenvalue of **A** not less than p_k times; that is, $m_k \geq p_k$. Hence, $r_{\mathbf{B}} = r(\mathbf{A} - \lambda_k\mathbf{I}) = n - p_k \geq n - m_k$, and (7.4) is established. ■

Corollary 7.1 *For λ_k being a simple root, $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1$.*

Proof. Lemma 7.2 gives $r(\mathbf{A} - \lambda_k\mathbf{I}) \geq n - 1$. But $|\mathbf{A} - \lambda_k\mathbf{I}| = 0$ and so $r(\mathbf{A} - \lambda_k\mathbf{I}) < n$. Therefore, $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - 1 = n - m_k$. ■

7.1.3 Proof of the Diagonability Theorem

For convenience, the theorem is restated.

Theorem 7.1 $\mathbf{A}_{n \times n}$, *having eigenvalues λ_k with multiplicity m_k for $k = 1, 2, \ldots, s$ and $\sum_{k=1}^{s} m_k = n$, has n eigenvectors that are linearly independent if and only if $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ for all $k = 1, 2, \ldots, s$; whereupon* **U** *of* $\mathbf{AU} = \mathbf{UD}$ *is nonsingular and* **A** *is diagonable as* $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$.

Proof. (We are indebted to Dr. B. L. Raktoe for the bulk of this proof.) Sufficiency: if $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$, then **U** is nonsingular.

Since $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$, the equation $(\mathbf{A} - \lambda_k\mathbf{I})\boldsymbol{x} = \mathbf{0}$ has exactly $n - (n - m_k) = m_k$ linearly independent non-null solutions. By definition, these solutions are eigenvectors of $\boldsymbol{A}$. Hence, associated with each λ_k there is a set of m_k linearly independent eigenvectors.

To show that the sets are linearly independent of each other, suppose they are not, and that one vector of the second set of vectors, $\mathbf{y}_2$ say, is a linear combination of vectors of the first set, $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{m_1}$. Then $\mathbf{y}_2 = \sum_{i=1}^{m_1} c_i\mathbf{z}_i$ for some scalars c_i not all zero. Multiplying this equation by **A** leads to $\mathbf{Ay}_2 = \sum c_i\mathbf{Az}_i$. Because $\mathbf{y}_2$ and the **z**'s are eigenvectors corresponding respectively to the different eigenvalues λ_2 and λ_1, this means $\lambda_2\mathbf{y}_2 = \sum c_i\lambda_1\mathbf{z}_i = \lambda_1\sum c_i\mathbf{z}_i = \lambda_1\mathbf{y}_2$, which cannot be true because $\lambda_2 \neq \lambda_1$ and they are not both zero. Therefore, the supposition is wrong, and we conclude that all *s* sets of m_k eigenvectors, for $k = 1, 2, \ldots, s$ are linearly independent; that is, $\boldsymbol{U}$ is nonsingular.

Necessity: that if $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$ exists, then $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$.

Because **D** is a diagonal matrix of all *n* eigenvalues, $(\mathbf{D} - \lambda_k\mathbf{I})$ has exactly m_k zeros in its diagonal and hence $r(\mathbf{D} - \lambda_k\mathbf{I}) = n - m_k$. But $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$ so that $\mathbf{A} = \mathbf{UDU}^{-1}$. Therefore, $(\mathbf{A} - \lambda_k\mathbf{I}) = (\mathbf{UDU}^{-1} - \lambda_k\mathbf{I}) = \mathbf{U}(\mathbf{D} - \lambda_k\mathbf{I})\mathbf{U}^{-1}$, and since multiplication by nonsingular matrices does not affect rank, $r(\mathbf{A} - \lambda_k\mathbf{I}) = r(\mathbf{D} - \lambda_k\mathbf{I}) = n - m_k$. ■

7.1.4 All Symmetric Matrices Are Diagonable

A symmetric matrix is diagonable because, as shown by the following lemma, the condition $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$ of the diagonability theorem is satisfied for every eigenvalue of $\mathbf{A} = \mathbf{A}'$.

Lemma 7.3 *If λ_k is an eigenvalue of $\mathbf{A} = \mathbf{A}'$, having multiplicity m_k, then $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$.*

Proof. Use $\mathbf{B} = \mathbf{A} - \lambda_k\mathbf{I}$. With $\mathbf{A}$ being symmetric, λ_k is real, and so $\mathbf{B}$ is real; and $\mathbf{B}$ is symmetric. But $r_{\mathbf{B}} = r_{\mathbf{B}'\mathbf{B}}$, and so $r_{\mathbf{B}} = r_{\mathbf{B}^2} = n - p = t$, say, so defining p and t. Therefore, [see Section 4.15, item (iv)] at least one submatrix of order $t \times t$ in $\mathbf{B}$ is nonsingular, and so the corresponding principal minor in $\mathbf{B}^2 = \mathbf{B}'\mathbf{B}$ is nonzero. And because (by Exercise 23 in Chapter 5) $\mathbf{B}'\mathbf{B}$ is positive semidefinite, that principal minor of $\mathbf{B}^2 = \mathbf{B}'\mathbf{B}$ is positive. Therefore the sum of such principal minors, $\mathrm{tr}_t(\mathbf{B}^2) \neq 0$. Also, because $r_{\mathbf{B}^2} = t$, all minors of order greater than t are zero, and so $\mathrm{tr}_i(\mathbf{B}^2) = 0$ for $i > t$. Hence, the characteristic equation of $\mathbf{B}^2$ is

$$(-\lambda)^n + (-\lambda)^{n-1}\,\mathrm{tr}_1(\mathbf{B}^2) + \cdots + (-\lambda)^{n-1}\,\mathrm{tr}_t\left(\mathbf{B}^2\right) = 0.$$

Because the last term of this equation is nonzero, λ^{n-t} factors out; that is, $\lambda = 0$ is a root $n - t$ times. Thus $z_{\mathbf{B}^2} = n - t = p$ and so $z_{\mathbf{B}} = p$. But for θ being an eigenvalue of $\mathbf{B}$, that of $\mathbf{A} = \mathbf{B} + \lambda_k\mathbf{I}$ is $\theta + \lambda_k$, and so $0 + \lambda_k$ is an eigenvalue of $\mathbf{A}$ with multiplicity p. Hence $p = m$. Thus $r_{\mathbf{B}} = r(\mathbf{A} - \lambda_k\mathbf{I}) = n - p = n - m_k$.

This lemma assures the diagonability of symmetric matrices; that is, for $\mathbf{A} = \mathbf{A}'$ the $\mathbf{U}$ in $\mathbf{AU} = \mathbf{UD}$ is nonsingular and so $\mathbf{U}^{-1}\mathbf{AU} = \mathbf{D}$; and as is further shown in Section 6.5.3, U is an orthogonal matrix so that for $\mathbf{A}$ being symmetric, $\mathbf{U}'\mathbf{AU} = \mathbf{D}$ and $\mathbf{A} = \mathbf{UDU}'$. Equivalent to diagonability is the fact that $r(\mathbf{A} - \lambda_k\mathbf{I}) = n - m_k$; this means that $(\mathbf{A} - \lambda_k\mathbf{I})\mathbf{u} = \mathbf{0}$ has $n - (n - m_k) = m_k$ linearly independent solutions for $\mathbf{u}$; that is, the number of linearly independent eigenvectors corresponding to λ_k equals its multiplicity.

7.2 Other Results for Symmetric Matrices

Theorem 6.2 in Section 6.5 presented a very useful decomposition of a symmetric matrix. We have also seen in Section 6.6.1 a characterization of the eigenvalues of a symmetric matrix that is positive definite (positive semidefinite). The next section gives other interesting results concerning symmetric matrices.

7.2.1 Non-Negative Definite (n.n.d.)

The definition of a non-negative definite matrix was given in Section 5.7. Since this definition is in terms of quadratic forms, n.n.d. matrices are usually taken as being symmetric, and so also have the following properties:

i. All eigenvalues are real (Section 6.5.1).
ii. They are diagonable (Section 6.5.2).
iii. Rank equals the number of nonzero eigenvalues (Section 6.5.4).

These lead, in turn, to further results, as follows.

Theorem 7.2 *The eigenvalues of a symmetric matrix are all non-negative if and only if the matrix is n.n.d.*

Proof. (Thanks go to J. C. Berry for this proof.) Let $\mathbf{A} = \mathbf{A}'$ be real. Then $\mathbf{A} = \boldsymbol{UDU}'$ for orthogonal $\boldsymbol{U}$ and $\boldsymbol{D} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ for λ_i, $i = 1, \ldots, n$, being the eigenvalues of $\boldsymbol{A}$. Hence, $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}'\boldsymbol{UDU}'\boldsymbol{x} = \boldsymbol{y}'\boldsymbol{D}\boldsymbol{y}$ for $\boldsymbol{y} = \boldsymbol{U}'\boldsymbol{x}$. Therefore, $\boldsymbol{A}$ is n.n.d. if, and only if, $\boldsymbol{D}$ is n.n.d. But, $\boldsymbol{D}$ is diagonal and so is n.n.d. if and only if $\lambda_i \geq 0$ for $i = 1, \ldots, n$. ■

7.2.2 Simultaneous Diagonalization of Two Symmetric Matrices

There are at least three situations in which it is possible to find a matrix **P** that will simultaneously diagonalize two symmetric matrices (of the same order, obviously). They are as follows.

i. **A** and **B** symmetric, with **AB** = **BA**: then there exists an orthogonal **P** such that **P′AP** = **D**, a diagonal matrix of the eigenvalues of **A**, and $\mathbf{P}'\mathbf{BP} = \boldsymbol{\Delta}$, some other diagonal matrix.
ii. **A** being positive definite as well as symmetric: then there exists a **P**, nonsingular but not necessarily orthogonal, such that **P′AP** = **I** and **P′BP** = **D** where **D** is a diagonal matrix whose diagonal elements are solutions for λ to $|\mathbf{B} - \lambda\mathbf{A}| = 0$. The latter equation is sometimes called the *characteristic equation of **B** with respect to **A***.
iii. **A** and **B** both non-negative definite: then there exists a nonsingular **P** such that **P′AP** and **P′BP** are both diagonal.

The three cases are stated as theorems and proved.

Theorem 7.3 *For symmetric* **A** *and* **B** *of the same order, there exists an orthogonal matrix* **P** *such that* **P′AP** *and* **P′BP** *are both diagonal if and only if* **AB** = **BA**.

Proof. This proof follows that of Graybill (1969, Theorem 12.2.12).

Sufficiency: that if **AB** = **BA**, an orthogonal **P** exists.

Because **A** is symmetric there exists **R** such that

$$\mathbf{R}'\mathbf{A}\mathbf{R} = \mathbf{D} = \mathrm{diag}\{\lambda_i \boldsymbol{I}_{m_i}\} \quad \text{for } \mathbf{R} \text{ orthogonal,} \tag{7.6}$$

where λ_i is one of the s distinct eigenvalues of **A** of multiplicity m_i. Let

$$\mathbf{R}'\mathbf{B}\mathbf{R} = \mathbf{C} = \mathbf{C}' = \{\boldsymbol{C}_{ij}\} \tag{7.7}$$

where **C** is partitioned conformably with **D** in (7.6).

With **AB** = **BA**, the orthogonality of **R**, and the symmetry of **A** and **B**, we have **CD** = **R′BRR′AR** = **R′BAR** = **R′ABR** = **R′ARR′BR** = **DC**. Hence, because **D** of (7.6) is diagonal, $\lambda_j \boldsymbol{C}_{ij} = \lambda_i \boldsymbol{C}_{ij}$ for $i \neq j$ and $\lambda_i \neq \lambda_j$, and so $\boldsymbol{C}_{ij} = \mathbf{0}$ for $i \neq j$. Thus **C** of (7.7) is block diagonal:

$$\mathbf{C} = \mathrm{diag}\{\boldsymbol{C}_{ii}\}. \tag{7.8}$$

Symmetry of **C** implies that C_{ii} is symmetric, so that there exists $\boldsymbol{Q}_i$ such that

$$\boldsymbol{Q}_i'\boldsymbol{C}_{ii}\boldsymbol{Q}_i = \Delta_{\mathbf{i}} \text{ is diagonal, with } \boldsymbol{Q}_i \text{ orthogonal.} \tag{7.9}$$

Let

$$\mathbf{Q} = \mathrm{diag}\{\boldsymbol{Q}_i\}, \quad \text{also orthogonal,} \tag{7.10}$$

and define **P** = **RQ**, which is orthogonal because **R** and **Q** are. Then $\mathbf{P}'\mathbf{A}\mathbf{P} = \mathbf{Q}'\mathbf{R}'\mathbf{A}\mathbf{R}\mathbf{Q} = \mathbf{Q}'\mathbf{D}\mathbf{Q} = \mathrm{diag}\{\lambda_i \boldsymbol{Q}_i'\boldsymbol{Q}_i\} = \mathrm{diag}\{\lambda_i \boldsymbol{I}_{m_i}\} = \mathbf{D}$, using (7.6) and (7.10); and **P′BP** = **Q′R′BRQ** = **Q′CQ** = diag{$\mathbf{\Delta_i}$} = **Δ**, using (7.7) and (7.9). Thus with **P** being orthogonal, **P′AP** and **P′BP** are diagonal.

Necessity: that if **P′AP** = **D** and **P′BP** = **Δ** are diagonal, then **AB** = **BA**.

This is easily proved: with **D** and **Δ** being diagonal, **DΔ** = **ΔD** and so **AB** = **PP′APP′BPP′** = **PDΔP′** = **PΔDP′** = **PP′BPP′APP′** = **BA**. ■

Theorem 7.4 *For symmetric matrices* **A** *and* **B** *of the same order, with* **A** *being positive definite, there exists a nonsingular matrix* **P** *(not necessarily orthogonal), such that* **P′AP** = **I** *and* **P′BP** *is a diagonal matrix of the solutions for λ to* |**B** − λ**A**| = 0.

Proof. When **A** is positive definite, it can be expressed as $\boldsymbol{A} = \boldsymbol{M}\boldsymbol{M}'$ where $\boldsymbol{M}$ is nonsingular (see Corollary 6.2). Also, $\boldsymbol{M}^{-1}\boldsymbol{A}\boldsymbol{M}'^{-1} = \boldsymbol{I}$. Therefore solutions $\lambda_i, i = 1, \ldots, n,$ for λ to $|\boldsymbol{M}^{-1}\boldsymbol{B}\boldsymbol{M}'^{-1} - \lambda\boldsymbol{I}| = 0$ are also solutions to

$|M^{-1}||B - \lambda A||M'^{-1}| = 0$, that is, to $|\mathbf{B} - \lambda\mathbf{A}| = 0$. Further, since $M^{-1}BM'^{-1}$ is symmetric there exists an orthogonal Q such that $Q'M^{-1}BM'^{-1}Q = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. Let $P = M'^{-1}Q$. Then, $P'BP = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $P'AP = Q'M^{-1}AM'^{-1}Q = Q'Q = I$. ■

Theorem 7.5 *For real symmetric matrices* **A** *and* **B** *of the same order and non-negative definite, there exists a nonsingular matrix* **P** *such that* **P′AP** *and* **P′BP** *are both diagonal.*

Proof. This proof is based on Newcombe (1960).

Let the order of **A** and **B** be $n \times n$, with ranks a and b, respectively, and without loss of generality suppose $b \geq a$. Then, because **A** is symmetric and non-negative definite, there exists, as in 6.42, a nonsingular matrix **T** such that on defining $\gamma = n - a$,

$$\mathbf{T}'\mathbf{A}\mathbf{T} = \mathbf{D} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix}, \qquad (7.11)$$

where, for clarity, a subscripted matrix is square, the subscript being its order. Denote **T′BT** by

$$\mathbf{T}'\mathbf{B}\mathbf{T} = \begin{bmatrix} \mathbf{E}_a & \mathbf{F} \\ \mathbf{F}' & \mathbf{G}_\gamma \end{bmatrix}.$$

It is non-negative definite of rank b. Suppose $\mathbf{E}_a$ has rank $r \leq a \leq b$. Then, using the last γ rows and columns of **T′BT**, perform elementary operations (of the type described in Section 3.4, such as E_{ij}, $P_{ij}(\lambda)$, and $R_{ii}(\lambda)$, to be represented by **S′** and **S**, respectively, so that for $\beta = b - r$ and $\delta = \gamma - \beta$ we reduce **T′BT** to

$$\mathbf{S}'\mathbf{T}'\mathbf{B}\mathbf{T}\mathbf{S} = \begin{bmatrix} \mathbf{C}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_\delta \end{bmatrix}, \qquad (7.12)$$

where $\mathbf{C}_a$ has rank r. The operations are to be done in such a way that the first a rows (columns) of **T′BT** are involved only by adding multiples of the other γ rows (columns) to them. Therefore, performing the operations represented by **S** on **T′AT** of (7.11) will have no effect on (7.11), since all the last γ rows and columns of (7.11) are null, that is,

$$\mathbf{S}'\mathbf{T}'\mathbf{A}\mathbf{T}\mathbf{S} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix}. \qquad (7.13)$$

Since $\mathbf{C}_a$ of (7.12) is symmetric of rank r, there is a $\mathbf{Q}_a$ such that

$$\mathbf{Q}_a'\mathbf{C}_a\mathbf{Q}_a = \mathbf{D}_a \qquad (7.14)$$
$$= D\{\lambda_1 \ldots \lambda_p 0 \ldots 0\} \quad \text{for } Q_a \text{ orthogonal.}$$

Therefore, on defining

$$\mathbf{R} = \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I}_\delta \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\gamma \end{bmatrix} \qquad (7.15)$$

we have from (7.12), (7.14), and (7.15)

$$\mathbf{R}'\mathbf{S}'\mathbf{T}'\mathbf{BTSR} = \begin{bmatrix} \mathbf{D}_a & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_\beta & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0}_\delta \end{bmatrix} = \Delta, \quad \text{diagonal;}$$

and from (7.13) and (7.15) we have

$$\mathbf{R}'\mathbf{S}'\mathbf{T}'\mathbf{ATSR} = \begin{bmatrix} \mathbf{Q}_a'\mathbf{Q}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix} = \begin{bmatrix} \mathbf{I}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_\gamma \end{bmatrix} = \mathbf{D}, \quad \text{diagonal.}$$

Hence, for $\mathbf{P} = \mathbf{TSR}$ we have $\mathbf{P}'\mathbf{AP} = \mathbf{D}$ and $\mathbf{P}'\mathbf{BP} = \boldsymbol{\Delta}$, both diagonal. ■

Theorem 7.3 can be extended to include the simultaneous diagonalization of several symmetric matrices. This extension is given in the following theorem whose proof is similar to that of Theorem 7.3:

Theorem 7.6 *Let* $A_1, A_2, \ldots, A_k$ *be symmetric matrices of the same order. Then there exists an orthogonal matrix* P *such that* $PA_1P', PA_2P', \ldots, PA_kP'$ *are all diagonal if and only if* $A_iA_j = A_jA_i$ *for all* $i, j = 1, 2, \ldots, k$.

Theorem 7.6 has interesting applications in linear models, as will be seen later.

7.3 The Cayley–Hamilton Theorem

Theorem 7.7 *Let* $\boldsymbol{A}$ *be a matrix of order* $n \times n$ *with the characteristic equation* $|\boldsymbol{A} - \lambda \boldsymbol{I}_n| = 0$, *which can be written as*

$$|\boldsymbol{A} - \lambda \boldsymbol{I}_n| = a_0 + a_1\lambda + a_2\lambda^2 + \cdots + a_n\lambda^n \\ = 0, \tag{7.16}$$

where $a_0, a_1, \ldots, a_n$ *are known coefficients that depend on the elements of* $\boldsymbol{A}$. *Then,* $\boldsymbol{A}$ *satisfies its characteristic equation, that is,*

$$a_0\boldsymbol{I}_n + a_1\boldsymbol{A} + a_2\boldsymbol{A}^2 + \cdots + a_n\boldsymbol{A}^n = \boldsymbol{0}. \tag{7.17}$$

Proof. Let $\boldsymbol{B} = \mathrm{adj}(\boldsymbol{A} - \lambda \boldsymbol{I}_n)$ be the adjugate, or adjoint, matrix of $\boldsymbol{A} - \lambda \boldsymbol{I}_n$ (see Section 4.12 for a definition of the adjugate matrix), where λ is not an eigenvalue of $\boldsymbol{A}$. By the definition of the adjugate we have

$$(\boldsymbol{A} - \lambda \boldsymbol{I}_n)^{-1} = \frac{adj(\boldsymbol{A} - \lambda \boldsymbol{I}_n)}{|\boldsymbol{A} - \lambda \boldsymbol{I}_n|}. \tag{7.18}$$

Since the elements of $\boldsymbol{B}$ are polynomials of degree $n - 1$ or less in λ, we can express $\boldsymbol{B}$ as

$$\boldsymbol{B} = \boldsymbol{B}_0 + \lambda \boldsymbol{B}_1 + \lambda^2 \boldsymbol{B}_2 + \cdots + \lambda^{n-1} B_{n-1}, \tag{7.19}$$

where $\boldsymbol{B}_0, \boldsymbol{B}_1, \ldots, \boldsymbol{B}_{n-1}$ are matrices of order $n \times n$ that do not depend on λ. Using (7.18) we can write

$$(\boldsymbol{A} - \lambda \boldsymbol{I}_n)\boldsymbol{B} = |\boldsymbol{A} - \lambda \boldsymbol{I}_n|\boldsymbol{I}_n \\ (a_0 + a_1\lambda + \cdots + a_n\lambda^n)\boldsymbol{I}_n. \tag{7.20}$$

Substituting the representation of $\boldsymbol{B}$ given by (7.19) in (7.20) we get

$$(\boldsymbol{A} - \lambda \boldsymbol{I}_n)(\boldsymbol{B}_0 + \lambda \boldsymbol{B}_1 + \cdots + \lambda^{n-1}\boldsymbol{B}_{n-1}) = (a_0 + a_1\lambda + \cdots + a_n\lambda^n)\boldsymbol{I}_n. \tag{7.21}$$

By comparing the coefficients of the powers of λ on both sides of (7.21), we obtain

$$\begin{aligned} \boldsymbol{A}\boldsymbol{B}_0 &= a_0\boldsymbol{I}_n \\ \boldsymbol{A}\boldsymbol{B}_1 - \boldsymbol{B}_0 &= a_1\boldsymbol{I}_n \\ \boldsymbol{A}\boldsymbol{B}_2 - \boldsymbol{B}_1 &= a_2\boldsymbol{I}_n \\ &\vdots \\ \boldsymbol{A}\boldsymbol{B}_{n-1} - \boldsymbol{B}_{n-2} &= a_{n-1}\boldsymbol{I}_n \\ -\boldsymbol{B}_{n-1} &= a_n\boldsymbol{I}_n. \end{aligned}$$

Multiplying on the left the second equality by $\boldsymbol{A}$, the third by $\boldsymbol{A}^2$, etc, the one before last by $\boldsymbol{A}^{n-1}$, and the last one by $\boldsymbol{A}^n$, then adding up all the resulting equalities, we finally conclude that (7.17) is true for any square matrix. This indicates that the matrix $\boldsymbol{A}$ satisfies its own characterisation equation. ■

A useful application of Cayley–Hamilton's theorem is that when the characteristic equation is known, even if the eigenvalues are not, the nth and successive powers of $\boldsymbol{A}$ can be obtained as polynomials of $\boldsymbol{A}$. This is also true for the inverse of $\boldsymbol{A}$, if $\boldsymbol{A}$ is nonsingular.

Example 7.2 *For*

$$\mathbf{A} = \begin{bmatrix} 4 & 3 \\ 2 & 5 \end{bmatrix}, \quad \mathbf{A}^2 = \begin{bmatrix} 22 & 27 \\ 18 & 31 \end{bmatrix}, \quad \mathbf{A}^{-1} = \frac{1}{14}\begin{bmatrix} 5 & -3 \\ -2 & 4 \end{bmatrix}$$

and the characteristic equation of $\boldsymbol{A}$ *is*

$$\lambda^2 - 9\lambda + 14 = 0.$$

The Cayley–Hamilton theorem is satisfied because

$$\mathbf{A}^2 - 9A + 14\mathbf{I} = \begin{bmatrix} 22 & 27 \\ 18 & 31 \end{bmatrix} - 9\begin{bmatrix} 4 & 3 \\ 2 & 5 \end{bmatrix} + 14\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Polynomials for obtaining powers of $\mathbf{A}$ *of degree 2 or more come from rewriting* $\mathbf{A}^2 - 9\mathbf{A} + 14\mathbf{I} = 0$ *as* $\mathbf{A}^2 = 9\mathbf{A} - 14\mathbf{I}$. *Hence,*

$$\mathbf{A}^3 = 9\mathbf{A}^2 - 14\mathbf{A} = 9\,(9\mathbf{A} - 14\mathbf{I}) - 14\mathbf{A} = 67\mathbf{A} - 126\mathbf{I}$$

and

$$\mathbf{A}^4 = 67\mathbf{A}^2 - 126\mathbf{A} = 67\,(9\mathbf{A} - 14\mathbf{I}) - 126\mathbf{A} = 477\mathbf{A} - 938\mathbf{I}.$$

This means that recurrence relations can be established between the coefficients in successive powers of $\mathbf{A}$, *so enabling* $\mathbf{A}^k$, *for* $k \geq n$, *to be written as a polynomial in* $\boldsymbol{A}$ *of degree* $n - 1$, *the coefficients being functions of* k. *Thus for the above example it can be shown that, for* $k \geq 2$,

$$\mathbf{A}^k = \left(\frac{1}{5}\right)\left(7^k - 2^k\right)\mathbf{A} - \left(\frac{14}{5}\right)\left(7^{k-1} - 2^{k-1}\right)\mathbf{I}.$$

7.4 The Singular-Value Decomposition

Theorem 7.8 *Let $\boldsymbol{A}$ be an $m \times n$ matrix ($m \le n$) of rank r. There exist orthogonal matrices $\boldsymbol{P}$, of order $m \times m$, and $\boldsymbol{Q}$, of order $n \times n$, such that*

$$\boldsymbol{A} = \boldsymbol{P}[\boldsymbol{D} \quad \boldsymbol{0}]\boldsymbol{Q}', \tag{7.22}$$

where $\boldsymbol{D}$ is an $m \times m$ diagonal matrix with non-negative diagonal elements d_i, $i = 1, 2, \ldots, m$, and $\boldsymbol{0}$ is a zero matrix of order $m \times (n - m)$. The positive diagonal elements of $\boldsymbol{D}$ are the square roots of the positive eigenvalues of $\boldsymbol{AA}'$ (or, equivalently, of $\boldsymbol{A}'\boldsymbol{A}$), and are called the singular values of $\boldsymbol{A}$. If $m = n$, then $\boldsymbol{P}'\boldsymbol{AQ} = \boldsymbol{D}$.

Proof. Assume that $m < n$. The matrix $\boldsymbol{AA}'$ is symmetric, positive semidefinite of rank r. Then, there exists an orthogonal matrix $\boldsymbol{P}$ such that

$$\begin{aligned}\boldsymbol{P}'\boldsymbol{AA}'\boldsymbol{P} &= \boldsymbol{D}^2 \\ &= \operatorname{diag}\left(d_1^2, d_2^2, \ldots, d_m^2\right) \\ &= \operatorname{diag}\left(d_1^2, d_2^2, \ldots, d_r^2, 0, 0, \ldots, 0\right)\end{aligned}$$

where $d^2_i \ne 0$, $i = 1, 2, \ldots, r$. Let us partition $\boldsymbol{P}$ as

$$\boldsymbol{P} = \left[\boldsymbol{P}_1 \quad \boldsymbol{P}_2\right],$$

where $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$ are of orders $m \times r$ and $m \times (m - r)$, respectively. We then have

$$\begin{aligned}\boldsymbol{P}'\boldsymbol{AA}'\boldsymbol{P} &= \left[\boldsymbol{P}_1'\boldsymbol{P}_2'\right]\boldsymbol{AA}'\left[\boldsymbol{P}_1 \quad \boldsymbol{P}_2\right] \\ &= \operatorname{diag}(\boldsymbol{D}_1^2, \boldsymbol{0}),\end{aligned}$$

where $\boldsymbol{D}_1^2 = \operatorname{diag}(d_1^1, d_2^2, \ldots, d_r^2)$. It follows that

$$\boldsymbol{P}_1'\boldsymbol{AA}'\boldsymbol{P}_1 = \boldsymbol{D}_1^2, \text{ or } (\boldsymbol{D}_1^{-1}\boldsymbol{P}_1'\boldsymbol{A})(\boldsymbol{D}_1^{-1}\boldsymbol{P}_1'\boldsymbol{A})' = \boldsymbol{I}_r, \tag{7.23}$$

and

$$\boldsymbol{P}_2'\boldsymbol{AA}'\boldsymbol{P}_2 = \boldsymbol{0}, \text{ or } (\boldsymbol{P}_2'\boldsymbol{A})(\boldsymbol{P}_2'\boldsymbol{A})' = \boldsymbol{0}.$$

Hence, $\boldsymbol{P}_2'\boldsymbol{A} = \boldsymbol{0}$. Let the matrix $\boldsymbol{Q}_1$ be defined as

$$\boldsymbol{Q}_1 = \boldsymbol{A}'\boldsymbol{P}_1\boldsymbol{D}_1^{-1}. \tag{7.24}$$

This matrix is of order $n \times r$ whose columns are orthonormal since $\boldsymbol{Q}_1'\boldsymbol{Q}_1 = \boldsymbol{I}_r$ by (7.23). Furthermore, there exits a matrix $\boldsymbol{Q}_2$ of order $n \times (n - r)$ whose columns are orthonormal and are orthogonal to the columns of $\boldsymbol{Q}_1$. Then, the matrix

$$Q = \begin{bmatrix} Q_1 & Q_2 \end{bmatrix}$$

is orthogonal of order $n \times n$. We then have

$$\begin{aligned} P'AQ &= \begin{bmatrix} P'_1 \\ P'_2 \end{bmatrix} A \begin{bmatrix} Q_1 & Q_2 \end{bmatrix} \\ &= \begin{bmatrix} P'_1AQ_1 & P'_1AQ_2 \\ P'_2AQ_1 & P'_2AQ_2 \end{bmatrix}. \end{aligned} \tag{7.25}$$

Note that P'_2AQ_1 and P'_2AQ_2 are each equal to a zero matrix since $P'_2A = \mathbf{0}$. Furthermore, $P'_1AQ_2 = \mathbf{0}$ since by the definition of the matrix Q,

$$Q'_1Q_2 = \mathbf{0}. \tag{7.26}$$

Using the expression for Q_1 given by (7.24) in (7.26) we get, $D_1^{-1}P'_1AQ_2 = \mathbf{0}$. This implies that $P'_1AQ_2 = \mathbf{0}$. Making the substitution in (7.25), we get

$$\begin{aligned} P'AQ &= \begin{bmatrix} P'_1AA'P_1D_1^{-1} & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{bmatrix} \\ &= \begin{bmatrix} D_1 & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{bmatrix} \\ &= \begin{bmatrix} D & \mathbf{0}_{m\times(n-m)} \end{bmatrix}, \end{aligned} \tag{7.27}$$

since $n - r = (m - r) + (n - m)$ and the fact that $P'_1AA'P_1 = D_1^2$ as shown in (7.23). We finally conclude that

$$A = P\begin{bmatrix} D & \mathbf{0}_{m\times(n-m)} \end{bmatrix}Q'. \quad \blacksquare$$

Corollary 7.2 *The columns of the matrix Q in Theorem 7.8 are orthonormal eigenvectors of $A'A$ which has the same nonzero eigenvalues as those of AA'.*

Proof. Using (7.27) we have

$$A = P\begin{bmatrix} D_1 & \mathbf{0}_{r\times(n-r)} \\ \mathbf{0}_{(m-r)\times r} & \mathbf{0}_{(m-r)\times(n-r)} \end{bmatrix}Q'. \tag{7.28}$$

Transposing both sides of (7.28) we get

$$A' = Q\begin{bmatrix} D_1 & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix}P'. \quad (7.29)$$

Using (7.28) and (7.29) and noting that $P'P = I_m$ we conclude

$$A'A = Q\begin{bmatrix} D_1 & 0_{r\times(m-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(m-r)} \end{bmatrix}\begin{bmatrix} D_1 & 0_{r\times(n-r)} \\ 0_{(m-r)\times r} & 0_{(m-r)\times(n-r)} \end{bmatrix}Q'$$
$$= Q\begin{bmatrix} D_1^2 & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix}Q'.$$

This clearly indicates that Q is an orthogonal matrix of eigenvectors of $A'A$ with nonzero eigenvalues given by the diagonal elements of D_1^2. ■

Example 7.3 *For*

$$\mathbf{A} = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 0 & 2 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}, \quad \mathbf{AA'} = \begin{bmatrix} 6 & 2 & 4 \\ 2 & 6 & 4 \\ 4 & 4 & 4 \end{bmatrix}$$

has nonzero eigenvalues 12 and 4. The rank of A *is therefore equal to 2. The matrix* P *of orthonormal eigenvectors of* AA' *is given by*

$$P = \begin{bmatrix} \sqrt{2} & \sqrt{3} & 1 \\ \sqrt{2} & -\sqrt{3} & 1 \\ \sqrt{2} & 0 & -2 \end{bmatrix}/\sqrt{6}.$$

The matrix $A'A$ *is*

$$A'A = \begin{bmatrix} 5 & 1 & 3 & 3 \\ 1 & 5 & 3 & 3 \\ 3 & 3 & 3 & 3 \\ 3 & 3 & 3 & 3 \end{bmatrix}$$

Its matrix Q *of orthonormal eigenvectors is*

$$Q = \begin{bmatrix} \sqrt{3} & \sqrt{6} & \sqrt{2} & 1 \\ \sqrt{3} & -\sqrt{6} & \sqrt{2} & 1 \\ \sqrt{3} & 0 & -2\sqrt{2} & 1 \\ \sqrt{3} & 0 & 0 & -3 \end{bmatrix} / \sqrt{12}.$$

In addition, the matrix $\boldsymbol{D}$ *is*

$$\boldsymbol{D} = \mathrm{diag}(\sqrt{12}, 2, 0).$$

Applying formula (7.22) we get

$$\boldsymbol{A} = \boldsymbol{P}[\boldsymbol{D} \quad \boldsymbol{0}_{3\times 1}]\boldsymbol{Q}'. \tag{7.30}$$

The reader can verify that the product of the three matrices on the right-hand side of (7.30) is equal to $\boldsymbol{A}$.

Good (1969) described several interesting applications of the singular-values decomposition in statistics, including the theory of least squares and the analysis of contingency tables.

Exercises

1. Show that if $\boldsymbol{A}$ is an idempotent matrix, then $\text{rank}(\boldsymbol{A}) = tr(\boldsymbol{A})$.

2. Find the singular-value decomposition of $\begin{bmatrix} 10 & -5 \\ 2 & -11 \\ 6 & -8 \end{bmatrix}$.

3. Find the rank of $\mathbf{M} = \begin{bmatrix} 1 & -3 & 5 \\ 0 & 1 & 7 \\ 2 & -4 & 28 \\ 3 & -13 & 0 \\ 4 & -14 & -3 \\ 5 & -9 & 30 \\ 6 & 2 & 173 \end{bmatrix}$.

4. Suppose

$$\mathbf{K} = \mathbf{K}', \quad \mathbf{K} = \mathbf{K}^3, \quad \mathbf{K1} = \mathbf{0}, \quad \text{and} \quad \mathbf{K}\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}.$$

Calculate the following values, in each case giving reasons as to why the value can be calculated *without calculating* **K**.

   a. The order of **K**.

   b. The rank of **K**.

   c. The trace of **K**.

   d. The determinant of **K**.

5. Let $\boldsymbol{A}$ be a matrix of order $n \times n$ that satisfies the equation

$$\boldsymbol{A}^2 + 2\boldsymbol{A} + \boldsymbol{I}_n = \mathbf{0}.$$

   a. Show that $\boldsymbol{A}$ is nonsingular.

b. Find the inverse of $\boldsymbol{A}$.

6. Suppose that there is an $n \times n$ matrix $\boldsymbol{A} \neq \boldsymbol{0}$ such that $\boldsymbol{A}^n = \boldsymbol{0}$. Show that such a matrix is not diagonable [see Banerjee and Roy (2014), Section 11.4].

7. Find a 2×2 matrix that is not diagonable.

8. Suppose that $\boldsymbol{A}$ is an $n \times n$ matrix and $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, where $\boldsymbol{P}$ is a nonsingular matrix. Find such a $\boldsymbol{P}$ matrix and the diagonal elements of $\boldsymbol{\Lambda}$ [see Banerjee and Roy (2014), Section 11.4].

9. Find a matrix $\boldsymbol{P}$ such that $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ is diagonal, where

$$\boldsymbol{A} = \begin{bmatrix} 6 & 3 & 2 \\ -5 & 0 & 3 \\ 7 & 5 & 1 \end{bmatrix}.$$

10. Show that the matrix $\boldsymbol{A}$ given by

$$\boldsymbol{A} = \begin{bmatrix} 5 & 3 \\ 0 & 5 \end{bmatrix}$$

is not diagonable.