

CHAPTER 2

Basic Concepts in Linear Algebra

In this chapter we present some fundamental concepts concerning vector spaces and matrix algebra. The purpose of the chapter is to familiarize the reader with these concepts, since they are essential to the understanding of some of the remaining chapters. For this reason, most of the theorems in this chapter will be stated without proofs. There are several excellent books on linear algebra that can be used for a more detailed study of this subject (see the bibliography at the end of this chapter).

In statistics, matrix algebra is used quite extensively, especially in linear models and multivariate analysis. The books by Basilevsky (1983), Graybill (1983), Magnus and Neudecker (1988), and Searle (1982) include many applications of matrices in these areas.

In this chapter, as well as in the remainder of the book, elements of the set of real numbers, R , are sometimes referred to as scalars. The Cartesian product $\times_{i=1}^n R$ is denoted by R^n , which is also known as the n -dimensional Euclidean space. Unless otherwise stated, all matrix elements are considered to be real numbers.

2.1. VECTOR SPACES AND SUBSPACES

A vector space over R is a set V of elements called vectors together with two operations, addition and scalar multiplication, that satisfy the following conditions:

1. $\mathbf{u} + \mathbf{v}$ is an element of V for all $\mathbf{u}, \mathbf{v}$ in V .
2. If α is a scalar and $\mathbf{u} \in V$, then $\alpha \mathbf{u} \in V$.
3. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ for all $\mathbf{u}, \mathbf{v}$ in V .
4. $\mathbf{u} + (\mathbf{v} + \mathbf{w}) = (\mathbf{u} + \mathbf{v}) + \mathbf{w}$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V .
5. There exists an element $\mathbf{0} \in V$ such that $\mathbf{0} + \mathbf{u} = \mathbf{u}$ for all $\mathbf{u}$ in V . This element is called the zero vector.

6. For each $\mathbf{u} \in V$ there exists a $\mathbf{v} \in V$ such that $\mathbf{u} + \mathbf{v} = \mathbf{0}$.
7. $\alpha(\mathbf{u} + \mathbf{v}) = \alpha\mathbf{u} + \alpha\mathbf{v}$ for any scalar α and any $\mathbf{u}$ and $\mathbf{v}$ in V .
8. $(\alpha + \beta)\mathbf{u} = \alpha\mathbf{u} + \beta\mathbf{u}$ for any scalars α and β and any $\mathbf{u}$ in V .
9. $\alpha(\beta\mathbf{u}) = (\alpha\beta)\mathbf{u}$ for any scalars α and β and any $\mathbf{u}$ in V .
10. $1\mathbf{u} = \mathbf{u}$ for any $\mathbf{u} \in V$.

EXAMPLE 2.1.1. A familiar example of a vector space is the n -dimensional Euclidean space R^n . Here, addition and multiplication are defined as follows: If $(u_1, u_2, \dots, u_n)$ and $(v_1, v_2, \dots, v_n)$ are two elements in R^n , then their sum is defined as $(u_1 + v_1, u_2 + v_2, \dots, u_n + v_n)$. If α is a scalar, then $\alpha(u_1, u_2, \dots, u_n) = (\alpha u_1, \alpha u_2, \dots, \alpha u_n)$.

EXAMPLE 2.1.2. Let V be the set of all polynomials in x of degree less than or equal to k . Then V is a vector space. Any element in V can be expressed as $\sum_{i=0}^k a_i x^i$, where the a_i 's are scalars.

EXAMPLE 2.1.3. Let V be the set of all functions defined on the closed interval $[-1, 1]$. Then V is a vector space. It can be seen that $f(x) + g(x)$ and $\alpha f(x)$ belong to V , where $f(x)$ and $g(x)$ are elements in V and α is any scalar.

EXAMPLE 2.1.4. The set V of all nonnegative functions defined on $[-1, 1]$ is not a vector space, since if $f(x) \in V$ and α is a negative scalar, then $\alpha f(x) \notin V$.

EXAMPLE 2.1.5. Let V be the set of all points (x, y) on a straight line given by the equation $2x - y + 1 = 0$. Then V is not a vector space. This is because if (x_1, y_1) and (x_2, y_2) belong to V , then $(x_1 + x_2, y_1 + y_2) \notin V$, since $2(x_1 + x_2) - (y_1 + y_2) + 1 = -1 \neq 0$. Alternatively, we can state that V is not a vector space because the zero element $(0, 0)$ does not belong to V . This violates condition 5 for a vector space.

A subset W of a vector space V is said to form a vector subspace if W itself is a vector space. Equivalently, W is a subspace if whenever $\mathbf{u}, \mathbf{v} \in W$ and α is a scalar, then $\mathbf{u} + \mathbf{v} \in W$ and $\alpha\mathbf{u} \in W$. For example, the set W of all continuous functions defined on $[-1, 1]$ is a vector subspace of V in Example 2.1.3. Also, the set of all points on the straight line $y - 2x = 0$ is a vector subspace of R^2 . However, the points on any straight line in R^2 not going through the origin $(0, 0)$ do not form a vector subspace, as was seen in Example 2.1.5.

Definition 2.1.1. Let V be a vector space, and $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ be a collection of n elements in V . These elements are said to be linearly dependent if there exist n scalars $\alpha_1, \alpha_2, \dots, \alpha_n$, not all equal to zero, such that $\sum_{i=1}^n \alpha_i \mathbf{u}_i = \mathbf{0}$. If, however, $\sum_{i=1}^n \alpha_i \mathbf{u}_i = \mathbf{0}$ is true only when all the α_i 's are zero, then

$\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ are linearly independent. It should be noted that if $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ are linearly independent, then none of them can be zero. If, for example, $\mathbf{u}_1 = \mathbf{0}$, then $\alpha \mathbf{u}_1 + 0\mathbf{u}_2 + \dots + 0\mathbf{u}_n = \mathbf{0}$ for any $\alpha \neq 0$, which implies that the $\mathbf{u}_i$'s are linearly dependent, a contradiction. $\square$

From the preceding definition we can say that a collection of n elements in a vector space are linearly dependent if at least one element in this collection can be expressed as a linear combination of the remaining $n - 1$ elements. If no element, however, can be expressed in this fashion, then the n elements are linearly independent. For example, in R^3 , $(1, 2, -2)$, $(-1, 0, 3)$, and $(1, 4, -1)$ are linearly dependent, since $2(1, 2, -2) + (-1, 0, 3) - (1, 4, -1) = \mathbf{0}$. On the other hand, it can be verified that $(1, 1, 0)$, $(1, 0, 2)$, and $(0, 1, 3)$ are linearly independent.

Definition 2.1.2. Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ be n elements in a vector space V . The collection of all linear combinations of the form $\sum_{i=1}^n \alpha_i \mathbf{u}_i$, where the α_i 's are scalars, is called a linear span of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ and is denoted by $L(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$. $\square$

It is easy to see from the preceding definition that $L(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$ is a vector subspace of V . This vector subspace is said to be spanned by $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$.

Definition 2.1.3. Let V be a vector space. If there exist linearly independent elements $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ in V such that $V = L(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$, then $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ are said to form a basis for V . The number n of elements in this basis is called the dimension of the vector space and is denoted by $\dim V$. $\square$

Note that a basis for a vector space is not unique. However, its dimension is unique. For example, the three vectors $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ form a basis for R^3 . Another basis for R^3 consists of $(1, 1, 0)$, $(1, 0, 1)$, and $(0, 1, 1)$.

If $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ form a basis for V and if $\mathbf{u}$ is a given element in V , then there exists a unique set of scalars, $\alpha_1, \alpha_2, \dots, \alpha_n$, such that $\mathbf{u} = \sum_{i=1}^n \alpha_i \mathbf{u}_i$. To show this, suppose that there exists another set of scalars, $\beta_1, \beta_2, \dots, \beta_n$, such that $\mathbf{u} = \sum_{i=1}^n \beta_i \mathbf{u}_i$. Then $\sum_{i=1}^n (\alpha_i - \beta_i) \mathbf{u}_i = \mathbf{0}$, which implies that $\alpha_i = \beta_i$ for all i , since the $\mathbf{u}_i$'s are linearly independent.

Let us now check the dimensions of the vector spaces for some of the examples described earlier. For Example 2.1.1, $\dim V = n$. In Example 2.1.2, $\{1, x, x^2, \dots, x^k\}$ is a basis for V ; hence $\dim V = k + 1$. As for Example 2.1.3, $\dim V$ is infinite, since there is no finite set of functions that can span V .

Definition 2.1.4. Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in R^n . The dot product (also called scalar product or inner product) of $\mathbf{u}$ and $\mathbf{v}$ is a scalar denoted by $\mathbf{u} \cdot \mathbf{v}$ and is given by

$$\mathbf{u} \cdot \mathbf{v} = \sum_{i=1}^n u_i v_i,$$

where u_i and v_i are the i th components of $\mathbf{u}$ and $\mathbf{v}$, respectively ($i = 1, 2, \dots, n$). In particular, if $\mathbf{u} = \mathbf{v}$, then $(\mathbf{u} \cdot \mathbf{u})^{1/2} = (\sum_{i=1}^n u_i^2)^{1/2}$ is called the Euclidean norm (or length) of $\mathbf{u}$ and is denoted by $\|\mathbf{u}\|_2$. The dot product of $\mathbf{u}$ and $\mathbf{v}$ is also equal to $\|\mathbf{u}\|_2 \|\mathbf{v}\|_2 \cos \theta$, where θ is the angle between $\mathbf{u}$ and $\mathbf{v}$. $\square$

Definition 2.1.5. Two vectors $\mathbf{u}$ and $\mathbf{v}$ in R^n are said to be orthogonal if their dot product is zero. $\square$

Definition 2.1.6. Let U be a vector subspace of R^n . The vectors $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ form an orthonormal basis for U if they satisfy the following properties:

1. $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ form a basis for U .
2. $\mathbf{e}_i \cdot \mathbf{e}_j = 0$ for all $i \neq j$ ($i, j = 1, 2, \dots, m$).
3. $\|\mathbf{e}_i\|_2 = 1$ for $i = 1, 2, \dots, m$.

Any collection of vectors satisfying just properties 2 and 3 are said to be orthonormal. $\square$

Theorem 2.1.1. Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ be a basis for a vector subspace U of R^n . Then there exists an orthonormal basis, $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$, for U , given by

$$\begin{aligned} \mathbf{e}_1 &= \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|_2}, & \text{where } \mathbf{v}_1 &= \mathbf{u}_1, \\ \mathbf{e}_2 &= \frac{\mathbf{v}_2}{\|\mathbf{v}_2\|_2}, & \text{where } \mathbf{v}_2 &= \mathbf{u}_2 - \frac{\mathbf{v}_1 \cdot \mathbf{u}_2}{\|\mathbf{v}_1\|_2^2} \mathbf{v}_1, \\ &\vdots \\ \mathbf{e}_m &= \frac{\mathbf{v}_m}{\|\mathbf{v}_m\|_2}, & \text{where } \mathbf{v}_m &= \mathbf{u}_m - \sum_{i=1}^{m-1} \frac{\mathbf{v}_i \cdot \mathbf{u}_m}{\|\mathbf{v}_i\|_2^2} \mathbf{v}_i. \end{aligned}$$

Proof. See Graybill (1983, Theorem 2.6.5). $\square$

The procedure of constructing an orthonormal basis from any given basis as described in Theorem 2.1.1 is known as the Gram-Schmidt orthonormalization procedure.

Theorem 2.1.2. Let $\mathbf{u}$ and $\mathbf{v}$ be two vectors in R^n . Then:

1. $|\mathbf{u} \cdot \mathbf{v}| \leq \|\mathbf{u}\|_2 \|\mathbf{v}\|_2$.
2. $\|\mathbf{u} + \mathbf{v}\|_2 \leq \|\mathbf{u}\|_2 + \|\mathbf{v}\|_2$.

Proof. See Marcus and Minc (1988, Theorem 3.4). $\square$

The inequality in part 1 of Theorem 2.1.2 is known as the *Cauchy–Schwarz inequality*. The one in part 2 is called the *triangle inequality*.

Definition 2.1.7. Let U be a vector subspace of R^n . The orthogonal complement of U , denoted by $U^\perp$, is the vector subspace of R^n which consists of all vectors $\mathbf{v}$ such that $\mathbf{u} \cdot \mathbf{v} = 0$ for all $\mathbf{u}$ in U . $\square$

Definition 2.1.8. Let $U_1, U_2, \dots, U_n$ be vector subspaces of the vector space U . The direct sum of these vector subspaces, denoted by $\bigoplus_{i=1}^n U_i$, consists of all vectors $\mathbf{u}$ that can be uniquely expressed as $\mathbf{u} = \sum_{i=1}^n \mathbf{u}_i$, where $\mathbf{u}_i \in U_i$, $i = 1, 2, \dots, n$. $\square$

Theorem 2.1.3. Let $U_1, U_2, \dots, U_n$ be vector subspaces of the vector space U . Then:

1. $\bigoplus_{i=1}^n U_i$ is a vector subspace of U .
2. If $U = \bigoplus_{i=1}^n U_i$, then $\bigcap_{i=1}^n U_i$ consists of just the zero element $\mathbf{0}$ of U .
3. $\dim \bigoplus_{i=1}^n U_i = \sum_{i=1}^n \dim U_i$.

Proof. The proof is left as an exercise. $\square$

Theorem 2.1.4. Let U be a vector subspace of R^n . Then $R^n = U \oplus U^\perp$.

Proof. See Marcus and Minc (1988, Theorem 3.3). $\square$

From Theorem 2.1.4 we conclude that any $\mathbf{v} \in R^n$ can be uniquely written as $\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$, where $\mathbf{v}_1 \in U$ and $\mathbf{v}_2 \in U^\perp$. In this case, $\mathbf{v}_1$ and $\mathbf{v}_2$ are called the projections of $\mathbf{v}$ on U and $U^\perp$, respectively.

2.2. LINEAR TRANSFORMATIONS

Let U and V be two vector spaces. A function $T: U \rightarrow V$ is called a linear transformation if $T(\alpha_1 \mathbf{u}_1 + \alpha_2 \mathbf{u}_2) = \alpha_1 T(\mathbf{u}_1) + \alpha_2 T(\mathbf{u}_2)$ for all $\mathbf{u}_1, \mathbf{u}_2$ in U and any scalars α_1 and α_2 . For example, let $T: R^3 \rightarrow R^3$ be defined as

$$T(x_1, x_2, x_3) = (x_1 - x_2, x_1 + x_3, x_3).$$

Then T is a linear transformation, since

$$\begin{aligned} & T[\alpha(x_1, x_2, x_3) + \beta(y_1, y_2, y_3)] \\ &= T(\alpha x_1 + \beta y_1, \alpha x_2 + \beta y_2, \alpha x_3 + \beta y_3) \\ &= (\alpha x_1 + \beta y_1 - \alpha x_2 - \beta y_2, \alpha x_1 + \beta y_1 + \alpha x_3 + \beta y_3, \alpha x_3 + \beta y_3) \\ &= \alpha(x_1 - x_2, x_1 + x_3, x_3) + \beta(y_1 - y_2, y_1 + y_3, y_3) \\ &= \alpha T(x_1, x_2, x_3) + \beta T(y_1, y_2, y_3). \end{aligned}$$

We note that the image of U under T , or the range of T , namely $T(U)$, is a vector subspace of V . This is true because if $\mathbf{v}_1, \mathbf{v}_2$ are in $T(U)$, then there exist $\mathbf{u}_1$ and $\mathbf{u}_2$ in U such that $\mathbf{v}_1 = T(\mathbf{u}_1)$ and $\mathbf{v}_2 = T(\mathbf{u}_2)$. Hence, $\mathbf{v}_1 + \mathbf{v}_2 = T(\mathbf{u}_1) + T(\mathbf{u}_2) = T(\mathbf{u}_1 + \mathbf{u}_2)$, which belongs to $T(U)$. Also, if α is a scalar, then $\alpha T(\mathbf{u}) = T(\alpha \mathbf{u}) \in T(U)$ for any $\mathbf{u} \in U$.

Definition 2.2.1. Let $T: U \rightarrow V$ be a linear transformation. The kernel of T , denoted by $\ker T$, is the collection of all vectors $\mathbf{u}$ in U such that $T(\mathbf{u}) = \mathbf{0}$, where $\mathbf{0}$ is the zero vector in V . The kernel of T is also called the null space of T .

As an example of a kernel, let $T: R^3 \rightarrow R^3$ be defined as $T(x_1, x_2, x_3) = (x_1 - x_2, x_1 - x_3)$. Then

$$\ker T = \{(x_1, x_2, x_3) | x_1 = x_2, x_1 = x_3\}$$

In this case, $\ker T$ consists of all points (x_1, x_2, x_3) in R^3 that lie on a straight line through the origin given by the equations $x_1 = x_2 = x_3$. $\square$

Theorem 2.2.1. Let $T: U \rightarrow V$ be a linear transformation. Then we have the following:

1. $\ker T$ is a vector subspace of U .
2. $\dim U = \dim(\ker T) + \dim[T(U)]$.

Proof. Part 1 is left as an exercise. To prove part 2 we consider the following. Let $\dim U = n$, $\dim(\ker T) = p$, and $\dim[T(U)] = q$. Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$ be a basis for $\ker T$, and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_q$ be a basis for $T(U)$. Then, there exist vectors $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q$ in U such that $T(\mathbf{w}_i) = \mathbf{v}_i$ ($i = 1, 2, \dots, q$). We need to show that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p; \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q$ form a basis for U , that is, they are linearly independent and span U .

Suppose that there exist scalars $\alpha_1, \alpha_2, \dots, \alpha_p; \beta_1, \beta_2, \dots, \beta_q$ such that

$$\sum_{i=1}^p \alpha_i \mathbf{u}_i + \sum_{i=1}^q \beta_i \mathbf{w}_i = \mathbf{0}. \quad (2.1)$$

Then

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

where $\mathbf{0}$ represents the zero vector in V

$$\begin{aligned} &= \sum_{i=1}^p \alpha_i T(\mathbf{u}_i) + \sum_{i=1}^q \beta_i T(\mathbf{w}_i) \\ &= \sum_{i=1}^q \beta_i T(\mathbf{w}_i), \quad \text{since } \mathbf{u}_i \in \ker T, i = 1, 2, \dots, p \\ &= \sum_{i=1}^q \beta_i \mathbf{v}_i. \end{aligned}$$

Since the $\mathbf{v}_i$'s are linearly independent, then $\beta_i = 0$ for $i = 1, 2, \dots, q$. From (2.1) it follows that $\alpha_i = 0$ for $i = 1, 2, \dots, p$, since the $\mathbf{u}_i$'s are also linearly independent. Thus the vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p; \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q$ are linearly independent.

Let us now suppose that $\mathbf{u}$ is any vector in U . To show that it belongs to $L(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p; \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q)$. Let $\mathbf{v} = T(\mathbf{u})$. Then there exist scalars $a_1, a_2, \dots, a_q$ such that $\mathbf{v} = \sum_{i=1}^q a_i \mathbf{v}_i$. It follows that

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

Thus,

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

and $\mathbf{u} - \sum_{i=1}^q a_i \mathbf{w}_i$ must then belong to $\ker T$. Hence,

$$\mathbf{u} - \sum_{i=1}^q a_i \mathbf{w}_i = \sum_{i=1}^p b_i \mathbf{u}_i \quad (2.2)$$

for some scalars, $b_1, b_2, \dots, b_p$. From (2.2) we then have

$$\mathbf{u} = \sum_{i=1}^p b_i \mathbf{u}_i + \sum_{i=1}^q a_i \mathbf{w}_i,$$

which shows that $\mathbf{u}$ belongs to the linear span of $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p; \mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_q$. We conclude that these vectors form a basis for U . Hence, $n = p + q$. $\square$

Corollary 2.2.1. $T: U \rightarrow V$ is a one-to-one linear transformation if and only if $\dim(\ker T) = 0$.

Proof. If T is a one-to-one linear transformation, then $\ker T$ consists of just one vector, namely, the zero vector. Hence, $\dim(\ker T) = 0$. Vice versa, if $\dim(\ker T) = 0$, or equivalently, if $\ker T$ consists of just the zero vector, then T must be a one-to-one transformation. This is true because if $\mathbf{u}_1$ and $\mathbf{u}_2$ are in U and such that $T(\mathbf{u}_1) = T(\mathbf{u}_2)$, then $T(\mathbf{u}_1 - \mathbf{u}_2) = \mathbf{0}$, which implies that $\mathbf{u}_1 - \mathbf{u}_2 \in \ker T$ and thus $\mathbf{u}_1 - \mathbf{u}_2 = \mathbf{0}$. $\square$

2.3. MATRICES AND DETERMINANTS

Matrix algebra was devised by the English mathematician Arthur Cayley (1821–1895). The use of matrices originated with Cayley in connection with

linear transformations of the form

$$ax_1 + bx_2 = y_1,$$

$$cx_1 + dx_2 = y_2,$$

where a, b, c , and d are scalars. This transformation is completely determined by the square array

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

which is called a matrix of order 2×2 . In general, let $T: U \rightarrow V$ be a linear transformation, where U and V are vector spaces of dimensions m and n , respectively. Let $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$ be a basis for U and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be a basis for V . For $i = 1, 2, \dots, m$, consider $T(\mathbf{u}_i)$, which can be uniquely represented as

$$T(\mathbf{u}_i) = \sum_{j=1}^n a_{ij} \mathbf{v}_j, \quad i = 1, 2, \dots, m,$$

where the a_{ij} 's are scalars. These scalars completely determine all possible values of T : If $\mathbf{u} \in U$, then $\mathbf{u} = \sum_{i=1}^m c_i \mathbf{u}_i$ for some scalars $c_1, c_2, \dots, c_m$. Then $T(\mathbf{u}) = \sum_{i=1}^m c_i T(\mathbf{u}_i) = \sum_{i=1}^m c_i (\sum_{j=1}^n a_{ij} \mathbf{v}_j)$. By definition, the rectangular array

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

is called a matrix of order $m \times n$, which indicates that $\mathbf{A}$ has m rows and n columns. The a_{ij} 's are called the elements of $\mathbf{A}$. In some cases it is more convenient to represent $\mathbf{A}$ using the notation $\mathbf{A} = (a_{ij})$. In particular, if $m = n$, then $\mathbf{A}$ is called a square matrix. Furthermore, if the off-diagonal elements of a square matrix $\mathbf{A}$ are zero, then $\mathbf{A}$ is called a diagonal matrix and is written as $\mathbf{A} = \text{Diag}(a_{11}, a_{22}, \dots, a_{nn})$. In this special case, if the diagonal elements are equal to 1, then $\mathbf{A}$ is called the identity matrix and is denoted by $\mathbf{I}_n$ to indicate that it is of order $n \times n$. A matrix of order $m \times 1$ is called a column vector. Likewise, a matrix of order $1 \times n$ is called a row vector.

2.3.1. Basic Operations on Matrices

1. *Equality of Matrices.* Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be two matrices of the same order. Then $\mathbf{A} = \mathbf{B}$ if and only if $a_{ij} = b_{ij}$ for all $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$.

2. *Addition of Matrices.* Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be two matrices of order $m \times n$. Then $\mathbf{A} + \mathbf{B}$ is a matrix $\mathbf{C} = (c_{ij})$ of order $m \times n$ such that $c_{ij} = a_{ij} + b_{ij}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$).
3. *Scalar Multiplication.* Let α be a scalar, and $\mathbf{A} = (a_{ij})$ be a matrix of order $m \times n$. Then $\alpha\mathbf{A} = (\alpha a_{ij})$.
4. *The Transpose of a Matrix.* Let $\mathbf{A} = (a_{ij})$ be a matrix of order $m \times n$. The transpose of $\mathbf{A}$, denoted by $\mathbf{A}'$, is a matrix of order $n \times m$ whose rows are the columns of $\mathbf{A}$. For example,

$$\text{if } \mathbf{A} = \begin{bmatrix} 2 & 3 & 1 \\ -1 & 0 & 7 \end{bmatrix}, \text{ then } \mathbf{A}' = \begin{bmatrix} 2 & -1 \\ 3 & 0 \\ 1 & 7 \end{bmatrix}.$$

A matrix $\mathbf{A}$ is symmetric if $\mathbf{A} = \mathbf{A}'$. It is skew-symmetric if $\mathbf{A}' = -\mathbf{A}$. A skew-symmetric matrix must necessarily have zero elements along its diagonal.

5. *Product of Matrices.* Let $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ be matrices of orders $m \times n$ and $n \times p$, respectively. The product $\mathbf{AB}$ is a matrix $\mathbf{C} = (c_{ij})$ of order $m \times p$ such that $c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, p$). It is to be noted that this product is defined only when the number of columns of $\mathbf{A}$ is equal to the number of rows of $\mathbf{B}$.

In particular, if $\mathbf{a}$ and $\mathbf{b}$ are column vectors of order $n \times 1$, then their dot product $\mathbf{a} \cdot \mathbf{b}$ can be expressed as a matrix product of the form $\mathbf{a}'\mathbf{b}$ or $\mathbf{b}'\mathbf{a}$.

6. *The Trace of a Matrix.* Let $\mathbf{A} = (a_{ij})$ be a square matrix of order $n \times n$. The trace of $\mathbf{A}$, denoted by $\text{tr}(\mathbf{A})$, is the sum of its diagonal elements, that is,

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

On the basis of this definition, it is easy to show that if $\mathbf{A}$ and $\mathbf{B}$ are matrices of order $n \times n$, then the following hold: (i) $\text{tr}(\mathbf{AB}) = \text{tr}(\mathbf{BA})$; (ii) $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$.

Definition 2.3.1. Let $\mathbf{A} = (a_{ij})$ be an $m \times n$ matrix. A submatrix $\mathbf{B}$ of $\mathbf{A}$ is a matrix which can be obtained from $\mathbf{A}$ by deleting a certain number of rows and columns.

In particular, if the i th row and j th column of $\mathbf{A}$ that contain the element a_{ij} are deleted, then the resulting matrix is denoted by $\mathbf{M}_{ij}$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$).

Let us now suppose that $\mathbf{A}$ is a square matrix of order $n \times n$. If rows $i_1, i_2, \dots, i_p$ and columns $i_1, i_2, \dots, i_p$ are deleted from $\mathbf{A}$, where $p < n$, then the resulting submatrix is called a principal submatrix of $\mathbf{A}$. In particular, if the deleted rows and columns are the last p rows and the last p columns, respectively, then such a submatrix is called a leading principal submatrix.

Definition 2.3.2. A partitioned matrix is a matrix that consists of several submatrices obtained by drawing horizontal and vertical lines that separate it into groups of rows and columns.

For example, the matrix

$$\mathbf{A} = \left[\begin{array}{cc|cc|c} 1 & \vdots & 0 & 3 & \vdots & 4 & -5 \\ 6 & \vdots & 2 & 10 & \vdots & 5 & 0 \\ \hline 3 & \vdots & 2 & 1 & \vdots & 0 & 2 \end{array} \right]$$

is partitioned into six submatrices by drawing one horizontal line and two vertical lines as shown above.

Definition 2.3.3. Let $\mathbf{A} = (a_{ij})$ be an $m_1 \times n_1$ matrix and $\mathbf{B}$ be an $m_2 \times n_2$ matrix. The direct (or Kronecker) product of $\mathbf{A}$ and $\mathbf{B}$, denoted by $\mathbf{A} \otimes \mathbf{B}$, is a matrix of order $m_1 m_2 \times n_1 n_2$ defined as a partitioned matrix of the form

$$\mathbf{A} \otimes \mathbf{B} = \left[\begin{array}{ccc|ccc} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n_1}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n_1}\mathbf{B} \\ \vdots & \vdots & & \vdots \\ a_{m_1 1}\mathbf{B} & a_{m_1 2}\mathbf{B} & \cdots & a_{m_1 n_1}\mathbf{B} \end{array} \right].$$

This matrix can be simplified by writing $\mathbf{A} \otimes \mathbf{B} = [a_{ij}\mathbf{B}]$. □

Properties of the direct product can be found in several matrix algebra books and papers. See, for example, Graybill (1983, Section 8.8), Henderson and Searle (1981), Magnus and Neudecker (1988, Chapter 2), and Searle (1982, Section 10.7). Some of these properties are listed below:

1. $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$.
2. $\mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C}$.
3. $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}$, if $\mathbf{AC}$ and $\mathbf{BD}$ are defined.
4. $\text{tr}(\mathbf{A} \otimes \mathbf{B}) = \text{tr}(\mathbf{A})\text{tr}(\mathbf{B})$, if $\mathbf{A}$ and $\mathbf{B}$ are square matrices.

The paper by Henderson, Pukelsheim, and Searle (1983) gives a detailed account of the history associated with direct products.

Definition 2.3.4. Let $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k$ be matrices of orders $m_i \times n_i$ ($i = 1, 2, \dots, k$). The direct sum of these matrices, denoted by $\bigoplus_{i=1}^k \mathbf{A}_i$, is a partitioned matrix of order $(\sum_{i=1}^k m_i) \times (\sum_{i=1}^k n_i)$ that has the block-diagonal form

$$\bigoplus_{i=1}^k \mathbf{A}_i = \text{Diag}(\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k).$$

The following properties can be easily shown on the basis of the preceding definition:

1. $\bigoplus_{i=1}^k \mathbf{A}_i + \bigoplus_{i=1}^k \mathbf{B}_i = \bigoplus_{i=1}^k (\mathbf{A}_i + \mathbf{B}_i)$, if $\mathbf{A}_i$ and $\mathbf{B}_i$ are of the same order for $i = 1, 2, \dots, k$.
2. $[\bigoplus_{i=1}^k \mathbf{A}_i][\bigoplus_{i=1}^k \mathbf{B}_i] = \bigoplus_{i=1}^k \mathbf{A}_i \mathbf{B}_i$, if $\mathbf{A}_i \mathbf{B}_i$ is defined for $i = 1, 2, \dots, k$.
3. $[\bigoplus_{i=1}^k \mathbf{A}_i]' = \bigoplus_{i=1}^k \mathbf{A}_i'$.
4. $\text{tr}(\bigoplus_{i=1}^k \mathbf{A}_i) = \sum_{i=1}^k \text{tr}(\mathbf{A}_i)$. $\square$

Definition 2.3.5. Let $\mathbf{A} = (a_{ij})$ be a square matrix of order $n \times n$. The determinant of $\mathbf{A}$, denoted by $\det(\mathbf{A})$, is a scalar quantity that can be computed iteratively as

$$\det(\mathbf{A}) = \sum_{j=1}^n (-1)^{j+1} a_{1j} \det(\mathbf{M}_{1j}), \quad (2.3)$$

where $\mathbf{M}_{1j}$ is a submatrix of $\mathbf{A}$ obtained by deleting row 1 and column j ($j = 1, 2, \dots, n$). For each j , the determinant of $\mathbf{M}_{1j}$ is obtained in terms of determinants of matrices of order $(n-2) \times (n-2)$ using a formula similar to (2.3). This process is repeated several times until the matrices on the right-hand side of (2.3) become of order 2×2 . The determinant of a 2×2 matrix such as $\mathbf{b} = (b_{ij})$ is given by $\det(\mathbf{b}) = b_{11}b_{22} - b_{12}b_{21}$. Thus by an iterative application of formula (2.3), the value of $\det(\mathbf{A})$ can be fully determined. For example, let $\mathbf{A}$ be the matrix

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

Then $\det(\mathbf{A}) = \det(\mathbf{A}_1) - 2\det(\mathbf{A}_2) - \det(\mathbf{A}_3)$, where $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ are 2×2 submatrices, namely

$$\mathbf{A}_1 = \begin{bmatrix} 0 & 3 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 5 & 3 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{A}_3 = \begin{bmatrix} 5 & 0 \\ 1 & 2 \end{bmatrix}.$$

It follows that $\det(\mathbf{A}) = -6 - 2(2) - 10 = -20$. $\square$

Definition 2.3.6. Let $\mathbf{A} = (a_{ij})$ be a square matrix order of $n \times n$. The determinant of $\mathbf{M}_{ij}$, the submatrix obtained by deleting row i and column j , is called a minor of $\mathbf{A}$ of order $n-1$. The quantity $(-1)^{i+j} \det(\mathbf{M}_{ij})$ is called a cofactor of the corresponding (i, j) th element of $\mathbf{A}$. More generally, if $\mathbf{A}$ is an $m \times n$ matrix and if we strike out all but p rows and the same number of columns from $\mathbf{A}$, where $p \leq \min(m, n)$, then the determinant of the resulting submatrix is called a minor of $\mathbf{A}$ of order p .

The determinant of a principal submatrix of a square matrix $\mathbf{A}$ is called a principal minor. If, however, we have a leading principal submatrix, then its determinant is called a leading principal minor. $\square$

NOTE 2.3.1. The determinant of a matrix $\mathbf{A}$ is defined only when $\mathbf{A}$ is a square matrix.

NOTE 2.3.2. The expansion of $\det(\mathbf{A})$ in (2.3) was carried out by multiplying the elements of the first row of $\mathbf{A}$ by their corresponding cofactors and then summing over j ($= 1, 2, \dots, n$). The same value of $\det(\mathbf{A})$ could have also been obtained by similar expansions according to the elements of any row of $\mathbf{A}$ (instead of the first row), or any column of $\mathbf{A}$. Thus if $\mathbf{M}_{ij}$ is a submatrix of $\mathbf{A}$ obtained by deleting row i and column j , then $\det(\mathbf{A})$ can be obtained by using any of the following expansions:

$$\text{By row } i: \quad \det(\mathbf{A}) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{M}_{ij}), \quad i = 1, 2, \dots, n.$$

$$\text{By column } j: \quad \det(\mathbf{A}) = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(\mathbf{M}_{ij}), \quad j = 1, 2, \dots, n.$$

NOTE 2.3.3. Some of the properties of determinants are the following:

- i. $\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B})$, if $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices.
- ii. If $\mathbf{A}'$ is the transpose of $\mathbf{A}$, then $\det(\mathbf{A}') = \det(\mathbf{A})$.
- iii. If $\mathbf{A}$ is an $n \times n$ matrix and α is a scalar, then $\det(\alpha\mathbf{A}) = \alpha^n \det(\mathbf{A})$.
- iv. If any two rows (or columns) of $\mathbf{A}$ are identical, then $\det(\mathbf{A}) = 0$.
- v. If any two rows (or columns) of $\mathbf{A}$ are interchanged, then $\det(\mathbf{A})$ is multiplied by -1 .
- vi. If $\det(\mathbf{A}) = 0$, then $\mathbf{A}$ is called a singular matrix. Otherwise, $\mathbf{A}$ is a nonsingular matrix.
- vii. If $\mathbf{A}$ and $\mathbf{B}$ are matrices of orders $m \times m$ and $n \times n$, respectively, then the following hold: (a) $\det(\mathbf{A} \otimes \mathbf{B}) = [\det(\mathbf{A})]^n [\det(\mathbf{B})]^m$; (b) $\det(\mathbf{A} \oplus \mathbf{B}) = [\det(\mathbf{A})][\det(\mathbf{B})]$.

NOTE 2.3.4. The history of determinants dates back to the fourteenth century. According to Smith (1958, page 273), the Chinese had some knowledge of determinants as early as about 1300 A.D. Smith (1958, page 440) also reported that the Japanese mathematician Seki Kōwa (1642–1708) had discovered the expansion of a determinant in solving simultaneous equations. In the West, the theory of determinants is believed to have originated with the German mathematician Gottfried Leibniz (1646–1716) in 1693, ten years

after the work of Seki Kōwa. However, the actual development of the theory of determinants did not begin until the publication of a book by Gabriel Cramer (1704–1752) (see Price, 1947, page 85) in 1750. Other mathematicians who contributed to this theory include Alexandre Vandermonde (1735–1796), Pierre-Simon Laplace (1749–1827), Carl Gauss (1777–1855), and Augustin-Louis Cauchy (1789–1857). Arthur Cayley (1821–1895) is credited with having been the first to introduce the common present-day notation of vertical bars enclosing a square matrix. For more interesting facts about the history of determinants, the reader is advised to read the article by Price (1947).

2.3.2. The Rank of a Matrix

Let $\mathbf{A} = (a_{ij})$ be a matrix of order $m \times n$. Let $\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_m$ denote the row vectors of $\mathbf{A}$, and let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ denote its column vectors. Consider the linear spans of the row and column vectors, namely, $V_1 = L(\mathbf{u}'_1, \mathbf{u}'_2, \dots, \mathbf{u}'_m)$, $V_2 = L(\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n)$, respectively.

Theorem 2.3.1. The vector spaces V_1 and V_2 have the same dimension.

Proof. See Lancaster (1969, Theorem 1.15.1), or Searle (1982, Section 6.6). $\square$

Thus, for any matrix $\mathbf{A}$, the number of linearly independent rows is the same as the number of linearly independent columns.

Definition 2.3.7. The rank of a matrix $\mathbf{A}$ is the number of its linearly independent rows (or columns). The rank of $\mathbf{A}$ is denoted by $r(\mathbf{A})$. $\square$

Theorem 2.3.2. If a matrix $\mathbf{A}$ has a nonzero minor of order r , and if all minors of order $r + 1$ and higher (if they exist) are zero, then $\mathbf{A}$ has rank r .

Proof. See Lancaster (1969, Lemma 1, Section 1.15). $\square$

For example, if $\mathbf{A}$ is the matrix

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

then $r(\mathbf{A}) = 2$. This is because $\det(\mathbf{A}) = 0$ and at least one minor of order 2 is different from zero.

There are several properties associated with the rank of a matrix. Some of these properties are the following:

1. $r(\mathbf{A}) = r(\mathbf{A}')$.
2. The rank of $\mathbf{A}$ is unchanged if $\mathbf{A}$ is multiplied by a nonsingular matrix. Thus if $\mathbf{A}$ is an $m \times n$ matrix and $\mathbf{P}$ is an $n \times n$ nonsingular matrix, then $r(\mathbf{A}) = r(\mathbf{AP})$.
3. $r(\mathbf{A}) = r(\mathbf{AA}') = r(\mathbf{A}'\mathbf{A})$.
4. If the matrix $\mathbf{A}$ is partitioned as $\mathbf{A} = [\mathbf{A}_1 : \mathbf{A}_2]$, where $\mathbf{A}_1$ and $\mathbf{A}_2$ are submatrices of the same order, then $r(\mathbf{A}_1 + \mathbf{A}_2) \leq r(\mathbf{A}) \leq r(\mathbf{A}_1) + r(\mathbf{A}_2)$. More generally, if the matrices $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k$ are of the same order and if $\mathbf{A}$ is partitioned as $\mathbf{A} = [\mathbf{A}_1 : \mathbf{A}_2 : \dots : \mathbf{A}_k]$, then

$$r\left(\sum_{i=1}^k \mathbf{A}_i\right) \leq r(\mathbf{A}) \leq \sum_{i=1}^k r(\mathbf{A}_i).$$

5. If the product $\mathbf{AB}$ is defined, then $r(\mathbf{A}) + r(\mathbf{B}) - n \leq r(\mathbf{AB}) \leq \min\{r(\mathbf{A}), r(\mathbf{B})\}$, where n is the number of columns of $\mathbf{A}$ (or the number of rows of $\mathbf{B}$).
6. $r(\mathbf{A} \otimes \mathbf{B}) = r(\mathbf{A})r(\mathbf{B})$.
7. $r(\mathbf{A} \oplus \mathbf{B}) = r(\mathbf{A}) + r(\mathbf{B})$.

Definition 2.3.8. Let $\mathbf{A}$ be a matrix of order $m \times n$ and rank r . Then we have the following:

1. $\mathbf{A}$ is said to have a full row rank if $r = m < n$.
2. $\mathbf{A}$ is said to have a full column rank if $r = n < m$.
3. $\mathbf{A}$ is of full rank if $r = m = n$. In this case, $\det(\mathbf{A}) \neq 0$, that is, $\mathbf{A}$ is a nonsingular matrix. $\square$

2.3.3. The Inverse of a Matrix

Let $\mathbf{A} = (a_{ij})$ be a nonsingular matrix of order $n \times n$. The inverse of $\mathbf{A}$, denoted by $\mathbf{A}^{-1}$, is an $n \times n$ matrix that satisfies the condition $\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_n$.

The inverse of $\mathbf{A}$ can be computed as follows: Let c_{ij} be the cofactor of a_{ij} (see Definition 2.3.6). Define the matrix $\mathbf{C}$ as $\mathbf{C} = (c_{ij})$. The transpose of $\mathbf{C}$ is called the adjugate or adjoint of $\mathbf{A}$ and is denoted by $\text{adj } \mathbf{A}$. The inverse of $\mathbf{A}$ is then given by

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

It can be verified that

$$\mathbf{A} \left[\frac{\text{adj } \mathbf{A}}{\det(\mathbf{A})} \right] = \left[\frac{\text{adj } \mathbf{A}}{\det(\mathbf{A})} \right] \mathbf{A} = \mathbf{I}_n.$$

For example, if $\mathbf{A}$ is the matrix

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

then $\det(\mathbf{A}) = -3$, and

$$\text{adj } \mathbf{A} = \begin{bmatrix} 2 & 1 & -2 \\ 3 & 0 & -3 \\ -7 & -2 & 4 \end{bmatrix}.$$

Hence,

$$\mathbf{A}^{-1} = \begin{bmatrix} -\frac{2}{3} & -\frac{1}{3} & \frac{2}{3} \\ -1 & 0 & 1 \\ \frac{7}{3} & \frac{2}{3} & -\frac{4}{3} \end{bmatrix}.$$

Some properties of the inverse operation are given below:

1. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
2. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$.
3. $\det(\mathbf{A}^{-1}) = 1/\det(\mathbf{A})$.
4. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.
5. $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$.
6. $(\mathbf{A} \oplus \mathbf{B})^{-1} = \mathbf{A}^{-1} \oplus \mathbf{B}^{-1}$.
7. If $\mathbf{A}$ is partitioned as

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

where $\mathbf{A}_{ij}$ is of order $n_i \times n_j$ ($i, j = 1, 2$), then

$$\det(\mathbf{A}) = \begin{cases} \det(\mathbf{A}_{11}) \cdot \det(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}) & \text{if } \mathbf{A}_{11} \text{ is nonsingular,} \\ \det(\mathbf{A}_{22}) \cdot \det(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}) & \text{if } \mathbf{A}_{22} \text{ is nonsingular.} \end{cases}$$

The inverse of $\mathbf{A}$ is partitioned as

$$\mathbf{A}^{-1} = \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix},$$

where

$$\mathbf{B}_{11} = (\mathbf{A}_{11} - \mathbf{A}_{12} \mathbf{A}_{22}^{-1} \mathbf{A}_{21})^{-1},$$

$$\mathbf{B}_{12} = -\mathbf{B}_{11} \mathbf{A}_{12} \mathbf{A}_{22}^{-1},$$

$$\mathbf{B}_{21} = -\mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{B}_{11},$$

$$\mathbf{B}_{22} = \mathbf{A}_{22}^{-1} + \mathbf{A}_{22}^{-1} \mathbf{A}_{21} \mathbf{B}_{11} \mathbf{A}_{12} \mathbf{A}_{22}^{-1}.$$

2.3.4. Generalized Inverse of a Matrix

This inverse represents a more general concept than the one discussed in the previous section. Let $\mathbf{A}$ be a matrix of order $m \times n$. Then, a generalized inverse of $\mathbf{A}$, denoted by $\mathbf{A}^-$, is a matrix of order $n \times m$ that satisfies the condition

$$\mathbf{A} \mathbf{A}^- \mathbf{A} = \mathbf{A}. \quad (2.4)$$

Note that $\mathbf{A}^-$ is defined even if $\mathbf{A}$ is not a square matrix. If $\mathbf{A}$ is a square matrix, it does not have to be nonsingular. Furthermore, condition (2.4) can be satisfied by infinitely many matrices (see, for example, Searle, 1982, Chapter 8). If $\mathbf{A}$ is nonsingular, then (2.4) is satisfied by only $\mathbf{A}^{-1}$. Thus $\mathbf{A}^{-1}$ is a special case of $\mathbf{A}^-$.

Theorem 2.3.3.

1. If $\mathbf{A}$ is a symmetric matrix, then $\mathbf{A}^-$ can be chosen to be symmetric.
2. $\mathbf{A}(\mathbf{A}'\mathbf{A})^- \mathbf{A}'\mathbf{A} = \mathbf{A}$ for any matrix $\mathbf{A}$.
3. $\mathbf{A}(\mathbf{A}'\mathbf{A})^- \mathbf{A}'$ is invariant to the choice of a generalized inverse of $\mathbf{A}'\mathbf{A}$.

Proof. See Searle (1982, pages 221–222). $\square$

2.3.5. Eigenvalues and Eigenvectors of a Matrix

Let $\mathbf{A}$ be a square matrix of order $n \times n$. By definition, a scalar λ is said to be an eigenvalue (or characteristic root) of $\mathbf{A}$ if $\mathbf{A} - \lambda \mathbf{I}_n$ is a singular matrix, that is,

$$\det(\mathbf{A} - \lambda \mathbf{I}_n) = 0. \quad (2.5)$$

Thus an eigenvalue of $\mathbf{A}$ satisfies a polynomial equation of degree n called the characteristic equation of $\mathbf{A}$. If λ is a multiple solution (or root) of equation (2.5), that is, (2.5) has several roots, say m , that are equal to λ , then λ is said to be an eigenvalue of multiplicity m .

Since $r(\mathbf{A} - \lambda \mathbf{I}_n) < n$ by the fact that $\mathbf{A} - \lambda \mathbf{I}_n$ is singular, the columns of $\mathbf{A} - \lambda \mathbf{I}_n$ must be linearly related. Hence, there exists a nonzero vector $\mathbf{v}$ such that

$$(\mathbf{A} - \lambda \mathbf{I}_n) \mathbf{v} = \mathbf{0}, \quad (2.6)$$

or equivalently,

$$\mathbf{A} \mathbf{v} = \lambda \mathbf{v}. \quad (2.7)$$

A vector satisfying (2.7) is called an eigenvector (or a characteristic vector) corresponding to the eigenvalue λ . From (2.7) we note that the linear transformation of $\mathbf{v}$ by the matrix $\mathbf{A}$ is a scalar multiple of $\mathbf{v}$.

The following theorems describe certain properties associated with eigenvalues and eigenvectors. The proofs of these theorems can be found in standard matrix algebra books (see the annotated bibliography).

Theorem 2.3.4. A square matrix $\mathbf{A}$ is singular if and only if at least one of its eigenvalues is equal to zero. In particular, if $\mathbf{A}$ is symmetric, then its rank is equal to the number of its nonzero eigenvalues.

Theorem 2.3.5. The eigenvalues of a symmetric matrix are real.

Theorem 2.3.6. Let $\mathbf{A}$ be a square matrix, and let $\lambda_1, \lambda_2, \dots, \lambda_k$ denote its distinct eigenvalues. If $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are eigenvectors of $\mathbf{A}$ corresponding to $\lambda_1, \lambda_2, \dots, \lambda_k$, respectively, then $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are linearly independent. In particular, if $\mathbf{A}$ is symmetric, then $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ are orthogonal to one another, that is, $\mathbf{v}_i' \mathbf{v}_j = 0$ for $i \neq j$ ($i, j = 1, 2, \dots, k$).

Theorem 2.3.7. Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices of orders $m \times m$ and $n \times n$, respectively. Let $\lambda_1, \lambda_2, \dots, \lambda_m$ be the eigenvalues of $\mathbf{A}$, and $\nu_1, \nu_2, \dots, \nu_n$ be the eigenvalues of $\mathbf{B}$. Then we have the following:

1. The eigenvalues of $\mathbf{A} \otimes \mathbf{B}$ are of the form $\lambda_i \nu_j$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$).
2. The eigenvalues of $\mathbf{A} \oplus \mathbf{B}$ are $\lambda_1, \lambda_2, \dots, \lambda_m; \nu_1, \nu_2, \dots, \nu_n$.

Theorem 2.3.8. Let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the eigenvalues of a matrix $\mathbf{A}$ of order $n \times n$. Then the following hold:

1. $\text{tr}(\mathbf{A}) = \sum_{i=1}^n \lambda_i$.
2. $\det(\mathbf{A}) = \prod_{i=1}^n \lambda_i$.

Theorem 2.3.9. Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices of orders $m \times n$ and $n \times m$ ($n \geq m$), respectively. The nonzero eigenvalues of $\mathbf{BA}$ are the same as those of $\mathbf{AB}$.

2.3.6. Some Special Matrices

1. The vector $\mathbf{1}_n$ is a column vector of ones of order $n \times 1$.
2. The matrix $\mathbf{J}_n$ is a matrix of ones of order $n \times n$.
3. *Idempotent Matrix.* A square matrix $\mathbf{A}$ for which $\mathbf{A}^2 = \mathbf{A}$ is called an idempotent matrix. For example, the matrix $\mathbf{A} = \mathbf{I}_n - (1/n)\mathbf{J}_n$ is idempotent of order $n \times n$. The eigenvalues of an idempotent matrix are equal to zeros and ones. It follows from Theorem 2.3.8 that the rank of an idempotent matrix, which is the same as the number of eigenvalues that are equal to 1, is also equal to its trace. Idempotent matrices are used in many applications in statistics (see Section 2.4).
4. *Orthogonal Matrix.* A square matrix $\mathbf{A}$ is orthogonal if $\mathbf{A}'\mathbf{A} = \mathbf{I}$. From this definition it follows that (i) $\mathbf{A}$ is orthogonal if and only if $\mathbf{A}' = \mathbf{A}^{-1}$; (ii) $|\det(\mathbf{A})| = 1$. A special orthogonal matrix is the Householder matrix, which is a symmetric matrix of the form

$$\mathbf{H} = \mathbf{I} - 2\mathbf{u}\mathbf{u}'/\mathbf{u}'\mathbf{u},$$

where $\mathbf{u}$ is a nonzero vector. Orthogonal matrices occur in many applications of matrix algebra and play an important role in statistics, as will be seen in Section 2.4.

2.3.7. The Diagonalization of a Matrix

Theorem 2.3.10 (The Spectral Decomposition Theorem). Let $\mathbf{A}$ be a symmetric matrix of order $n \times n$. There exists an orthogonal matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$, where $\mathbf{\Lambda} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ is a diagonal matrix whose diagonal elements are the eigenvalues of $\mathbf{A}$. The columns of $\mathbf{P}$ are the corresponding orthonormal eigenvectors of $\mathbf{A}$.

Proof. See Basilevsky (1983, Theorem 5.8, page 200). □

If $\mathbf{P}$ is partitioned as $\mathbf{P} = [\mathbf{p}_1 : \mathbf{p}_2 : \dots : \mathbf{p}_n]$, where $\mathbf{p}_i$ is an eigenvector of $\mathbf{A}$ with eigenvalue λ_i ($i = 1, 2, \dots, n$), then $\mathbf{A}$ can be written as

$$\mathbf{A} = \sum_{i=1}^n \lambda_i \mathbf{p}_i \mathbf{p}_i'.$$

For example, if

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

then $\mathbf{A}$ has two distinct eigenvalues, $\lambda_1 = 0$ of multiplicity 2 and $\lambda_2 = 5$. For $\lambda_1 = 0$ we have two orthonormal eigenvectors, $\mathbf{p}_1 = (2, 0, 1)' / \sqrt{5}$ and $\mathbf{p}_2 = (0, 1, 0)'$. Note that $\mathbf{p}_1$ and $\mathbf{p}_2$ span the kernel (null space) of the linear transformation represented by $\mathbf{A}$. For $\lambda_2 = 5$ we have the normal eigenvector $\mathbf{p}_3 = (1, 0, -2)' / \sqrt{5}$, which is orthogonal to both $\mathbf{p}_1$ and $\mathbf{p}_2$. Hence, $\mathbf{P}$ and $\mathbf{\Lambda}$ in Theorem 2.3.10 for the matrix $\mathbf{A}$ are

$$\mathbf{P} = \begin{bmatrix} \frac{2}{\sqrt{5}} & 0 & \frac{1}{\sqrt{5}} \\ 0 & 1 & 0 \\ \frac{1}{\sqrt{5}} & 0 & \frac{-2}{\sqrt{5}} \end{bmatrix},$$

$$\mathbf{\Lambda} = \text{Diag}(0, 0, 5).$$

The next theorem gives a more general form of the spectral decomposition theorem.

Theorem 2.3.11 (The Singular-Value Decomposition Theorem). Let $\mathbf{A}$ be a matrix of order $m \times n$ ($m \leq n$) and rank r . There exist orthogonal matrices $\mathbf{P}$ and $\mathbf{Q}$ such that $\mathbf{A} = \mathbf{P}[\mathbf{D} : \mathbf{0}]\mathbf{Q}'$, where $\mathbf{D} = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ is a diagonal matrix with nonnegative diagonal elements called the singular values of $\mathbf{A}$, and $\mathbf{0}$ is a zero matrix of order $m \times (n - m)$. The diagonal elements of $\mathbf{D}$ are the square roots of the eigenvalues of $\mathbf{A}\mathbf{A}'$.

Proof. See, for example, Searle (1982, pages 316–317). $\square$

2.3.8. Quadratic Forms

Let $\mathbf{A} = (a_{ij})$ be a symmetric matrix of order $n \times n$, and let $\mathbf{x} = (x_1, x_2, \dots, x_n)'$ be a column vector of order $n \times 1$. The function

$$\begin{aligned} q(\mathbf{x}) &= \mathbf{x}'\mathbf{A}\mathbf{x} \\ &= \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_i x_j \end{aligned}$$

is called a quadratic form in $\mathbf{x}$.

A quadratic form $\mathbf{x}'\mathbf{A}\mathbf{x}$ is said to be the following:

1. Positive definite if $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$ and is zero only if $\mathbf{x} = \mathbf{0}$.
2. Positive semidefinite if $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x}$ and $\mathbf{x}'\mathbf{A}\mathbf{x} = 0$ for at least one nonzero value of $\mathbf{x}$.
3. Nonnegative definite if $\mathbf{A}$ is either positive definite or positive semidefinite.

Theorem 2.3.12. Let $\mathbf{A} = (a_{ij})$ be a symmetric matrix of order $n \times n$. Then $\mathbf{A}$ is positive definite if and only if either of the following two conditions is satisfied:

1. The eigenvalues of $\mathbf{A}$ are all positive.
2. The leading principal minors of $\mathbf{A}$ are all positive, that is,

$$a_{11} > 0, \quad \det \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} > 0, \dots, \quad \det(\mathbf{A}) > 0.$$

Proof. The proof of part 1 follows directly from the spectral decomposition theorem. For the proof of part 2, see Lancaster (1969, Theorem 2.14.4). $\square$

Theorem 2.3.13. Let $\mathbf{A} = (a_{ij})$ be a symmetric matrix of order $n \times n$. Then $\mathbf{A}$ is positive semidefinite if and only if its eigenvalues are nonnegative with at least one of them equal to zero.

Proof. See Basilevsky (1983, Theorem 5.10, page 203). $\square$

2.3.9. The Simultaneous Diagonalization of Matrices

By simultaneous diagonalization we mean finding a matrix, say $\mathbf{Q}$, that can reduce several square matrices to a diagonal form. In many situations there may be a need to diagonalize several matrices simultaneously. This occurs frequently in statistics, particularly in analysis of variance.

The proofs of the following theorems can be found in Graybill (1983, Chapter 12).

Theorem 2.3.14. Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrices of order $n \times n$.

1. If $\mathbf{A}$ is positive definite, then there exists a nonsingular matrix $\mathbf{Q}$ such that $\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{I}_n$ and $\mathbf{Q}'\mathbf{B}\mathbf{Q} = \mathbf{D}$, where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are the roots of the polynomial equation $\det(\mathbf{B} - \lambda\mathbf{A}) = 0$.

2. If $\mathbf{A}$ and $\mathbf{B}$ are positive semidefinite, then there exists a nonsingular matrix $\mathbf{Q}$ such that

$$\mathbf{Q}'\mathbf{A}\mathbf{Q} = \mathbf{D}_1,$$

$$\mathbf{Q}'\mathbf{B}\mathbf{Q} = \mathbf{D}_2,$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrices (for a detailed proof of this result, see Newcomb, 1960).

Theorem 2.3.15. Let $\mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_k$ be symmetric matrices of order $n \times n$. Then there exists an orthogonal matrix $\mathbf{P}$ such that

$$\mathbf{A}_i = \mathbf{P}\mathbf{\Lambda}_i\mathbf{P}', \quad i = 1, 2, \dots, k,$$

where $\mathbf{\Lambda}_i$ is a diagonal matrix, if and only if $\mathbf{A}_i\mathbf{A}_j = \mathbf{A}_j\mathbf{A}_i$ for all $i \neq j$ ($i, j = 1, 2, \dots, k$).

2.3.10. Bounds on Eigenvalues

Let $\mathbf{A}$ be a symmetric matrix of order $n \times n$. We denote the i th eigenvalue of $\mathbf{A}$ by $e_i(\mathbf{A})$, $i = 1, 2, \dots, n$. The smallest and largest eigenvalues of $\mathbf{A}$ are denoted by $e_{\min}(\mathbf{A})$ and $e_{\max}(\mathbf{A})$, respectively.

Theorem 2.3.16. $e_{\min}(\mathbf{A}) \leq \mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{x} \leq e_{\max}(\mathbf{A})$.

Proof. This follows directly from the spectral decomposition theorem. $\square$

The ratio $\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{x}$ is called Rayleigh's quotient for $\mathbf{A}$. The lower and upper bounds in Theorem 2.3.16 can be achieved by choosing $\mathbf{x}$ to be an eigenvector associated with $e_{\min}(\mathbf{A})$ and $e_{\max}(\mathbf{A})$, respectively. Thus Theorem 2.3.16 implies that

$$\inf_{\mathbf{x} \neq \mathbf{0}} \left[\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right] = e_{\min}(\mathbf{A}), \quad (2.8)$$

$$\sup_{\mathbf{x} \neq \mathbf{0}} \left[\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right] = e_{\max}(\mathbf{A}). \quad (2.9)$$

Theorem 2.3.17. If $\mathbf{A}$ is a symmetric matrix and $\mathbf{B}$ is a positive definite matrix, both of order $n \times n$, then

$$e_{\min}(\mathbf{B}^{-1}\mathbf{A}) \leq \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} \leq e_{\max}(\mathbf{B}^{-1}\mathbf{A})$$

Proof. The proof is left to the reader. $\square$

Note that the above lower and upper bounds are equal to the infimum and supremum, respectively, of the ratio $\mathbf{x}'\mathbf{A}\mathbf{x}/\mathbf{x}'\mathbf{B}\mathbf{x}$ for $\mathbf{x} \neq \mathbf{0}$.

Theorem 2.3.18. If $\mathbf{A}$ is a positive semidefinite matrix and $\mathbf{B}$ is a positive definite matrix, both of order $n \times n$, then for any i ($i = 1, 2, \dots, n$),

$$e_i(\mathbf{A})e_{\min}(\mathbf{B}) \leq e_i(\mathbf{AB}) \leq e_i(\mathbf{A})e_{\max}(\mathbf{B}). \quad (2.10)$$

Furthermore, if $\mathbf{A}$ is positive definite, then for any i ($i = 1, 2, \dots, n$),

$$\frac{e_i^2(\mathbf{AB})}{e_{\max}(\mathbf{A})e_{\max}(\mathbf{B})} \leq e_i(\mathbf{A})e_i(\mathbf{B}) \leq \frac{e_i^2(\mathbf{AB})}{e_{\min}(\mathbf{A})e_{\min}(\mathbf{B})}$$

Proof. See Anderson and Gupta (1963, Corollary 2.2.1). □

A special case of the double inequality in (2.10) is

$$e_{\min}(\mathbf{A})e_{\min}(\mathbf{B}) \leq e_i(\mathbf{AB}) \leq e_{\max}(\mathbf{A})e_{\max}(\mathbf{B}),$$

for all i ($i = 1, 2, \dots, n$).

Theorem 2.3.19. Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrices of order $n \times n$. Then, the following hold:

1. $e_i(\mathbf{A}) \leq e_i(\mathbf{A} + \mathbf{B})$, $i = 1, 2, \dots, n$, if $\mathbf{B}$ is nonnegative definite.
2. $e_i(\mathbf{A}) < e_i(\mathbf{A} + \mathbf{B})$, $i = 1, 2, \dots, n$, if $\mathbf{B}$ is positive definite.

Proof. See Bellman (1970, Theorem 3, page 117). □

Theorem 2.3.20 (Schur's Theorem). Let $\mathbf{A} = (a_{ij})$ be a symmetric matrix of order $n \times n$, and let $\|\mathbf{A}\|_2$ denote its Euclidean norm, defined as

$$\|\mathbf{A}\|_2 = \left(\sum_{i=1}^n \sum_{j=1}^n a_{ij}^2 \right)^{1/2}.$$

Then

$$\sum_{i=1}^n e_i^2(\mathbf{A}) = \|\mathbf{A}\|_2^2.$$

Proof. See Lancaster (1969, Theorem 7.3.1). □

Since $\|\mathbf{A}\|_2 \leq n \max_{i,j} |a_{ij}|$, then from Theorem 2.3.20 we conclude that

$$|e_{\max}(\mathbf{A})| \leq n \max_{i,j} |a_{ij}|.$$

Theorem 2.3.21. Let $\mathbf{A}$ be a symmetric matrix of order $n \times n$, and let m and s be defined as

$$m = \frac{\text{tr}(\mathbf{A})}{n}, \quad s = \left(\frac{\text{tr}(\mathbf{A}^2)}{n} - m^2 \right)^{1/2}.$$

Then

$$\begin{aligned} m - s(n-1)^{1/2} &\leq e_{\min}(\mathbf{A}) \leq m - \frac{s}{(n-1)^{1/2}}, \\ m + \frac{s}{(n-1)^{1/2}} &\leq e_{\max}(\mathbf{A}) \leq m + s(n-1)^{1/2}, \\ e_{\max}(\mathbf{A}) - e_{\min}(\mathbf{A}) &\leq s(2n)^{1/2}. \end{aligned}$$

Proof. See Wolkowicz and Styan (1980, Theorems 2.1 and 2.5). $\square$

2.4. APPLICATIONS OF MATRICES IN STATISTICS

The use of matrix algebra is quite prevalent in statistics. In fact, in the areas of experimental design, linear models, and multivariate analysis, matrix algebra is considered the most frequently used branch of mathematics. Applications of matrices in these areas are well documented in several books, for example, Basilevsky (1983), Graybill (1983), Magnus and Neudecker (1988), and Searle (1982). We shall therefore not attempt to duplicate the material given in these books.

Let us consider the following applications:

2.4.1. The Analysis of the Balanced Mixed Model

In analysis of variance, a linear model associated with a given experimental situation is said to be balanced if the numbers of observations in the subclasses of the data are the same. For example, the two-way crossed-classification model with interaction,

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}, \quad (2.11)$$

$i = 1, 2, \dots, a$; $j = 1, 2, \dots, b$; $k = 1, 2, \dots, n$, is balanced, since there are n observations for each combination of i and j . Here, α_i and β_j represent the main effects of the factors under consideration, $(\alpha\beta)_{ij}$ denotes the interaction effect, and ϵ_{ijk} is a random error term. Model (2.11) can be written in vector form as

$$\mathbf{y} = \mathbf{H}_0\tau_0 + \mathbf{H}_1\tau_1 + \mathbf{H}_2\tau_2 + \mathbf{H}_3\tau_3 + \mathbf{H}_4\tau_4, \quad (2.12)$$

where $\mathbf{y}$ is the vector of observations, $\tau_0 = \mu$, $\tau_1 = (\alpha_1, \alpha_2, \dots, \alpha_a)'$, $\tau_2 = (\beta_1, \beta_2, \dots, \beta_b)'$, $\tau_3 = [(\alpha\beta)_{11}, (\alpha\beta)_{12}, \dots, (\alpha\beta)_{ab}]'$, and $\tau_4 = (\epsilon_{111}, \epsilon_{112}, \dots, \epsilon_{abn})'$. The matrices $\mathbf{H}_i$ ($i = 0, 1, 2, 3, 4$) can be expressed as direct products of the form

$$\begin{aligned}\mathbf{H}_0 &= \mathbf{1}_a \otimes \mathbf{1}_b \otimes \mathbf{1}_n, \\ \mathbf{H}_1 &= \mathbf{I}_a \otimes \mathbf{1}_b \otimes \mathbf{1}_n, \\ \mathbf{H}_2 &= \mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_n, \\ \mathbf{H}_3 &= \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_n, \\ \mathbf{H}_4 &= \mathbf{I}_a \otimes \mathbf{I}_b \otimes \mathbf{I}_n.\end{aligned}$$

In general, any balanced linear model can be written in vector form as

$$\mathbf{y} = \sum_{l=0}^{\nu} \mathbf{H}_l \tau_l, \quad (2.13)$$

where $\mathbf{H}_l$ ($l = 0, 1, \dots, \nu$) is a direct product of identity matrices and vectors of ones (see Khuri, 1982). If $\tau_0, \tau_1, \dots, \tau_{\theta}$ ($\theta < \nu - 1$) are fixed unknown parameter vectors (fixed effects), and $\tau_{\theta+1}, \tau_{\theta+2}, \dots, \tau_{\nu}$ are random vectors (random effects), then model (2.11) is called a balanced mixed model. Furthermore, if we assume that the random effects are independent and have the normal distributions $N(\mathbf{0}, \sigma_l^2 \mathbf{I}_{c_l})$, where c_l is the number of columns of $\mathbf{H}_l$, $l = \theta + 1, \theta + 2, \dots, \nu$, then, because model (2.11) is balanced, its statistical analysis becomes very simple. Here, the σ_l^2 's are called the model's variance components. A balanced mixed model can be written as

$$\mathbf{y} = \mathbf{Xg} + \mathbf{Zh} \quad (2.14)$$

where $\mathbf{Xg} = \sum_{l=0}^{\theta} \mathbf{H}_l \tau_l$ is the fixed portion of the model, and $\mathbf{Zh} = \sum_{l=\theta+1}^{\nu} \mathbf{H}_l \tau_l$ is its random portion. The variance-covariance matrix of $\mathbf{y}$ is given by

$$\Sigma = \sum_{l=\theta+1}^{\nu} \mathbf{A}_l \sigma_l^2,$$

where $\mathbf{A}_l = \mathbf{H}_l \mathbf{H}_l'$ ($l = \theta + 1, \theta + 2, \dots, \nu$). Note that $\mathbf{A}_l \mathbf{A}_p = \mathbf{A}_p \mathbf{A}_l$ for all $l \neq p$. Hence, the matrices $\mathbf{A}_l$ can be diagonalized simultaneously (see Theorem 2.3.15).

If $\mathbf{y}'\mathbf{A}\mathbf{y}$ is a quadratic form in $\mathbf{y}$, then $\mathbf{y}'\mathbf{A}\mathbf{y}$ is distributed as a noncentral chi-squared variate $\chi_m'^2(\eta)$ if and only if $\mathbf{A}\Sigma$ is idempotent of rank m , where η is the noncentrality parameter and is given by $\eta = \mathbf{g}'\mathbf{X}'\mathbf{A}\mathbf{Xg}$ (see Searle, 1971, Section 2.5).

The total sum of squares, $\mathbf{y}'\mathbf{y}$, can be uniquely partitioned as

$$\mathbf{y}'\mathbf{y} = \sum_{l=0}^{\nu} \mathbf{y}'\mathbf{P}_l \mathbf{y},$$

where the $\mathbf{P}_l$'s are idempotent matrices such that $\mathbf{P}_l \mathbf{P}_s = 0$ for all $l \neq s$ (see Khuri, 1982). The quadratic form $\mathbf{y}'\mathbf{P}_l\mathbf{y}$ ($l = 0, 1, \dots, \nu$) is positive semidefinite and represents the sum of squares for the l th effect in model (2.13).

Theorem 2.4.1. Consider the balanced mixed model (2.14), where the random effects are assumed to be independently and normally distributed with zero means and variance-covariance matrices $\sigma_l^2 \mathbf{I}_{c_l}$ ($l = \theta + 1, \theta + 2, \dots, \nu$). Then we have the following:

1. $\mathbf{y}'\mathbf{P}_0\mathbf{y}, \mathbf{y}'\mathbf{P}_1\mathbf{y}, \dots, \mathbf{y}'\mathbf{P}_\nu\mathbf{y}$ are statistically independent.
2. $\mathbf{y}'\mathbf{P}_l\mathbf{y}/\delta_l$ is distributed as a noncentral chi-squared variate with degrees of freedom equal to the rank of $\mathbf{P}_l$ and noncentrality parameter given by $\eta_l = \mathbf{g}'\mathbf{X}'\mathbf{P}_l\mathbf{X}\mathbf{g}/\delta_l$ for $l = 0, 1, \dots, \theta$, where δ_l is a particular linear combination of the variance components $\sigma_{\theta+1}^2, \sigma_{\theta+2}^2, \dots, \sigma_\nu^2$. However, for $l = \theta + 1, \theta + 2, \dots, \nu$, that is, for the random effects, $\mathbf{y}'\mathbf{P}_l\mathbf{y}/\delta_l$ is distributed as a central chi-squared variate with m_l degrees of freedom, where $m_l = r(\mathbf{P}_l)$.

Proof. See Theorem 4.1 in Khuri (1982). $\square$

Theorem 2.4.1 provides the basis for a complete analysis of any balanced mixed model, as it can be used to obtain exact tests for testing the significance of the fixed effects and the variance components.

A linear function $\mathbf{a}'\mathbf{g}$, of $\mathbf{g}$ in model (2.14), is estimable if there exists a linear function, $\mathbf{c}'\mathbf{y}$, of the observations such that $E(\mathbf{c}'\mathbf{y}) = \mathbf{a}'\mathbf{g}$. In Searle (1971, Section 5.4) it is shown that $\mathbf{a}'\mathbf{g}$ is estimable if and only if $\mathbf{a}'$ belongs to the linear span of the rows of $\mathbf{X}$. In Khuri (1984) we have the following theorem:

Theorem 2.4.2. Consider the balanced mixed model in (2.14). Then we have the following:

1. $r(\mathbf{P}_l\mathbf{X}) = r(\mathbf{P}_l)$, $l = 0, 1, \dots, \theta$.
2. $r(\mathbf{X}) = \sum_{l=0}^{\theta} r(\mathbf{P}_l\mathbf{X})$.
3. $\mathbf{P}_0\mathbf{X}\mathbf{g}, \mathbf{P}_1\mathbf{X}\mathbf{g}, \dots, \mathbf{P}_\theta\mathbf{X}\mathbf{g}$ are linearly independent and span the space of all estimable linear functions of $\mathbf{g}$.

Theorem 2.4.2 is useful in identifying a basis of estimable linear functions of the fixed effects in model (2.14).

2.4.2. The Singular-Value Decomposition

The singular-value decomposition of a matrix is far more useful, both in statistics and in matrix algebra, than is commonly realized. For example, it

plays a significant role in regression analysis. Let us consider the linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (2.15)$$

where $\mathbf{y}$ is a vector of n observations, $\mathbf{X}$ is an $n \times p$ ($n \geq p$) matrix consisting of known constants, $\boldsymbol{\beta}$ is an unknown parameter vector, and $\boldsymbol{\epsilon}$ is a random error vector. Using Theorem 2.3.11, the matrix $\mathbf{X}'$ can be expressed as

$$\mathbf{X}' = \mathbf{P}[\mathbf{D} : \mathbf{0}]\mathbf{Q}', \quad (2.16)$$

where $\mathbf{P}$ and $\mathbf{Q}$ are orthogonal matrices of orders $p \times p$ and $n \times n$, respectively, and $\mathbf{D}$ is a diagonal matrix of order $p \times p$ consisting of nonnegative diagonal elements. These are the singular values of $\mathbf{X}$ (or of $\mathbf{X}'$) and are the positive square roots of the eigenvalues of $\mathbf{X}'\mathbf{X}$. From (2.16) we get

$$\mathbf{X} = \mathbf{Q} \begin{bmatrix} \mathbf{D} \\ \mathbf{0}' \end{bmatrix} \mathbf{P}'. \quad (2.17)$$

If the columns of $\mathbf{X}$ are linearly related, then they are said to be multicollinear. In this case, $\mathbf{X}$ has rank r ($< p$), and the columns of $\mathbf{X}$ belong to a vector subspace of dimension r . At least one of the eigenvalues of $\mathbf{X}'\mathbf{X}$, and hence at least one of the singular values of $\mathbf{X}$, will be equal to zero. In practice, such exact multicollinearities rarely occur in statistical applications. Rather, the columns of $\mathbf{X}$ may be “nearly” linearly related. In this case, the rank of $\mathbf{X}$ is p , but some of the singular values of $\mathbf{X}$ will be “near zero.” We shall use the term multicollinearity in a broader sense to describe the latter situation. It is also common to use the term “ill conditioning” to refer to the same situation.

The presence of multicollinearities in $\mathbf{X}$ can have adverse effects on the least-squares estimate, $\hat{\boldsymbol{\beta}}$, of $\boldsymbol{\beta}$ in (2.15). This can be easily seen from the fact that $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ and $\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$, where σ^2 is the error variance. Large variances associated with the elements of $\hat{\boldsymbol{\beta}}$ can therefore be expected when the columns of $\mathbf{X}$ are multicollinear. This causes $\hat{\boldsymbol{\beta}}$ to become an unreliable estimate of $\boldsymbol{\beta}$. For a detailed study of multicollinearity and its effects, see Belsley, Kuh, and Welsch (1980, Chapter 3), Montgomery and Peck (1982, Chapter 8), and Myers (1990, Chapter 3).

The singular-value decomposition of $\mathbf{X}$ can provide useful information for detecting multicollinearity, as we shall now see. Let us suppose that the columns of $\mathbf{X}$ are multicollinear. Because of this, some of the singular values of $\mathbf{X}$, say p_2 ($< p$) of them, will be “near zero.” Let us partition $\mathbf{D}$ in (2.17) as

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \end{bmatrix},$$

where $\mathbf{D}_1$ and $\mathbf{D}_2$ are of orders $p_1 \times p_1$ and $p_2 \times p_2$ ($p_1 = p - p_2$), respectively. The diagonal elements of $\mathbf{D}_2$ consist of those singular values of $\mathbf{X}$ labeled as “near zero.” Let us now write (2.17) as

$$\mathbf{XP} = \mathbf{Q} \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_2 \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \quad (2.18)$$

Let us next partition $\mathbf{P}$ and $\mathbf{Q}$ as $\mathbf{P} = [\mathbf{P}_1 : \mathbf{P}_2]$, $\mathbf{Q} = [\mathbf{Q}_1 : \mathbf{Q}_2]$, where $\mathbf{P}_1$ and $\mathbf{P}_2$ have p_1 and p_2 columns, respectively, and $\mathbf{Q}_1$ and $\mathbf{Q}_2$ have p_1 and $n - p_1$ columns, respectively. From (2.18) we conclude that

$$\mathbf{XP}_1 = \mathbf{Q}_1 \mathbf{D}_1, \quad (2.19)$$

$$\mathbf{XP}_2 \approx \mathbf{0}, \quad (2.20)$$

where $\approx$ represents approximate equality. The matrix $\mathbf{XP}_2$ is “near zero” because of the smallness of the diagonal elements of $\mathbf{D}_2$.

We note from (2.20) that each column of $\mathbf{P}_2$ provides a “near”-linear relationship among the columns of $\mathbf{X}$. If (2.20) were an exact equality, then the columns of $\mathbf{P}_2$ would provide an orthonormal basis for the null space of $\mathbf{X}$.

We have mentioned that the presence of multicollinearity is indicated by the “smallness” of the singular values of $\mathbf{X}$. The problem now is to determine what “small” is. For this purpose it is common in statistics to use the condition number of $\mathbf{X}$, denoted by $\kappa(\mathbf{X})$. By definition

$$\kappa(\mathbf{X}) = \frac{\lambda_{\max}}{\lambda_{\min}},$$

where $\lambda_{\max}$ and $\lambda_{\min}$ are, respectively, the largest and smallest singular values of $\mathbf{X}$. Since the singular values of $\mathbf{X}$ are the positive square roots of the eigenvalues of $\mathbf{X}'\mathbf{X}$, then $\kappa(\mathbf{X})$ can also be written as

$$\kappa(\mathbf{X}) = \sqrt{\frac{e_{\max}(\mathbf{X}'\mathbf{X})}{e_{\min}(\mathbf{X}'\mathbf{X})}}.$$

If $\kappa(\mathbf{X})$ is less than 10, then there is no serious problem with multicollinearity. Values of $\kappa(\mathbf{X})$ between 10 and 30 indicate moderate to strong multicollinearity, and if $\kappa > 30$, severe multicollinearity is implied.

More detailed discussions concerning the use of the singular-value decomposition in regression can be found in Mandel (1982). See also Lowerre (1982). Good (1969) described several applications of this decomposition in statistics and in matrix algebra.

2.4.3. Extrema of Quadratic Forms

In many statistical problems there is a need to find the extremum (maximum or minimum) of a quadratic form or a ratio of quadratic forms. Let us, for example, consider the following problem:

Let $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$ be a collection of random vectors, all having the same number of elements. Suppose that these vectors are independently and identically distributed (i.i.d.) as $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where both $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are unknown. Consider testing the hypothesis $H_0: \boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus its alternative $H_a: \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$, where $\boldsymbol{\mu}_0$ is some hypothesized value of $\boldsymbol{\mu}$. We need to develop a test statistic for testing H_0 .

The multivariate hypothesis H_0 is true if and only if the univariate hypotheses

$$H_0(\boldsymbol{\lambda}): \boldsymbol{\lambda}'\boldsymbol{\mu} = \boldsymbol{\lambda}'\boldsymbol{\mu}_0$$

are true for all $\boldsymbol{\lambda} \neq \mathbf{0}$. A test statistic for testing $H_0(\boldsymbol{\lambda})$ is the following:

$$t(\boldsymbol{\lambda}) = \frac{\boldsymbol{\lambda}'(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)\sqrt{n}}{\sqrt{\boldsymbol{\lambda}'\mathbf{S}\boldsymbol{\lambda}}},$$

where $\bar{\mathbf{X}} = \sum_{i=1}^n \mathbf{X}_i/n$ and $\mathbf{S}$ is the sample variance-covariance matrix, which is an unbiased estimator of $\boldsymbol{\Sigma}$, and is given by

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$$

Large values of $t^2(\boldsymbol{\lambda})$ indicate falsehood of $H_0(\boldsymbol{\lambda})$. Since H_0 is rejected if and only if $H_0(\boldsymbol{\lambda})$ is rejected for at least one $\boldsymbol{\lambda}$, then the condition to reject H_0 at the α -level is $\sup_{\boldsymbol{\lambda} \neq \mathbf{0}} [t^2(\boldsymbol{\lambda})] > c_\alpha$, where c_α is the upper $100\alpha\%$ point of the distribution of $\sup_{\boldsymbol{\lambda} \neq \mathbf{0}} [t^2(\boldsymbol{\lambda})]$. But

$$\begin{aligned} \sup_{\boldsymbol{\lambda} \neq \mathbf{0}} [t^2(\boldsymbol{\lambda})] &= \sup_{\boldsymbol{\lambda} \neq \mathbf{0}} \frac{n|\boldsymbol{\lambda}'(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)|^2}{\boldsymbol{\lambda}'\mathbf{S}\boldsymbol{\lambda}} \\ &= n \sup_{\boldsymbol{\lambda} \neq \mathbf{0}} \frac{\boldsymbol{\lambda}'(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'\boldsymbol{\lambda}}{\boldsymbol{\lambda}'\mathbf{S}\boldsymbol{\lambda}} \\ &= n e_{\max} [\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'], \end{aligned}$$

by Theorem 2.3.17.

Now,

$$\begin{aligned} e_{\max} [\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'] &= e_{\max} [(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)], \\ &= (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{X}} - \boldsymbol{\mu}_0). \end{aligned}$$

by Theorem 2.3.9.

Hence,

$$\sup_{\lambda \neq 0} [t^2(\lambda)] = n(\bar{\mathbf{X}} - \boldsymbol{\mu}_0)' \mathbf{S}^{-1} (\bar{\mathbf{X}} - \boldsymbol{\mu}_0)$$

is the test statistic for the multivariate hypothesis H_0 . This is called Hotelling's T^2 -statistic. Its critical values are obtained in terms of the critical values of the F -distribution (see, for example, Morrison, 1967, Chapter 4).

Another example of using the extremum of a ratio of quadratic forms is in the determination of the canonical correlation coefficient between two random vectors (see Exercise 2.26). The article by Bush and Olkin (1959) lists several similar statistical applications.

2.4.4. The Parameterization of Orthogonal Matrices

Orthogonal matrices are used frequently in statistics, especially in linear models and multivariate analysis (see, for example, Graybill, 1961, Chapter 11; James, 1954).

The n^2 elements of an $n \times n$ orthogonal matrix $\mathbf{Q}$ are subject to $n(n+1)/2$ constraints because $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_n$. These elements can therefore be represented by $n^2 - n(n+1)/2 = n(n-1)/2$ independent parameters. The need for such a representation arises in several situations. For example, in the design of experiments, there may be a need to search for an orthogonal matrix that satisfies a certain optimality criterion. Using the independent parameters of an orthogonal matrix can facilitate this search. Khuri and Myers (1981) followed this approach in their construction of a response surface design that is robust to nonnormality of the error distribution associated with the response function. Another example is the generation of random orthogonal matrices for carrying out simulation experiments. This was used by Heiberger, Velleman, and Ypelaar (1983) to construct test data with special properties for multivariate linear models. Anderson, Olkin, and Underhill (1987) proposed a procedure to generate random orthogonal matrices.

Methods to parameterize an orthogonal matrix were reviewed in Khuri and Good (1989). One such method is to use the relationship between an orthogonal matrix and a skew-symmetric matrix. If $\mathbf{Q}$ is an orthogonal matrix with determinant equal to one, then it can be written in the form

$$\mathbf{Q} = e^{\mathbf{T}},$$

where $\mathbf{T}$ is a skew-symmetric matrix (see, for example, Gantmacher, 1959). The elements of $\mathbf{T}$ above its main diagonal can be used to parameterize $\mathbf{Q}$. This exponential mapping is defined by the infinite series

$$e^{\mathbf{T}} = \mathbf{I} + \mathbf{T} + \frac{\mathbf{T}^2}{2!} + \frac{\mathbf{T}^3}{3!} + \cdots.$$

The exponential parameterization was used in a theorem concerning the asymptotic joint density function of the eigenvalues of the sample variance–covariance matrix (Muirhead, 1982, page 394).

Another parameterization of $\mathbf{Q}$ is given by

$$\mathbf{Q} = (\mathbf{I} - \mathbf{U})(\mathbf{I} + \mathbf{U})^{-1},$$

where $\mathbf{U}$ is a skew-symmetric matrix. This relationship is valid provided that $\mathbf{Q}$ does not have the eigenvalue -1 . Otherwise, $\mathbf{Q}$ can be written as

$$\mathbf{Q} = \mathbf{L}(\mathbf{I} - \mathbf{U})(\mathbf{I} + \mathbf{U})^{-1},$$

where $\mathbf{L}$ is a diagonal matrix in which each element on the diagonal is either 1 or -1 . Arthur Cayley (1821–1895) is credited with having introduced the relationship between $\mathbf{Q}$ and $\mathbf{U}$.

Finally, the recent article by Olkin (1990) illustrates the strong interplay between statistics and linear algebra. The author listed several areas of statistics with a strong linear algebra component.

FURTHER READING AND ANNOTATED BIBLIOGRAPHY

- Anderson, T. W., and S. D. Gupta (1963). “Some inequalities on characteristic roots of matrices,” *Biometrika*, **50**, 522–524.
- Anderson, T. W., I. Olkin, and L. G. Underhill (1987). “Generation of random orthogonal matrices,” *SIAM J. Sci. Statist. Comput.*, **8**, 625–629.
- Basilevsky, A. (1983). *Applied Matrix Algebra in the Statistical Sciences*. North-Holland, New York. (This book addresses topics in matrix algebra that are useful in both applied and theoretical branches of the statistical sciences.)
- Bellman, R. (1970). *Introduction to Matrix Analysis*, 2nd ed. McGraw-Hill, New York. (An excellent reference book on matrix algebra. The minimum–maximum characterization of eigenvalues is discussed in Chap. 7. Kronecker products are studied in Chap. 12. Some applications of matrices to stochastic processes and probability theory are given in Chap. 14.)
- Belsley, D. A., E. Kuh, and R. E. Welsch (1980). *Regression Diagnostics*. Wiley, New York. (This is a good reference for learning about multicollinearity in linear statistical models that was discussed in Section 2.4.2. Examples are provided based on actual econometric data.)
- Bush, K. A., and I. Olkin (1959). “Extrema of quadratic forms with applications to statistics,” *Biometrika*, **46**, 483–486.
- Gantmacher, F. R. (1959). *The Theory of Matrices*, Vols. I and II. Chelsea, New York. (These two volumes provide a rather more advanced study of matrix algebra than standard introductory texts. Methods to parameterize an orthogonal matrix, which were mentioned in Section 2.4.4, are discussed in Vol. I.)

- Golub, G. H., and C. F. Van Loan (1983). *Matrix Computations*. Johns Hopkins University Press, Baltimore, Maryland.
- Good, I. J. (1969). "Some applications of the singular decomposition of a matrix." *Technometrics*, **11**, 823–831.
- Graybill, F. A. (1961). *An Introduction to Linear Statistical Models*, Vol. I. McGraw-Hill, New York. (This is considered a classic textbook in experimental statistics. It is concerned with the mathematical treatment, using matrix algebra, of linear statistical models.)
- Graybill, F. A. (1983). *Matrices with Applications in Statistics*, 2nd ed. Wadsworth, Belmont, California. (This frequently referenced textbook contains a great number of theorems in matrix algebra, and describes many properties of matrices that are pertinent to linear model and mathematical statistics.)
- Healy, M. J. R. (1986). *Matrices for Statistics*. Clarendon Press, Oxford, England. (This is a short book that provides a brief coverage of some basic concepts in matrix algebra. Some applications in statistics are also mentioned.)
- Heiberger, R. M., P. F. Velleman, and M. A. Ypelaar (1983). "Generating test data with independently controllable features for multivariate general linear forms." *J. Amer. Statist. Assoc.*, **78**, 585–595.
- Henderson, H. V., F. Pukelsheim, and S. R. Searle (1983). "On the history of the Kronecker product." *Linear and Multilinear Algebra*, **14**, 113–120.
- Henderson, H. V., and S. R. Searle (1981). "The vec-permutation matrix, the vec operator and Kronecker products: A review." *Linear and Multilinear Algebra*, **9**, 271–288.
- Hoerl, A. E., and R. W. Kennard (1970). "Ridge regression: Applications to nonorthogonal problems." *Technometrics*, **12**, 69–82.
- James, A. T. (1954). "Normal multivariate analysis and the orthogonal group." *Ann. Math. Statist.*, **25**, 40–75.
- Khuri, A. I. (1982). "Direct products: A powerful tool for the analysis of balanced data." *Comm. Statist. Theory Methods*, **11**, 2903–2920.
- Khuri, A. I. (1984). "Interval estimation of fixed effects and of functions of variance components in balanced mixed models." *Sankhyā, Series B*, **46**, 10–28. (Section 5 in this article gives a procedure for the construction of exact simultaneous confidence intervals on estimable linear functions of the fixed effects in a balanced mixed model.)
- Khuri, A. I., and I. J. Good (1989). "The parameterization of orthogonal matrices: A review mainly for statisticians." *South African Statist. J.*, **23**, 231–250.
- Khuri, A. I., and R. H. Myers (1981). "Design related robustness of tests in regression models." *Comm. Statist. Theory Methods*, **10**, 223–235.
- Lancaster, P. (1969). *Theory of Matrices*. Academic Press, New York. (This book is written primarily for students of applied mathematics, engineering, or science who want to acquire a good knowledge of the theory of matrices. Chap. 7 has an interesting discussion concerning the behavior of matrix eigenvalues under perturbation of the elements of the matrix.)
- Lowerre, J. M. (1982). "An introduction to modern matrix methods and statistics." *Amer. Statist.*, **36**, 113–115. (An application of the singular-value decomposition is given in Section 2 of this article.)

- Magnus, J. R., and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York. (This book consists of six parts. Part one deals with the basics of matrix algebra. The remaining parts are devoted to the development of matrix differential calculus and its applications to statistics and econometrics. Part four has a chapter on inequalities concerning eigenvalues that pertains to Section 2.3.10 in this chapter.)
- Mandel, J. (1982). "Use of the singular-value decomposition in regression analysis." *Amer. Statist.*, **36**, 15–24.
- Marcus, M., and H. Minc (1988). *Introduction to Linear Algebra*. Dover, New York. (This book presents an introduction to the fundamental concepts of linear algebra and matrix theory.)
- Marsaglia, G., and G. P. H. Styan (1974). "Equalities and inequalities for ranks of matrices." *Linear and Multilinear Algebra*, **2**, 269–292. (This is an interesting collection of results on ranks of matrices. It includes a wide variety of equalities and inequalities for ranks of products, of sums, and of partitioned matrices.)
- May, W. G. (1970). *Linear Algebra*. Scott, Foresman and Company, Glenview, Illinois.
- Montgomery, D. C., and E. A. Peck (1982). *Introduction to Linear Regression Analysis*. Wiley, New York. (Chap. 8 in this book has an interesting discussion concerning multicollinearity. It includes the sources of multicollinearity, its harmful effects in regression, available diagnostics, and a survey of remedial measures. This chapter provides useful additional information to the material in Section 2.4.2.)
- Morrison, D. F. (1967). *Multivariate Statistical Methods*. McGraw-Hill, New York. (This book can serve as an introductory text to multivariate analysis.)
- Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. (This book is designed as a text for a graduate-level course in multivariate analysis.)
- Myers, R. H. (1990). *Classical and Modern Regression with Applications*, 2nd ed. PWS-Kent, Boston. (Chap. 8 in this book should be useful reading concerning multicollinearity and its effects.)
- Newcomb, R. W. (1960). "On the simultaneous diagonalization of two semidefinite matrices." *Quart. Appl. Math.*, **19**, 144–146.
- Olkin, I. (1990). "Interface between statistics and linear algebra." In *Matrix Theory and Applications*, Vol. 40, C. R. Johnson, ed., American Mathematical Society, Providence, Rhode Island, pp. 233–256.
- Price, G. B. (1947). "Some identities in the theory of determinants." *Amer. Math. Monthly*, **54**, 75–90. (Section 10 in this article gives some history of the theory of determinants.)
- Rogers, G. S. (1984). "Kronecker products in ANOVA—a first step." *Amer. Statist.*, **38**, 197–202.
- Searle, S. R. (1971). *Linear Models*. Wiley, New York.
- Searle, S. R. (1982). *Matrix Algebra Useful for Statistics*. Wiley, New York. (This is a useful book introducing matrix algebra in a manner that is helpful in the statistical analysis of data and in statistics in general. Chaps. 13, 14, and 15 present applications of matrices in regression and linear models.)
- Seber, G. A. F. (1984). *Multivariate Observations*. Wiley, New York. (This is a good reference on applied multivariate analysis that is suited for a graduate-level course.)

Smith, D. E. (1958). *History of Mathematics*. Vol. I. Dover, New York. (This interesting book contains, among other things, some history concerning the development of the theory of determinants and matrices.)

Wolkowicz, H., and G. P. H. Styan (1980). "Bounds for eigenvalues using traces." *Linear Algebra Appl.*, **29**, 471–506.

EXERCISES

In Mathematics

- 2.1. Show that a set of $n \times 1$ vectors, $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_m$, is always linearly dependent if $m > n$.
- 2.2. Let W be a vector subspace of V such that $W = L(\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n)$, where the $\mathbf{u}_i$'s ($i = 1, 2, \dots, n$) are linearly independent. If $\mathbf{v}$ is any vector in V that is not in W , then the vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n, \mathbf{v}$ are linearly independent.
- 2.3. Prove Theorem 2.1.3.
- 2.4. Prove part 1 of Theorem 2.2.1.
- 2.5. Let $T: U \rightarrow V$ be a linear transformation. Show that T is one-to-one if and only if whenever $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n$ are linearly independent in U , then $T(\mathbf{u}_1), T(\mathbf{u}_2), \dots, T(\mathbf{u}_n)$ are linearly independent in V .
- 2.6. Let $T: R_n \rightarrow R_m$ be represented by an $n \times m$ matrix of rank ρ .
 - (a) Show that $\dim[T(R_n)] = \rho$.
 - (b) Show that if $n \leq m$ and $\rho = n$, then T is one-to-one.
- 2.7. Show that $\text{tr}(\mathbf{A}'\mathbf{A}) = 0$ if and only if $\mathbf{A} = \mathbf{0}$.
- 2.8. Let $\mathbf{A}$ be a symmetric positive semidefinite matrix of order $n \times n$. Show that $\mathbf{v}'\mathbf{A}\mathbf{v} = 0$ if and only if $\mathbf{A}\mathbf{v} = \mathbf{0}$.
- 2.9. The matrices $\mathbf{A}$ and $\mathbf{B}$ are symmetric and positive semidefinite of order $n \times n$ such that $\mathbf{AB} = \mathbf{BA}$. Show that $\mathbf{AB}$ is positive semidefinite.
- 2.10. If $\mathbf{A}$ is a symmetric $n \times n$ matrix, and $\mathbf{B}$ is an $n \times n$ skew-symmetric matrix, then show that $\text{tr}(\mathbf{AB}) = 0$.
- 2.11. Suppose that $\text{tr}(\mathbf{PA}) = 0$ for every skew-symmetric matrix $\mathbf{P}$. Show that the matrix $\mathbf{A}$ is symmetric.

- 2.12.** Let $\mathbf{A}$ be an $n \times n$ matrix and $\mathbf{C}$ be a nonsingular matrix of order $n \times n$. Show that $\mathbf{A}$, $\mathbf{C}^{-1}\mathbf{A}\mathbf{C}$, and $\mathbf{C}\mathbf{A}\mathbf{C}^{-1}$ have the same set of eigenvalues.
- 2.13.** Let $\mathbf{A}$ be an $n \times n$ symmetric matrix, and let λ be an eigenvalue of $\mathbf{A}$ of multiplicity k . Then $\mathbf{A} - \lambda\mathbf{I}_n$ has rank $n - k$.
- 2.14.** Let $\mathbf{A}$ be a nonsingular matrix of order $n \times n$, and let $\mathbf{c}$ and $\mathbf{d}$ be $n \times 1$ vectors. If $\mathbf{d}'\mathbf{A}^{-1}\mathbf{c} \neq -1$, then

$$(\mathbf{A} + \mathbf{c}\mathbf{d}')^{-1} = \mathbf{A}^{-1} - \frac{(\mathbf{A}^{-1}\mathbf{c})(\mathbf{d}'\mathbf{A}^{-1})}{1 + \mathbf{d}'\mathbf{A}^{-1}\mathbf{c}}.$$

This is known as the Sherman-Morrison formula.

- 2.15.** Show that if $\mathbf{A}$ and $\mathbf{I}_k + \mathbf{V}'\mathbf{A}^{-1}\mathbf{U}$ are nonsingular, then

$$(\mathbf{A} + \mathbf{U}\mathbf{V}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I}_k + \mathbf{V}'\mathbf{A}^{-1}\mathbf{U})^{-1}\mathbf{V}'\mathbf{A}^{-1},$$

where $\mathbf{A}$ is of order $n \times n$, and $\mathbf{U}$ and $\mathbf{V}$ are of order $n \times k$. This result is known as the Sherman-Morrison-Woodbury formula and is a generalization of the result in Exercise 2.14.

- 2.16.** Prove Theorem 2.3.17.
- 2.17.** Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ idempotent matrices. Show that $\mathbf{A} - \mathbf{B}$ is idempotent if and only if $\mathbf{AB} = \mathbf{BA} = \mathbf{B}$.
- 2.18.** Let $\mathbf{A}$ be an orthogonal matrix. What can be said about the eigenvalues of $\mathbf{A}$?
- 2.19.** Let $\mathbf{A}$ be a symmetric matrix of order $n \times n$, and let $\mathbf{L}$ be a matrix of order $n \times m$. Show that

$$e_{\min}(\mathbf{A})\text{tr}(\mathbf{L}'\mathbf{L}) \leq \text{tr}(\mathbf{L}'\mathbf{A}\mathbf{L}) \leq e_{\max}(\mathbf{A})\text{tr}(\mathbf{L}'\mathbf{L})$$

- 2.20.** Let $\mathbf{A}$ be a nonnegative definite matrix of order $n \times n$, and let $\mathbf{L}$ be a matrix of order $n \times m$. Show that
- (a) $e_{\min}(\mathbf{L}'\mathbf{A}\mathbf{L}) \geq e_{\min}(\mathbf{A})e_{\min}(\mathbf{L}'\mathbf{L})$,
- (b) $e_{\max}(\mathbf{L}'\mathbf{A}\mathbf{L}) \leq e_{\max}(\mathbf{A})e_{\max}(\mathbf{L}'\mathbf{L})$.
- 2.21.** Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ symmetric matrices with $\mathbf{A}$ nonnegative definite. Show that

$$e_{\min}(\mathbf{B})\text{tr}(\mathbf{A}) \leq \text{tr}(\mathbf{AB}) \leq e_{\max}(\mathbf{B})\text{tr}(\mathbf{A}).$$

2.22. Let $\mathbf{A}^-$ be a g -inverse of $\mathbf{A}$. Show that

- (a) $\mathbf{A}^- \mathbf{A}$ is idempotent,
- (b) $r(\mathbf{A}^-) \geq r(\mathbf{A})$,
- (c) $r(\mathbf{A}) = r(\mathbf{A}^- \mathbf{A})$.

In Statistics

2.23. Let $\mathbf{y} = (y_1, y_2, \dots, y_n)'$ be a normal random vector $N(\mathbf{0}, \sigma^2 \mathbf{I}_n)$. Let $\bar{y}$ and s^2 be the sample mean and sample variance given by

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i,$$

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n} \right].$$

- (a) Show that $\mathbf{A}$ is an idempotent matrix of rank $n-1$, where $\mathbf{A}$ is an $n \times n$ matrix such that $\mathbf{y}' \mathbf{A} \mathbf{y} = (n-1)s^2$.
- (b) What distribution does $(n-1)s^2/\sigma^2$ have?
- (c) Show that $\bar{y}$ and $\mathbf{A} \mathbf{y}$ are uncorrelated; then conclude that $\bar{y}$ and s^2 are statistically independent.

2.24. Consider the one-way classification model

$$y_{ij} = \mu + \alpha_i + \epsilon_{ij}, \quad i = 1, 2, \dots, a; \quad j = 1, 2, \dots, n_i,$$

where μ and α_i ($i = 1, 2, \dots, a$) are unknown parameters and ϵ_{ij} is a random error with a zero mean. Show that

- (a) $\alpha_i - \alpha_{i'}$ is an estimable linear function for all $i \neq i'$ ($i, i' = 1, 2, \dots, a$),
- (b) μ is nonestimable.

2.25. Consider the linear model

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{X}$ is a known matrix of order $n \times p$ and rank r ($\leq p$), $\boldsymbol{\beta}$ is an unknown parameter vector, and $\boldsymbol{\epsilon}$ is a random error vector such that $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and $\text{Var}(\boldsymbol{\epsilon}) = \sigma^2 \mathbf{I}_n$.

- (a) Show that $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ is an idempotent matrix.
- (b) Let $\mathbf{I}'\mathbf{y}$ be an unbiased linear estimator of $\boldsymbol{\lambda}'\boldsymbol{\beta}$. Show that

$$\text{Var}(\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}}) \leq \text{Var}(\mathbf{I}'\mathbf{y}),$$

where $\boldsymbol{\lambda}'\hat{\boldsymbol{\beta}} = \boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{y}$.

The result given in part (b) is known as the *Gauss–Markov theorem*.

- 2.26. Consider the linear model in Exercise 2.25, and suppose that $r(\mathbf{X}) = p$. Hoerl and Kennard (1970) introduced an estimator of $\boldsymbol{\beta}$ called the ridge estimator $\boldsymbol{\beta}^*$:

$$\boldsymbol{\beta}^* = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y},$$

where k is a “small” fixed number. For an appropriate value of k , $\boldsymbol{\beta}^*$ provides improved accuracy in the estimation of $\boldsymbol{\beta}$ over the least-squares estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Let $\mathbf{X}'\mathbf{X} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}'$ be the spectral decomposition of $\mathbf{X}'\mathbf{X}$. Show that $\boldsymbol{\beta}^* = \mathbf{P}\mathbf{D}\mathbf{P}'\hat{\boldsymbol{\beta}}$, where $\mathbf{D}$ is a diagonal matrix whose i th diagonal element is $\lambda_i/(\lambda_i + k)$, $i = 1, 2, \dots, p$, and where $\lambda_1, \lambda_2, \dots, \lambda_p$ are the diagonal elements of $\mathbf{\Lambda}$.

- 2.27. Consider the ratio

$$\rho^2 = \frac{(\mathbf{x}'\mathbf{A}\mathbf{y})^2}{(\mathbf{x}'\mathbf{B}_1\mathbf{x})(\mathbf{y}'\mathbf{B}_2\mathbf{y})},$$

where $\mathbf{A}$ is a matrix of order $m \times n$ and $\mathbf{B}_1, \mathbf{B}_2$ are positive definite of orders $m \times m$ and $n \times n$, respectively. Show that

$$\sup_{\mathbf{x}, \mathbf{y}} \rho^2 = e_{\max}(\mathbf{B}_1^{-1}\mathbf{A}\mathbf{B}_2^{-1}\mathbf{A}').$$

[Hint: Define $\mathbf{C}_1$ and $\mathbf{C}_2$ as symmetric nonsingular matrices such that $\mathbf{C}_1^2 = \mathbf{B}_1, \mathbf{C}_2^2 = \mathbf{B}_2$. Let $\mathbf{C}_1\mathbf{x} = \mathbf{u}, \mathbf{C}_2\mathbf{y} = \mathbf{v}$. Then ρ^2 can be written as

$$\rho^2 = \frac{(\mathbf{u}'\mathbf{C}_1^{-1}\mathbf{A}\mathbf{C}_2^{-1}\mathbf{v})^2}{(\mathbf{u}'\mathbf{u})(\mathbf{v}'\mathbf{v})} = (\mathbf{v}'\mathbf{C}_1^{-1}\mathbf{A}\mathbf{C}_2^{-1}\boldsymbol{\tau})^2,$$

where $\mathbf{v} = \mathbf{u}/(\mathbf{u}'\mathbf{u})^{1/2}, \boldsymbol{\tau} = \mathbf{v}/(\mathbf{v}'\mathbf{v})^{1/2}$ are unit vectors. Verify the result of this problem after noting that ρ^2 is now the square of a dot product.]

Note: This exercise has the following application in multivariate analysis: Let $\mathbf{z}_1$ and $\mathbf{z}_2$ be random vectors with zero means and variance-covariance matrices $\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{22}$, respectively. Let $\boldsymbol{\Sigma}_{12}$ be the covariance matrix of $\mathbf{z}_1$ and $\mathbf{z}_2$. On choosing $\mathbf{A} = \boldsymbol{\Sigma}_{12}, \mathbf{B}_1 = \boldsymbol{\Sigma}_{11}, \mathbf{B}_2 = \boldsymbol{\Sigma}_{22}$, the positive square root of the supremum of ρ^2 is called the *canonical correlation coefficient* between $\mathbf{z}_1$ and $\mathbf{z}_2$. It is a measure of the linear association between $\mathbf{z}_1$ and $\mathbf{z}_2$ (see, for example, Seber, 1984, Section 5.7).