2 Matrix Algebra

If we write a linear model such as (1.2) for each of n observations in a dataset, the n resulting models can be expressed in a single compact matrix expression. Then the estimation and testing results can be more easily obtained using matrix theory.

In the present chapter, we review the elements of matrix theory needed in the remainder of the book. Proofs that seem instructive are included or called for in the problems. For other proofs, see Graybill (1969), Searle (1982), Harville (1997), Schott (1997), or any general text on matrix theory. We begin with some basic definitions in Section 2.1.

2.1 MATRIX AND VECTOR NOTATION

2.1.1 Matrices, Vectors, and Scalars

A *matrix* is a rectangular or square array of numbers or variables. We use uppercase boldface letters to represent matrices. In this book, all elements of matrices will be real numbers or variables representing real numbers. For example, the height (in inches) and weight (in pounds) for three students are listed in the following matrix:

$$\mathbf{A} = \begin{pmatrix} 65 & 154 \\ 73 & 182 \\ 68 & 167 \end{pmatrix}. \tag{2.1}$$

To represent the elements of $\mathbf{A}$ as variables, we use

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{pmatrix}. \tag{2.2}$$

The first subscript in a_{ij} indicates the row; the second identifies the column. The notation $\mathbf{A} = (a_{ij})$ represents a matrix by means of a typical element.

The matrix $\mathbf{A}$ in (2.1) or (2.2) has three rows and two columns, and we say that $\mathbf{A}$ is 3×2, or that the *size* of $\mathbf{A}$ is 3×2.

A *vector* is a matrix with a single row or column. Elements in a vector are often identified by a single subscript; for example

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

As a convention, we use lowercase boldface letters for column vectors and lowercase boldface letters followed by the prime symbol ($'$) for row vectors; for example

$$\mathbf{x}' = (x_1, x_2, x_3) = (x_1 \; x_2 \; x_3).$$

(Row vectors are regarded as *transposes* of column vectors. The transpose is defined in Section 2.1.3 below). We use either commas or spaces to separate elements of a row vector.

Geometrically, a row or column vector with p elements can be associated with a point in a p-dimensional space. The elements in the vector are the coordinates of the point. Sometimes we are interested in the distance from the origin to the point (vector), the distance between two points (vectors), or the angle between the arrows drawn from the origin to the two points.

In the context of matrices and vectors, a single real number is called a *scalar*. Thus 2.5, -9, and 7.26 are scalars. A variable representing a scalar will be denoted by a lightface letter (usually lowercase), such as c. A scalar is technically distinct from a 1×1 matrix in terms of its uses and properties in matrix algebra. The same notation is often used to represent a scalar and a 1×1 matrix, but the meaning is usually obvious from the context.

2.1.2 Matrix Equality

Two matrices or two vectors are equal if they are of the same size and if the elements in corresponding positions are equal; for example

$$\begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix} = \begin{pmatrix} 3 & -2 & 4 \\ 1 & 3 & 7 \end{pmatrix},$$

but

$$\begin{pmatrix} 5 & 2 & -9 \\ 8 & -4 & 6 \end{pmatrix} \neq \begin{pmatrix} 5 & 3 & -9 \\ 8 & -4 & 6 \end{pmatrix}.$$

2.1.3 Transpose

If we interchange the rows and columns of a matrix $\mathbf{A}$, the resulting matrix is known as the *transpose* of $\mathbf{A}$ and is denoted by $\mathbf{A}'$; for example

$$\mathbf{A} = \begin{pmatrix} 6 & -2 \\ 4 & 7 \\ 1 & 3 \end{pmatrix}, \qquad \mathbf{A}' = \begin{pmatrix} 6 & 4 & 1 \\ -2 & 7 & 3 \end{pmatrix}.$$

Formally, if $\mathbf{A}$ is denoted by $\mathbf{A} = (a_{ij})$, then $\mathbf{A}'$ is defined as

$$\mathbf{A}' = (a_{ij})' = (a_{ji}). \tag{2.3}$$

This notation indicates that the element in the ith row and jth column of $\mathbf{A}$ is found in the jth row and ith column of $\mathbf{A}'$. If the matrix $\mathbf{A}$ is $n \times p$, then $\mathbf{A}'$ is $p \times n$.

If a matrix is transposed twice, the result is the original matrix.

Theorem 2.1. If $\mathbf{A}$ is any matrix, then

$$(\mathbf{A}')' = \mathbf{A}. \tag{2.4}$$

Proof. By (2.3), $\mathbf{A}' = (a_{ij})' = (a_{ji})$. Then $(\mathbf{A}')' = (a_{ji})' = (a_{ij}) = \mathbf{A}$. $\square$

(The notation $\square$ is used to indicate the end of a theorem proof, corollary proof or example.)

2.1.4 Matrices of Special Form

If the transpose of a matrix $\mathbf{A}$ is the same as the original matrix, that is, if $\mathbf{A}' = \mathbf{A}$ or equivalently $(a_{ji}) = (a_{ij})$, then the matrix $\mathbf{A}$ is said to be *symmetric*. For example

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 6 \\ 2 & 10 & -7 \\ 6 & -7 & 9 \end{pmatrix}$$

is symmetric. Clearly, all symmetric matrices are square.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{p1} & a_{22} & \cdots & a_{pp} \end{pmatrix}$$

The *diagonal* of a $p \times p$ square matrix $\mathbf{A} = (a_{ij})$ consists of the elements $a_{11}, a_{22}, \ldots, a_{pp}$. If a matrix contains zeros in all off-diagonal positions, it is said to be a *diagonal matrix*; for example, consider the matrix

$$\mathbf{D} = \begin{pmatrix} 8 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \end{pmatrix},$$

which can also be denoted as

$$\mathbf{D} = \text{diag}(8, \ -3, 0, \ 4).$$

We also use the notation diag($\mathbf{A}$) to indicate a diagonal matrix with the same diagonal elements as $\mathbf{A}$; for example

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 6 \\ 2 & 10 & -7 \\ 6 & -7 & 9 \end{pmatrix}, \quad \text{diag}(\mathbf{A}) = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

A diagonal matrix with a 1 in each diagonal position is called an *identity* matrix, and is denoted by $\mathbf{I}$; for example

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.5}$$

An *upper triangular matrix* is a square matrix with zeros below the diagonal; for example,

$$\mathbf{T} = \begin{pmatrix} 7 & 2 & 3 & -5 \\ 0 & 0 & -2 & 6 \\ 0 & 0 & 4 & 1 \\ 0 & 0 & 0 & 8 \end{pmatrix}.$$

A *lower triangular matrix* is defined similarly.

A vector of 1s is denoted by $\mathbf{j}$:

$$\mathbf{j} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{2.6}$$

A square matrix of 1s is denoted by $\mathbf{J}$; for example

$$\mathbf{J} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \tag{2.7}$$

We denote a vector of zeros by $\mathbf{0}$ and a matrix of zeros by $\mathbf{O}$; for example

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{O} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \tag{2.8}$$

2.2 OPERATIONS

We now define sums and products of matrices and vectors and consider some properties of these sums and products.

2.2.1 Sum of Two Matrices or Two Vectors

If two matrices or two vectors are the same size, they are said to be *conformal for addition*. Their *sum* is found by adding corresponding elements. Thus, if $\mathbf{A}$ is $n \times p$ and $\mathbf{B}$ is $n \times p$, then $\mathbf{C} = \mathbf{A} + \mathbf{B}$ is also $n \times p$ and is found as $\mathbf{C} = (c_{ij}) = (a_{ij} + b_{ij})$; for example

$$\begin{pmatrix} 7 & -3 & 4 \\ 2 & 8 & -5 \end{pmatrix} + \begin{pmatrix} 11 & 5 & -6 \\ 3 & 4 & 2 \end{pmatrix} = \begin{pmatrix} 18 & 2 & -2 \\ 5 & 12 & -3 \end{pmatrix}.$$

The *difference* $\mathbf{D} = \mathbf{A} - \mathbf{B}$ between two conformal matrices $\mathbf{A}$ and $\mathbf{B}$ is defined similarly: $\mathbf{D} = (d_{ij}) = (a_{ij} - b_{ij})$.

Two properties of matrix addition are given in the following theorem.

Theorem 2.2a. If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times m$, then

(i) $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$. (2.9)

(ii) $(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'$. (2.10)

□

2.2.2 Product of a Scalar and a Matrix

Any scalar can be multiplied by any matrix. The product of a scalar and a matrix is defined as the product of each element of the matrix and the scalar:

$$c\mathbf{A} = (ca_{ij}) = \begin{pmatrix} ca_{11} & ca_{12} & \cdots & ca_{1m} \\ ca_{21} & ca_{22} & \cdots & ca_{2m} \\ \vdots & \vdots & & \vdots \\ ca_{n1} & ca_{n2} & \cdots & ca_{nm} \end{pmatrix}. \tag{2.11}$$

Since $ca_{ij} = a_{ij}c$, the product of a scalar and a matrix is commutative:

$$c\mathbf{A} = \mathbf{A}c. \tag{2.12}$$

2.2.3 Product of Two Matrices or Two Vectors

In order for the product $\mathbf{AB}$ to be defined, the number of columns in $\mathbf{A}$ must equal the number of rows in $\mathbf{B}$, in which case $\mathbf{A}$ and $\mathbf{B}$ are said to be *conformal for multiplication*. Then the (ij)th element of the product $\mathbf{C} = \mathbf{AB}$ is defined as

$$c_{ij} = \sum_k a_{ik} b_{kj}, \tag{2.13}$$

which is the sum of products of the elements in the ith row of $\mathbf{A}$ and the elements in the jth column of $\mathbf{B}$. Thus we multiply every row of $\mathbf{A}$ by every column of $\mathbf{B}$. If $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$, then $\mathbf{C} = \mathbf{AB}$ is $n \times p$. We illustrate matrix multiplication in the following example.

Example 2.2.3. Let

$$\mathbf{A} = \begin{pmatrix} 2 & 1 & 3 \\ 4 & 6 & 5 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 1 & 4 \\ 2 & 6 \\ 3 & 8 \end{pmatrix}.$$

Then

$$\mathbf{AB} = \begin{pmatrix} 2 \cdot 1 + 1 \cdot 2 + 3 \cdot 3 & 2 \cdot 4 + 1 \cdot 6 + 3 \cdot 8 \\ 4 \cdot 1 + 6 \cdot 2 + 5 \cdot 3 & 4 \cdot 4 + 6 \cdot 6 + 5 \cdot 8 \end{pmatrix} = \begin{pmatrix} 13 & 38 \\ 31 & 92 \end{pmatrix},$$

$$\mathbf{BA} = \begin{pmatrix} 18 & 25 & 23 \\ 28 & 38 & 36 \\ 38 & 51 & 49 \end{pmatrix}.$$

□

Note that a 1×1 *matrix* $\mathbf{A}$ can only be multiplied on the right by a $1 \times n$ matrix $\mathbf{B}$ or on the left by an $n \times 1$ matrix $\mathbf{C}$, whereas a *scalar* can be multiplied on the right or left by a matrix of any size.

If $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$, where $n \neq p$, then $\mathbf{AB}$ is defined, but $\mathbf{BA}$ is not defined. If $\mathbf{A}$ is $n \times p$ and $\mathbf{B}$ is $p \times n$, then $\mathbf{AB}$ is $n \times n$ and $\mathbf{BA}$ is $p \times p$. In this case, of course, $\mathbf{AB} \neq \mathbf{BA}$, as illustrated in Example 2.2.3. If $\mathbf{A}$ and $\mathbf{B}$ are both $n \times n$, then $\mathbf{AB}$ and $\mathbf{BA}$ are the same size, but, in general

$$\mathbf{AB} \neq \mathbf{BA}. \tag{2.14}$$

[There are a few exceptions to (2.14), for example, two diagonal matrices or a square matrix and an identity.] Thus matrix multiplication is not commutative, and certain familiar manipulations with real numbers cannot be done with matrices. However, matrix multiplication is distributive over addition or subtraction:

$$\mathbf{A}(\mathbf{B} \pm \mathbf{C}) = \mathbf{AB} \pm \mathbf{AC}, \tag{2.15}$$

$$(\mathbf{A} \pm \mathbf{B})\mathbf{C} = \mathbf{AC} \pm \mathbf{BC}. \tag{2.16}$$

Using (2.15) and (2.16), we can expand products such as $(\mathbf{A} - \mathbf{B})(\mathbf{C} - \mathbf{D})$:

$$\begin{aligned} (\mathbf{A} - \mathbf{B})(\mathbf{C} - \mathbf{D}) &= (\mathbf{A} - \mathbf{B})\mathbf{C} - (\mathbf{A} - \mathbf{B})\mathbf{D} \qquad [\text{by (2.15)}] \\ &= \mathbf{AC} - \mathbf{BC} - \mathbf{AD} + \mathbf{BD} \qquad [\text{by (2.16)}]. \end{aligned} \tag{2.17}$$

Multiplication involving vectors follows the same rules as for matrices. Suppose that $\mathbf{A}$ is $n \times p$, $\mathbf{b}$ is $p \times 1$, $\mathbf{c}$ is $p \times 1$, and $\mathbf{d}$ is $n \times 1$. Then $\mathbf{Ab}$ is a column vector of size $n \times 1$, $\mathbf{d}'\mathbf{A}$ is a row vector of size $1 \times p$, $\mathbf{b}'\mathbf{c}$ is a sum of products (1×1), $\mathbf{bc}'$ is a $p \times p$ matrix, and $\mathbf{cd}'$ is a $p \times n$ matrix. Since $\mathbf{b}'\mathbf{c}$ is a 1×1 sum of products, it is equal to $\mathbf{c}'\mathbf{b}$:

$$\begin{aligned} \mathbf{b}'\mathbf{c} &= b_1c_1 + b_2c_2 + \cdots + b_pc_p, \\ \mathbf{c}'\mathbf{b} &= c_1b_1 + c_2b_2 + \cdots + c_pb_p, \\ \mathbf{b}'\mathbf{c} &= \mathbf{c}'\mathbf{b}. \end{aligned} \tag{2.18}$$

The matrix $\mathbf{cd}'$ is given by

$$\mathbf{cd}' = \begin{pmatrix} c_1d_1 & c_1d_2 & \cdots & c_1d_n \\ c_2d_1 & c_2d_2 & \cdots & c_2d_n \\ \vdots & \vdots & & \vdots \\ c_pd_1 & c_pd_2 & \cdots & c_pd_n \end{pmatrix}. \tag{2.19}$$

Similarly

$$\mathbf{b}'\mathbf{b} = b_1^2 + b_2^2 + \cdots + b_p^2, \tag{2.20}$$

$$\mathbf{b}\mathbf{b}' = \begin{pmatrix} b_1^2 & b_1b_2 & \cdots & b_1b_p \\ b_2b_1 & b_2^2 & \cdots & b_2b_p \\ \vdots & \vdots & & \vdots \\ b_pb_1 & b_pb_2 & \cdots & b_p^2 \end{pmatrix}. \tag{2.21}$$

Thus, $\mathbf{b}'\mathbf{b}$ is a sum of squares and $\mathbf{b}\mathbf{b}'$ is a (symmetric) square matrix.

The square root of the sum of squares of the elements of a $p \times 1$ vector $\mathbf{b}$ is the distance from the origin to the point $\mathbf{b}$ and is also referred to as the *length* of $\mathbf{b}$:

$$\text{Length of } \mathbf{b} = \sqrt{\mathbf{b}'\mathbf{b}} = \sqrt{\sum_{i=1}^{p} b_i^2}. \tag{2.22}$$

If $\mathbf{j}$ is an $n \times 1$ vector of 1s as defined in (2.6), then by (2.20) and (2.21), we have

$$\mathbf{j}'\mathbf{j} = n, \quad \mathbf{j}\mathbf{j}' = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \cdots & 1 \end{pmatrix} = \mathbf{J}, \tag{2.23}$$

where $\mathbf{J}$ is an $n \times n$ square matrix of 1s as illustrated in (2.7). If $\mathbf{a}$ is $n \times 1$ and $\mathbf{A}$ is $n \times p$, then

$$\mathbf{a}'\mathbf{j} = \mathbf{j}'\mathbf{a} = \sum_{i=1}^{n} a_i, \tag{2.24}$$

$$\mathbf{j}'\mathbf{A} = \left(\sum_i a_{i1}, \sum_i a_{i2}, \ldots, \sum_i a_{ip} \right), \qquad \mathbf{A}\mathbf{j} = \begin{pmatrix} \sum_j a_{1j} \\ \sum_j a_{2j} \\ \vdots \\ \sum_j a_{nj} \end{pmatrix}. \tag{2.25}$$

Thus $\mathbf{a}'\mathbf{j}$ is the sum of the elements in $\mathbf{a}$, $\mathbf{j}'\mathbf{A}$ contains the column sums of $\mathbf{A}$, and $\mathbf{A}\mathbf{j}$ contains the row sums of $\mathbf{A}$. Note that in $\mathbf{a}'\mathbf{j}$, the vector $\mathbf{j}$ is $n \times 1$; in $\mathbf{j}'\mathbf{A}$, the vector $\mathbf{j}$ is $n \times 1$; and in $\mathbf{A}\mathbf{j}$, the vector $\mathbf{j}$ is $p \times 1$.

The transpose of the product of two matrices is the product of the transposes in reverse order.

Theorem 2.2b. If $\mathbf{A}$ is $n \times p$ and $\mathbf{B}$ is $p \times m$, then

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'. \tag{2.26}$$

Proof. Let $\mathbf{C} = \mathbf{AB}$. Then by (2.13)

$$\mathbf{C} = (c_{ij}) = \left(\sum_{k=1}^{p} a_{ik} b_{kj} \right).$$

By (2.3), the transpose of $\mathbf{C} = \mathbf{AB}$ becomes

$$\begin{aligned}(\mathbf{AB})' &= \mathbf{C}' = (c_{ij})' = (c_{ji}) \\ &= \left(\sum_{k=1}^{p} a_{jk} b_{ki} \right) = \left(\sum_{k=1}^{p} b_{ki} a_{jk} \right) = \mathbf{B}'\mathbf{A}'. \qquad \square\end{aligned}$$

We illustrate the steps in the proof of Theorem 2.2b using a 2×3 matrix $\mathbf{A}$ and a 3×2 matrix $\mathbf{B}$:

$$\begin{aligned}
\mathbf{AB} &= \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix} \\
&= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}, \\
(\mathbf{AB})' &= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} \\ a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix} \\
&= \begin{pmatrix} b_{11}a_{11} + b_{21}a_{12} + b_{31}a_{13} & b_{11}a_{21} + b_{21}a_{22} + b_{31}a_{23} \\ b_{12}a_{11} + b_{22}a_{12} + b_{32}a_{13} & b_{12}a_{21} + b_{22}a_{22} + b_{32}a_{23} \end{pmatrix} \\
&= \begin{pmatrix} b_{11} & b_{21} & b_{31} \\ b_{12} & b_{22} & b_{32} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix} \\
&= \mathbf{B}'\mathbf{A}'.
\end{aligned}$$

The following corollary to Theorem 2.2b gives the transpose of the product of three matrices.

Corollary 1. If $\mathbf{A}, \mathbf{B}$, and $\mathbf{C}$ are conformal so that $\mathbf{ABC}$ is defined, then $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$. □

Suppose that $\mathbf{A}$ is $n \times m$ and $\mathbf{B}$ is $m \times p$. Let $\mathbf{a}_i'$ be the ith *row* of $\mathbf{A}$ and $\mathbf{b}_j$ be the jth *column* of $\mathbf{B}$, so that

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_n' \end{pmatrix}, \qquad \mathbf{B} = (\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p).$$

Then, by definition, the (ij)th element of $\mathbf{AB}$ is $\mathbf{a}_i'\mathbf{b}_j$:

$$\mathbf{AB} = \begin{pmatrix} \mathbf{a}_1'\mathbf{b}_1 & \mathbf{a}_1'\mathbf{b}_2 & \cdots & \mathbf{a}_1'\mathbf{b}_p \\ \mathbf{a}_2'\mathbf{b}_1 & \mathbf{a}_2'\mathbf{b}_2 & \cdots & \mathbf{a}_2'\mathbf{b}_p \\ \vdots & \vdots & & \vdots \\ \mathbf{a}_n'\mathbf{b}_1 & \mathbf{a}_n'\mathbf{b}_2 & \cdots & \mathbf{a}_n'\mathbf{b}_p \end{pmatrix}.$$

This product can be written in terms of the rows of $\mathbf{A}$:

$$\mathbf{AB} = \begin{pmatrix} \mathbf{a}_1'(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p) \\ \mathbf{a}_2'(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p) \\ \vdots \\ \mathbf{a}_n'(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p) \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1'\mathbf{B} \\ \mathbf{a}_2'\mathbf{B} \\ \vdots \\ \mathbf{a}_n'\mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_n' \end{pmatrix}\mathbf{B}. \tag{2.27}$$

The first column of $\mathbf{AB}$ can be expressed in terms of $\mathbf{A}$ as

$$\begin{pmatrix} \mathbf{a}_1'\mathbf{b}_1 \\ \mathbf{a}_2'\mathbf{b}_1 \\ \vdots \\ \mathbf{a}_n'\mathbf{b}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \vdots \\ \mathbf{a}_n' \end{pmatrix}\mathbf{b}_1 = \mathbf{A}\mathbf{b}_1.$$

Likewise, the second column is $\mathbf{A}\mathbf{b}_2$, and so on. Thus $\mathbf{AB}$ can be written in terms of the columns of $\mathbf{B}$:

$$\mathbf{AB} = \mathbf{A}(\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_p) = (\mathbf{A}\mathbf{b}_1, \mathbf{A}\mathbf{b}_2, \ldots, \mathbf{A}\mathbf{b}_p). \tag{2.28}$$

Any matrix $\mathbf{A}$ can be multiplied by its transpose to form $\mathbf{A}'\mathbf{A}$ or $\mathbf{A}\mathbf{A}'$. Some properties of these two products are given in the following theorem.

Theorem 2.2c. Let $\mathbf{A}$ be any $n \times p$ matrix. Then $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ have the following properties.

(i) $\mathbf{A}'\mathbf{A}$ is $p \times p$ and its elements are products of the *columns* of $\mathbf{A}$.
(ii) $\mathbf{A}\mathbf{A}'$ is $n \times n$ and its elements are products of the *rows* of $\mathbf{A}$.
(iii) Both $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$ are symmetric.
(iv) If $\mathbf{A}'\mathbf{A} = \mathbf{O}$, then $\mathbf{A} = \mathbf{O}$. □

Let $\mathbf{A}$ be an $n \times n$ matrix and let $\mathbf{D} = \text{diag}(d_1, d_2, \ldots, d_n)$. In the product $\mathbf{DA}$, the ith row of $\mathbf{A}$ is multiplied by d_i, and in $\mathbf{AD}$, the jth column of $\mathbf{A}$ is multiplied by d_j. For example, if $n = 3$, we have

$$\mathbf{DA} = \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} = \begin{pmatrix} d_1a_{11} & d_1a_{12} & d_1a_{13} \\ d_2a_{21} & d_2a_{22} & d_2a_{23} \\ d_3a_{31} & d_3a_{32} & d_3a_{33} \end{pmatrix}, \tag{2.29}$$

$$\mathbf{AD} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \begin{pmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{pmatrix} = \begin{pmatrix} d_1a_{11} & d_2a_{12} & d_3a_{13} \\ d_1a_{21} & d_2a_{22} & d_3a_{23} \\ d_1a_{31} & d_2a_{32} & d_3a_{33} \end{pmatrix}, \tag{2.30}$$

$$\mathbf{DAD} = \begin{pmatrix} d_1^2a_{11} & d_1d_2a_{12} & d_1d_3a_{13} \\ d_2d_1a_{21} & d_2^2a_{22} & d_2d_3a_{23} \\ d_3d_1a_{31} & d_3d_2a_{32} & d_3^2a_{33} \end{pmatrix}. \tag{2.31}$$

Note that $\mathbf{DA} \neq \mathbf{AD}$. However, in the special case where the diagonal matrix is the identity, (2.29) and (2.30) become

$$\mathbf{IA} = \mathbf{AI} = \mathbf{A}. \tag{2.32}$$

If $\mathbf{A}$ is rectangular, (2.32) still holds, but the two identities are of different sizes.

If $\mathbf{A}$ is a symmetric matrix and $\mathbf{y}$ is a vector, the product

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \sum_i a_{ii}y_i^2 + \sum_{i \neq j} a_{ij}y_i y_j \tag{2.33}$$

is called a *quadratic form*. If $\mathbf{x}$ is $n \times 1$, $\mathbf{y}$ is $p \times 1$, and $\mathbf{A}$ is $n \times p$, the product

$$\mathbf{x}'\mathbf{A}\mathbf{y} = \sum_{ij} a_{ij}x_i y_j \tag{2.34}$$

is called a *bilinear form*.

2.2.4 Hadamard Product of Two Matrices or Two Vectors

Sometimes a third type of product, called the *elementwise* or *Hadamard product*, is useful. If two matrices or two vectors are of the same size (conformal for addition), the Hadamard product is found by simply multiplying corresponding elements:

$$(a_{ij}b_{ij}) = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots & a_{1p}b_{1p} \\ a_{21}b_{21} & a_{22}b_{22} & \cdots & a_{2p}b_{2p} \\ \vdots & \vdots & & \vdots \\ a_{n1}b_{n1} & a_{n2}b_{n2} & \cdots & a_{np}b_{np} \end{pmatrix}.$$

2.3 PARTITIONED MATRICES

It is sometimes convenient to partition a matrix into submatrices. For example, a partitioning of a matrix $\mathbf{A}$ into four (square or rectangular) submatrices of appropriate sizes can be indicated symbolically as follows:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}.$$

To illustrate, let the 4×5 matrix $\mathbf{A}$ be partitioned as

$$\mathbf{A} = \left(\begin{array}{ccc|cc} 7 & 2 & 5 & 8 & 4 \\ -3 & 4 & 0 & 2 & 7 \\ \hline 9 & 3 & 6 & 5 & -2 \\ 3 & 1 & 2 & 1 & 6 \end{array}\right) = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

where

$$\mathbf{A}_{11} = \begin{pmatrix} 7 & 2 & 5 \\ -3 & 4 & 0 \end{pmatrix}, \quad \mathbf{A}_{12} = \begin{pmatrix} 8 & 4 \\ 2 & 7 \end{pmatrix},$$

$$\mathbf{A}_{21} = \begin{pmatrix} 9 & 3 & 6 \\ 3 & 1 & 2 \end{pmatrix}, \quad \mathbf{A}_{22} = \begin{pmatrix} 5 & -2 \\ 1 & 6 \end{pmatrix}.$$

If two matrices $\mathbf{A}$ and $\mathbf{B}$ are conformal for multiplication, and if $\mathbf{A}$ and $\mathbf{B}$ are partitioned so that the submatrices are appropriately conformal, then the product $\mathbf{AB}$ can be found using the usual pattern of row by column multiplication with the submatrices as if they were single elements; for example

$$\begin{aligned} \mathbf{AB} &= \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \begin{pmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix}. \end{aligned} \tag{2.35}$$

If $\mathbf{B}$ is replaced by a vector $\mathbf{b}$ partitioned into two sets of elements, and if $\mathbf{A}$ is correspondingly partitioned into two sets of columns, then (2.35) becomes

$$\mathbf{Ab} = (\mathbf{A}_1, \mathbf{A}_2) \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \end{pmatrix} = \mathbf{A}_1\mathbf{b}_1 + \mathbf{A}_2\mathbf{b}_2, \tag{2.36}$$

where the number of columns of $\mathbf{A}_1$ is equal to the number of elements of $\mathbf{b}_1$, and $\mathbf{A}_2$ and $\mathbf{b}_2$ are similarly conformal. Note that the partitioning in $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$ is indicated by a comma.

The partitioned multiplication in (2.36) can be extended to individual columns of $\mathbf{A}$ and individual elements of $\mathbf{b}$:

$$\mathbf{Ab} = (\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} = b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \cdots + b_p\mathbf{a}_p. \tag{2.37}$$

Thus **Ab** is expressible as a linear combination of the columns of **A**, in which the coefficients are elements of **b**. We illustrate (2.37) in the following example.

Example 2.3. Let

$$\mathbf{A} = \begin{pmatrix} 6 & -2 & 3 \\ 2 & 1 & 0 \\ 4 & 3 & 2 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 4 \\ 2 \\ -1 \end{pmatrix}.$$

Then

$$\mathbf{Ab} = \begin{pmatrix} 17 \\ 10 \\ 20 \end{pmatrix}.$$

Using a linear combination of columns of **A** as in (2.37), we obtain

$$\begin{aligned} \mathbf{Ab} &= b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + b_2\mathbf{a}_3 \\ &= 4\begin{pmatrix} 6 \\ 2 \\ 4 \end{pmatrix} + 2\begin{pmatrix} -2 \\ 1 \\ 3 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} \\ &= \begin{pmatrix} 24 \\ 8 \\ 16 \end{pmatrix} + \begin{pmatrix} -4 \\ 2 \\ 6 \end{pmatrix} - \begin{pmatrix} 3 \\ 0 \\ 2 \end{pmatrix} = \begin{pmatrix} 17 \\ 10 \\ 20 \end{pmatrix}. \end{aligned}$$

□

By (2.28) and (2.37), the columns of the product **AB** are linear combinations of the columns of **A**. The coefficients for the *j*th column of **AB** are the elements of the *j*th column of **B**.

The product of a row vector and a matrix, $\mathbf{a}'\mathbf{B}$, can be expressed as a linear combination of the rows of **B**, in which the coefficients are elements of $\mathbf{a}'$:

$$\mathbf{a}'\mathbf{B} = (a_1, a_2, \ldots, a_n)\begin{pmatrix} \mathbf{b}'_1 \\ \mathbf{b}'_2 \\ \vdots \\ \mathbf{b}'_n \end{pmatrix} = a_1\mathbf{b}'_1 + a_2\mathbf{b}'_2 + \cdots + a_n\mathbf{b}'_n. \tag{2.38}$$

By (2.27) and (2.38), the rows of the matrix product **AB** are linear combinations of the rows of **B**. The coefficients for the *i*th row of **AB** are the elements of the *i*th row of **A**.

Finally, we note that if a matrix $\mathbf{A}$ is partitioned as $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)$, then

$$\mathbf{A}' = (\mathbf{A}_1, \mathbf{A}_2)' = \begin{pmatrix} \mathbf{A}_1' \\ \mathbf{A}_2' \end{pmatrix}. \tag{2.39}$$

2.4 RANK

Before defining the rank of a matrix, we first introduce the notion of linear independence and dependence. A set of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ is said to be *linearly dependent* if scalars $c_1, c_2, \ldots, c_n$ (not all zero) can be found such that

$$c_1\mathbf{a}_1 + c_2\mathbf{a}_2 + \cdots + c_n\mathbf{a}_n = \mathbf{0}. \tag{2.40}$$

If no coefficients $c_1, c_2, \ldots, c_n$ can be found that satisfy (2.40), the set of vectors $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ is said to be *linearly independent*. By (2.37) this can be restated as follows. The columns of $\mathbf{A}$ are linearly independent if $\mathbf{Ac} = \mathbf{0}$ implies $\mathbf{c} = \mathbf{0}$. (If a set of vectors includes $\mathbf{0}$, the set is linearly dependent.) If (2.40) holds, then at least one of the vectors $\mathbf{a}_i$ can be expressed as a linear combination of the other vectors in the set. Among linearly independent vectors there is no redundancy of this type.

The *rank* of any square or rectangular matrix $\mathbf{A}$ is defined as

$$\begin{aligned} \text{rank}(\mathbf{A}) &= \text{number of linearly independent columns of } \mathbf{A} \\ &= \text{number of linearly independent rows of } \mathbf{A}. \end{aligned}$$

It can be shown that the number of linearly independent columns of any matrix is always equal to the number of linearly independent rows.

If a matrix $\mathbf{A}$ has a single nonzero element, with all other elements equal to 0, then $\text{rank}(\mathbf{A}) = 1$. The vector $\mathbf{0}$ and the matrix $\mathbf{O}$ have rank 0.

Suppose that a rectangular matrix $\mathbf{A}$ is $n \times p$ of rank p, where $p < n$. (We typically shorten this statement to "$\mathbf{A}$ is $n \times p$ of rank $p < n$.") Then $\mathbf{A}$ has maximum possible rank and is said to be of *full rank*. In general, the maximum possible rank of an $n \times p$ matrix $\mathbf{A}$ is $\min(n, p)$. Thus, in a rectangular matrix, the rows or columns (or both) are linearly dependent. We illustrate this in the following example.

Example 2.4a. The rank of

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 3 \\ 5 & 2 & 4 \end{pmatrix}$$

is 2 because the two rows are linearly independent (neither row is a multiple of the other). Hence, by the definition of rank, the number of linearly independent columns is also 2. Therefore, the columns are linearly dependent, and by (2.40) there exist constants c_1, c_2, and c_3 such that

$$c_1\begin{pmatrix}1\\5\end{pmatrix} + c_2\begin{pmatrix}-2\\2\end{pmatrix} + c_3\begin{pmatrix}3\\4\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix}. \tag{2.41}$$

By (2.37), we can write (2.41) in the form

$$\begin{pmatrix}1 & -2 & 3\\5 & 2 & 4\end{pmatrix}\begin{pmatrix}c_1\\c_2\\c_3\end{pmatrix} = \begin{pmatrix}0\\0\end{pmatrix} \quad \text{or} \quad \mathbf{Ac} = \mathbf{0}. \tag{2.42}$$

The solution to (2.42) is given by any multiple of $\mathbf{c} = (14, -11, -12)'$. In this case, the product $\mathbf{Ac}$ is equal to $\mathbf{0}$, even though $\mathbf{A} \neq \mathbf{O}$ and $\mathbf{c} \neq \mathbf{0}$. This is possible because of the linear dependence of the column vectors of $\mathbf{A}$. □

We can extend (2.42) to products of matrices. It is possible to find $\mathbf{A} \neq \mathbf{O}$ and $\mathbf{B} \neq \mathbf{O}$ such that

$$\mathbf{AB} = \mathbf{O}; \tag{2.43}$$

for example

$$\begin{pmatrix}1 & 2\\2 & 4\end{pmatrix}\begin{pmatrix}2 & 6\\-1 & -3\end{pmatrix} = \begin{pmatrix}0 & 0\\0 & 0\end{pmatrix}.$$

We can also exploit the linear dependence of rows or columns of a matrix to create expressions such as $\mathbf{AB} = \mathbf{CB}$, where $\mathbf{A} \neq \mathbf{C}$. Thus in a matrix equation, we cannot, in general, cancel a matrix from both sides of the equation. There are two exceptions to this rule: (1) if $\mathbf{B}$ is a full-rank square matrix, then $\mathbf{AB} = \mathbf{CB}$ implies $\mathbf{A} = \mathbf{C}$; (2) the other special case occurs when the expression holds for all possible values of the matrix common to both sides of the equation; for example

$$\text{if } \mathbf{Ax} = \mathbf{Bx} \text{ for all possible values of } \mathbf{x}, \tag{2.44}$$

then $\mathbf{A} = \mathbf{B}$. To see this, let $\mathbf{x} = (1, 0, \ldots, 0)'$. Then, by (2.37) the first column of $\mathbf{A}$ equals the first column of $\mathbf{B}$. Now let $\mathbf{x} = (0, 1, 0, \ldots, 0)'$, and the second column of $\mathbf{A}$ equals the second column of $\mathbf{B}$. Continuing in this fashion, we obtain $\mathbf{A} = \mathbf{B}$.

Example 2.4b. We illustrate the existence of matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ such that $\mathbf{AB} = \mathbf{CB}$, where $\mathbf{A} \neq \mathbf{C}$. Let

$$\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 0 & -1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 2 & 1 & 1 \\ 5 & -6 & -4 \end{pmatrix}.$$

Then

$$\mathbf{AB} = \mathbf{CB} = \begin{pmatrix} 3 & 5 \\ 1 & 4 \end{pmatrix}.$$

□

The following theorem gives a general case and two special cases for the rank of a product of two matrices.

Theorem 2.4

(i) If the matrices $\mathbf{A}$ and $\mathbf{B}$ are conformal for multiplication, then $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$ and $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B})$.

(ii) Multiplication by a full–rank square matrix does not change the rank; that is, if $\mathbf{B}$ and $\mathbf{C}$ are full–rank square matrices, $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{CA}) = \text{rank}(\mathbf{A})$.

(iii) For any matrix $\mathbf{A}$, $\text{rank}(\mathbf{A}'\mathbf{A}) = \text{rank}(\mathbf{AA}') = \text{rank}(\mathbf{A}') = \text{rank}(\mathbf{A})$.

PROOF

(i) All the columns of $\mathbf{AB}$ are linear combinations of the columns of $\mathbf{A}$ (see a comment following Example 2.3). Consequently, the number of linearly independent columns of $\mathbf{AB}$ is less than or equal to the number of linearly independent columns of $\mathbf{A}$, and $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{A})$. Similarly, all the rows of $\mathbf{AB}$ are linear combinations of the rows of $\mathbf{B}$ [see a comment following (2.38)], and therefore $\text{rank}(\mathbf{AB}) \leq \text{rank}(\mathbf{B})$.

(ii) This will be proved later.

(iii) This will also be proved later.

□

2.5 INVERSE

A full-rank square matrix is said to be *nonsingular*. A nonsingular matrix $\mathbf{A}$ has a unique *inverse*, denoted by $\mathbf{A}^{-1}$, with the property that

$$\mathbf{AA}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}. \tag{2.45}$$

If $\mathbf{A}$ is square and less than full rank, then it does not have an inverse and is said to be *singular*. Note that full-rank rectangular matrices do not have inverses as in (2.45). From the definition in (2.45), it is clear that $\mathbf{A}$ is the inverse of $\mathbf{A}^{-1}$:

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}. \tag{2.46}$$

Example 2.5. Let

$$\mathbf{A} = \begin{pmatrix} 4 & 7 \\ 2 & 6 \end{pmatrix}.$$

Then

$$\mathbf{A}^{-1} = \begin{pmatrix} .6 & -.7 \\ -.2 & .4 \end{pmatrix}$$

and

$$\begin{pmatrix} 4 & 7 \\ 2 & 6 \end{pmatrix}\begin{pmatrix} .6 & -.7 \\ -.2 & .4 \end{pmatrix} = \begin{pmatrix} .6 & -.7 \\ -.2 & .4 \end{pmatrix}\begin{pmatrix} 4 & 7 \\ 2 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \qquad \square$$

We can now prove Theorem 2.4(ii).

PROOF. If $\mathbf{B}$ is a full-rank square (nonsingular) matrix, there exists a matrix $\mathbf{B}^{-1}$ such that $\mathbf{B}\mathbf{B}^{-1} = \mathbf{I}$. Then, by Theorem 2.4(i), we have

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{B}\mathbf{B}^{-1}) \leq \text{rank}(\mathbf{A}\mathbf{B}) \leq \text{rank}(\mathbf{A}).$$

Thus both inequalities become equalities, and $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{B})$. Similarly, $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{C}\mathbf{A})$ for $\mathbf{C}$ nonsingular. $\square$

In applications, inverses are typically found by computer. Many calculators also compute inverses. Algorithms for hand calculation of inverses of small matrices can be found in texts on matrix algebra.

If $\mathbf{B}$ is nonsingular and $\mathbf{A}\mathbf{B} = \mathbf{C}\mathbf{B}$, then we can multiply on the right by $\mathbf{B}^{-1}$ to obtain $\mathbf{A} = \mathbf{C}$. (If $\mathbf{B}$ is singular or rectangular, we can't cancel it from both sides of $\mathbf{A}\mathbf{B} = \mathbf{C}\mathbf{B}$; see Example 2.4b and the paragraph preceding the example.) Similarly, if $\mathbf{A}$ is nonsingular, the system of equations $\mathbf{A}\mathbf{x} = \mathbf{c}$ has the unique solution

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}, \tag{2.47}$$

since we can multiply on the left by $\mathbf{A}^{-1}$ to obtain

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{A}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{c} \\ \mathbf{I}\mathbf{x} &= \mathbf{A}^{-1}\mathbf{c}. \end{aligned}$$

Two properties of inverses are given in the next two theorems.

Theorem 2.5a. If $\mathbf{A}$ is nonsingular, then $\mathbf{A}'$ is nonsingular and its inverse can be found as

$$(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'. \tag{2.48}$$

□

Theorem 2.5b. If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular matrices of the same size, then $\mathbf{AB}$ is nonsingular and

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{2.49}$$

□

We now give the inverses of some special matrices. If $\mathbf{A}$ is symmetric and nonsingular and is partitioned as

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

and if $\mathbf{B} = \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}$, then, provided $\mathbf{A}_{11}^{-1}$ and $\mathbf{B}^{-1}$ exist, the inverse of $\mathbf{A}$ is given by

$$\mathbf{A}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1} \\ -\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{B}^{-1} \end{pmatrix}. \tag{2.50}$$

As a special case of (2.50), consider the symmetric nonsingular matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{a}_{12} \\ \mathbf{a}_{12}' & a_{22} \end{pmatrix},$$

in which $\mathbf{A}_{11}$ is square, a_{22} is a 1×1 matrix, and $\mathbf{a}_{12}$ is a vector. Then if $\mathbf{A}_{11}^{-1}$ exists, $\mathbf{A}^{-1}$ can be expressed as

$$\mathbf{A}^{-1} = \frac{1}{b}\begin{pmatrix} b\mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{a}_{12}\mathbf{a}_{12}'\mathbf{A}_{11}^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{a}_{12} \\ -\mathbf{a}_{12}'\mathbf{A}_{11}^{-1} & 1 \end{pmatrix}, \tag{2.51}$$

where $b = a_{22} - \mathbf{a}'_{12}\mathbf{A}_{11}^{-1}\mathbf{a}_{12}$. As another special case of (2.50), we have

$$\begin{pmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22}^{-1} \end{pmatrix}. \tag{2.52}$$

If a square matrix of the form $\mathbf{B} + \mathbf{cc}'$ is nonsingular, where $\mathbf{c}$ is a vector and $\mathbf{B}$ is a nonsingular matrix, then

$$(\mathbf{B} + \mathbf{cc}')^{-1} = \mathbf{B}^{-1} - \frac{\mathbf{B}^{-1}\mathbf{cc}'\mathbf{B}^{-1}}{1 + \mathbf{c}'\mathbf{B}^{-1}\mathbf{c}}. \tag{2.53}$$

In more generality, if $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{A} + \mathbf{PBQ}$ are nonsingular, then

$$(\mathbf{A} + \mathbf{PBQ})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{PB}(\mathbf{B} + \mathbf{BQA}^{-1}\mathbf{PB})^{-1}\mathbf{BQA}^{-1}. \tag{2.54}$$

Both (2.53) and (2.54) can be easily verified (Problems 2.33 and 2.34).

2.6 POSITIVE DEFINITE MATRICES

Quadratic forms were introduced in (2.33). For example, the quadratic form $3y_1^2 + y_2^2 + 2y_3^2 + 4y_1y_2 + 5y_1y_3 - 6y_2y_3$ can be expressed as

$$3y_1^2 + y_2^2 + 2y_3^2 + 4y_1y_2 + 5y_1y_3 - 6y_2y_3 = \mathbf{y}'\mathbf{A}\mathbf{y},$$

where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}, \qquad \mathbf{A} = \begin{pmatrix} 3 & 4 & 5 \\ 0 & 1 & -6 \\ 0 & 0 & 2 \end{pmatrix}.$$

However, the same quadratic form can also be expressed in terms of the symmetric matrix

$$\frac{1}{2}(\mathbf{A} + \mathbf{A}') = \begin{pmatrix} 3 & 2 & \frac{5}{2} \\ 2 & 1 & -3 \\ \frac{5}{2} & -3 & 2 \end{pmatrix}.$$

In general, any quadratic form $\mathbf{y}'\mathbf{A}\mathbf{y}$ can be expressed as

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \mathbf{y}'\left(\frac{\mathbf{A} + \mathbf{A}'}{2}\right)\mathbf{y}, \tag{2.55}$$

and thus the matrix of a quadratic form can always be chosen to be symmetric (and thereby unique).

The sums of squares we will encounter in regression (Chapters 6–11) and analysis–of–variance (Chapters 12–15) can be expressed in the form $\mathbf{y}'\mathbf{A}\mathbf{y}$, where $\mathbf{y}$ is an observation vector. Such quadratic forms remain positive (or at least nonnegative) for all possible values of $\mathbf{y}$. We now consider quadratic forms of this type.

If the symmetric matrix $\mathbf{A}$ has the property $\mathbf{y}'\mathbf{A}\mathbf{y} > 0$ for all possible $\mathbf{y}$ except $\mathbf{y} = \mathbf{0}$, then the quadratic form $\mathbf{y}'\mathbf{A}\mathbf{y}$ is said to be *positive definite*, and $\mathbf{A}$ is said to be a *positive definite* matrix. Similarly, if $\mathbf{y}'\mathbf{A}\mathbf{y} \geq 0$ for all $\mathbf{y}$ and there is at least one $\mathbf{y} \neq \mathbf{0}$ such that $\mathbf{y}'\mathbf{A}\mathbf{y} = 0$, then $\mathbf{y}'\mathbf{A}\mathbf{y}$ and $\mathbf{A}$ are said to be *positive semidefinite*. Both types of matrices are illustrated in the following example.

Example 2.6. To illustrate a positive definite matrix, consider

$$\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 3 \end{pmatrix}$$

and the associated quadratic form

$$\mathbf{y}'\mathbf{A}\mathbf{y} = 2y_1^2 - 2y_1y_2 + 3y_2^2 = 2(y_1 - \tfrac{1}{2}y_2)^2 + \tfrac{5}{2}y_2^2,$$

which is clearly positive as long as y_1 and y_2 are not both zero.

To illustrate a positive semidefinite matrix, consider

$$(2y_1 - y_2)^2 + (3y_1 - y_3)^2 + (3y_2 - 2y_3)^2,$$

which can be expressed as $\mathbf{y}'\mathbf{A}\mathbf{y}$, with

$$\mathbf{A} = \begin{pmatrix} 13 & -2 & -3 \\ -2 & 10 & -6 \\ -3 & -6 & 5 \end{pmatrix}.$$

If $2y_1 = y_2, 3y_1 = y_3,$ and $3y_2 = 2y_3,$ then $(2y_1 - y_2)^2 + (3y_1 - y_3)^2 + (3y_2 - 2y_3)^2 = 0$. Thus $\mathbf{y}'\mathbf{A}\mathbf{y} = 0$ for any multiple of $\mathbf{y} = (1, 2, 3)'$. Otherwise $\mathbf{y}'\mathbf{A}\mathbf{y} > 0$ (except for $\mathbf{y} = \mathbf{0}$). □

In the matrices in Example 2.6, the diagonal elements are positive. For positive definite matrices, this is true in general.

Theorem 2.6a

(i) If $\mathbf{A}$ is positive definite, then all its diagonal elements a_{ii} are positive.
(ii) If $\mathbf{A}$ is positive semidefinite, then all $a_{ii} \geq 0$.

PROOF

(i) Let $\mathbf{y}' = (0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the ith position and 0's elsewhere. Then $\mathbf{y}'\mathbf{A}\mathbf{y} = a_{ii} > 0$.
(ii) Let $\mathbf{y}' = (0, \ldots, 0, 1, 0, \ldots, 0)$ with a 1 in the ith position and 0's elsewhere. Then $\mathbf{y}'\mathbf{A}\mathbf{y} = a_{ii} \geq 0$. □

Some additional properties of positive definite and positive semidefinite matrices are given in the following theorems.

Theorem 2.6b. Let $\mathbf{P}$ be a nonsingular matrix.

(i) If $\mathbf{A}$ is positive definite, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is positive definite.
(ii) If $\mathbf{A}$ is positive semidefinite, then $\mathbf{P}'\mathbf{A}\mathbf{P}$ is positive semidefinite.

PROOF

(i) To show that $\mathbf{y}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{y} > 0$ for $\mathbf{y} \neq \mathbf{0}$, note that $\mathbf{y}'(\mathbf{P}'\mathbf{A}\mathbf{P})\mathbf{y} = (\mathbf{P}\mathbf{y})'\mathbf{A}(\mathbf{P}\mathbf{y})$. Since $\mathbf{A}$ is positive definite, $(\mathbf{P}\mathbf{y})'\mathbf{A}(\mathbf{P}\mathbf{y}) > 0$ provided that $\mathbf{P}\mathbf{y} \neq \mathbf{0}$. By (2.47), $\mathbf{P}\mathbf{y} = \mathbf{0}$ only if $\mathbf{y} = \mathbf{0}$, since $\mathbf{P}^{-1}\mathbf{P}\mathbf{y} = \mathbf{P}^{-1}\mathbf{0} = \mathbf{0}$. Thus $\mathbf{y}'\mathbf{P}'\mathbf{A}\mathbf{P}\mathbf{y} > 0$ if $\mathbf{y} \neq \mathbf{0}$.
(ii) See problem 2.36. □

Corollary 1. Let $\mathbf{A}$ be a $p \times p$ positive definite matrix and let $\mathbf{B}$ be a $k \times p$ matrix of rank $k \leq p$. Then $\mathbf{B}\mathbf{A}\mathbf{B}'$ is positive definite. □

Corollary 2. Let $\mathbf{A}$ be a $p \times p$ positive definite matrix and let $\mathbf{B}$ be a $k \times p$ matrix. If $k > p$ or if $\text{rank}(\mathbf{B}) = r$, where $r < k$ and $r < p$, then $\mathbf{B}\mathbf{A}\mathbf{B}'$ is positive semidefinite. □

Theorem 2.6c. A symmetric matrix $\mathbf{A}$ is positive definite if and only if there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{P}'\mathbf{P}$.

PROOF. We prove the "if" part only. Suppose $\mathbf{A} = \mathbf{P}'\mathbf{P}$ for nonsingular $\mathbf{P}$. Then

$$\mathbf{y}'\mathbf{A}\mathbf{y} = \mathbf{y}'\mathbf{P}'\mathbf{P}\mathbf{y} = (\mathbf{P}\mathbf{y})'(\mathbf{P}\mathbf{y}).$$

This is a sum of squares [see (2.20)] and is positive unless $\mathbf{P}\mathbf{y} = \mathbf{0}$. By (2.47), $\mathbf{P}\mathbf{y} = \mathbf{0}$ only if $\mathbf{y} = \mathbf{0}$. □

Corollary 1. A positive definite matrix is nonsingular. □

One method of factoring a positive definite matrix $\mathbf{A}$ into a product $\mathbf{P}'\mathbf{P}$ as in Theorem 2.6c is provided by the Cholesky decomposition (Seber and Lee 2003, pp. 335–337), by which $\mathbf{A}$ can be factored uniquely into $\mathbf{A} = \mathbf{T}'\mathbf{T}$, where $\mathbf{T}$ is a nonsingular upper triangular matrix.

For any square or rectangular matrix $\mathbf{B}$, the matrix $\mathbf{B}'\mathbf{B}$ is positive definite or positive semidefinite.

Theorem 2.6d. Let $\mathbf{B}$ be an $n \times p$ matrix.

(i) If $\text{rank}(\mathbf{B}) = p$, then $\mathbf{B}'\mathbf{B}$ is positive definite.
(ii) If $\text{rank}(\mathbf{B}) < p$, then $\mathbf{B}'\mathbf{B}$ is positive semidefinite.

PROOF

(i) To show that $\mathbf{y}'\mathbf{B}'\mathbf{B}\mathbf{y} > 0$ for $\mathbf{y} \neq \mathbf{0}$, we note that

$$\mathbf{y}'\mathbf{B}'\mathbf{B}\mathbf{y} = (\mathbf{B}\mathbf{y})'(\mathbf{B}\mathbf{y}),$$

which is a sum of squares and is thereby positive unless $\mathbf{B}\mathbf{y} = \mathbf{0}$. By (2.37), we can express $\mathbf{B}\mathbf{y}$ in the form

$$\mathbf{B}\mathbf{y} = y_1\mathbf{b}_1 + y_2\mathbf{b}_2 + \cdots + y_p\mathbf{b}_p.$$

This linear combination is not $\mathbf{0}$ (for any $\mathbf{y} \neq \mathbf{0}$) because $\text{rank}(\mathbf{B}) = p$, and the columns of $\mathbf{B}$ are therefore linearly independent [see (2.40)].

(ii) If $\text{rank}(\mathbf{B}) < p$, then we can find $\mathbf{y} \neq \mathbf{0}$ such that

$$\mathbf{B}\mathbf{y} = y_1\mathbf{b}_1 + y_2\mathbf{b}_2 + \cdots + y_p\mathbf{b}_p = \mathbf{0}$$

since the columns of $\mathbf{B}$ are linearly dependent [see (2.40)]. Hence $\mathbf{y}'\mathbf{B}'\mathbf{B}\mathbf{y} \geq 0$.

□

Note that if $\mathbf{B}$ is a square matrix, the matrix $\mathbf{BB} = \mathbf{B}^2$ is not necessarily positive semidefinite. For example, let

$$\mathbf{B} = \begin{pmatrix} 1 & -2 \\ 1 & -2 \end{pmatrix}.$$

Then

$$\mathbf{B}^2 = \begin{pmatrix} -1 & 2 \\ -1 & 2 \end{pmatrix}, \quad \mathbf{B}'\mathbf{B} = \begin{pmatrix} 2 & -4 \\ -4 & 8 \end{pmatrix}.$$

In this case, $\mathbf{B}^2$ is not positive semidefinite, but $\mathbf{B}'\mathbf{B}$ is positive semidefinite, since $\mathbf{y}'\mathbf{B}'\mathbf{B}\mathbf{y} = 2(y_1 - 2y_2)^2$.

Two additional properties of positive definite matrices are given in the following theorems.

Theorem 2.6e. If $\mathbf{A}$ is positive definite, then $\mathbf{A}^{-1}$ is positive definite.

Proof. By Theorem 2.6c, $\mathbf{A} = \mathbf{P}'\mathbf{P}$, where $\mathbf{P}$ is nonsingular. By Theorems 2.5a and 2.5b, $\mathbf{A}^{-1} = (\mathbf{P}'\mathbf{P})^{-1} = \mathbf{P}^{-1}(\mathbf{P}')^{-1} = \mathbf{P}^{-1}(\mathbf{P}^{-1})'$, which is positive definite by Theorem 2.6c. □

Theorem 2.6f. If $\mathbf{A}$ is positive definite and is partitioned in the form

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

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square, then $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are positive definite.

Proof. We can write $\mathbf{A}_{11}$, for example, as $\mathbf{A}_{11} = (\mathbf{I}, \mathbf{O})\mathbf{A}\begin{pmatrix} \mathbf{I} \\ \mathbf{O} \end{pmatrix}$, where $\mathbf{I}$ is the same size as $\mathbf{A}_{11}$. Then by Corollary 1 to Theorem 2.6b, $\mathbf{A}_{11}$ is positive definite. □

2.7 SYSTEMS OF EQUATIONS

The system of n (linear) equations in p unknowns

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p &= c_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p &= c_2 \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{np}x_p &= c_n \end{aligned} \tag{2.56}$$

can be written in matrix form as

$$\mathbf{Ax} = \mathbf{c}, \tag{2.57}$$

where $\mathbf{A}$ is $n \times p$, $\mathbf{x}$ is $p \times 1$, and $\mathbf{c}$ is $n \times 1$. Note that if $n \neq p$, $\mathbf{x}$ and $\mathbf{c}$ are of different sizes. If $n = p$ and $\mathbf{A}$ is nonsingular, then by (2.47), there exists a unique solution vector $\mathbf{x}$ obtained as $\mathbf{x} = \mathbf{A}^{-1}\mathbf{c}$. If $n > p$, so that $\mathbf{A}$ has more rows than columns, then $\mathbf{Ax} = \mathbf{c}$ typically has no solution. If $n < p$, so that $\mathbf{A}$ has fewer rows than columns, then $\mathbf{Ax} = \mathbf{c}$ typically has an infinite number of solutions.

If the system of equations $\mathbf{Ax} = \mathbf{c}$ has one or more solution vectors, it is said to be *consistent*. If the system has no solution, it is said to be *inconsistent*.

To illustrate the structure of a consistent system of equations $\mathbf{Ax} = \mathbf{c}$, suppose that $\mathbf{A}$ is $p \times p$ of rank $r < p$. Then the rows of $\mathbf{A}$ are linearly dependent, and there exists some $\mathbf{b}$ such that [see (2.38)]

$$\mathbf{b}'\mathbf{A} = b_1\mathbf{a}_1' + b_2\mathbf{a}_2' + \cdots + b_p\mathbf{a}_p' = \mathbf{0}'.$$

Then we must also have $\mathbf{b}'\mathbf{c} = b_1c_1 + b_2c_2 + \cdots + b_pc_p = 0$, since multiplication of $\mathbf{Ax} = \mathbf{c}$ by $\mathbf{b}'$ gives $\mathbf{b}'\mathbf{Ax} = \mathbf{b}'\mathbf{c}$, or $\mathbf{0}'\mathbf{x} = \mathbf{b}'\mathbf{c}$. Otherwise, if $\mathbf{b}'\mathbf{c} \neq 0$, there is no $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{c}$. Hence, in order for $\mathbf{Ax} = \mathbf{c}$ to be consistent, the same linear relationships, if any, that exist among the rows of $\mathbf{A}$ must exist among the elements (rows) of $\mathbf{c}$. This is formalized by comparing the rank of $\mathbf{A}$ with the rank of the *augmented matrix* ($\mathbf{A}$, $\mathbf{c}$). The notation ($\mathbf{A}$, $\mathbf{c}$) indicates that $\mathbf{c}$ has been appended to $\mathbf{A}$ as an additional column.

Theorem 2.7 The system of equations $\mathbf{Ax} = \mathbf{c}$ has at least one solution vector $\mathbf{x}$ if and only if rank($\mathbf{A}$) = rank($\mathbf{A}$, $\mathbf{c}$).

PROOF. Suppose that rank($\mathbf{A}$) = rank($\mathbf{A}$, $\mathbf{c}$), so that appending $\mathbf{c}$ does not change the rank. Then $\mathbf{c}$ is a linear combination of the columns of $\mathbf{A}$; that is, there exists some $\mathbf{x}$ such that

$$x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_p\mathbf{a}_p = \mathbf{c},$$

which, by (2.37), can be written as $\mathbf{Ax} = \mathbf{c}$. Thus $\mathbf{x}$ is a solution.

Conversely, suppose that there exists a solution vector $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{c}$. In general, rank ($\mathbf{A}$) $\leq$ rank($\mathbf{A}, \mathbf{c}$) (Harville 1997, p. 41). But since there exists an $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{c}$, we have

$$\begin{aligned} \text{rank}(\mathbf{A}, \mathbf{c}) &= \text{rank}(\mathbf{A}, \mathbf{Ax}) = \text{rank}[\mathbf{A}(\mathbf{I}, \mathbf{x})] \\ &\leq \text{rank}(\mathbf{A}) \qquad \text{[by Theorem 2.4(i)]}. \end{aligned}$$

Hence

$$\text{rank}(\mathbf{A}) \leq \text{rank}(\mathbf{A}, \mathbf{c}) \leq \text{rank}(\mathbf{A}),$$

and we have $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}, \mathbf{c})$. □

A consistent system of equations can be solved by the usual methods given in elementary algebra courses for eliminating variables, such as adding a multiple of one equation to another or solving for a variable and substituting into another equation. In the process, one or more variables may end up as arbitrary constants, thus generating an infinite number of solutions. A method of solution involving generalized inverses is given in Section 2.8.2. Some illustrations of systems of equations and their solutions are given in the following examples.

Example 2.7a. Consider the system of equations

$$\begin{aligned} x_1 + 2x_2 &= 4 \\ x_1 - x_2 &= 1 \\ x_1 + x_2 &= 3 \end{aligned}$$

or

$$\begin{pmatrix} 1 & 2 \\ 1 & -1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}.$$

The augmented matrix is

$$(\mathbf{A}, \mathbf{c}) = \begin{pmatrix} 1 & 2 & 4 \\ 1 & -1 & 1 \\ 1 & 1 & 3 \end{pmatrix},$$

which has rank = 2 because the third column is equal to twice the first column plus the second:

$$2\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 3 \end{pmatrix}.$$

Since $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}, \mathbf{c}) = 2$, there is at least one solution. If we add twice the first equation to the second, the result is a multiple of the third equation. Thus the third equation is redundant, and the first two can readily be solved to obtain the unique solution $\mathbf{x} = (2, 1)'$.

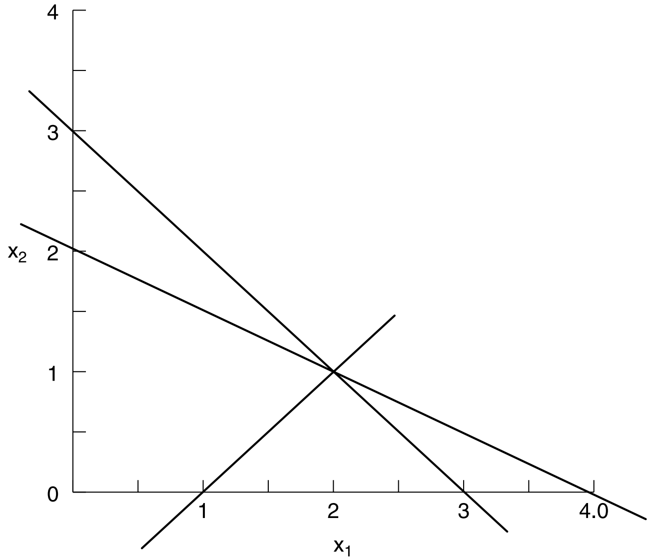

Figure 2.1 Three lines representing the three equations in Example 2.7a.

The three lines representing the three equations are plotted in Figure 2.1. Notice that the three lines intersect at the point (2, 1), which is the unique solution of the three equations. □

Example 2.7b. If we change the 3 to 2 in the third equation in Example 2.7, the augmented matrix becomes

$$(\mathbf{A}, \mathbf{c}) = \begin{pmatrix} 1 & 2 & 4 \\ 1 & -1 & 1 \\ 1 & 1 & 2 \end{pmatrix},$$

which has rank = 3, since no linear combination of columns is $\mathbf{0}$. [Alternatively, $|(\mathbf{A}, \mathbf{c})| \neq 0$, and $(\mathbf{A}, \mathbf{c})$ is nonsingular; see Theorem 2.9(iii)] Hence rank $(\mathbf{A}, \mathbf{c}) = 3 \neq \text{rank}(\mathbf{A}) = 2$, and the system is inconsistent.

The three lines representing the three equations are plotted in Figure 2.2, in which we see that the three lines do not have a common point of intersection. [For the "best" approximate solution, one approach is to use least squares; that is, we find the values of x_1 and x_2 that minimize $(x_1 + 2x_2 - 4)^2 + (x_1 - x_2 - 1)^2 + (x_1 + x_2 - 2)^2$.] □

Example 2.7c. Consider the system

$$\begin{aligned} x_1 + x_2 + x_3 &= 1 \\ 2x_1 + x_2 + 3x_3 &= 5 \\ 3x_1 + 2x_2 + 4x_3 &= 6. \end{aligned}$$

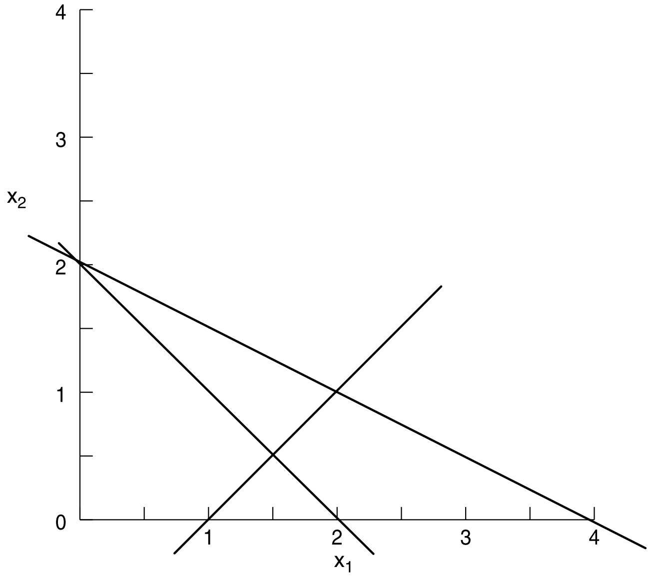

Figure 2.2 Three lines representing the three equations in Example 2.7b.

The third equation is the sum of the first two, but the second is not a multiple of the first. Thus, $\text{rank}(\mathbf{A}, \mathbf{c}) = \text{rank}(\mathbf{A}) = 2$, and the system is consistent.

By solving the first two equations for x_1 and x_2 in terms of x_3, we obtain

$$\begin{aligned} x_1 &= -2x_3 + 4 \\ x_2 &= x_3 - 3. \end{aligned}$$

The solution vector can be expressed as

$$\mathbf{x} = \begin{pmatrix} -2x_3 + 4 \\ x_3 - 3 \\ x_3 \end{pmatrix} = x_3 \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} 4 \\ -3 \\ 0 \end{pmatrix},$$

where x_3 is an arbitrary constant. Geometrically, $\mathbf{x}$ is the line representing the intersection of the two planes corresponding to the first two equations. □

2.8 GENERALIZED INVERSE

We now consider generalized inverses of those matrices that do not have inverses in the usual sense [see (2.45)]. A solution of a consistent system of equations $\mathbf{Ax} = \mathbf{c}$ can be expressed in terms of a generalized inverse of $\mathbf{A}$.

2.8.1 Definition and Properties

A *generalized inverse* of an $n \times p$ matrix $\mathbf{A}$ is any matrix $\mathbf{A}^-$ that satisfies

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}. \tag{2.58}$$

A generalized inverse is not unique except when $\mathbf{A}$ is nonsingular, in which case $\mathbf{A}^- = \mathbf{A}^{-1}$. A generalized inverse is also called a *conditional inverse*.

Every matrix, whether square or rectangular, has a generalized inverse. This holds even for vectors. For example, let

$$\mathbf{x} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Then $\mathbf{x}_1^- = (1, 0, 0, 0)$ is a generalized inverse of $\mathbf{x}$ satisfying (2.58). Other examples are $\mathbf{x}_2^- = (0, \frac{1}{2}, 0, 0), \mathbf{x}_3^- = (0, 0, \frac{1}{3}, 0)$, and $\mathbf{x}_4^- = (0, 0, 0, \frac{1}{4})$. For each $\mathbf{x}_i^-$, we have

$$\mathbf{x}\mathbf{x}_i^-\mathbf{x} = \mathbf{x}1 = \mathbf{x}, \quad i = 1, 2, 3, 4.$$

In this illustration, $\mathbf{x}$ is a column vector and $\mathbf{x}_i^-$ is a row vector. This pattern is generalized in the following theorem.

Theorem 2.8a. If $\mathbf{A}$ is $n \times p$, any generalized inverse $\mathbf{A}^-$ is $p \times n$. □

In the following example we give two illustrations of generalized inverses of a singular matrix.

Example 2.8.1. Let

$$\mathbf{A} = \begin{pmatrix} 2 & 2 & 3 \\ 1 & 0 & 1 \\ 3 & 2 & 4 \end{pmatrix}. \tag{2.59}$$

The third row of $\mathbf{A}$ is the sum of the first two rows, and the second row is not a multiple of the first; hence $\mathbf{A}$ has rank 2. Let

$$\mathbf{A}_1^- = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{A}_2^- = \begin{pmatrix} 0 & 1 & 0 \\ 0 & -\frac{3}{2} & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}. \tag{2.60}$$

It is easily verified that $\mathbf{A}\mathbf{A}_1^-\mathbf{A} = \mathbf{A}$ and $\mathbf{A}\mathbf{A}_2^-\mathbf{A} = \mathbf{A}$. □

The methods used to obtain $\mathbf{A}_1^-$ and $\mathbf{A}_2^-$ in (2.60) are described in Theorem 2.8b and the five-step algorithm following the theorem.

Theorem 2.8b. Suppose $\mathbf{A}$ is $n \times p$ of rank r and that $\mathbf{A}$ is partitioned as

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

where $\mathbf{A}_{11}$ is $r \times r$ of rank r. Then a generalized inverse of $\mathbf{A}$ is given by

$$\mathbf{A}^- = \begin{pmatrix} \mathbf{A}_{11}^{-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix},$$

where the three $\mathbf{O}$ matrices are of appropriate sizes so that $\mathbf{A}^-$ is $p \times n$.

PROOF. By multiplication of partitioned matrices, as in (2.35), we obtain

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \begin{pmatrix} \mathbf{I} & \mathbf{O} \\ \mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{O} \end{pmatrix}\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix}.$$

To show that $\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{A}_{22}$, multiply $\mathbf{A}$ by

$$\mathbf{B} = \begin{pmatrix} \mathbf{I} & \mathbf{O} \\ -\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{I} \end{pmatrix},$$

where $\mathbf{O}$ and $\mathbf{I}$ are of appropriate sizes, to obtain

$$\mathbf{B}\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{O} & \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix}.$$

The matrix $\mathbf{B}$ is nonsingular, and the rank of $\mathbf{BA}$ is therefore $r = \text{rank}(\mathbf{A})$ [see Theorem 2.4(ii)]. In $\mathbf{BA}$, the submatrix $\begin{pmatrix} \mathbf{A}_{11} \\ \mathbf{O} \end{pmatrix}$ is of rank r, and the columns headed by $\mathbf{A}_{12}$ are therefore linear combinations of the columns headed by $\mathbf{A}_{11}$. By a comment following Example 2.3, this relationship can be expressed as

$$\begin{pmatrix} \mathbf{A}_{12} \\ \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} \\ \mathbf{O} \end{pmatrix}\mathbf{Q} \tag{2.61}$$

for some matrix $\mathbf{Q}$. By (2.27), the right side of (2.61) becomes

$$\begin{pmatrix} \mathbf{A}_{11} \\ \mathbf{O} \end{pmatrix} \mathbf{Q} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{Q} \\ \mathbf{O}\mathbf{Q} \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{Q} \\ \mathbf{O} \end{pmatrix}.$$

Thus $\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12} = \mathbf{O}$, or

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

□

Corollary 1. Suppose that $\mathbf{A}$ is $n \times p$ of rank r and that $\mathbf{A}$ is partitioned as in Theorem 2.8b, where $\mathbf{A}_{22}$ is $r \times r$ of rank r. Then a generalized inverse of $\mathbf{A}$ is given by

$$\mathbf{A}^{-} = \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22}^{-1} \end{pmatrix},$$

where the three $\mathbf{O}$ matrices are of appropriate sizes so that $\mathbf{A}^{-}$ is $p \times n$. □

The nonsingular submatrix need not be in the $\mathbf{A}_{11}$ or $\mathbf{A}_{22}$ position, as in Theorem 2.8b or its corollary. Theorem 2.8b can be extended to the following algorithm for finding a conditional inverse $\mathbf{A}^{-}$ for any $n \times p$ matrix $\mathbf{A}$ of rank r (Searle 1982, p. 218):

1. Find any nonsingular $r \times r$ submatrix $\mathbf{C}$. It is not necessary that the elements of $\mathbf{C}$ occupy adjacent rows and columns in $\mathbf{A}$.
2. Find $\mathbf{C}^{-1}$ and $(\mathbf{C}^{-1})'$.
3. Replace the elements of $\mathbf{C}$ by the elements of $(\mathbf{C}^{-1})'$.
4. Replace all other elements in $\mathbf{A}$ by zeros.
5. Transpose the resulting matrix.

Some properties of generalized inverses are given in the following theorem, which is the theoretical basis for many of the results in Chapter 11.

Theorem 2.8c. Let $\mathbf{A}$ be $n \times p$ of rank r, let $\mathbf{A}^{-}$ be any generalized inverse of $\mathbf{A}$, and let $(\mathbf{A}'\mathbf{A})^{-}$ be any generalized inverse of $\mathbf{A}'\mathbf{A}$. Then

(i) $\text{rank}(\mathbf{A}^{-}\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}^{-}) = \text{rank}(\mathbf{A}) = r$.
(ii) $(\mathbf{A}^{-})'$ is a generalized inverse of $\mathbf{A}'$; that is, $(\mathbf{A}')^{-} = (\mathbf{A}^{-})'$.
(iii) $\mathbf{A} = \mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'\mathbf{A}$ and $\mathbf{A}' = \mathbf{A}'\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$.
(iv) $(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is a generalized inverse of $\mathbf{A}$; that is, $\mathbf{A}^{-} = (\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$.

(v) $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ is symmetric, has rank $= r$, and is invariant to the choice of $(\mathbf{A}'\mathbf{A})^{-}$; that is, $\mathbf{A}(\mathbf{A}'\mathbf{A})^{-}\mathbf{A}'$ remains the same, no matter what value of $(\mathbf{A}'\mathbf{A})^{-}$ is used. □

A generalized inverse of a symmetric matrix is not necessarily symmetric. However, it is also true that a symmetric generalized inverse can always be found for a symmetric matrix; see Problem 2.46. In this book, we will assume that generalized inverses of symmetric matrices are symmetric.

2.8.2 Generalized Inverses and Systems of Equations

Generalized inverses can be used to find solutions to a system of equations.

Theorem 2.8d. If the system of equations $\mathbf{Ax} = \mathbf{c}$ is consistent and if $\mathbf{A}^{-}$ is any generalized inverse for $\mathbf{A}$, then $\mathbf{x} = \mathbf{A}^{-}\mathbf{c}$ is a solution.

PROOF. Since $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$, we have

$$\mathbf{AA}^{-}\mathbf{Ax} = \mathbf{Ax}.$$

Substituting $\mathbf{Ax} = \mathbf{c}$ on both sides, we obtain

$$\mathbf{AA}^{-}\mathbf{c} = \mathbf{c}.$$

Writing this in the form $\mathbf{A}(\mathbf{A}^{-}\mathbf{c}) = \mathbf{c}$, we see that $\mathbf{A}^{-}\mathbf{c}$ is a solution to $\mathbf{Ax} = \mathbf{c}$. □

Different choices of $\mathbf{A}^{-}$ will result in different solutions for $\mathbf{Ax} = \mathbf{c}$.

Theorem 2.8e. If the system of equations $\mathbf{Ax} = \mathbf{c}$ is consistent, then all possible solutions can be obtained in the following two ways:

(i) Use a specific $\mathbf{A}^{-}$ in $\mathbf{x} = \mathbf{A}^{-}\mathbf{c} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{h}$, and use all possible values of the arbitrary vector $\mathbf{h}$.

(ii) Use all possible values of $\mathbf{A}^{-}$ in $\mathbf{x} = \mathbf{A}^{-}\mathbf{c}$ if $\mathbf{c} \neq \mathbf{0}$.

PROOF. See Searle (1982, p. 238). □

A necessary and sufficient condition for the system of equations $\mathbf{Ax} = \mathbf{c}$ to be consistent can be given in terms of a generalized inverse of $\mathbf{A}$ (Graybill 1976, p. 36).

Theorem 2.8f. The system of equations $\mathbf{Ax} = \mathbf{c}$ has a solution if and only if for any generalized inverse $\mathbf{A}^-$ of $\mathbf{A}$

$$\mathbf{AA}^-\mathbf{c} = \mathbf{c}.$$

Proof. Suppose that $\mathbf{Ax} = \mathbf{c}$ is consistent. Then, by Theorem 2.8d, $\mathbf{x} = \mathbf{A}^-\mathbf{c}$ is a solution. Multiply $\mathbf{c} = \mathbf{Ax}$ by $\mathbf{AA}^-$ to obtain

$$\mathbf{AA}^-\mathbf{c} = \mathbf{AA}^-\mathbf{Ax} = \mathbf{Ax} = \mathbf{c}.$$

Conversely, suppose $\mathbf{AA}^-\mathbf{c} = \mathbf{c}$. Multiply $\mathbf{x} = \mathbf{A}^-\mathbf{c}$ by $\mathbf{A}$ to obtain

$$\mathbf{Ax} = \mathbf{AA}^-\mathbf{c} = \mathbf{c}.$$

Hence, a solution exists, namely, $\mathbf{x} = \mathbf{A}^-\mathbf{c}$. □

Theorem 2.8f provides an alternative to Theorem 2.7a for determining whether a system of equations is consistent.

2.9 DETERMINANTS

The *determinant* of an $n \times n$ matrix $\mathbf{A}$ is a scalar function of $\mathbf{A}$ defined as the sum of all $n!$ possible products of n elements such that

1. each product contains one element from every row and every column of $\mathbf{A}$.
2. the factors in each product are written so that the column subscripts appear in order of magnitude and each product is then preceded by a plus or minus sign according to whether the number of inversions in the row subscripts is even or odd. (An *inversion* occurs whenever a larger number precedes a smaller one.)

The determinant of $\mathbf{A}$ is denoted by $|\mathbf{A}|$ or $\det(\mathbf{A})$. The preceding definition is not very useful in evaluating determinants, except in the case of 2×2 or 3×3 matrices. For larger matrices, determinants are typically found by computer. Some calculators also evaluate determinants.

The determinants of some special square matrices are given in the following theorem.

Theorem 2.9a.

(i) If $\mathbf{D} = \text{diag}(d_1, d_2, \ldots, d_n)$, $\quad |\mathbf{D}| = \prod_{i=1}^n d_i$.

(ii) The determinant of a triangular matrix is the product of the diagonal elements.
(iii) If $\mathbf{A}$ is singular, $|\mathbf{A}| = 0$.
(iv) If $\mathbf{A}$ is nonsingular, $|\mathbf{A}| \neq 0$.
(v) If $\mathbf{A}$ is positive definite, $|\mathbf{A}| > 0$.
(vi) $|\mathbf{A}'| = |\mathbf{A}|$.
(vii) If $\mathbf{A}$ is nonsingular, $|\mathbf{A}^{-1}| = \dfrac{1}{|\mathbf{A}|}$.

□

Example 2.9a. We illustrate each of the properties in Theorem 2.9a.

(i) diagonal: $\begin{vmatrix} 2 & 0 \\ 0 & 3 \end{vmatrix} = (2)(3) - (0)(0) = (2)(3)$.

(ii) triangular: $\begin{vmatrix} 2 & 1 \\ 0 & 3 \end{vmatrix} = (2)(3) - (0)(1) = (2)(3)$.

(iii) singular: $\begin{vmatrix} 1 & 2 \\ 3 & 6 \end{vmatrix} = (1)(6) - (3)(2) = 0$,

nonsingular: $\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = (1)(4) - (3)(2) = -2$.

(iv) positive definite: $\begin{vmatrix} 3 & -2 \\ -2 & 4 \end{vmatrix} = (3)(4) - (-2)(-2) = 8 > 0$.

(v) transpose: $\begin{vmatrix} 3 & -7 \\ 2 & 1 \end{vmatrix} = (3)(1) - (2)(-7) = 17$,

$\begin{vmatrix} 3 & 2 \\ -7 & 1 \end{vmatrix} = (3)(1) - (-7)(2) = 17$.

(vi) inverse:

$$\begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} .4 & -.2 \\ -.1 & .3 \end{pmatrix}, \quad \begin{vmatrix} 3 & 2 \\ 1 & 4 \end{vmatrix} = 10, \quad \begin{vmatrix} .4 & -.2 \\ -.1 & .3 \end{vmatrix} = .1.$$

□

As a special case of (62), suppose that all diagonal elements are equal, say, $\mathbf{D} = \text{diag}(c, c, \ldots, c) = c\mathbf{I}$. Then

$$|\mathbf{D}| = |c\mathbf{I}| = \prod_{i=1}^{n} c = c^n. \tag{2.68}$$

By extension, if an $n \times n$ matrix is multiplied by a scalar, the determinant becomes

$$|c\mathbf{A}| = c^n |\mathbf{A}|. \tag{2.69}$$

The determinant of certain partitioned matrices is given in the following theorem.

Theorem 2.9b. If the square matrix $\mathbf{A}$ is partitioned as

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}, \tag{2.70}$$

and if $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square and nonsingular (but not necessarily the same size), then

$$|\mathbf{A}| = |\mathbf{A}_{11}|\,|\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}| \tag{2.71}$$

$$= |\mathbf{A}_{22}|\,|\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}|. \tag{2.72}$$

□

Note the analogy of (2.71) and (2.72) to the case of the determinant of a 2×2 matrix:

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} &= a_{11}a_{22} - a_{21}a_{12} \\ &= a_{11}\left(a_{22} - \frac{a_{21}a_{12}}{a_{11}}\right) \\ &= a_{22}\left(a_{11} - \frac{a_{12}a_{21}}{a_{22}}\right). \end{aligned}$$

Corollary 1. Suppose

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix} \quad \text{or} \quad \mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{O} & \mathbf{A}_{22} \end{pmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square (but not necessarily the same size). Then in either case

$$|\mathbf{A}| = |\mathbf{A}_{11}|\,|\mathbf{A}_{22}|. \tag{2.73}$$

□

Corollary 2. Let

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22} \end{pmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square (but not necessarily the same size). Then

$$|\mathbf{A}| = |\mathbf{A}_{11}|\,|\mathbf{A}_{22}|. \tag{2.74}$$

□

Corollary 3. If $\mathbf{A}$ has the form $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{a}_{12} \\ \mathbf{a}'_{12} & a_{22} \end{pmatrix}$, where $\mathbf{A}_{11}$ is a nonsingular matrix, $\mathbf{a}_{12}$ is a vector, and a_{22} is a 1×1 matrix, then

$$|\mathbf{A}| = \begin{vmatrix} \mathbf{A}_{11} & \mathbf{a}_{12} \\ \mathbf{a}'_{12} & a_{22} \end{vmatrix} = |\mathbf{A}_{11}|(a_{22} - \mathbf{a}'_{12}\mathbf{A}_{11}^{-1}\mathbf{a}_{12}). \tag{2.75}$$

□

Corollary 4. If $\mathbf{A}$ has the form $\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{c} \\ -\mathbf{c}' & 1 \end{pmatrix}$, where $\mathbf{c}$ is a vector and $\mathbf{B}$ is a nonsingular matrix, then

$$|\mathbf{B} + \mathbf{c}\mathbf{c}'| = |\mathbf{B}|(1 + \mathbf{c}'\mathbf{B}^{-1}\mathbf{c}). \tag{2.76}$$

□

The determinant of the product of two square matrices is given in the following theorem.

Theorem 2.9c. If $\mathbf{A}$ and $\mathbf{B}$ are square and the same size, then the determinant of the product is the product of the determinants:

$$|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}|. \tag{2.77}$$

□

Corollary 1

$$|\mathbf{AB}| = |\mathbf{BA}|. \tag{2.78}$$

□

Corollary 2

$$|\mathbf{A}^2| = |\mathbf{A}|^2. \tag{2.79}$$

□

Example 2.9b. To illustrate Theorem 2.9c, let

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \begin{pmatrix} 3 & -2 \\ 1 & 2 \end{pmatrix}.$$

Then

$$\mathbf{AB} = \begin{pmatrix} 5 & 2 \\ 13 & 2 \end{pmatrix}, \qquad |\mathbf{AB}| = -16,$$

$$|\mathbf{A}| = -2, \quad |\mathbf{B}| = 8, \quad |\mathbf{A}|\,|\mathbf{B}| = -16.$$

□

2.10 ORTHOGONAL VECTORS AND MATRICES

Two $n \times 1$ vectors **b** and **b** are said to be *orthogonal* if

$$\mathbf{a}'\mathbf{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n = 0. \tag{2.80}$$

Note that the term *orthogonal* applies to *two* vectors, not to a single vector.

Geometrically, two orthogonal vectors are perpendicular to each other. This is illustrated in Figure 2.3 for the vectors $\mathbf{x}_1 = (4, 2)'$ and $\mathbf{x}_2 = (-1, 2)'$. Note that $\mathbf{x}_1'\mathbf{x}_2 = (4)(-1) + (2)(2) = 0$.

To show that two orthogonal vectors are perpendicular, let θ be the angle between vectors **a** and **b** in Figure 2.4. The vector from the terminal point of **a** to the terminal point of **b** can be represented as $\mathbf{c} = \mathbf{b} - \mathbf{a}$. The law of cosines for the relationship of

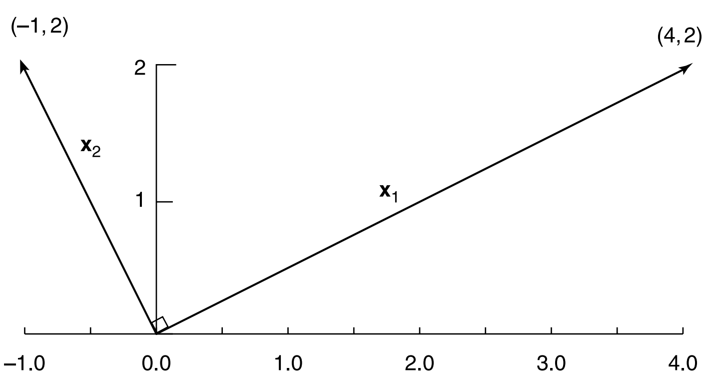

Figure 2.3 Two orthogonal (perpendicular) vectors.

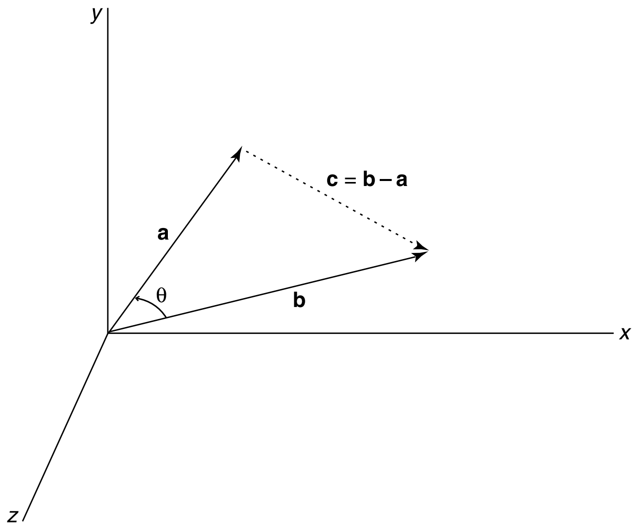

Figure 2.4 Vectors **a** and **b** in 3-space.

θ to the sides of the triangle can be stated in vector form as

$$\cos\theta = \frac{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b} - (\mathbf{b}-\mathbf{a})'(\mathbf{b}-\mathbf{a})}{2\sqrt{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b})}}$$
$$= \frac{\mathbf{a}'\mathbf{a} + \mathbf{b}'\mathbf{b} - (\mathbf{b}'\mathbf{b} + \mathbf{a}'\mathbf{a} - 2\mathbf{a}'\mathbf{b})}{2\sqrt{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b})}}$$
$$= \frac{\mathbf{a}'\mathbf{b}}{\sqrt{(\mathbf{a}'\mathbf{a})(\mathbf{b}'\mathbf{b})}}. \tag{2.81}$$

When $\theta = 90^\circ$, $\mathbf{a}'\mathbf{b} = 0$ since $\cos(90^\circ) = 0$. Thus **a** and **b** are *perpendicular* when $\mathbf{a}'\mathbf{b} = 0$.

If $\mathbf{a}'\mathbf{a} = 1$, the vector **a** is said to be *normalized*. A vector **b** can be normalized by dividing by its length, $\sqrt{\mathbf{b}'\mathbf{b}}$. Thus

$$\mathbf{c} = \frac{\mathbf{b}}{\sqrt{\mathbf{b}'\mathbf{b}}} \tag{2.82}$$

is normalized so that $\mathbf{c}'\mathbf{c} = 1$.

A set of $p \times 1$ vectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_p$ that are normalized ($\mathbf{c}_i'\mathbf{c}_i = 1$ for all i) and mutually orthogonal ($\mathbf{c}_i'\mathbf{c}_j = 0$ for all $i \neq j$) is said to be an *orthonormal set* of vectors. If the $p \times p$ matrix $\mathbf{C} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_p)$ has orthonormal columns, $\mathbf{C}$ is called an *orthogonal* matrix. Since the elements of $\mathbf{C}'\mathbf{C}$ are products of columns of

$\mathbf{C}$ [see Theorem 2.2c(i)], an orthogonal matrix $\mathbf{C}$ has the property

$$\mathbf{C}'\mathbf{C} = \mathbf{I}. \tag{2.83}$$

It can be shown that an orthogonal matrix $\mathbf{C}$ also satisfies

$$\mathbf{C}\mathbf{C}' = \mathbf{I}. \tag{2.84}$$

Thus an orthogonal matrix $\mathbf{C}$ has orthonormal rows as well as orthonormal columns. It is also clear from (2.83) and (2.84) that $\mathbf{C}' = \mathbf{C}^{-1}$ if $\mathbf{C}$ is orthogonal.

Example 2.10. To illustrate an orthogonal matrix, we start with

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

whose columns are mutually orthogonal but not orthonormal. To normalize the three columns, we divide by their respective lengths, $\sqrt{3}$, $\sqrt{6}$, and $\sqrt{2}$, to obtain the matrix

$$\mathbf{C} = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{pmatrix},$$

whose columns are orthonormal. Note that the rows of $\mathbf{C}$ are also orthonormal, so that $\mathbf{C}$ satisfies (2.84) as well as (2.83). □

Multiplication of a vector by an orthogonal matrix has the effect of rotating axes; that is, if a point $\mathbf{x}$ is transformed to $\mathbf{z} = \mathbf{C}\mathbf{x}$, where $\mathbf{C}$ is orthogonal, then the distance from the origin to $\mathbf{z}$ is the same as the distance to $\mathbf{x}$:

$$\mathbf{z}'\mathbf{z} = (\mathbf{C}\mathbf{x})'(\mathbf{C}\mathbf{x}) = \mathbf{x}'\mathbf{C}'\mathbf{C}\mathbf{x} = \mathbf{x}'\mathbf{I}\mathbf{x} = \mathbf{x}'\mathbf{x}. \tag{2.85}$$

Hence, the transformation from $\mathbf{x}$ to $\mathbf{z}$ is a rotation.

Some properties of orthogonal matrices are given in the following theorem.

Theorem 2.10. If the $p \times p$ matrix $\mathbf{C}$ is orthogonal and if $\mathbf{A}$ is any $p \times p$ matrix, then

(i) $|\mathbf{C}| = +1$ or -1.

(ii) $|\mathbf{C}'\mathbf{A}\mathbf{C}| = |\mathbf{A}|$.
(iii) $-1 \leq c_{ij} \leq 1$, where c_{ij} is any element of $\mathbf{C}$.

2.11 TRACE

The *trace* of an $n \times n$ matrix $\mathbf{A} = (a_{ij})$ is a scalar function defined as the sum of the diagonal elements of $\mathbf{A}$; that is, $\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}$. For example, suppose

$$\mathbf{A} = \begin{pmatrix} 8 & 4 & 2 \\ 2 & -3 & 6 \\ 3 & 5 & 9 \end{pmatrix}.$$

Then

$$\mathrm{tr}(\mathbf{A}) = 8 - 3 + 9 = 14.$$

Some properties of the trace are given in the following theorem.

Theorem 2.11

(i) If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$, then

$$\mathrm{tr}(\mathbf{A} \pm \mathbf{B}) = \mathrm{tr}(\mathbf{A}) \pm \mathrm{tr}(\mathbf{B}). \tag{2.86}$$

(ii) If $\mathbf{A}$ is $n \times p$ and $\mathbf{B}$ is $p \times n$, then

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}). \tag{2.87}$$

Note that in (2.87) n can be less than, equal to, or greater than p.

(iii) If $\mathbf{A}$ is $n \times p$, then

$$\mathrm{tr}(\mathbf{A}'\mathbf{A}) = \sum_{i=1}^{p} \mathbf{a}_i'\mathbf{a}_i, \tag{2.88}$$

where $\mathbf{a}_i$ is the ith *column* of $\mathbf{A}$.

(iv) If $\mathbf{A}$ is $n \times p$, then

$$\mathrm{tr}(\mathbf{A}\mathbf{A}') = \sum_{i=1}^{n} \mathbf{a}_i'\mathbf{a}_i, \tag{2.89}$$

where $\mathbf{a}_i'$ is the ith *row* of $\mathbf{A}$.

(v) If $\mathbf{A} = (a_{ij})$ is an $n \times p$ matrix with representative element a_{ij}, then

$$\mathrm{tr}(\mathbf{A}'\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{A}') = \sum_{i=1}^{n}\sum_{j=1}^{p} a_{ij}^2. \tag{2.90}$$

(vi) If $\mathbf{A}$ is any $n \times n$ matrix and $\mathbf{P}$ is any $n \times n$ nonsingular matrix, then

$$\mathrm{tr}(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) = \mathrm{tr}(\mathbf{A}). \tag{2.91}$$

(vii) If $\mathbf{A}$ is any $n \times n$ matrix and $\mathbf{C}$ is any $n \times n$ orthogonal matrix, then

$$\mathrm{tr}(\mathbf{C}'\mathbf{A}\mathbf{C}) = \mathrm{tr}(\mathbf{A}). \tag{2.92}$$

(viii) If $\mathbf{A}$ is $n \times p$ of rank r and $\mathbf{A}^-$ is a generalized inverse of $\mathbf{A}$, then

$$\mathrm{tr}(\mathbf{A}^-\mathbf{A}) = \mathrm{tr}(\mathbf{A}\mathbf{A}^-) = r. \tag{2.93}$$

Proof. We prove parts (ii), (iii), and (vi).

(ii) By (2.13), the ith diagonal element of $\mathbf{E} = \mathbf{AB}$ is $e_{ii} = \sum_k a_{ik}b_{ki}$. Then

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{E}) = \sum_i e_{ii} = \sum_i \sum_k a_{ik}b_{ki}.$$

Similarly, the ith diagonal element of $\mathbf{F} = \mathbf{BA}$ is $f_{ii} = \sum_k b_{ik}a_{ki}$, and

$$\begin{aligned}\mathrm{tr}(\mathbf{BA}) = \mathrm{tr}(\mathbf{F}) = \sum_i f_{ii} &= \sum_i \sum_k b_{ik}a_{ki} \\ &= \sum_k \sum_i a_{ki}b_{ik} = \mathrm{tr}(\mathbf{E}) = \mathrm{tr}(\mathbf{AB}).\end{aligned}$$

(iii) By Theorem 2.2c(i), $\mathbf{A}'\mathbf{A}$ is obtained as products of columns of $\mathbf{A}$. If $\mathbf{a}_i$ is the ith column of $\mathbf{A}$, then the ith diagonal element of $\mathbf{A}'\mathbf{A}$ is $\mathbf{a}_i'\mathbf{a}_i$.

(vi) By (2.87) we obtain

$$\mathrm{tr}(\mathbf{P}^{-1}\mathbf{A}\mathbf{P}) = \mathrm{tr}(\mathbf{A}\mathbf{P}\mathbf{P}^{-1}) = \mathrm{tr}(\mathbf{A}).$$

□

Example 2.11. We illustrate parts (ii) and (viii) of Theorem 2.11.

(ii) Let

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

Then

$$\mathbf{AB} = \begin{pmatrix} 9 & 10 & 16 \\ 4 & -8 & -3 \\ 24 & 16 & 34 \end{pmatrix}, \quad \mathbf{BA} = \begin{pmatrix} 3 & 17 \\ 30 & 32 \end{pmatrix},$$

$$\mathrm{tr}(\mathbf{AB}) = 9 - 8 + 34 = 35, \quad \mathrm{tr}(\mathbf{BA}) = 3 + 32 = 35.$$

(viii) Using $\mathbf{A}$ in (2.59) and $\mathbf{A}_1^-$ in (2.60), we obtain

$$\mathbf{A}^-\mathbf{A} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{AA}^- = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix},$$

$$\mathrm{tr}(\mathbf{A}^-\mathbf{A}) = 1 + 1 + 0 = 2 = \mathrm{rank}(\mathbf{A}),$$
$$\mathrm{tr}(\mathbf{AA}^-) = 1 + 1 + 0 = 2 = \mathrm{rank}(\mathbf{A}).$$

□

2.12 EIGENVALUES AND EIGENVECTORS

2.12.1 Definition

For every square matrix $\mathbf{A}$, a scalar λ and a nonzero vector $\mathbf{x}$ can be found such that

$$\mathbf{Ax} = \lambda\mathbf{x}, \tag{2.94}$$

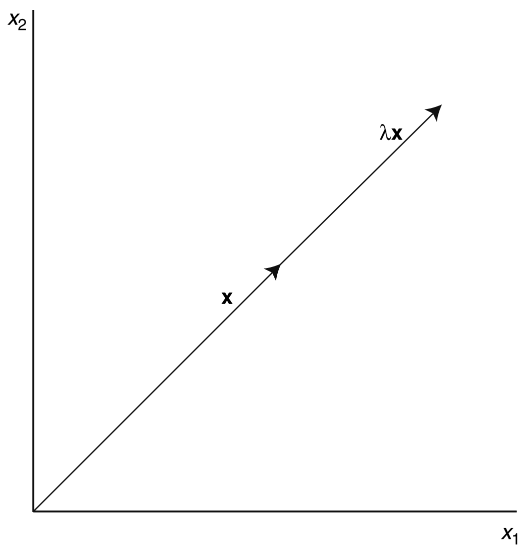

Figure 2.5 An eigenvector $\mathbf{x}$ is transformed to $\lambda\mathbf{x}$.

where λ is an *eigenvalue* of $\mathbf{A}$ and $\mathbf{x}$ is an *eigenvector*. (These terms are sometimes referred to as characteristic root and characteristic vector, respectively.) Note that in (2.94), the vector $\mathbf{x}$ is transformed by $\mathbf{A}$ onto a multiple of itself, so that the point $\mathbf{Ax}$ is on the line passing through $\mathbf{x}$ and the origin. This is illustrated in Figure 2.5.

To find λ and $\mathbf{x}$ for a matrix $\mathbf{A}$, we write (2.94) as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}. \tag{2.95}$$

By (2.37), $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x}$ is a linear combination of the columns of $\mathbf{A} - \lambda\mathbf{I}$, and by (2.40) and (2.95), these columns are linearly dependent. Thus the square matrix $(\mathbf{A} - \lambda\mathbf{I})$ is singular, and by Theorem 2.9a(iii), we can solve for λ using

$$|\mathbf{A} - \lambda\mathbf{I}| = 0, \tag{2.96}$$

which is known as the *characteristic equation*.

If $\mathbf{A}$ is $n \times n$, the characteristic equation (2.96) will have n roots; that is, $\mathbf{A}$ will have n eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. The λ's will not necessarily all be distinct, or all nonzero, or even all real. (However, the eigenvalues of a symmetric matrix are real; see Theorem 2.12c.) After finding $\lambda_1, \lambda_2, \ldots, \lambda_n$ using (2.96), the accompanying eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ can be found using (2.95).

If an eigenvalue is 0, the corresponding eigenvector is not $\mathbf{0}$. To see this, note that if $\lambda = 0$, then $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$ becomes $\mathbf{Ax} = \mathbf{0}$, which has solutions for $\mathbf{x}$ because $\mathbf{A}$ is singular, and the columns are therefore linearly dependent. [The matrix $\mathbf{A}$ is singular because it has a zero eigenvalue; see (63) and (2.107).]

If we multiply both sides of (2.95) by a scalar k, we obtain

$$k(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = k\mathbf{0} = \mathbf{0},$$

which can be rewritten as

$$(\mathbf{A} - \lambda\mathbf{I})k\mathbf{x} = \mathbf{0} \qquad \text{[by (2.12)]}.$$

Thus if $\mathbf{x}$ is an eigenvector of $\mathbf{A}$, $k\mathbf{x}$ is also an eigenvector. Eigenvectors are therefore unique only up to multiplication by a scalar. (There are many solution vectors $\mathbf{x}$ because $\mathbf{A} - \lambda\mathbf{I}$ is singular; see Section 2.8) Hence, the length of $\mathbf{x}$ is arbitrary, but its direction from the origin is unique; that is, the relative values of (ratios of) the elements of $\mathbf{x} = (x_1, x_2, \ldots, x_n)'$ are unique. Typically, an eigenvector $\mathbf{x}$ is scaled to normalized form as in (2.82), $\mathbf{x}'\mathbf{x} = 1$.

Example 2.12.1. To illustrate eigenvalues and eigenvectors, consider the matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 2 \\ -1 & 4 \end{pmatrix}.$$

By (2.96), the characteristic equation is

$$|\mathbf{A} - \lambda\mathbf{I}| = \begin{vmatrix} 1-\lambda & 2 \\ -1 & 4-\lambda \end{vmatrix} = (1-\lambda)(4-\lambda) + 2 = 0,$$

which becomes

$$\lambda^2 - 5\lambda + 6 = (\lambda - 3)(\lambda - 2) = 0,$$

with roots $\lambda_1 = 3$ and $\lambda_2 = 2$.

To find the eigenvector $\mathbf{x}_1$ corresponding to $\lambda_1 = 3$, we use (2.95)

$$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{x}_1 = \mathbf{0},$$

$$\begin{pmatrix} 1-3 & 2 \\ -1 & 4-3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

which can be written as

$$\begin{aligned} -2x_1 + 2x_2 &= 0 \\ -x_1 + x_2 &= 0. \end{aligned}$$

The second equation is a multiple of the first, and either equation yields $x_1 = x_2$. The solution vector can be written with $x_1 = x_2 = c$ as an arbitrary constant:

$$\mathbf{x}_1 = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_1 \end{pmatrix} = x_1\begin{pmatrix} 1 \\ 1 \end{pmatrix} = c\begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

If c is set equal to $1/\sqrt{2}$ to normalize the eigenvector, we obtain

$$\mathbf{x}_1 = \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}.$$

Similarly, corresponding to $\lambda_2 = 2$, we obtain

$$\mathbf{x}_2 = \begin{pmatrix} 2/\sqrt{5} \\ 1/\sqrt{5} \end{pmatrix}.$$

□

2.12.2 Functions of a Matrix

If λ is an eigenvalue of $\mathbf{A}$ with corresponding eigenvector $\mathbf{x}$, then for certain functions $g(\mathbf{A})$, an eigenvalue is given by $g(\lambda)$ and $\mathbf{x}$ is the corresponding eigenvector of $g(\mathbf{A})$ as well as of $\mathbf{A}$. We illustrate some of these cases:

1. If λ is an eigenvalue of $\mathbf{A}$, then $c\lambda$ is an eigenvalue of $c\mathbf{A}$, where c is an arbitrary constant such that $c \neq 0$. This is easily demonstrated by multiplying the defining relationship $\mathbf{Ax} = \lambda\mathbf{x}$ by c:

$$c\mathbf{Ax} = c\lambda\mathbf{x}. \tag{2.97}$$

   Note that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ corresponding to λ, and $\mathbf{x}$ is also an eigenvector of $c\mathbf{A}$ corresponding to $c\lambda$.

2. If λ is an eigenvalue of the $\mathbf{A}$ and $\mathbf{x}$ is the corresponding eigenvector of $\mathbf{A}$, then $c\lambda + k$ is an eigenvalue of the matrix $c\mathbf{A} + k\mathbf{I}$ and $\mathbf{x}$ is an eigenvector of $c\mathbf{A} + k\mathbf{I}$, where c and k are scalars. To show this, we add $k\mathbf{x}$ to (2.97):

$$\begin{aligned} c\mathbf{Ax} + k\mathbf{x} &= c\lambda\mathbf{x} + k\mathbf{x}, \\ (c\mathbf{A} + k\mathbf{I})\mathbf{x} &= (c\lambda + k)\mathbf{x}. \end{aligned} \tag{2.98}$$

   Thus $c\lambda + k$ is an eigenvalue of $c\mathbf{A} + k\mathbf{I}$ and $\mathbf{x}$ is the corresponding eigenvector of $c\mathbf{A} + k\mathbf{I}$. Note that (2.98) does not extend to $\mathbf{A} + \mathbf{B}$ for arbitrary $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$; that is, $\mathbf{A} + \mathbf{B}$ does not have $\lambda_A + \lambda_B$ for an eigenvalue, where λ_A is an eigenvalue of $\mathbf{A}$ and λ_B is an eigenvalue of $\mathbf{B}$.

3. If λ is an eigenvalue of $\mathbf{A}$, then λ^2 is an eigenvalue of $\mathbf{A}^2$. This can be demonstrated by multiplying the defining relationship $\mathbf{Ax} = \lambda\mathbf{x}$ by $\mathbf{A}$:

$$\begin{aligned} \mathbf{A}(\mathbf{Ax}) &= \mathbf{A}(\lambda\mathbf{x}), \\ \mathbf{A}^2\mathbf{x} &= \lambda\mathbf{Ax} = \lambda(\lambda\mathbf{x}) = \lambda^2\mathbf{x}. \end{aligned} \tag{2.99}$$

   Thus λ^2 is an eigenvalue of $\mathbf{A}^2$, and $\mathbf{x}$ is the corresponding eigenvector of $\mathbf{A}^2$. This can be extended to any power of $\mathbf{A}$:

$$\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}; \tag{2.100}$$

   that is, λ^k is an eigenvalue of $\mathbf{A}^k$, and $\mathbf{x}$ is the corresponding eigenvector.

4. If λ is an eigenvalue of the nonsingular matrix $\mathbf{A}$, then $1/\lambda$ is an eigenvalue of $\mathbf{A}^{-1}$. To demonstrate this, we multiply $\mathbf{Ax} = \lambda\mathbf{x}$ by $\mathbf{A}^{-1}$ to obtain

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{Ax} &= \mathbf{A}^{-1}\lambda\mathbf{x}, \\ \mathbf{x} &= \lambda\mathbf{A}^{-1}\mathbf{x}, \\ \mathbf{A}^{-1}\mathbf{x} &= \frac{1}{\lambda}\mathbf{x}. \end{aligned} \tag{2.101}$$

   Thus $1/\lambda$ is an eigenvalue of $\mathbf{A}^{-1}$, and $\mathbf{x}$ is an eigenvector of both $\mathbf{A}$ and $\mathbf{A}^{-1}$.

5. The results in (2.97) and (2.100) can be used to obtain eigenvalues and eigenvectors of a polynomial in $\mathbf{A}$. For example, if λ is an eigenvalue of $\mathbf{A}$, then

$$\begin{aligned} (\mathbf{A}^3 + 4\mathbf{A}^2 - 3\mathbf{A} + 5\mathbf{I})\mathbf{x} &= \mathbf{A}^3\mathbf{x} + 4\mathbf{A}^2\mathbf{x} - 3\mathbf{Ax} + 5\mathbf{x} \\ &= \lambda^3\mathbf{x} + 4\lambda^2\mathbf{x} - 3\lambda\mathbf{x} + 5\mathbf{x} \\ &= (\lambda^3 + 4\lambda^2 - 3\lambda + 5)\mathbf{x}. \end{aligned}$$

   Thus $\lambda^3 + 4\lambda^2 - 3\lambda + 5$ is an eigenvalue of $\mathbf{A}^3 + 4\mathbf{A}^2 - 3\mathbf{A} + 5\mathbf{I}$, and $\mathbf{x}$ is the corresponding eigenvector.

For certain matrices, property 5 can be extended to an infinite series. For example, if λ is an eigenvalue of $\mathbf{A}$, then, by (2.98), $1 - \lambda$ is an eigenvalue of $\mathbf{I} - \mathbf{A}$. If $\mathbf{I} - \mathbf{A}$ is nonsingular, then, by (2.101), $1/(1 - \lambda)$ is an eigenvalue of $(\mathbf{I} - \mathbf{A})^{-1}$. If $-1 < \lambda < 1$, then $1/(1 - \lambda)$ can be represented by the series

$$\frac{1}{1 - \lambda} = 1 + \lambda + \lambda^2 + \lambda^3 + \cdots.$$

Correspondingly, if all eigenvalues of $\mathbf{A}$ satisfy $-1 < \lambda < 1$, then

$$(\mathbf{I} - \mathbf{A})^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \cdots. \tag{2.102}$$

2.12.3 Products

It was noted in a comment following (2.98) that the eigenvalues of $\mathbf{A} + \mathbf{B}$ are not of the form $\lambda_A + \lambda_B$, where λ_A is an eigenvalue of $\mathbf{A}$ and λ_B is an eigenvalue of $\mathbf{B}$. Similarly, the eigenvalues of $\mathbf{AB}$ are not products of the form $\lambda_A\lambda_B$. However, the eigenvalues of $\mathbf{AB}$ are the same as those of $\mathbf{BA}$.

Theorem 2.12a. If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ or if $\mathbf{A}$ is $n \times p$ and $\mathbf{B}$ is $p \times n$, then the (nonzero) eigenvalues of $\mathbf{AB}$ are the same as those of $\mathbf{BA}$. If $\mathbf{x}$ is an eigenvector of $\mathbf{AB}$, then $\mathbf{Bx}$ is an eigenvector of $\mathbf{BA}$. □

Two additional results involving eigenvalues of products are given in the following theorem.

Theorem 2.12b. Let $\mathbf{A}$ be any $n \times n$ matrix.

(i) If $\mathbf{P}$ is any $n \times n$ nonsingular matrix, then $\mathbf{A}$ and $\mathbf{P}^{-1}\mathbf{AP}$ have the same eigenvalues.

(ii) If $\mathbf{C}$ is any $n \times n$ orthogonal matrix, then $\mathbf{A}$ and $\mathbf{C}'\mathbf{AC}$ have the same eigenvalues. □

2.12.4 Symmetric Matrices

Two properties of the eigenvalues and eigenvectors of a symmetric matrix are given in the following theorem.

Theorem 2.12c. Let $\mathbf{A}$ be an $n \times n$ symmetric matrix.

(i) The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of $\mathbf{A}$ are real.

(ii) The eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ of $\mathbf{A}$ corresponding to distinct eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_k$ are mutually orthogonal; the eigenvectors $\mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \ldots, \mathbf{x}_n$ corresponding to the nondistinct eigenvalues can be chosen to be mutually orthogonal to each other and to the other eigenvectors; that is, $\mathbf{x}_i'\mathbf{x}_j = 0$ for $i \neq j$. □

If the eigenvectors of a symmetric matrix $\mathbf{A}$ are normalized and placed as columns of a matrix $\mathbf{C}$, then by Theorem 2.12c(ii), $\mathbf{C}$ is an orthogonal matrix. This orthogonal matrix can be used to express $\mathbf{A}$ in terms of its eigenvalues and eigenvectors.

Theorem 2.12d. If $\mathbf{A}$ is an $n \times n$ symmetric matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ and normalized eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, then $\mathbf{A}$ can be expressed as

$$\mathbf{A} = \mathbf{CDC}' \tag{2.103}$$

$$= \sum_{i=1}^{n} \lambda_i \mathbf{x}_i \mathbf{x}_i', \tag{2.104}$$

where $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ and $\mathbf{C}$ is the orthogonal matrix $\mathbf{C} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$. The result in either (2.103) or (2.104) is often called the *spectral decomposition* of $\mathbf{A}$.

Proof. By Theorem 2.12c(ii), $\mathbf{C}$ is orthogonal. Then by (2.84), $\mathbf{I} = \mathbf{CC}'$, and multiplication by $\mathbf{A}$ gives

$$\mathbf{A} = \mathbf{ACC}'.$$

We now substitute $\mathbf{C} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ to obtain

$$\begin{aligned}\mathbf{A} &= \mathbf{A}(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)\mathbf{C}' \\ &= (\mathbf{A}\mathbf{x}_1, \mathbf{A}\mathbf{x}_2, \ldots, \mathbf{A}\mathbf{x}_n)\mathbf{C}' \qquad \text{[by (2.28)]} \\ &= (\lambda_1\mathbf{x}_1, \lambda_2\mathbf{x}_2, \ldots, \lambda_n\mathbf{x}_n)\mathbf{C}' \qquad \text{[by (2.94)]} \\ &= \mathbf{CDC}', \end{aligned} \tag{2.105}$$

since multiplication on the right by $\mathbf{D} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ multiplies columns of $\mathbf{C}$ by elements of $\mathbf{D}$ [see (2.30)]. Now writing $\mathbf{C}'$ in the form

$$\mathbf{C}' = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)' = \begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} \qquad \text{[by (2.39)]},$$

(2.105) becomes

$$\begin{aligned}\mathbf{A} &= (\lambda_1\mathbf{x}_1, \lambda_2\mathbf{x}_2, \ldots, \lambda_n\mathbf{x}_n)\begin{pmatrix} \mathbf{x}_1' \\ \mathbf{x}_2' \\ \vdots \\ \mathbf{x}_n' \end{pmatrix} \\ &= \lambda_1\mathbf{x}_1\mathbf{x}_1' + \lambda_2\mathbf{x}_2\mathbf{x}_2' + \cdots + \lambda_n\mathbf{x}_n\mathbf{x}_n'. \end{aligned}$$

□

Corollary 1. If $\mathbf{A}$ is symmetric and $\mathbf{C}$ and $\mathbf{D}$ are defined as in Theorem 2.12d, then $\mathbf{C}$ *diagonalizes* $\mathbf{A}$:

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D}. \tag{2.106}$$

□

We can express the determinant and trace of a square matrix $\mathbf{A}$ in terms of its eigenvalues.

Theorem 2.12e. If $\mathbf{A}$ is any $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then

(i)
$$|\mathbf{A}| = \prod_{i=1}^{n} \lambda_i. \tag{2.107}$$

(ii)
$$\mathrm{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i. \tag{2.108}$$

□

We have included Theorem 2.12e here because it is easy to prove for a symmetric matrix $\mathbf{A}$ using Theorem 2.12d (see Problem 2.72). However, the theorem is true for any square matrix (Searle 1982, p. 278).

Example 2.12.4. To illustrate Theorem 2.12e, consider the matrix $\mathbf{A}$ in Example 2.12.1

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

which has eigenvalues $\lambda_1 = 3$ and $\lambda_2 = 2$. The product $\lambda_1\lambda_2 = 6$ is the same as $|\mathbf{A}| = 4 - (-1)(2) = 6$. The sum $\lambda_1 + \lambda_2 = 3 + 2 = 5$ is the same as $\mathrm{tr}(\mathbf{A}) = 1 + 4 = 5$. □

2.12.5 Positive Definite and Semidefinite Matrices

The eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ of positive definite and positive semidefinite matrices (Section 2.6) are positive and nonnegative, respectively.

Theorem 2.12f. Let $\mathbf{A}$ be $n \times n$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$.

(i) If $\mathbf{A}$ is positive definite, then $\lambda_i > 0$ for $i = 1, 2, \ldots, n$.
(ii) If $\mathbf{A}$ is positive semidefinite, then $\lambda_i \geq 0$ for $i = 1, 2, \ldots, n$. The number of eigenvalues λ_i for which $\lambda_i > 0$ is the rank of $\mathbf{A}$.

PROOF.

(i) For any λ_i, we have $\mathbf{Ax}_i = \lambda_i \mathbf{x}_i$. Multiplying by $\mathbf{x}_i'$, we obtain

$$\mathbf{x}_i'\mathbf{Ax}_i = \lambda_i \mathbf{x}_i'\mathbf{x}_i,$$

$$\lambda_i = \frac{\mathbf{x}_i'\mathbf{Ax}_i}{\mathbf{x}_i'\mathbf{x}_i} > 0.$$

In the second expression, $\mathbf{x}_i'\mathbf{Ax}_i$ is positive because $\mathbf{A}$ is positive definite, and $\mathbf{x}_i'\mathbf{x}_i$ is positive because $\mathbf{x}_i \neq \mathbf{0}$. □

If a matrix $\mathbf{A}$ is positive definite, we can find a *square root matrix* $\mathbf{A}^{1/2}$ as follows. Since the eigenvalues of $\mathbf{A}$ are positive, we can substitute the square roots $\sqrt{\lambda_i}$ for λ_i in the spectral decomposition of $\mathbf{A}$ in (2.103), to obtain

$$\mathbf{A}^{1/2} = \mathbf{CD}^{1/2}\mathbf{C}', \tag{2.109}$$

where $\mathbf{D}^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_n})$. The matrix $\mathbf{A}^{1/2}$ is symmetric and has the property

$$\mathbf{A}^{1/2}\mathbf{A}^{1/2} = (\mathbf{A}^{1/2})^2 = \mathbf{A}. \tag{2.110}$$

2.13 IDEMPOTENT MATRICES

A square matrix $\mathbf{A}$ is said to be *idempotent* if $\mathbf{A}^2 = \mathbf{A}$. Most idempotent matrices in this book are symmetric. Many of the sums of squares in regression (Chapters 6–11) and analysis of variance (Chapters 12–15) can be expressed as quadratic forms $\mathbf{y}'\mathbf{A}\mathbf{y}$. The idempotence of $\mathbf{A}$ or of a product involving $\mathbf{A}$ will be used to establish that $\mathbf{y}'\mathbf{A}\mathbf{y}$ (or a multiple of $\mathbf{y}'\mathbf{A}\mathbf{y}$) has a chi-square distribution.

An example of an idempotent matrix is the identity matrix $\mathbf{I}$.

Theorem 2.13a. The only nonsingular idempotent matrix is the identity matrix $\mathbf{I}$.

Proof. If $\mathbf{A}$ is idempotent and nonsingular, then $\mathbf{A}^2 = \mathbf{A}$ and the inverse $\mathbf{A}^{-1}$ exists. If we multiply $\mathbf{A}^2 = \mathbf{A}$ by $\mathbf{A}^{-1}$, we obtain

$$\mathbf{A}^{-1}\mathbf{A}^2 = \mathbf{A}^{-1}\mathbf{A},$$
$$\mathbf{A} = \mathbf{I}.$$

□

Many of the matrices of quadratic forms we will encounter in later chapters are singular idempotent matrices. We now give some properties of such matrices.

Theorem 2.13b. If $\mathbf{A}$ is singular, symmetric, and idempotent, then $\mathbf{A}$ is positive semidefinite.

Proof. Since $\mathbf{A} = \mathbf{A}'$ and $\mathbf{A} = \mathbf{A}^2$, we have

$$\mathbf{A} = \mathbf{A}^2 = \mathbf{A}\mathbf{A} = \mathbf{A}'\mathbf{A},$$

which is positive semidefinite by Theorem 2.6d(ii). □

If a is a real number such that $a^2 = a$, then a is either 0 or 1. The analogous property for matrices is that if $\mathbf{A}^2 = \mathbf{A}$, then the eigenvalues of $\mathbf{A}$ are 0s and 1s.

Theorem 2.13c. If $\mathbf{A}$ is an $n \times n$ symmetric idempotent matrix of rank r, then $\mathbf{A}$ has r eigenvalues equal to 1 and $n - r$ eigenvalues equal to 0.

Proof. By (2.99), if $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$, then $\mathbf{A}^2\mathbf{x} = \lambda^2\mathbf{x}$. Since $\mathbf{A}^2 = \mathbf{A}$, we have $\mathbf{A}^2\mathbf{x} = \mathbf{A}\mathbf{x} = \lambda\mathbf{x}$. Equating the right sides of $\mathbf{A}^2\mathbf{x} = \lambda^2\mathbf{x}$ and $\mathbf{A}^2\mathbf{x} = \lambda\mathbf{x}$, we have

$$\lambda\mathbf{x} = \lambda^2\mathbf{x} \quad \text{or} \quad (\lambda - \lambda^2)\mathbf{x} = \mathbf{0}.$$

But $\mathbf{x} \neq \mathbf{0}$, and therefore $\lambda - \lambda^2 = 0$, from which, λ is either 0 or 1.

By Theorem 2.13b, $\mathbf{A}$ is positive semidefinite, and therefore by Theorem 2.12f(ii), the number of nonzero eigenvalues is equal to rank($\mathbf{A}$). Thus r eigenvalues of $\mathbf{A}$ are equal to 1 and the remaining $n - r$ eigenvalues are equal to 0. □

We can use Theorems 2.12e and 2.13c to find the rank of a symmetric idempotent matrix.

Theorem 2.13d. If $\mathbf{A}$ is symmetric and idempotent of rank r, then $\text{rank}(\mathbf{A}) = \text{tr}(\mathbf{A}) = r$.

PROOF. By Theorem 2.12e(ii), $\text{tr}(\mathbf{A}) = \sum_{i=1}^{n} \lambda_i$, and by Theorem 2.13c, $\sum_{i=1}^{n} \lambda_i = r$. □

Some additional properties of idempotent matrices are given in the following four theorems.

Theorem 2.13e. If $\mathbf{A}$ is an $n \times n$ idempotent matrix, $\mathbf{P}$ is an $n \times n$ nonsingular matrix, and $\mathbf{C}$ is an $n \times n$ orthogonal matrix, then

(i) $\mathbf{I} - \mathbf{A}$ is idempotent.
(ii) $\mathbf{A}(\mathbf{I} - \mathbf{A}) = \mathbf{O}$ and $(\mathbf{I} - \mathbf{A})\mathbf{A} = \mathbf{O}$.
(iii) $\mathbf{P}^{-1}\mathbf{AP}$ is idempotent.
(iv) $\mathbf{C}'\mathbf{AC}$ is idempotent. (If $\mathbf{A}$ is symmetric, $\mathbf{C}'\mathbf{AC}$ is a symmetric idempotent matrix.) □

Theorem 2.13f. Let $\mathbf{A}$ be $n \times p$ of rank r, let $\mathbf{A}^-$ be any generalized inverse of $\mathbf{A}$, and let $(\mathbf{A}'\mathbf{A})^-$ be any generalized inverse of $\mathbf{A}'\mathbf{A}$. Then $\mathbf{A}^-\mathbf{A}, \mathbf{AA}^-$, and $\mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$ are all idempotent. □

Theorem 2.13g. Suppose that the $n \times n$ symmetric matrix $\mathbf{A}$ can be written as $\mathbf{A} = \sum_{i=1}^{k} \mathbf{A}_i$ for some k, where each $\mathbf{A}_i$ is an $n \times n$ symmetric matrix. Then any two of the following conditions implies the third condition.

(i) $\mathbf{A}$ is idempotent.
(ii) Each of $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ is idempotent.
(iii) $\mathbf{A}_i\mathbf{A}_j = \mathbf{O}$ for $i \neq j$.

□

Theorem 2.13h. If $\mathbf{I} = \sum_{i=1}^{k} \mathbf{A}_i$, where each $n \times n$ matrix $\mathbf{A}_i$ is symmetric of rank r_i, and if $n = \sum_{i=1}^{k} r_i$, then both of the following are true:

(i) Each of $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ is idempotent.
(ii) $\mathbf{A}_i\mathbf{A}_j = \mathbf{O}$ for $i \neq j$.

□

2.14 VECTOR AND MATRIX CALCULUS

2.14.1 Derivatives of Functions of Vectors and Matrices

Let $u = f(\mathbf{x})$ be a function of the variables $x_1, x_2, \ldots, x_p$ in $\mathbf{x} = (x_1, x_2, \ldots, x_p)'$, and let $\partial u/\partial x_1, \partial u/\partial x_2, \ldots, \partial u/\partial x_p$ be the partial derivatives. We define $\partial u/\partial \mathbf{x}$ as

$$\frac{\partial u}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial u}{\partial x_1} \\ \dfrac{\partial u}{\partial x_2} \\ \vdots \\ \dfrac{\partial u}{\partial x_p} \end{pmatrix}. \tag{2.111}$$

Two specific functions of interest are $u = \mathbf{a}'\mathbf{x}$ and $u = \mathbf{x}'\mathbf{A}\mathbf{x}$. Their derivatives with respect to $\mathbf{x}$ are given in the following two theorems.

Theorem 2.14a. Let $u = \mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a}$, where $\mathbf{a}' = (a_1, a_2, \ldots, a_p)$ is a vector of constants. Then

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{a})}{\partial \mathbf{x}} = \mathbf{a}. \tag{2.112}$$

PROOF

$$\frac{\partial u}{\partial x_i} = \frac{\partial(a_1x_1 + a_2x_2 + \cdots + a_px_p)}{\partial x_i} = a_i.$$

Thus by (2.111) we obtain

$$\frac{\partial u}{\partial \mathbf{x}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \mathbf{a}.$$

□

Theorem 2.14b. Let $u = \mathbf{x}'\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a symmetric matrix of constants. Then

$$\frac{\partial u}{\partial \mathbf{x}} = \frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x}. \tag{2.113}$$

PROOF. We demonstrate that (2.113) holds for the special case in which $\mathbf{A}$ is 3×3. The illustration could be generalized to a symmetric $\mathbf{A}$ of any size. Let

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad \text{and} \quad \mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix} = \begin{pmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \mathbf{a}_3' \end{pmatrix}.$$

Then $\mathbf{x}'\mathbf{A}\mathbf{x} = x_1^2 a_{11} + 2x_1x_2a_{12} + 2x_1x_3a_{13} + x_2^2 a_{22} + 2x_2x_3a_{23} + x_3^2 a_{33}$, and we have

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_1} = 2x_1a_{11} + 2x_2a_{12} + 2x_3a_{13} = 2\mathbf{a}_1'\mathbf{x}$$
$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_2} = 2x_1a_{12} + 2x_2a_{22} + 2x_3a_{23} = 2\mathbf{a}_2'\mathbf{x}$$
$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_3} = 2x_1a_{13} + 2x_2a_{23} + 2x_3a_{33} = 2\mathbf{a}_3'\mathbf{x}.$$

Thus by (2.11), (2.27), and (2.111), we obtain

$$\frac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_1} \\ \dfrac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_2} \\ \dfrac{\partial(\mathbf{x}'\mathbf{A}\mathbf{x})}{\partial x_3} \end{pmatrix} = 2\begin{pmatrix} \mathbf{a}_1'\mathbf{x} \\ \mathbf{a}_2'\mathbf{x} \\ \mathbf{a}_3'\mathbf{x} \end{pmatrix} = 2\mathbf{A}\mathbf{x}.$$

□

Now let $u = f(\mathbf{X})$ be a function of the variables $x_{11}, x_{12}, \ldots, x_{pp}$ in the $p \times p$ matrix $\mathbf{X}$, and let $(\partial u/\partial x_{11}), (\partial u/\partial x_{12}), \ldots, (\partial u/\partial x_{pp})$ be the partial derivatives. Similarly to (2.111), we define $\partial u/\partial \mathbf{X}$ as

$$\frac{\partial u}{\partial \mathbf{X}} = \begin{pmatrix} \dfrac{\partial u}{\partial x_{11}} & \cdots & \dfrac{\partial u}{\partial x_{1p}} \\ \vdots & & \vdots \\ \dfrac{\partial u}{\partial x_{p1}} & \cdots & \dfrac{\partial u}{\partial x_{pp}} \end{pmatrix}. \tag{2.114}$$

Two functions of interest of this type are $u = \text{tr}(\mathbf{XA})$ and $u = \ln|\mathbf{X}|$ for a positive definite matrix $\mathbf{X}$.

Theorem 2.14c. Let $u = \text{tr}(\mathbf{XA})$, where $\mathbf{X}$ is a $p \times p$ positive definite matrix and $\mathbf{A}$ is a $p \times p$ matrix of constants. Then

$$\frac{\partial u}{\partial \mathbf{X}} = \frac{\partial[\text{tr}(\mathbf{XA})]}{\partial \mathbf{X}} = \mathbf{A} + \mathbf{A}' - \text{diag}\,\mathbf{A}. \tag{2.115}$$

PROOF. Note that $\text{tr}(\mathbf{XA}) = \sum_{i=1}^{p}\sum_{j=1}^{p} x_{ij}a_{ji}$ [see the proof of Theorem 2.11(ii)]. Since $x_{ij} = x_{ji}$, $[\partial\text{tr}(\mathbf{XA})]/\partial x_{ij} = a_{ji} + a_{ij}$ if $i \neq j$, and $[\partial\text{tr}(\mathbf{XA})]/\partial x_{ii} = a_{ii}$. The result follows. □

Theorem 2.14d. Let $u = \ln|\mathbf{X}|$ where $\mathbf{X}$ is a $p \times p$ positive definite matrix. Then

$$\frac{\partial \ln|\mathbf{X}|}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \text{diag}(\mathbf{X}^{-1}). \tag{2.116}$$

PROOF. See Harville (1997, p. 306). See Problem 2.83 for a demonstration that this theorem holds for 2×2 matrices. □

2.14.2 Derivatives Involving Inverse Matrices and Determinants

Let $\mathbf{A}$ be an $n \times n$ nonsingular matrix with elements a_{ij} that are functions of a scalar x. We define $\partial\mathbf{A}/\partial x$ as the $n \times n$ matrix with elements $\partial a_{ij}/\partial x$. The related derivative $\partial\mathbf{A}^{-1}/\partial x$ is often of interest. If $\mathbf{A}$ is positive definite, the derivative $(\partial/\partial x)\log|\mathbf{A}|$ is also often of interest.

Theorem 2.14e. Let $\mathbf{A}$ be nonsingular of order n with derivative $\partial\mathbf{A}/\partial x$. Then

$$\frac{\partial\mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x}\mathbf{A}^{-1} \tag{2.117}$$

PROOF. Because $\mathbf{A}$ is nonsingular, we have

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Thus

$$\frac{\partial\mathbf{A}^{-1}}{\partial x}\mathbf{A} + \mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x} = \mathbf{O}.$$

Hence

$$\frac{\partial\mathbf{A}^{-1}}{\partial x}\mathbf{A} = -\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x},$$

and so

$$\frac{\partial\mathbf{A}^{-1}}{\partial x} = -\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x}\mathbf{A}^{-1}.$$

□

Theorem 2.14f. Let $\mathbf{A}$ be an $n \times n$ positive define matrix. Then

$$\frac{\partial\log|\mathbf{A}|}{\partial x} = \text{tr}\left(\mathbf{A}^{-1}\frac{\partial\mathbf{A}}{\partial x}\right). \tag{2.118}$$

PROOF. Since **A** is positive definite, its spectral decomposition (Theorem 2.12d) can be written as $\mathbf{CDC}'$, where **C** is an orthogonal matrix and **D** is a diagonal matrix of positive eigenvalues, λ_i. Using Theorem 2.12e, we obtain

$$\begin{aligned}
\frac{\partial \log|\mathbf{A}|}{\partial x} &= \frac{\partial \log \prod_{i=1}^n \lambda_i}{\partial x} \\
&= \frac{\partial \sum_{i=1}^n \log \lambda_i}{\partial x} \\
&= \sum_{i=1}^n \frac{1}{\lambda_i} \frac{\partial \lambda_i}{\partial x} \\
&= \mathrm{tr}\left(\mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial x}\right).
\end{aligned}$$

Now

$$\begin{aligned}
\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial x} &= \mathbf{CD}^{-1}\mathbf{C}' \frac{\partial \mathbf{CDC}'}{\partial x} \\
&= \mathbf{CD}^{-1}\mathbf{C}' \left[\mathbf{C} \frac{\partial \mathbf{DC}'}{\partial x} + \frac{\partial \mathbf{C}}{\partial x} \mathbf{DC}'\right] \\
&= \mathbf{CD}^{-1}\mathbf{C}' \left[\mathbf{C} \frac{\partial \mathbf{D}}{\partial x} \mathbf{C}' + \mathbf{CD} \frac{\partial \mathbf{C}'}{\partial x} + \frac{\partial \mathbf{C}}{\partial x} \mathbf{DC}'\right] \\
&= \mathbf{CD}^{-1} \frac{\partial \mathbf{D}}{\partial x} \mathbf{C}' + \mathbf{C} \frac{\partial \mathbf{C}'}{\partial x} + \mathbf{CD}^{-1}\mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} \mathbf{DC}'.
\end{aligned}$$

Using Theorem 2.11(i) and (ii), we have

$$\mathrm{tr}\left(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial x}\right) = \mathrm{tr}\left(\mathbf{D}^{-1} \frac{\partial \mathbf{D}}{\partial x} + \mathbf{C} \frac{\partial \mathbf{C}'}{\partial x} + \mathbf{C}' \frac{\partial \mathbf{C}}{\partial x}\right).$$

Since **C** is orthogonal, $\mathbf{C}'\mathbf{C} = \mathbf{I}$ which implies that

$$\frac{\partial \mathbf{C}'\mathbf{C}}{\partial x} = \mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} + \frac{\partial \mathbf{C}'}{\partial x} \mathbf{C} = \mathbf{O}$$

and

$$\mathrm{tr}\left(\mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} + \frac{\partial \mathbf{C}'\mathbf{C}}{\partial x}\right) = \mathrm{tr}\left(\mathbf{C}' \frac{\partial \mathbf{C}}{\partial x} + \mathbf{C} \frac{\partial \mathbf{C}'}{\partial x}\right) = 0.$$

Thus $\mathrm{tr}[\mathbf{A}^{-1}(\partial \mathbf{A}/\partial x)] = \mathrm{tr}[\mathbf{D}^{-1}(\partial \mathbf{D}/\partial x)]$ and the result follows. □

2.14.3 Maximization or Minimization of a Function of a Vector

Consider a function $u = f(\mathbf{x})$ of the p variables in $\mathbf{x}$. In many cases we can find a maximum or minimum of u by solving the system of p equations

$$\frac{\partial u}{\partial \mathbf{x}} = \mathbf{0}. \tag{2.119}$$

Occasionally the situation requires the maximization or minimization of the function u, subject to q constraints on $\mathbf{x}$. We denote the constraints as $h_1(\mathbf{x}) = 0,\ h_2(\mathbf{x}) = 0, \ldots, h_q(\mathbf{x}) = 0$ or, more succinctly, $\mathbf{h}(\mathbf{x}) = \mathbf{0}$. Maximization or minimization of u subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ can often be carried out by the method of Lagrange multipliers. We denote a vector of q unknown constants (the *Lagrange multipliers*) by $\boldsymbol{\lambda}$ and let $\mathbf{y}' = (\mathbf{x}', \boldsymbol{\lambda}')$. We then let $v = u + \boldsymbol{\lambda}'\mathbf{h}(\mathbf{x})$. The maximum or minimum of u subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ is obtained by solving the equations

$$\frac{\partial v}{\partial \mathbf{y}} = \mathbf{0}$$

or, equivalently

$$\frac{\partial u}{\partial \mathbf{x}} + \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\boldsymbol{\lambda} = \mathbf{0} \quad \text{and} \quad \mathbf{h}(\mathbf{x}) = \mathbf{0}, \tag{2.120}$$

where

$$\frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \begin{pmatrix} \dfrac{\partial h_1}{\partial x_1} & \cdots & \dfrac{\partial h_q}{\partial x_1} \\ \vdots & & \vdots \\ \dfrac{\partial h_1}{\partial x_p} & \cdots & \dfrac{\partial h_q}{\partial x_p} \end{pmatrix}.$$

PROBLEMS

2.1 Prove Theorem 2.2a.

2.2 Let $\mathbf{A} = \begin{pmatrix} 7 & -3 & 2 \\ 4 & 9 & 5 \end{pmatrix}$.

(**a**) Find $\mathbf{A}'$.

(**b**) Verify that $(\mathbf{A}')' = \mathbf{A}$, thus illustrating Theorem 2.1.

(**c**) Find $\mathbf{A}'\mathbf{A}$ and $\mathbf{A}\mathbf{A}'$.

2.3 Let $\mathbf{A} = \begin{pmatrix} 2 & 4 \\ -1 & 3 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 1 & 3 \\ 2 & -1 \end{pmatrix}$.

(**a**) Find $\mathbf{AB}$ and $\mathbf{BA}$.

(**b**) Find $|\mathbf{A}|$, $|\mathbf{B}|$, and $|\mathbf{AB}|$, and verify that Theorem 2.9c holds in this case.

(**c**) Find $|\mathbf{BA}|$ and compare to $|\mathbf{AB}|$.

(**d**) Find $(\mathbf{AB})'$ and compare to $\mathbf{B}'\mathbf{A}'$.

(**e**) Find $\text{tr}(\mathbf{AB})$ and compare to $\text{tr}(\mathbf{BA})$.

(**f**) Find the eigenvalues of $\mathbf{AB}$ and of $\mathbf{BA}$, thus illustrating Theorem 2.12a.

2.4 Let $\mathbf{A} = \begin{pmatrix} 1 & 3 & -4 \\ 5 & -7 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 3 & -2 & 5 \\ 6 & 9 & 7 \end{pmatrix}$.

(**a**) Find $\mathbf{A} + \mathbf{B}$ and $\mathbf{A} - \mathbf{B}$.

(**b**) Find $\mathbf{A}'$ and $\mathbf{B}'$.

(**c**) Find $(\mathbf{A} + \mathbf{B})'$ and $\mathbf{A}' + \mathbf{B}'$, thus illustrating Theorem 2.2a(ii).

2.5 Verify the distributive law in (2.15), $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$.

2.6 Let $\mathbf{A} = \begin{pmatrix} 8 & 3 & 7 \\ -2 & 5 & -3 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} -2 & 5 \\ 3 & 7 \\ 6 & -4 \end{pmatrix}$, $\mathbf{C} = \begin{pmatrix} 1 & 2 \\ -3 & 1 \\ 2 & 4 \end{pmatrix}$.

(**a**) Find $\mathbf{AB}$ and $\mathbf{BA}$.

(**b**) Find $\mathbf{B} + \mathbf{C}$, $\mathbf{AC}$, and $\mathbf{A}(\mathbf{B} + \mathbf{C})$. Compare $\mathbf{A}(\mathbf{B} + \mathbf{C})$ with $\mathbf{AB} + \mathbf{AC}$, thus illustrating (2.15).

(**c**) Compare $(\mathbf{AB})'$ with $\mathbf{B}'\mathbf{A}'$, thus illustrating Theorem 2.2b.

(**d**) Compare $\text{tr}(\mathbf{AB})$ with $\text{tr}(\mathbf{BA})$ and confirm that (2.87) holds in this case.

(**e**) Let $\mathbf{a}_1'$ and $\mathbf{a}_2'$ be the two rows of $\mathbf{A}$. Find $\begin{pmatrix} \mathbf{a}_1'\mathbf{B} \\ \mathbf{a}_2'\mathbf{B} \end{pmatrix}$ and compare with $\mathbf{AB}$ in part (a), thus illustrating (2.27).

(**f**) Let $\mathbf{b}_1$ and $\mathbf{b}_2$ be the two columns of $\mathbf{B}$. Find $(\mathbf{Ab}_1, \mathbf{Ab}_2)$ and compare with $\mathbf{AB}$ in part (a), thus illustrating (2.28).

2.7 Let $\mathbf{A} = \begin{pmatrix} 3 & 2 & 1 \\ 6 & 4 & 2 \\ 12 & 8 & 4 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} 1 & -1 & 2 \\ -1 & 1 & -2 \\ -1 & 1 & -2 \end{pmatrix}$.

(**a**) Show that $\mathbf{AB} = \mathbf{O}$.

(**b**) Find a vector $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{0}$.

(**c**) What is the rank of $\mathbf{A}$ and the rank of $\mathbf{B}$?

2.8 If $\mathbf{j}$ is a vector of 1s, as defined in (2.6), show that

(**a**) $\mathbf{j}'\mathbf{a} = \mathbf{a}'\mathbf{j} = \sum_i a_i$, as in (2.24).

(**b**) $\mathbf{Aj}$ is a column vector whose elements are the row sums of $\mathbf{A}$, as in (2.25).

(**c**) $\mathbf{j}'\mathbf{A}$ is a row vector whose elements are the column sums of $\mathbf{A}$, as in (2.25).

2.9 Prove Corollary 1 to Theorem 2.2b; that is, assuming that **A**, **B**, and **C** are conformal, show that $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$.

2.10 Prove Theorem 2.2c.

2.11 Use matrix **A** in Problem 2.6 and let

$$\mathbf{D}_1 = \begin{pmatrix} 3 & 0 \\ 0 & -2 \end{pmatrix}, \qquad \mathbf{D}_2 = \begin{pmatrix} 5 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 6 \end{pmatrix}.$$

Find $\mathbf{D}_1\mathbf{A}$ and $\mathbf{AD}_2$, thus illustrating (2.29) and (2.30).

2.12 Let $\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}, \qquad \mathbf{D} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}.$

Find **DA**, **AD**, and **DAD**.

2.13 For $\mathbf{y}' = (y_1, y_2, y_3)$ and the symmetric matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{12} & a_{22} & a_{23} \\ a_{13} & a_{23} & a_{33} \end{pmatrix},$$

express $\mathbf{y}'\mathbf{Ay}$ in the form given in (2.33).

2.14 Let $\mathbf{A} = \begin{pmatrix} 5 & -1 & 3 \\ -1 & 1 & 2 \\ 3 & 2 & 7 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 6 & -2 & 3 \\ 7 & 1 & 0 \\ 2. & -3 & 5 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 2 & -3 \\ -1 & 4 \\ 3 & 1 \end{pmatrix},$

$$\mathbf{x} = \begin{pmatrix} 3 \\ -1 \\ 2 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 3 \\ 2 \\ 4 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} 2 \\ 5 \end{pmatrix}.$$

Find the following:

(a) $\mathbf{Bx}$	**(h)** $\mathbf{xy}'$
(b) $\mathbf{y}'\mathbf{B}$	**(i)** $\mathbf{B}'\mathbf{B}$
(c) $\mathbf{x}'\mathbf{Ax}$	**(j)** $\mathbf{yz}'$
(d) $\mathbf{x}'\mathbf{Cz}$	**(k)** $\mathbf{zy}'$
(e) $\mathbf{x}'\mathbf{x}$	**(l)** $\sqrt{\mathbf{y}'\mathbf{y}}$
(f) $\mathbf{x}'\mathbf{y}$	**(m)** $\mathbf{C}'\mathbf{C}$
(g) $\mathbf{xx}'$	

2.15 Use **x**, **y**, **A**, and **B** as defined in Problem 2.14.

(a) Find $\mathbf{x} + \mathbf{y}$ and $\mathbf{x} - \mathbf{y}$.

(**b**) Find tr($\mathbf{A}$), tr($\mathbf{B}$), $\mathbf{A}+\mathbf{B}$, and tr($\mathbf{A}+\mathbf{B}$).

(**c**) Find $\mathbf{AB}$ and $\mathbf{BA}$.

(**d**) Find tr($\mathbf{AB}$) and tr($\mathbf{BA}$).

(**e**) Find $|\mathbf{AB}|$ and $|\mathbf{BA}|$.

(**f**) Find $(\mathbf{AB})'$ and $\mathbf{B}'\mathbf{A}'$.

2.16 Using $\mathbf{B}$ and $\mathbf{x}$ in Problem 2.14, find $\mathbf{Bx}$ as a linear combination of the columns of $\mathbf{B}$, as in (2.37), and compare with $\mathbf{Bx}$ as found in Problem 2.14(a).

2.17 Let $\mathbf{A}=\begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$, $\quad \mathbf{B}=\begin{pmatrix} 1 & -6 & 2 \\ 5 & 0 & 3 \end{pmatrix}$, $\quad \mathbf{I}=\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

(**a**) Show that $(\mathbf{AB})'=\mathbf{B}'\mathbf{A}'$ as in (2.26).

(**b**) Show that $\mathbf{AI}=\mathbf{A}$ and that $\mathbf{IB}=\mathbf{B}$.

(**c**) Find $|\mathbf{A}|$.

(**d**) Find $\mathbf{A}^{-1}$.

(**e**) Find $(\mathbf{A}^{-1})^{-1}$ and compare with $\mathbf{A}$, thus verifying (2.46).

(**f**) Find $(\mathbf{A}')^{-1}$ and verify that it is equal to $(\mathbf{A}^{-1})'$ as in Theorem 2.5a.

2.18 Let $\mathbf{A}$ and $\mathbf{B}$ be defined and partitioned as follows:

$$\mathbf{A}=\left(\begin{array}{cc|c} 2 & 1 & 2 \\ 3 & 2 & 0 \\ \hline 1 & 0 & 1 \end{array}\right), \qquad \mathbf{B}=\left(\begin{array}{ccc|c} 1 & 1 & 1 & 0 \\ 2 & 1 & 1 & 2 \\ \hline 2 & 3 & 1 & 2 \end{array}\right).$$

(**a**) Find $\mathbf{AB}$ as in (2.35), using the indicated partitioning.

(**b**) Check by finding $\mathbf{AB}$ in the usual way, ignoring the partitioning.

2.19 Partition the matrices $\mathbf{A}$ and $\mathbf{B}$ in Problem 2.18 as follows:

$$\mathbf{A}=\left(\begin{array}{c|cc} 2 & 1 & 2 \\ 3 & 2 & 0 \\ 1 & 0 & 1 \end{array}\right)=(\mathbf{a}_1, \mathbf{A}_2),$$

$$\mathbf{B}=\left(\begin{array}{cccc} 1 & 1 & 1 & 0 \\ \hline 2 & 1 & 1 & 2 \\ 2 & 3 & 1 & 2 \end{array}\right)=\begin{pmatrix} \mathbf{b}_1' \\ \mathbf{B}_2 \end{pmatrix}.$$

Repeat parts (a) and (b) of Problem 2.18. Note that in this case, (2.35) becomes $\mathbf{AB}=\mathbf{a}_1\mathbf{b}_1'+\mathbf{A}_2\mathbf{B}_2$.

2.20 Let $\mathbf{A} = \begin{pmatrix} 5 & -2 & 3 \\ 7 & 3 & 1 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 2 \\ 4 \\ -3 \end{pmatrix}.$

Find $\mathbf{Ab}$ as a linear combination of the columns of $\mathbf{A}$ as in (2.37) and check the result by finding $\mathbf{Ab}$ in the usual way.

2.21 Show that each column of the product $\mathbf{AB}$ can be expressed as a linear combination of the columns of $\mathbf{A}$, with coefficients arising from the corresponding column of $\mathbf{B}$, as noted following Example 2.3.

2.22 Let $\mathbf{A} = \begin{pmatrix} 3 & 0 & 2 \\ 1 & -1 & 1 \\ 2 & 1 & 0 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} -2 & -1 \\ 3 & 1 \\ 1 & -1 \end{pmatrix}.$

Express the columns of $\mathbf{AB}$ as linear combinations of the columns of $\mathbf{A}$.

2.23 Show that if a set of vectors includes $\mathbf{0}$, the set is linearly dependent, as noted following (2.40).

2.24 Suppose that $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ and that $\mathbf{AB} = \mathbf{O}$ as in (2.43). Show that $\mathbf{A}$ and $\mathbf{B}$ are both singular or one of them is $\mathbf{O}$.

2.25 Let $\mathbf{A} = \begin{pmatrix} 1 & 3 & 2 \\ 2 & 0 & -1 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} 2 & 1 & 1 \\ 5 & -6 & -4 \end{pmatrix}.$

Find $\mathbf{AB}$ and $\mathbf{CB}$. Are they equal? What are the ranks of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$?

2.26 Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 2 \\ 1 & 0 & -1 \end{pmatrix}, \qquad \mathbf{B} = \begin{pmatrix} 2 & 1 \\ 0 & 2 \\ 1 & 0 \end{pmatrix}.$

(**a**) Find a matrix $\mathbf{C}$ such that $\mathbf{AB} = \mathbf{CB}$. Is $\mathbf{C}$ unique?

(**b**) Find a vector $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{0}$. Can you do this for $\mathbf{B}$?

2.27 Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 2 \\ 4 & -2 & 3 \\ 1 & 0 & -1 \end{pmatrix}, \qquad \mathbf{x} = \begin{pmatrix} 5 \\ 2 \\ 3 \end{pmatrix}.$

(**a**) Find a matrix $\mathbf{B} \neq \mathbf{A}$ such that $\mathbf{Ax} = \mathbf{Bx}$. Why is this possible? Can $\mathbf{A}$ and $\mathbf{B}$ be nonsingular? Can $\mathbf{A} - \mathbf{B}$ be nonsingular?

(**b**) Find a matrix $\mathbf{C} \neq \mathbf{O}$ such that $\mathbf{Cx} = \mathbf{0}$. Can $\mathbf{C}$ be nonsingular?

2.28 Prove Theorem 2.5a.

2.29 Prove Theorem 2.5b.

2.30 Use the matrix $\mathbf{A}$ in Problem 2.17, and let $\mathbf{B} = \begin{pmatrix} 4 & -2 \\ 3 & 1 \end{pmatrix}$. Find $\mathbf{AB}$, $\mathbf{B}^{-1}$, and $(\mathbf{AB})^{-1}$. Verify that Theorem 2.5b holds in this case.

2.31 Show that the partitioned matrix $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$ has the inverse indicated in (2.50).

2.32 Show that the partitioned matrix $\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{a}_{12} \\ \mathbf{a}'_{12} & a_{22} \end{pmatrix}$ has the inverse given in (2.51).

2.33 Show that $\mathbf{B} + \mathbf{cc}'$ has the inverse indicated in (2.53).

2.34 Show that $\mathbf{A} + \mathbf{PBQ}$ has the inverse indicated in (2.54).

2.35 Show that $\mathbf{y}'\mathbf{Ay} = \mathbf{y}'\left[\frac{1}{2}(\mathbf{A} + \mathbf{A}')\right]\mathbf{y}$ as in (2.55).

2.36 Prove Theorem 2.6b(ii).

2.37 Prove Corollaries 1 and 2 of Theorem 2.6b.

2.38 Prove the "only if" part of Theorem 2.6c.

2.39 Prove Corollary 1 to Theorem 2.6c.

2.40 Compare the rank of the augmented matrix with the rank of the coefficient matrix for each of the following systems of equations. Find solutions where they exist.

(**a**)
$$\begin{aligned} x_1 + 2x_2 + 3x_3 &= 6 \\ x_1 - x_2 &= 2 \\ x_1 - x_3 &= -1 \end{aligned}$$

(**b**)
$$\begin{aligned} x_1 - x_2 + 2x_3 &= 2 \\ x_1 - x_2 - x_3 &= -1 \\ 2x_1 - 2x_2 + x_3 &= 2 \end{aligned}$$

(**c**)
$$\begin{aligned} x_1 + x_2 + x_3 + x_4 &= 8 \\ x_1 - x_2 - x_3 - x_4 &= 6 \\ 3x_1 + x_2 + x_3 + x_4 &= 22 \end{aligned}$$

2.41 Prove Theorem 2.8a.

2.42 For the matrices $\mathbf{A}, \mathbf{A}_1^-$, and $\mathbf{A}_2^-$ in (2.59) and (2.60), show that $\mathbf{AA}_1^-\mathbf{A} = \mathbf{A}$ and $\mathbf{AA}_2^-\mathbf{A} = \mathbf{A}$.

2.43 Show that $\mathbf{A}_1^-$ in (2.60) can be obtained using Theorem 2.8b.

2.44 Show that $\mathbf{A}_2^-$ in (2.60) can be obtained using the five-step algorithm following Theorem 2.8b.

2.45 Prove Theorem 2.8c.

2.46 Show that if $\mathbf{A}$ is symmetric, there exists a symmetric generalized inverse for $\mathbf{A}$, as noted following Theorem 2.8c.

2.47 Let $\mathbf{A} = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 2 & 0 \\ 2 & 0 & 2 \end{pmatrix}$.

(**a**) Find a symmetric generalized inverse for $\mathbf{A}$.

(**b**) Find a nonsymmetric generalized inverse for $\mathbf{A}$.

2.48 (**a**) Show that if $\mathbf{A}$ is nonsingular, then $\mathbf{A}^- = \mathbf{A}^{-1}$.

(**b**) Show that if $\mathbf{A}$ is $n \times p$ of rank $p < n$, then $\mathbf{A}^-$ is a "left inverse" of $\mathbf{A}$, that is, $\mathbf{A}^-\mathbf{A} = \mathbf{I}$.

2.49 Prove Theorem 2.9a parts (iv) and (vi).

2.50 Use $\mathbf{A} = \begin{pmatrix} 2 & 5 \\ 1 & 3 \end{pmatrix}$ from Problem 2.17 to illustrate (64), (2.66), and (2.67) in Theorem 2.9a.

2.51 (**a**) Multiply $\mathbf{A}$ in Problem 2.50 by 10 and verify that (2.69) holds in this case.
(**b**) Verify that (2.69) holds in general.

2.52 Prove Corollaries 1, 2, 3, and 4 of Theorem 2.9b.

2.53 Prove Corollaries 1 and 2 of Theorem 2.9c.

2.54 Use $\mathbf{A}$ in Problem 2.50 and let $\mathbf{B} = \begin{pmatrix} 4 & -2 \\ 3 & 1 \end{pmatrix}$.

(**a**) Find $|\mathbf{A}|, |\mathbf{B}|, \mathbf{AB}$, and $|\mathbf{AB}|$ and illustrate (2.77).
(**b**) Find $|\mathbf{A}|^2$ and $|\mathbf{A}^2|$ and illustrate (2.79).

2.55 Use Theorem 2.9c and Corollary 1 of Theorem 2.9b to prove Theorem 2.9b.

2.56 Show that if $\mathbf{C}'\mathbf{C} = \mathbf{I}$, then $\mathbf{CC}' = \mathbf{I}$ as in (2.84).

2.57 The columns of the following matrix are mutually orthogonal:

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 0 & 2 \\ 1 & 1 & 1 \end{pmatrix}.$$

(**a**) Normalize the columns of $\mathbf{A}$ by dividing each column by its length; denote the resulting matrix by $\mathbf{C}$.
(**b**) Show that $\mathbf{C}'\mathbf{C} = \mathbf{CC}' = \mathbf{I}$.

2.58 Prove Theorem 2.10a.

2.59 Prove Theorem 2.11 parts (i), (iv), (v), and (vii).

2.60 Use matrix $\mathbf{B}$ in Problem 2.26 to illustrate Theorem 2.11 parts (iii) and (iv).

2.61 Use matrix $\mathbf{A}$ in Problem 2.26 to illustrate Theorem 2.11(v), that is, $\mathrm{tr}(\mathbf{A}'\mathbf{A}) = \mathrm{tr}(\mathbf{AA}') = \sum_{ij} a_{ij}^2$.

2.62 Show that $\mathrm{tr}(\mathbf{A}^-\mathbf{A}) = \mathrm{tr}(\mathbf{AA}^-) = r = \mathrm{rank}(\mathbf{A})$, as in (2.93).

2.63 Use $\mathbf{A}$ in (2.59) and $\mathbf{A}_2^-$ in (2.60) to illustrate Theorem 2.11(viii), that is, $\mathrm{tr}(\mathbf{A}^-\mathbf{A}) = \mathrm{tr}(\mathbf{AA}^-) = r = \mathrm{rank}(\mathbf{A})$.

2.64 Obtain $\mathbf{x}_2 = (2/\sqrt{5}, 1/\sqrt{5})'$ in Example 2.12.1.

2.65 For $k = 3$, show that $\mathbf{A}^k\mathbf{x} = \lambda^k\mathbf{x}$ as in (2.100).

2.66 Show that $\lim_{k\to\infty} \mathbf{A}^k = \mathbf{O}$ in (2.102) if $\mathbf{A}$ is symmetric and if all eigenvalues of $\mathbf{A}$ satisfy $-1 < \lambda < 1$.

2.67 Prove Theorem 2.12a.

2.68 Prove Theorem 2.12b.

2.69 Prove Theorem 2.12c(ii) for the case where the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct.

2.70 Prove Corollary 1 to Theorem 2.12d.

2.71 Let $\mathbf{A} = \begin{pmatrix} 3 & 1 & 1 \\ 1 & 0 & 2 \\ 1 & 2 & 0 \end{pmatrix}$.

(**a**) The eigenvalues of $\mathbf{A}$ are $1, 4, -2$. Find the normalized eigenvectors and use them as columns in an orthogonal matrix $\mathbf{C}$.

(**b**) Show that $\mathbf{A} = \mathbf{CDC}'$, as in (2.103), where $\mathbf{D} = \text{diag}(1, 4, -2)$.

(**c**) Show that $\mathbf{C}'\mathbf{AC} = \mathbf{D}$ as in (2.106).

2.72 Prove Theorem 2.12e for a symmetric matrix $\mathbf{A}$.

2.73 Let $\mathbf{A} = \begin{pmatrix} 1 & 1 & -2 \\ -1 & 2 & 1 \\ 0 & 1 & -1 \end{pmatrix}$.

(**a**) Find the eigenvalues and associated normalized eigenvectors.

(**b**) Find $\text{tr}(\mathbf{A})$ and $|\mathbf{A}|$ and verify that $\text{tr}(\mathbf{A}) = \sum_{i=1}^{3} \lambda_i$ and $|\mathbf{A}| = \prod_{i=1}^{3} \lambda_i$, as in Theorem 2.12e.

2.74 Prove Theorem 2.12f(ii).

2.75 Let $\mathbf{A} = \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \\ -1 & -1 & 3 \end{pmatrix}$.

(**a**) Show that $|\mathbf{A}| > 0$.

(**b**) Find the eigenvalues of $\mathbf{A}$. Are they all positive?

2.76 Let $\mathbf{A}^{1/2}$ be defined as in (2.109).

(**a**) Show that $\mathbf{A}^{1/2}$ is symmetric.

(**b**) Show that $(\mathbf{A}^{1/2})^2 = \mathbf{A}$ as in (2.110).

2.77 For the positive definite matrix $\mathbf{A} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}$, calculate the eigenvalues and eigenvectors and find the square root matrix $\mathbf{A}^{1/2}$ as in (2.109). Check by showing $(\mathbf{A}^{1/2})^2 = \mathbf{A}$.

2.78 Prove Theorem 2.13e.

2.79 Prove Theorem 2.13f.

2.80 Let $\mathbf{A} = \begin{pmatrix} \frac{2}{3} & 0 & \frac{\sqrt{2}}{3} \\ 0 & 1 & 0 \\ \frac{\sqrt{2}}{3} & 0 & \frac{1}{3} \end{pmatrix}$.

(**a**) Find the rank of $\mathbf{A}$.

(**b**) Show that $\mathbf{A}$ is idempotent.

(**c**) Show that $\mathbf{I} - \mathbf{A}$ is dempotent.

(**d**) Show that $\mathbf{A}(\mathbf{I} - \mathbf{A}) = \mathbf{O}$.

(**e**) Find tr($\mathbf{A}$).

(**f**) Find the eigenvalues of $\mathbf{A}$.

2.81 Consider a $p \times p$ matrix $\mathbf{A}$ with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_p$. Show that $[\mathrm{tr}(\mathbf{A})]^2 = \mathrm{tr}(\mathbf{A}^2) + 2\sum\sum_{i \neq j} \lambda_i \lambda_j$.

2.82 Consider a nonsingular $n \times n$ matrix $\mathbf{A}$ whose elements are functions of the scalar x. Also consider the full-rank $p \times n$ matrix $\mathbf{B}$. Let $\mathbf{H} = \mathbf{B}'(\mathbf{BAB}')^{-1}\mathbf{B}$. Show that

$$\frac{\partial \mathbf{H}}{\partial \mathbf{x}} = -\mathbf{H}\frac{\partial \mathbf{A}}{\partial \mathbf{x}}\mathbf{H}.$$

2.83 Show that

$$\frac{\partial \ln |\mathbf{X}|}{\partial \mathbf{X}} = 2\mathbf{X}^{-1} - \mathrm{diag}\,\mathbf{X}^{-1}$$

for a 2×2 positive definite matrix $\mathbf{X}$.

2.84 Let $u = \mathbf{x}'\mathbf{A}\mathbf{x}$ where $\mathbf{x}$ is a 3×1 vector and $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Use the Lagrange multiplier method to find the vector $\mathbf{x}$ that minimizes u subject to the constraints $x_1 + x_2 = 2$, and $x_2 + x_3 = 3$.