Chapter 1

MATRIX ALGEBRA

We will present in this chapter some aspects of matrix algebra that are needed in this book. Most results presented here can be found in standard books on matrix algebra. Proofs are provided for those results that are either easily derived or not readily available elsewhere. For readers who have no knowledge of matrix algebra, this chapter is essential for the rest of the book. They may find the one-chapter treatment of matrix algebra in the book by Hildebrand (1954) helpful and informative. For readers who have some knowledge of matrix algebra this chapter can be skimmed or skipped altogether, depending on how familiar they are with the subject. If they want to find the proofs omitted in this chapter or want to devote more time on the subject, the books by Hohn (1965) and Pease (1965) are recommended. The notations employed in this chapter have no relations, in most cases, with the notations adopted in the rest of the book. This point should be kept in mind in referring back to this chapter.

1.1 Notations, Definitions, and Identities

A matrix is a rectangular array of numbers, variables, or functions. The *order* of a matrix is defined by the number of rows and columns in the matrix. If a matrix has n rows and m columns it is called an n *by* m, or $n \times m$ matrix. With few exceptions a matrix with more than one row and one column is denoted by a bold-face upper case letters **A**, **B**,... Thus the $n \times m$ matrix **A** is written as

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1m} \\ A_{21} & A_{22} & \cdots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nm} \end{bmatrix} \tag{1.1-1}$$

in which A_{ij} denotes an element of the matrix. The first index i of A_{ij} specifies the row in which the element is located while the second index j specifies the column in which the element is located. When the rows and the columns of the matrix **A** are interchanged, we have the *transpose* of the matrix which is denoted by $\mathbf{A}^T$. Thus the transpose of (1.1-1) is

$$\mathbf{A}^T = \begin{bmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1m} & A_{2m} & \cdots & A_{nm} \end{bmatrix}, \tag{1.1-2}$$

In contrast to **A** that has the order $n \times m$, $\mathbf{A}^T$ has the order $m \times n$.

A vector with n components can be represented by an $n \times 1$ matrix. Since it contains only one column, it is called a *column matrix*. A vector or a column matrix will be denoted by a bold-face lower case letters **a**, **b**,... If

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix},$$

its transpose

$$\mathbf{a}^T = [a_1, \ a_2, \ \ldots, \ a_n]$$

is an one-row matrix.

The addition of matrices **A** and **B** is written as

$$\mathbf{C} = \mathbf{A} + \mathbf{B}.$$

The elements C_{ij} are related to A_{ij} and B_{ij} by

$$C_{ij} = A_{ij} + B_{ij}. \tag{1.1-3}$$

Therefore two matrices must be of the same order before they can be added.

If **C** is the product of **A** and **B**, we write

$$\mathbf{C} = \mathbf{AB}.$$

The elements C_{ij} are related to A_{ij} and B_{ij} by

$$C_{ij} = A_{ik} B_{kj} \tag{1.1-4}$$

in which repeated indices imply summation. The product makes sense only when the number of columns in **A** is identical to the number of rows in **B**. If **A** is an $n \times m$ matrix and **B** is an $m \times n$ matrix, **AB** and **BA** have the order $n \times n$ and $m \times m$, respectively. Hence the matrix products **AB** and **BA** are different. When **A** and **B** are both of order $n \times n$, so are **AB** and **BA**. Nevertheless the product of **A** and **B** in general does not commute, i.e.,

$$\mathbf{AB} \neq \mathbf{BA}. \tag{1.1-5}$$

Therefore

$$\begin{aligned}(\mathbf{A}+\mathbf{B})^2 &= (\mathbf{A}+\mathbf{B})(\mathbf{A}+\mathbf{B}) \\ &= \mathbf{A}^2+\mathbf{AB}+\mathbf{BA}+\mathbf{A}^2 \\ &\neq \mathbf{A}^2+2\mathbf{AB}+\mathbf{B}^2,\end{aligned}$$

in general. We will study in Section 1.5 conditions under which the product of two $n\times n$ matrices commutes.

It can be shown that the transpose of the product of two matrices is equal to the product of the transposes of the two matrices in the reverse order, i.e.,

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T. \tag{1.1-6}$$

The *scalar* product of two vectors $\mathbf{a}$ and $\mathbf{b}$ is denoted by

$$\mathbf{a}\cdot\mathbf{b} \quad \text{or} \quad \mathbf{a}^T\mathbf{b}.$$

The former is in vector notation and the latter in matrix notation. The *length* $|\mathbf{a}|$ of a vector $\mathbf{a}$ is

$$|\mathbf{a}| = (\mathbf{a}\cdot\mathbf{a})^{1/2} \quad \text{or} \quad |\mathbf{a}| = \left(\mathbf{a}^T\mathbf{a}\right)^{1/2}. \tag{1.1-7}$$

The length so obtained may not be real when $\mathbf{a}$ is complex. In fact the length may be zero when $\mathbf{a}$ is complex. An example is $\mathbf{a}=\mathbf{e}_1+i\mathbf{e}_2$ where $\mathbf{e}_1$ and $\mathbf{e}_2$ are real unit vectors orthogonal to each other. A complex vector $\mathbf{a}$ is called *isotropic* if $\mathbf{a}\cdot\mathbf{a}=0$ (Hayes, 1984). If $\mathbf{a}$ is not isotropic, and if $\mathbf{a}$ is an eigenvector which is unique up to a constant multiplier, the length of $\mathbf{a}$ can be made unity by choosing the constant multiplier properly. In some literature the length of a complex vector is defined as

$$(\bar{\mathbf{a}}\cdot\mathbf{a})^{1/2}$$

where the overbar denotes the complex conjugate. We will not employ this definition here.

An $n\times n$ matrix is called a *square* matrix of order n. Its transpose is also a square matrix of order n. The matrix is *symmetric* if

$$\mathbf{A} = \mathbf{A}^T. \tag{1.1-8}$$

A one-element matrix, i.e., an 1×1 matrix is obviously symmetric. A matrix is *skew-symmetric*, or *antisymmetric*, if

$$\mathbf{A} = -\mathbf{A}^T. \tag{1.1-9}$$

The elements $A_{11}, A_{22},\ldots A_{66}$ of the matrix are called the *diagonal*

elements. It is readily shown that the diagonal elements of a skew-symmetric matrix are zeros.

A *diagonal matrix* is a square matrix whose non-diagonal elements are zeros. It is a symmetric matrix. If

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{1.1-10a}$$

we may write it as

$$\Lambda = \operatorname{diag}[\lambda_1,\ \lambda_2,\ \dots\ \lambda_n] \tag{1.1-10b}$$

or

$$\Lambda = \langle \lambda_* \rangle. \tag{1.1-10c}$$

An *identity matrix* $\mathbf{I}$ is a diagonal matrix whose diagonal elements are ones. Thus

$$\mathbf{I} = \operatorname{diag}[1,\ 1,\ \dots,\ 1].$$

It has the property that

$$\mathbf{AI} = \mathbf{A}, \qquad \mathbf{IA} = \mathbf{A}, \tag{1.1-11}$$

for any matrix $\mathbf{A}$. If $\mathbf{A}$ is an $n \times m$ matrix, the $\mathbf{I}$ in $(1.1\text{-}11)_1$ is a diagonal matrix of order m while the $\mathbf{I}$ in $(1.1\text{-}11)_2$ is a diagonal matrix of order n. Thus the order of the unit matrix $\mathbf{I}$ is self-determined from the equation it is in.

The *trace* of a square matrix $\mathbf{A}$, written as $\mathrm{tr}\mathbf{A}$, is the sum of all its diagonal elements, i.e.,

$$\mathrm{tr}\mathbf{A} = A_{11} + A_{22} + \dots + A_{nn}. \tag{1.1-12}$$

It is clear that

$$\mathrm{tr}(\mathbf{A}^T) = \mathrm{tr}\mathbf{A} \tag{1.1-13}$$

and, from the definition of the matrix product (1.1-4),

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}). \tag{1.1-14}$$

The matrices $\mathbf{A}$ and $\mathbf{B}$ in (1.1-14) need not be square matrices as long as $\mathbf{AB}$ and $\mathbf{BA}$ are.

The *determinant* $|\mathbf{A}|$ of a square matrix $\mathbf{A}$ is

$$|\mathbf{A}| = \delta_{ij\dots s} A_{i1} A_{j2} \dots A_{sn},$$

where again repeated indices imply summation. The *permutation symbol* $\delta_{ij\dots s}$ vanishes if any two subscripts are identical and

assumes the value of +1 or −1 if the subscripts $i,j,\ldots,s$ form an even or odd permutation, respectively. The two-index permutation symbol δ_{ij} is the *Kronecker delta*. It can be shown that

$$|\mathbf{A}^T| = |\mathbf{A}|, \tag{1.1-15}$$

and that

$$|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|. \tag{1.1-16}$$

If **A** is a 3×3 matrix,

$$|\mathbf{A}| = \delta_{ijk} A_{i1} A_{j2} A_{k3} \tag{1.1-17a}$$

or

$$\delta_{pqr} |\mathbf{A}| = \delta_{ijk} A_{ip} A_{jq} A_{kr}.$$

Since

$$\delta_{pqr} \delta_{pqr} = 6,$$

we have an alternate expression

$$|\mathbf{A}| = \tfrac{1}{6} \delta_{ijk} \delta_{pqr} A_{ip} A_{jq} A_{kr}. \tag{1.1-17b}$$

The *cofactor* $\hat{A}_{i1}$ is the coefficient of A_{i1} in (1.1-17a), i.e.,

$$\hat{A}_{i1} = \delta_{ijk} A_{j2} A_{k3} = \delta_{ijk} \delta_{123} A_{j2} A_{k3} = \tfrac{1}{2} \delta_{ijk} \delta_{1qr} A_{jq} A_{kr}. \tag{1.1-18}$$

Hence

$$\hat{A}_{ip} = \tfrac{1}{2} \delta_{ijk} \delta_{pqr} A_{jq} A_{kr}. \tag{1.1-19}$$

We also have the relation

$$\mathbf{A}\hat{\mathbf{A}}^T = |\mathbf{A}|\mathbf{I} \tag{1.1-20}$$

where the matrix $\hat{\mathbf{A}}^T$ is called the *adjoint matrix* of **A**.

A useful identity is (Ting, 1970)

$$|\mathbf{A} + \alpha \mathbf{g}\mathbf{h}^T| = |\mathbf{A}| + \alpha \mathbf{g}^T \check{\mathbf{A}} \mathbf{h} \tag{1.1-21}$$

where **A** is a square matrix of order n, **g** and **h** are n×1 column matrices, and α is a scalar.

A square matrix **A** is *singular* or *nonsingular* when $|\mathbf{A}|$ is zero or nonzero. If **A** is nonsingular, there exist matrices **C** and **D** such that

$$\mathbf{AC} = \mathbf{I}, \qquad \mathbf{DA} = \mathbf{I}.$$

If we post-multiply the second equation by **C** and make use of the first equation and (1.1-11), we have

$$\mathbf{D} = \mathbf{C}.$$

Therefore one has

$$\mathbf{AC} = \mathbf{I} = \mathbf{CA}. \tag{1.1-22}$$

This is a special case in which the matrix product commutes. The matrix $\mathbf{C}$ is the *inverse* of $\mathbf{A}$, and is written as

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

Equations (1.1-22) are equivalent to

$$\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}. \tag{1.1-23}$$

It is easily shown that $\mathbf{A}^{-1}$ is symmetric if $\mathbf{A}$ is. When $\mathbf{A}$, $\mathbf{B}$ are both nonsingular, so is $\mathbf{AB}$. Let

$$\mathbf{U} = (\mathbf{AB})^{-1}.$$

From

$$(\mathbf{AB})\mathbf{U} = \mathbf{I}, \quad \mathbf{BU} = \mathbf{A}^{-1}, \quad \mathbf{U} = \mathbf{B}^{-1}\mathbf{A}^{-1},$$

we have

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{1.1-24}$$

A square matrix $\mathbf{A}$ is *orthogonal*, or *rotational*, if

$$\mathbf{A}^T = \mathbf{A}^{-1}. \tag{1.1-25}$$

Therefore an orthogonal matrix satisfies the relations

$$\mathbf{AA}^T = \mathbf{I} = \mathbf{A}^T\mathbf{A}. \tag{1.1-26}$$

Taking the determinant of the matrices on both sides of the equalities we obtain

$$|\mathbf{A}|^2 = 1, \quad \text{or} \quad |\mathbf{A}| = \pm 1. \tag{1.1-27}$$

An orthogonal matrix is called *proper* orthogonal when its determinant is $+1$, and *improper* orthogonal when its determinant is -1.

When an orthogonal matrix $\mathbf{A}$ is real it has the following geometrical interpretation. Consider the equation

$$\mathbf{x}^* = \mathbf{Ax}$$

which transforms a real vector $\mathbf{x}$ into a new vector $\mathbf{x}^*$ through the matrix $\mathbf{A}$. A different vector $\mathbf{y}$ will be transformed into $\mathbf{y}^*$ by

$$\mathbf{y}^* = \mathbf{Ay}.$$

When $\mathbf{A}$ is orthogonal, it is readily shown that

$$(\mathbf{x}^*)^T\mathbf{y}^* = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{y}) = \mathbf{x}^T\mathbf{A}^T\mathbf{A}\mathbf{y} = \mathbf{x}^T\mathbf{y}. \tag{1.1-28}$$

In the particular case when $\mathbf{y} = \mathbf{x}$, (1.1-28) tells us that the length of a vector is unchanged after the transformation. Using this result, (1.1-28) implies that the angle between any two vectors is unchanged after the transformation. Therefore a real orthogonal matrix transforms a space into another space without distorting the space. The transformation is a rigid body rotation if $\mathbf{A}$ is proper orthogonal. It is a rigid body rotation and a mirror reflection about a plane when $\mathbf{A}$ is improper orthogonal. If three vectors $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$ form a right-hand triad before the transformation, the three vectors $\mathbf{a}^*$, $\mathbf{b}^*$, $\mathbf{c}^*$ after the transformation form a right-hand triad if $\mathbf{A}$ is proper orthogonal, and a left-hand triad if $\mathbf{A}$ is improper orthogonal.

1.2 Eigenvalues and Eigenvectors

The standard *eigenrelation* for a square matrix $\mathbf{A}$ which may be complex is

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}. \tag{1.2-1}$$

λ is an eigenvalue and $\mathbf{x}$ is an eigenvector. Rewriting it as

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0} \tag{1.2-2}$$

we must have, for a nontrivial solution of $\mathbf{x}$,

$$|\mathbf{A} - \lambda\mathbf{I}| = 0. \tag{1.2-3}$$

Expansion of the determinant leads to the *eigenequation* or the *characteristic equation* for λ:

$$\lambda^n - c_1\lambda^{n-1} + \ldots + (-1)^n c_n = 0, \tag{1.2-4}$$

which provides n roots for the eigenvalues $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_n$. The two coefficients c_1 and c_n have particularly simple relations with λ_i and $\mathbf{A}$. They are

$$\begin{aligned} c_1 &= \lambda_1 + \lambda_2 + \ldots + \lambda_n = \operatorname{tr}\mathbf{A}, \\ c_n &= \lambda_1\lambda_2\ldots\lambda_n = |\mathbf{A}|. \end{aligned} \tag{1.2-5}$$

For $n = 3$ the eigenequation is

$$\lambda^3 - c_1\lambda^2 + c_2\lambda - c_3 = 0, \tag{1.2-6a}$$

where

$$c_1 = \lambda_1 + \lambda_2 + \lambda_3 = \text{tr}\,\mathbf{A},$$
$$c_2 = \lambda_1\lambda_2 + \lambda_2\lambda_3 + \lambda_3\lambda_1 = \tfrac{1}{2}\left\{(\text{tr}\mathbf{A})^2 - \text{tr}(\mathbf{A}^2)\right\}, \qquad (1.2\text{-}6b)$$
$$c_3 = \lambda_1\lambda_2\lambda_3 = |\mathbf{A}|.$$

The vector $\mathbf{x}$ in (1.2-1) is a *right eigenvector*. The *left eigenvector*, or the *reciprocal eigenvector*, $\mathbf{y}$ is defined by

$$\mathbf{y}^T\mathbf{A} = \lambda\mathbf{y}^T \quad \text{or} \quad \mathbf{A}^T\mathbf{y} = \lambda\mathbf{y}. \qquad (1.2\text{-}7)$$

The eigenvalue λ of $(1.2\text{-}7)_2$ satisfies the equation

$$|\mathbf{A}^T - \lambda\mathbf{I}| = 0$$

which is identical to (1.2-3) in view of (1.1-15).

Let $\mathbf{x}_i$ and $\mathbf{y}_i$ $(i = 1,2,\ldots,n)$ be the right and left eigenvectors associated with the eigenvalue λ_i. From

$$\mathbf{A}\mathbf{x}_2 = \lambda_2\mathbf{x}_2, \quad \mathbf{y}_1^T\mathbf{A} = \lambda_1\mathbf{y}_1^T,$$

and pre-multiplying the first equation by $\mathbf{y}_1^T$ and post-multiplying the second equation by $\mathbf{x}_2$, we obtain

$$0 = (\lambda_1 - \lambda_2)\mathbf{y}_1^T\mathbf{x}_2.$$

Therefore

$$\mathbf{y}_1 \cdot \mathbf{x}_2 = 0, \quad \text{if } \lambda_1 \neq \lambda_2,$$

or, more generally,

$$\mathbf{y}_i \cdot \mathbf{x}_j = 0, \quad \text{if } \lambda_i \neq \lambda_j. \qquad (1.2\text{-}8)$$

The left eigenvectors and the right eigenvectors associated with different eigenvalues are *orthogonal* to each other. If λ_i is a simple root, i.e., it is not a repeated root, $\mathbf{x}_i$ and $\mathbf{y}_i$ are unique up to a constant multiplier. We may *normalize* $\mathbf{x}_i$ and $\mathbf{y}_i$ such that

$$\mathbf{y}_i \cdot \mathbf{x}_i = 1, \quad i \text{ not summed}. \qquad (1.2\text{-}9)$$

Since $\mathbf{x}_i$ and $\mathbf{y}_i$ each has a constant multiplier, (1.2-9) does not provide a unique solution for $\mathbf{x}_i$ and $\mathbf{y}_i$.

When $\mathbf{A}$ is symmetric the left eigenvector $\mathbf{y}$ and the right eigenvector $\mathbf{x}$ can be taken identical. If $\mathbf{A}$ is real and symmetric the eigenvalues λ's and the associated eigenvectors $\mathbf{x}$'s can be shown to be real. If λ and $\mathbf{x}$ were complex we have by taking the complex conjugate of (1.2-1),

$$\mathbf{A}\bar{\mathbf{x}} = \bar{\lambda}\bar{\mathbf{x}}.$$

Hence

$$\mathbf{x}^T\mathbf{A}\bar{\mathbf{x}} = \bar{\lambda}\mathbf{x}^T\bar{\mathbf{x}}$$

or, after transposing both sides and noticing that $\mathbf{A}^T=\mathbf{A}$,

$$\bar{\mathbf{x}}^T\mathbf{A}\mathbf{x} = \bar{\lambda}\bar{\mathbf{x}}^T\mathbf{x}. \tag{1.2-10}$$

If we pre-multiply (1.2-1) by $\bar{\mathbf{x}}^T$ and subtract (1.2-10) from the result we have

$$(\lambda - \bar{\lambda})\bar{\mathbf{x}}^T\mathbf{x} = 0.$$

Since $\bar{\mathbf{x}}^T\mathbf{x}$ is real and nonzero, $\lambda = \bar{\lambda}$, which means that λ is real.

When $\mathbf{A}$ is singular $\lambda = 0$ is an eigenvalue. The eigenvector associated with $\lambda = 0$ is called the *null vector*. If the matrix is not symmetric, the right and left null vectors are in general different.

Let $\mathbf{W}$ be a skew-symmetric matrix of order n, i.e.,

$$\mathbf{W} = -\mathbf{W}^T .$$

When n is an odd integer,

$$|\mathbf{W}| = (-1)^n|\mathbf{W}^T| = -|\mathbf{W}|$$

leads to

$$|\mathbf{W}| = 0.$$

Thus $\mathbf{W}$ is singular. The right null vector $\mathbf{w}$ satisfies the equation

$$\mathbf{W}\mathbf{w} = \mathbf{0}. \tag{1.2-11}$$

For $n = 3$, $\mathbf{W}$ can be written as

$$\mathbf{W} = \begin{bmatrix} 0 & w_3 & -w_2 \\ -w_3 & 0 & w_1 \\ w_2 & -w_1 & 0 \end{bmatrix} \tag{1.2-12}$$

and the right and left null vectors have the solution

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}. \tag{1.2-13}$$

The null vector $\mathbf{w}$ is the *axial vector* of the skew-symmetric matrix $\mathbf{W}$ (Chadwick, 1976a). $\mathbf{W}$ and $\mathbf{w}$ are related by

$$W_{ij} = \delta_{ijk}w_k, \quad w_k = \tfrac{1}{2}\delta_{kij}W_{ij}. \tag{1.2-14}$$

Unless $\mathbf{W} = \mathbf{0}$ the axial vector $\mathbf{w}$ given by (1.2-13) is unique except a constant multiplier.

Theorem 1.2-1 If two 3×3 skew-symmetric matrices $\mathbf{W}_1$ and $\mathbf{W}_2$ share the same axial vector, they differ by a constant multiplier k, i.e.,

$$\mathbf{W}_1 = k\mathbf{W}_2.$$

1.3 Diagonalization of Simple and Semisimple Matrices

The right eigenvectors $\mathbf{x}$'s associated with distinct eigenvalues λ's are independent of each other. The same statement applies to the left eigenvectors $\mathbf{y}$'s. A matrix is *simple* if all its eigenvalues are distinct, i.e., all eigenvalues are simple roots of (1.2-3). The n right eigenvectors $\mathbf{x}_i$ are independent of each other and span an n-dimensional space, so do the n left eigenvectors $\mathbf{y}_i$. When normalized, (1.2-8) and (1.2-9) can be combined as

$$\mathbf{y}_i \cdot \mathbf{x}_j = \delta_{ij} \tag{1.3-1}$$

where δ_{ij} is the *Kronecker* delta. Introducing the $n \times n$ matrices

$$\begin{aligned} \mathbf{X} &= [\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_n], \\ \mathbf{Y} &= [\mathbf{y}_1,\ \mathbf{y}_2,\ \ldots,\ \mathbf{y}_n], \end{aligned} \tag{1.3-2}$$

which are nonsingular, (1.3-1) is written as

$$\mathbf{Y}^T\mathbf{X} = \mathbf{I} = \mathbf{X}\mathbf{Y}^T. \tag{1.3-3}$$

Thus $\mathbf{Y}^T$ is the inverse, or the *reciprocal*, of $\mathbf{X}$ which provides a justification for calling $\mathbf{y}_i$ the *reciprocal* vector. The second equality in (1.3-3) follows from (1.1-23). Employing the notation of (1.1-10c), the eigenrelation (1.2-1) written for $\lambda = \lambda_1, \lambda_2, \ldots, \lambda_n$ can be combined into one equation as

$$\mathbf{A}\mathbf{X} = \mathbf{X}\langle\lambda_*\rangle. \tag{1.3-4}$$

Application of $(1.3\text{-}3)_{1,2}$ leads to the *diagonalization*

$$\mathbf{Y}^T\mathbf{A}\mathbf{X} = \langle\lambda_*\rangle \quad \text{and} \quad \mathbf{A} = \mathbf{X}\langle\lambda_*\rangle\mathbf{Y}^T. \tag{1.3-5}$$

When $\mathbf{A}$ is symmetric, $\mathbf{y}_i = \mathbf{x}_i$ and (1.3-1) becomes

$$\mathbf{x}_i \cdot \mathbf{x}_j = \delta_{ij}. \tag{1.3-6}$$

The eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are all unit vectors and form an *orthonormal* set. Equation (1.3-3) is replaced by

$$\mathbf{X}^T\mathbf{X} = \mathbf{I} = \mathbf{X}\mathbf{X}^T, \tag{1.3-7}$$

and (1.3-5) by

$$\mathbf{X}^T\mathbf{A}\mathbf{X} = \langle \lambda_* \rangle \quad \text{and} \quad \mathbf{A} = \mathbf{X}\langle \lambda_* \rangle \mathbf{X}^T. \tag{1.3-8}$$

The results obtained so far assume that the matrix **A** is simple, i.e., the eigenvalues λ's are distinct. When the eigenvalues λ's have a repeated root of multiplicity q, the eigenvectors associated with the repeated root may not have q independent eigenvectors. If they have q independent eigenvectors, and if the same situation happens to all other repeated roots, we have n independent eigenvectors for the matrix **A** of order n. A matrix of order n that has repeated eigenvalues and possesses n independent eigenvectors is called *semisimple*. If the matrix has less than n independent eigenvectors, it is called *nonsemisimple*.

For semisimple matrices the eigenvectors associated with a repeated eigenvalue are not unique. The nonuniqueness is more than a constant multiplier on the eigenvector. The right eigenvector $\mathbf{x}_i$ and the left eigenvector $\mathbf{y}_i$ associated with a repeated eigenvalue are not necessarily orthogonal to each other. However, due to the nonuniqueness, it is possible to choose a set of eigenvectors such that the orthogonality relations hold. We could also normalize these eigenvectors according to (1.2-9) (see also Section 12.11). Therefore (1.3-1)–(1.3-5) apply to semisimple matrices also. We will discuss the diagonalization of nonsemisimple matrices in the next section.

It can be shown that if **A** is real and symmetric, it is either simple or semisimple, i.e., there exit n independent eigenvectors regardless of whether **A** has a repeated eigenvalue or not. More generally, if **A** is complex and is *Hermitian* (Section 1.7), it is either simple or semisimple. However if **A** is complex and symmetric, it can be nonsemisimple. Scott (1993a,b) has shown that a complex symmetric matrix **A** has an isotropic eigenvector (Section 1.1) if and only if **A** is nonsemisimple.

1.4 Nonsemisimple Matrices

When a matrix **A** is nonsemisimple, it has a repeated eigenvalue λ of multiplicity q but does not have q independent eigenvectors associated with the repeated eigenvalue. In this case the missing eigenvectors are provided by the *generalized eigenvectors*.

Even though **A** does not have q independent eigenvectors, it must have at least one eigenvector, say $\mathbf{x}_1$, associated with a repeated eigenvalue λ. The generalized eigenvectors $\mathbf{x}_2$, $\mathbf{x}_3$,..., are determined by

$$\begin{aligned}
\mathbf{A}\mathbf{x}_1 &= \lambda\mathbf{x}_1,\\
\mathbf{A}\mathbf{x}_2 &= \lambda\mathbf{x}_2 + \mathbf{x}_1,\\
\mathbf{A}\mathbf{x}_3 &= \lambda\mathbf{x}_3 + \mathbf{x}_2,\\
&\vdots\\
\mathbf{A}\mathbf{x}_k &= \lambda\mathbf{x}_k + \mathbf{x}_{k-1}.
\end{aligned} \tag{1.4-1}$$

These equations can be rewritten as

$$\begin{aligned}
(\mathbf{A}-\lambda\mathbf{I})\mathbf{x}_1 &= \mathbf{0},\\
(\mathbf{A}-\lambda\mathbf{I})^2\mathbf{x}_2 &= \mathbf{0},\\
(\mathbf{A}-\lambda\mathbf{I})^3\mathbf{x}_3 &= \mathbf{0},\\
&\vdots\\
(\mathbf{A}-\lambda\mathbf{I})^k\mathbf{x}_k &= \mathbf{0}.
\end{aligned} \tag{1.4-2}$$

Therefore $\mathbf{x}_2$ is a generalized eigenvector of *rank* 2 and $\mathbf{x}_k$ is a generalized eigenvector of *rank k*. Together with $\mathbf{x}_1$, they form a *chain of length* $k \leq q$ and are independent of each other. They are not unique. If $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ are solutions of (1.4-2), so are

$$\begin{aligned}
\hat{\mathbf{x}}_1 &= \mathbf{x}_1,\\
\hat{\mathbf{x}}_2 &= \mathbf{x}_2 + a_1\mathbf{x}_1,\\
\hat{\mathbf{x}}_3 &= \mathbf{x}_3 + a_1\mathbf{x}_2 + a_2\mathbf{x}_1\\
&\vdots\\
\hat{\mathbf{x}}_k &= \mathbf{x}_k + a_1\mathbf{x}_{k-1} + a_2\mathbf{x}_{k-2} + \ldots + a_{k-1}\mathbf{x}_1,
\end{aligned}$$

where $a_1, a_2, \ldots, a_{k-1}$ are arbitrary constants. The subspace spanned by $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \ldots, \hat{\mathbf{x}}_k$ are identical to the space spanned by $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$. We see that only the highest rank $\hat{\mathbf{x}}_k$ contains all the arbitrary constants. Once $\hat{\mathbf{x}}_k$ is chosen, the whole chain is determined. This suggests that we determine (1.4-2) in the reverse order from $\mathbf{x}_k$.

To find $\mathbf{x}_k$, we look for an integer k for which (Pease, 1965)

$$(\mathbf{A}-\lambda\mathbf{I})^k\mathbf{x}_k = \mathbf{0}, \quad (\mathbf{A}-\lambda\mathbf{I})^{k-1}\mathbf{x}_k \neq \mathbf{0}. \tag{1.4-3}$$

When an $\mathbf{x}_k$ exists, we determine the remaining vectors of the chain from (1.4-1) as

$$\begin{aligned}
\mathbf{x}_{k-1} &= (\mathbf{A}-\lambda\mathbf{I})\mathbf{x}_k,\\
\mathbf{x}_{k-2} &= (\mathbf{A}-\lambda\mathbf{I})\mathbf{x}_{k-1},\\
&\vdots\\
\mathbf{x}_1 &= (\mathbf{A}-\lambda\mathbf{I})\mathbf{x}_2.
\end{aligned} \tag{1.4-4}$$

A repeated eigenvalue of multiplicity q has more than one chain when $k < q$. If (1.4-3) are satisfied by $k = k_1$ and k_2, we have two chains of length k_1 and k_2, respectively, associated with the repeated eigenvalue λ. A chain of length one is the one for which $\mathbf{x}_1$ of $(1.4\text{-}1)_1$ exists but $\mathbf{x}_2$ of $(1.4\text{-}1)_2$ does not exist. If all repeated eigenvalues have chains of length one, the matrix is semisimple.

The k equations (1.4-1) in the chain can be written in one equation as

$$\mathbf{A}[\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_k]=[\mathbf{x}_1,\ \mathbf{x}_2,\ \ldots,\ \mathbf{x}_k]\mathbf{J}_k(\lambda), \tag{1.4-5}$$

where $\mathbf{J}_k(\lambda)$ is a $k\times k$ matrix with the repeated eigenvalue λ on the diagonal, 1 on the first super diagonal, and zeros elsewhere. Thus

$$\mathbf{J}_1(\lambda)=[\lambda],$$

$$\mathbf{J}_2(\lambda)=\begin{bmatrix}\lambda & 1\\ 0 & \lambda\end{bmatrix},$$

$$\mathbf{J}_3(\lambda)=\begin{bmatrix}\lambda & 1 & 0\\ 0 & \lambda & 1\\ 0 & 0 & \lambda\end{bmatrix}, \tag{1.4-6}$$

$$\mathbf{J}_4(\lambda)=\begin{bmatrix}\lambda & 1 & 0 & 0\\ 0 & \lambda & 1 & 0\\ 0 & 0 & \lambda & 1\\ 0 & 0 & 0 & \lambda\end{bmatrix},$$

$$\vdots$$

The matrix $\mathbf{J}_k(\lambda)$ is called the *upper Jordan block of order k.* When all chains associated with a repeated eigenvalue λ of multiplicity q are found, we should have q independent eigenvectors and generalized eigenvectors. With other repeated eigenvalues treated in the same manner, (1.4-5) for the full set can be written as

$$\mathbf{AX}=\mathbf{XJ} \tag{1.4-7}$$

where the $\mathbf{X}$ is defined in $(1.3\text{-}2)_1$ in which $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are now the eigenvectors and generalized eigenvectors and

$$\mathbf{J}=\begin{bmatrix}\mathbf{J}_{k_1}(\lambda_\alpha) & \mathbf{0} & \mathbf{0} & \ldots\\ \mathbf{0} & \mathbf{J}_{k_2}(\lambda_\beta) & \mathbf{0} & \ldots\\ \mathbf{0} & \mathbf{0} & \mathbf{J}_{k_3}(\lambda_\gamma) & \ldots\\ \vdots & \vdots & \vdots & \ddots\end{bmatrix}. \tag{1.4-8}$$

This is the *Jordan canonical form*. There is no loss in generality in arranging k_1, k_2, ... in the order of ascending rank, i.e.,

$$1 \le k_1 \le k_2 \le \ldots,$$
$$k_1 + k_2 + \ldots = n.$$

It should be pointed out that λ_α, λ_β, λ_γ,..., need not be distinct. One could have $\lambda_\alpha = \lambda_\gamma \neq \lambda_\beta$. It can be shown that $\mathbf{X}$ is nonsingular. Therefore its inverse, denoted by $\mathbf{Y}^T$, exists and we obtain from (1.4-7)

$$\mathbf{Y}^T\mathbf{A}\mathbf{X} = \mathbf{J} \quad \text{and} \quad \mathbf{A} = \mathbf{X}\mathbf{J}\mathbf{Y}^T. \tag{1.4-9}$$

These are the *diagonalization* of a nonsemisimple matrix $\mathbf{A}$.

The use of notation $\mathbf{Y}^T$ as the inverse of $\mathbf{X}$ is intentional. It can be shown that the columns of $\mathbf{Y}$ are the left eigenvectors and generalized left eigenvectors. The generalized left eigenvectors can be defined similar to the generalized right eigenvectors in (1.4-1) except that the order of the rank is reversed, i.e.,

$$\begin{aligned}
\mathbf{A}^T\mathbf{y}_k &= \lambda\mathbf{y}_k, \\
\mathbf{A}^T\mathbf{y}_{k-1} &= \lambda\mathbf{y}_{k-1} + \mathbf{y}_k, \\
\mathbf{A}^T\mathbf{y}_{k-2} &= \lambda\mathbf{y}_{k-2} + \mathbf{y}_{k-1}, \\
&\vdots \\
\mathbf{A}^T\mathbf{y}_1 &= \lambda\mathbf{y}_1 + \mathbf{y}_2.
\end{aligned} \tag{1.4-10}$$

Thus $\mathbf{y}_{k-1}$ is of rank 2 while $\mathbf{y}_1$ is of rank k.

It should be noted that the Jordan canonical form $\mathbf{J}$ is not a symmetric matrix. When $\mathbf{A}$ is complex, symmetric, and nonsemisimple, a symmetric canonical form can be obtained by a different approach (Scott, 1993a,b).

As an illustration consider the real matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & -2 & 2 & 1 & 0 \\ 0 & 2 & -1 & 1 & 1 \\ 0 & -2 & 1 & 0 & 0 \end{bmatrix}. \tag{1.4-11}$$

It is readily shown that

$$|\mathbf{A} - \lambda\mathbf{I}| = (1-\lambda)^5$$

and hence $\lambda = 1$ is a repeated eigenvalue of multiplicity 5. With λ

= 1 we have

$$(\mathbf{A}-\lambda\mathbf{I})=\begin{bmatrix}0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & -2 & 1 & 1 & 0\\ 0 & 2 & -1 & 0 & 1\\ 0 & -2 & 1 & 0 & -1\end{bmatrix},$$

$$(\mathbf{A}-\lambda\mathbf{I})^2=\begin{bmatrix}0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 1\\ 0 & 0 & 0 & -1 & -1\\ 0 & 0 & 0 & 1 & 1\end{bmatrix},$$

and

$$(\mathbf{A}-\lambda\mathbf{I})^3=\mathbf{0}.$$

We see that $(1.4\text{-}3)_1$ is satisfied for $k = 3$ with

$$\mathbf{x}_3^T=(a,\ b,\ c,\ d,\ e)$$

where a, b, c, d, and e are arbitrary constants. Equation $(1.4\text{-}3)_2$ is satisfied if $d + e \neq 0$. Let

$$\mathbf{x}_3^T=(0,\ 0,\ 0,\ 1,\ 0)$$

be the highest ranking member of the chain. The remaining members of the chain are obtained from (1.4-4) as

$$\mathbf{x}_2^T=(0,\ 0,\ 1,\ 0,\ 0),$$
$$\mathbf{x}_1^T=(0,\ 0,\ 1,\ -1,\ 1).$$

This completes the chain of length 3. Equations $(1.4\text{-}3)_{1,2}$ are also satisfied for $k = 2$ with

$$\mathbf{x}_2^{*T}=(0,\ 1,\ 0,\ 0,\ 0),$$

and (1.4-4) gives

$$\mathbf{x}_1^{*T}=(1,\ 0,\ -2,\ 2,\ -2).$$

This completes the second chain of length 2. In conformity with the ascending order of k in (1.4-8), we rename $\mathbf{x}_1^*$ and $\mathbf{x}_2^*$ of the second chain as $\mathbf{x}_1$ and $\mathbf{x}_2$, respectively, and $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ of the first chain as $\mathbf{x}_3$, $\mathbf{x}_4$, $\mathbf{x}_5$, respectively. Therefore,

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ -2 & 0 & 1 & 1 & 0 \\ 2 & 0 & -1 & 0 & 1 \\ -2 & 0 & 1 & 0 & 0 \end{bmatrix}$$

and

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_2(1) & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_3(1) \end{bmatrix}.$$

The inverse $\mathbf{Y}^T$ of $\mathbf{X}$ can be shown to be

$$\mathbf{Y}^T = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}.$$

It is readily verified that

$$\mathbf{y}_1^{*T} = (1,\ 0,\ 0,\ 0,\ 0),$$
$$\mathbf{y}_2^{*T} = (0,\ 1,\ 0,\ 0,\ 0),$$

and

$$\mathbf{y}_1^{T} = (2,\ 0,\ 0,\ 0,\ 1),$$
$$\mathbf{y}_2^{T} = (0,\ 0,\ 1,\ 0,\ -1),$$
$$\mathbf{y}_3^{T} = (0,\ 0,\ 0,\ 1,\ 1),$$

are, respectively, the chains of lengths 2 and 3 which satisfy (1.4-10) for $\lambda = 1$.

1.5 Commutative Matrices

Two square matrices are called *commutative* if

$$\mathbf{AB} = \mathbf{BA}.$$

It is obvious that all diagonal matrices are commutative. We also saw in (1.1-23) that $\mathbf{A}$ and $\mathbf{A}^{-1}$ are commutative.

Let $\mathbf{A}$ and $\mathbf{B}$ be simple or semisimple. Furthermore, let $\mathbf{A}$ and $\mathbf{B}$ have the same set of eigenvectors. By $(1.3\text{-}5)_2$ $\mathbf{A}$ and $\mathbf{B}$ can be diagonalized as

$$\mathbf{A} = \mathbf{X}\langle\lambda_*\rangle_A \mathbf{Y}^T, \quad \mathbf{B} = \mathbf{X}\langle\lambda_*\rangle_B \mathbf{Y}^T,$$

where $\langle\lambda_*\rangle_A$ and $\langle\lambda_*\rangle_B$ are diagonal matrices whose diagonal elements are the eigenvalues of **A** and **B**, respectively. We then have

$$\begin{aligned}\mathbf{AB} &= \mathbf{X}\langle\lambda_*\rangle_A \mathbf{Y}^T\mathbf{X}\langle\lambda_*\rangle_B \mathbf{Y}^T = \mathbf{X}\langle\lambda_*\rangle_A\langle\lambda_*\rangle_B \mathbf{Y}^T, \\ \mathbf{BA} &= \mathbf{X}\langle\lambda_*\rangle_B \mathbf{Y}^T\mathbf{X}\langle\lambda_*\rangle_A \mathbf{Y}^T = \mathbf{X}\langle\lambda_*\rangle_B\langle\lambda_*\rangle_A \mathbf{Y}^T = \mathbf{X}\langle\lambda_*\rangle_A\langle\lambda_*\rangle_B \mathbf{Y}^T.\end{aligned} \tag{1.5-1}$$

Therefore **AB** = **BA**. It can be shown that the converse is also true, i.e., if **AB** = **BA**, **A** and **B** share the same set of eigenvectors.

> **Theorem 1.5–1** Let **A** and **B** be simple or semisimple. Then **AB=BA** if and only if they share the same set of eigenvectors.

If **A** and **B** are real and symmetric, they are necessarily simple or semisimple. Therefore a special case of Theorem 1.5-1 is:

> **Theorem 1.5–2** Let **A** and **B** be real and symmetric. Then **AB=BA** if and only if they share the same set of eigenvectors.

We now turn to the case when **A** and **B** are both nonsemisimple. First, from the definition of $\mathbf{J}_k(\lambda)$ in (1.4-6) we have

$$\mathbf{J}_1(a)\mathbf{J}_1(b) = [ab],$$

$$\mathbf{J}_2(a)\mathbf{J}_2(b) = \begin{bmatrix} ab & a+b \\ 0 & ab \end{bmatrix},$$

$$\mathbf{J}_3(a)\mathbf{J}_3(b) = \begin{bmatrix} ab & a+b & 1 \\ 0 & ab & a+b \\ 0 & 0 & ab \end{bmatrix},$$

$$\mathbf{J}_4(a)\mathbf{J}_4(b) = \begin{bmatrix} ab & a+b & 1 & 0 \\ 0 & ab & a+b & 1 \\ 0 & 0 & ab & a+b \\ 0 & 0 & 0 & ab \end{bmatrix},$$

$$\vdots$$

It is seen that $\mathbf{J}_k(a)\mathbf{J}_k(b)$ has ab on its diagonal, $a+b$ on its first

super diagonal, 1 on its second super diagonal, and zeros elsewhere. The important point to observe is that $\mathbf{J}_k(a)$ and $\mathbf{J}_k(b)$ commute, as is evident from the expressions on the right sides of the equalities which are symmetric with a and b. This means that if $\mathbf{J}_A$ and $\mathbf{J}_B$ are the Jordan canonical forms for $\mathbf{A}$ and $\mathbf{B}$, respectively, and if $\mathbf{J}_A$ and $\mathbf{J}_B$ have the same structure, i.e., the k_1, k_2,..., in (1.4-8) are the same for $\mathbf{J}_A$ and $\mathbf{J}_B$,

$$\mathbf{J}_A\mathbf{J}_B = \mathbf{J}_B\mathbf{J}_A. \tag{1.5-2}$$

Following the derivation of (1.5-1) we have:

> **Theorem 1.5–3** Let $\mathbf{A}$ and $\mathbf{B}$ be nonsemisimple. If the Jordan canonical forms for $\mathbf{A}$ and $\mathbf{B}$ have the same structure, (i.e., the k_1, k_2,..., in (1.4-8) are the same for $\mathbf{A}$ and $\mathbf{B}$), and if $\mathbf{A}$ and $\mathbf{B}$ share the same set of eigenvectors and generalized eigenvectors, $\mathbf{A}$ and $\mathbf{B}$ commute.

1.6 Positive Definite Real Matrices

In this section we will consider real matrices only, even though some of the results obtained can be extended to complex matrices.

The *quadratic form* of a real square matrix $\mathbf{C}$ is

$$\mathbf{u}^T\mathbf{C}\mathbf{u} \tag{1.6-1}$$

where $\mathbf{u}$ is any nonzero vector or column matrix. The quadratic form is an 1×1 matrix. Since the transpose of an 1×1 matrix is identical to itself,

$$\mathbf{u}^T\mathbf{C}\mathbf{u} = (\mathbf{u}^T\mathbf{C}\mathbf{u})^T = \mathbf{u}^T\mathbf{C}^T\mathbf{u}.$$

From this it is clear that the quadratic form of a skew-symmetric matrix $\mathbf{W}$ vanishes, i.e.,

$$\mathbf{u}^T\mathbf{W}\mathbf{u} = 0, \quad \text{if } \mathbf{W} = -\mathbf{W}^T. \tag{1.6-2}$$

Therefore if the $\mathbf{C}$ in (1.6-1) is not symmetric, only the symmetric part of $\mathbf{C}$ contributes to the quadratic form.

A real symmetric matrix $\mathbf{C}$ is *positive definite* if

$$\mathbf{u}^T\mathbf{C}\mathbf{u} > 0 \tag{1.6-3}$$

for any nonzero real vector $\mathbf{u}$. It is *positive semidefinite* if

$$\mathbf{u}^T\mathbf{C}\mathbf{u} \geq 0. \tag{1.6-4}$$

Since $\mathbf{C}$ is real and symmetric, its eigenvalues are real, and the right and left eigenvectors that are identical are also real (Section 1.2).

Let λ_i and $\mathbf{x}_i$ $(i = 1,2,\ldots,n)$ be the eigenvalues and eigenvectors of $\mathbf{C}$, respectively. Employing the diagonalization $(1.3\text{-}8)_2$ we have

$$\mathbf{u}^T\mathbf{C}\mathbf{u} = \mathbf{u}^T\mathbf{X}\langle\lambda_*\rangle\mathbf{X}^T\mathbf{u} = (\mathbf{X}^T\mathbf{u})^T\langle\lambda_*\rangle(\mathbf{X}^T\mathbf{u}),$$

or

$$\mathbf{u}^T\mathbf{C}\mathbf{u} = \mathbf{z}^T\langle\lambda_*\rangle\mathbf{z} = \lambda_1 z_1^2 + \lambda_2 z_2^2 + \ldots + \lambda_n z_n^2, \qquad \mathbf{z} = \mathbf{X}^T\mathbf{u}. \tag{1.6-5}$$

That $\mathbf{X}^T$ is nonsingular means that if $\mathbf{u}$ is nonzero, so is $\mathbf{z}$. We therefore have the following theorem:

> **Theorem 1.6–1** A real symmetric matrix is positive definite if and only if all its eigenvalues are positive and nonzero. If some of the eigenvalues are zero, the matrix is positive semidefinite.

A consequence of the theorem is that, from $(1.2\text{-}5)_2$, the determinant of a real symmetric matrix must be positive and nonzero if the matrix is positive definite.

When we remove an equal number of rows and columns from a symmetric matrix $\mathbf{C}$ we obtain a square submatrix. The determinant of such a submatrix is called a *minor*. If the rows and columns removed have the same indices, the diagonal elements of the submatrix are the diagonal elements of the original matrix. The submatrix so obtained is called a *principal submatrix* and the determinant of the principal submatrix is a *principal minor*. The diagonal elements C_{11}, $C_{22},\ldots,C_{nn}$ and the determinant $|\mathbf{C}|$ are principal minors. If the $r\times r$ principal submatrix is from the first r rows and the first r columns of $\mathbf{C}$, we have a *leading* principal submatrix. The determinant of a leading principal submatrix is a *leading principal minor*.

The condition (1.6-3) for a positive definite matrix $\mathbf{C}$ implies that all principal submatrices of $\mathbf{C}$ must be positive definite. This can be proved easily as follows. Consider the principal submatrix

$$\begin{bmatrix} C_{11} & C_{13} \\ C_{13} & C_{33} \end{bmatrix}. \tag{1.6-6}$$

By choosing

$$\mathbf{u}^T = [u_1,\ 0,\ u_3,\ \ldots,\ 0],$$

(1.6-3) can be written as

$$[u_1,\ u_3]\begin{bmatrix} C_{11} & C_{13} \\ C_{13} & C_{33} \end{bmatrix}\begin{bmatrix} u_1 \\ u_3 \end{bmatrix} > 0.$$

which shows that the 2×2 matrix is positive definite. A necessary condition for the 2×2 matrix to be positive definite is that its determinant is positive. Therefore (1.6-3) also implies that all principal minors of **C** must be positive..

There are 7 principal minors for a 3×3 symmetric matrix and 15 for a 4×4 symmetric matrix. The following theorem (Hohn, 1965) tells us that we do not have to inspect all principal minors to see if a real symmetric matrix is positive definite.

> **Theorem 1.6–2** A real symmetric matrix is positive definite if and only if the *leading* principal minors of the matrix are positive.

This means that a real symmetric matrix **C** of order n is positive definite if and only if the following n conditions are satisfied.

$$C_{11} > 0,$$

$$\begin{vmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{vmatrix} > 0,$$

$$\begin{vmatrix} C_{11} & C_{12} & C_{13} \\ C_{12} & C_{22} & C_{23} \\ C_{13} & C_{23} & C_{33} \end{vmatrix} > 0, \tag{1.6-7}$$

$$\vdots$$

$$\begin{vmatrix} C_{11} & C_{12} & \cdots & C_{1n} \\ C_{12} & C_{22} & \cdots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{1n} & C_{2n} & \cdots & C_{nn} \end{vmatrix} > 0.$$

Theorem 1.6–2 provides an alternate to Theorem 1.6–1 for a set of necessary and sufficient conditions for a matrix to be positive definite. The determination of the eigenvalues of a matrix in general requires a numerical solution. Theorem 1.6–1 is therefore not very useful if the eigenvalues cannot be obtained explicitly in terms of the elements of the matrix.

The bounds on the quadratic form $\mathbf{u}^T\mathbf{C}\mathbf{u}$ of (1.6-5) for an arbitrarily given **u** can be set as follows. Let

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n. \tag{1.6-8}$$

Equation (1.6-5) can be written as

$$\mathbf{u}^T\mathbf{C}\mathbf{u} = \lambda_1|\mathbf{z}|^2 - (\lambda_1 - \lambda_2)z_2^2 - \ldots - (\lambda_1 - \lambda_n)z_n^2$$
$$= (\lambda_1 - \lambda_n)z_1^2 + \ldots + (\lambda_{n-1} - \lambda_n)z_{n-1}^2 + \lambda_n|\mathbf{z}|^2.$$

Since

$$|\mathbf{z}|^2 = (\mathbf{X}^T\mathbf{u})^T(\mathbf{X}^T\mathbf{u}) = \mathbf{u}^T\mathbf{u} = |\mathbf{u}|^2$$

we have

$$\lambda_n|\mathbf{u}|^2 \le \mathbf{u}^T\mathbf{C}\mathbf{u} \le \lambda_1|\mathbf{u}|^2. \tag{1.6-9}$$

Therefore, if $\mathbf{u}$ is normalized to a unit vector, (1.6-9) tells us that the quadratic form $\mathbf{u}^T\mathbf{C}\mathbf{u}$ for any unit vector $\mathbf{u}$ is bounded between the largest and the smallest eigenvalues of $\mathbf{C}$. The matrix $\mathbf{C}$ need not be positive definite. Thus some of the λ's in (1.6-8) maybe negative.

When an $n\times n$ matrix $\mathbf{A}$ is singular its determinant as well as its minor may vanish. The *rank* of $\mathbf{A}$ is the size of the largest submatrix whose determinant is nonzero. The rank of a singular matrix is at most $n-1$. If the rank is $n-k$ there are k independent null vectors.

1.7 Hermitian Matrices

A complex matrix $\mathbf{M}$ is *Hermitian* if

$$\mathbf{M} = \overline{\mathbf{M}}^T \tag{1.7-1}$$

where the overbar denotes the complex conjugate. Writing $\mathbf{M}$ as

$$\mathbf{M} = \mathbf{D} + i\mathbf{W} \tag{1.7-2}$$

in which $\mathbf{D}$ and $\mathbf{W}$ are real and inserting into (1.7-1) yields

$$\mathbf{D} = \mathbf{D}^T, \quad \mathbf{W} = -\mathbf{W}^T. \tag{1.7-3}$$

Therefore the real part of a Hermitian matrix is symmetric while the imaginary part is skew-symmetric. A real symmetric matrix is a special case of Hermitian matrices. It is easily shown that if $\mathbf{M}$ is Hermitian, so are $-\mathbf{M}$, $\mathbf{M}^T$, $\overline{\mathbf{M}}$, and $\mathbf{M}^{-1}$, provided the inverse exists.

A complex matrix is *skew-Hermitian* if

$$\mathbf{M} = -\overline{\mathbf{M}}^T. \tag{1.7-4}$$

A real skew-symmetric matrix is a special case of skew-Hermitian matrices. Equation (1.7-4) can be rewritten as

$$(i\mathbf{M}) = (\overline{i\mathbf{M}})^T$$

which says that $i\mathbf{M}$ is Hermitian. Therefore a skew-Hermitian matrix is converted to a Hermitian matrix by a multiplication of i.

We define the quadratic form of a Hermitian as

$$\bar{\mathbf{u}}^T\mathbf{M}\mathbf{u} = (\bar{\mathbf{u}}^T\mathbf{M}\mathbf{u})^T = \mathbf{u}^T\mathbf{M}^T\bar{\mathbf{u}} = \mathbf{u}^T\bar{\mathbf{M}}\bar{\mathbf{u}}. \tag{1.7-5}$$

The last expression is the complex conjugate of the first expression. Therefore the quadratic form $\bar{\mathbf{u}}^T\mathbf{M}\mathbf{u}$ is real for any Hermitian matrix $\mathbf{M}$.

The eigenrelation for a Hermitian matrix $\mathbf{M}$ is

$$\mathbf{M}\mathbf{x} = \lambda\mathbf{x}. \tag{1.7-6}$$

Premultiplied by $\bar{\mathbf{x}}^T$ we have

$$\bar{\mathbf{x}}^T\mathbf{M}\mathbf{x} = \lambda\bar{\mathbf{x}}^T\mathbf{x} \tag{1.7-7}$$

which indicates that the eigenvalue λ must be real because $\bar{\mathbf{x}}^T\mathbf{M}\mathbf{x}$ and $\bar{\mathbf{x}}^T\mathbf{x}$ are. Taking the complex conjugate of (1.7-6) and using (1.7-1) we have

$$\mathbf{M}^T\bar{\mathbf{x}} = \lambda\bar{\mathbf{x}} \quad \text{or} \quad \bar{\mathbf{x}}^T\mathbf{M} = \lambda\bar{\mathbf{x}}^T.$$

Therefore $\bar{\mathbf{x}}$ is the left eigenvector, which is the complex conjugate of the right eigenvector. From

$$\mathbf{M}\mathbf{x}_2 = \lambda_2\mathbf{x}_2, \qquad \bar{\mathbf{x}}_1^T\mathbf{M} = \lambda_1\bar{\mathbf{x}}_1^T, \tag{1.7-8}$$

we obtain, after premultiplying $(1.7\text{-}8)_1$ by $\bar{\mathbf{x}}_1^T$ and postmultiplying $(1.7\text{-}8)_2$ by $\mathbf{x}_2$,

$$0 = (\lambda_1 - \lambda_2)\bar{\mathbf{x}}_1^T\mathbf{x}_2$$

confirming the orthogonality relation (1.2-8).

A Hermitian matrix $\mathbf{M}$ is *positive definite* if

$$\bar{\mathbf{u}}^T\mathbf{M}\mathbf{u} > 0,$$

and positive semidefinite if

$$\bar{\mathbf{u}}^T\mathbf{M}\mathbf{u} \geq 0$$

for any nonzero complex vector $\mathbf{u}$. Writing $\mathbf{u}$ as

$$\mathbf{u} = \mathbf{a} + i\mathbf{b}$$

where $\mathbf{a}$ and $\mathbf{b}$ are real and imaginary parts of $\mathbf{u}$, it is readily shown that

$$\bar{\mathbf{u}}^T\mathbf{M}\mathbf{u} = \mathbf{a}^T\mathbf{D}\mathbf{a} + \mathbf{b}^T\mathbf{D}\mathbf{b} - 2\mathbf{a}^T\mathbf{W}\mathbf{b}. \tag{1.7-9}$$

Equation (1.7-9) tells us that a necessary condition for **M** to be positive definite is that the real part **D** be positive definite. This condition however is not sufficient.

We have shown in (1.7-7) that the eigenvalues of a Hermitian matrix are real. A necessary and sufficient condition for a Hermitian matrix to be positive definite is that its eigenvalues are positive and nonzero. If λ is an eigenvalue of a positive definite Hermitian matrix **M**, it is also an eigenvalue of $\overline{\mathbf{M}}$ and $\mathbf{M}^T$, and λ^{-1} is an eigenvalue of $\mathbf{M}^{-1}$ and $\overline{\mathbf{M}}^{-1}$. Hence $\mathbf{M}^T$, $\overline{\mathbf{M}}$, $\mathbf{M}^{-1}$, and $\overline{\mathbf{M}}^{-1}$ are positive definite if **M** is.

In the remainder of this section we focus our attention on 3×3 Hermitian matrices. Let the skew-symmetric matrix **W** in (1.7-2) and its axial vector **w** be given by

$$\mathbf{W} = \begin{bmatrix} 0 & w_3 & -w_2 \\ -w_3 & 0 & w_1 \\ w_2 & -w_1 & 0 \end{bmatrix}, \qquad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}. \tag{1.7-10}$$

It was shown in (1.2-14) that **W** and **w** are related by

$$W_{ij} = \delta_{ijk} w_k, \quad w_k = \tfrac{1}{2}\delta_{kij} W_{ij}. \tag{1.7-11}$$

The last term in (1.7-9) can be written as

$$\mathbf{a}^T\mathbf{W}\mathbf{b} = a_i \delta_{ijk} \omega_k b_j = \begin{vmatrix} a_1 & b_1 & w_1 \\ a_2 & b_2 & w_2 \\ a_3 & b_3 & w_3 \end{vmatrix}.$$

This term vanishes if the three vectors **a**, **b**, and **w** are coplanar.

We next show that

$$\operatorname{tr}\mathbf{M} = \operatorname{tr}\mathbf{D}, \quad \operatorname{tr}(\mathbf{M}^2) = \operatorname{tr}(\mathbf{D}^2) + 2\mathbf{w}^T\mathbf{w}, \quad |\mathbf{M}| = |\mathbf{D}| - \mathbf{w}^T\mathbf{D}\mathbf{w}. \tag{1.7-12}$$

Equations $(1.7\text{-}12)_{1,2}$ follow from (1.7-2) and the identities

$$\operatorname{tr}\mathbf{W} = 0, \quad \operatorname{tr}(\mathbf{W}\mathbf{D}) = \mathbf{0} = \operatorname{tr}(\mathbf{D}\mathbf{W}), \quad \operatorname{tr}(\mathbf{W}^2) = -2\mathbf{w}^T\mathbf{w},$$

which can be established easily. As to $(1.7\text{-}12)_3$ we use (1.1-17a) to write

$$|\mathbf{M}| = \delta_{ijk}(D_{i1} + iW_{i1})(D_{j2} + iW_{j2})(D_{k3} + iW_{k3}).$$

Since $|\mathbf{M}|$ is real, keeping the real terms gives

$$|\mathbf{M}| = |\mathbf{D}| - \delta_{ijk} D_{i1} W_{j2} W_{k3} - \delta_{ijk} W_{i1} D_{j2} W_{k3} - \delta_{ijk} W_{i1} W_{j2} D_{k3}.$$

The second term can be written as

$$\delta_{ijk}D_{i1}W_{j2}W_{k3}=\begin{vmatrix} D_{11} & w_3 & -w_2 \\ D_{21} & 0 & w_1 \\ D_{31} & -w_1 & 0 \end{vmatrix} = D_{11}w_1^2 + D_{21}w_2w_1 + D_{31}w_3w_1.$$

The remaining two terms can be determined similarly. Hence

$$|\mathbf{M}| = |\mathbf{D}| - D_{ik}w_iw_k \tag{1.7-13}$$

which is $(1.7\text{-}12)_3$.

Identities (1.7-12) provide relations between the eigenvalues of **M** and **D**. Let

$$d_1 \geq d_2 \geq d_3$$

be the eigenvalues of **D** and

$$\lambda_1 \geq \lambda_2 \geq \lambda_3$$

be the eigenvalues of **M**. It can be shown that

$$\lambda_1 + \lambda_2 + \lambda_3 = d_1 + d_2 + d_3, \tag{1.7-14a}$$

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 > d_1^2 + d_2^2 + d_3^2, \tag{1.7-14b}$$

$$\lambda_1 \geq d_1 \geq \lambda_2 \geq d_3 \geq \lambda_3. \tag{1.7-14c}$$

Equations (1.7-14a,b) follow directly from $(1.7\text{-}12)_{1,2}$ and (1.2-6b). To prove (1.7-14c) consider

$$f(\lambda) = |\mathbf{M} - \lambda\mathbf{I}| = |\mathbf{D} - \lambda\mathbf{I} + i\mathbf{W}| = |\mathbf{D} - \lambda\mathbf{I}| - \mathbf{w}^T(\mathbf{D} - \lambda\mathbf{I})\mathbf{w} \tag{1.7-15}$$

where use has been made of the identity $(1.7\text{-}12)_3$. By $(1.3\text{-}8)_2$ the diagonalization of **D** is

$$\mathbf{D} = \mathbf{X}\langle d_* \rangle \mathbf{X}^T$$

where the columns of **X** are the eigenvectors of **D**. Hence

$$f(\lambda) = |\mathbf{D} - \lambda\mathbf{I}| - \hat{\mathbf{w}}^T\langle d_* - \lambda \rangle \hat{\mathbf{w}}, \quad \hat{\mathbf{w}} = \mathbf{X}^T\mathbf{w},$$

or

$$f(\lambda) = (d_1 - \lambda)(d_2 - \lambda)(d_3 - \lambda) - [(d_1 - \lambda)\hat{w}_1^2 + (d_2 - \lambda)\hat{w}_2^2 + (d_3 - \lambda)\hat{w}_3^2].$$

This leads to the inequalities in (1.7-14c) because

$$f(\infty) < 0, \quad f(d_1) \geq 0, \quad f(d_3) \leq 0, \quad f(-\infty) > 0.$$

From (1.7-14c) and $(1.7\text{-}12)_3$ we have the following theorem.

Theorem 1.7–1 A necessary and sufficient condition for a 3×3 Hermitian matrix $\mathbf{D}+i\mathbf{W}$ to be positive definite is that $\mathbf{D}$ be positive definite and

$$|\mathbf{D}| > \mathbf{w}^T\mathbf{D}\mathbf{w}.$$

1.8 Eigenplane

When an eigenvector $\mathbf{x}$ is complex, it can be written as

$$\mathbf{x} = \mathbf{e}_1 + i\mathbf{e}_2$$

where $\mathbf{e}_1$ and $\mathbf{e}_2$ are, respectively, the real and imaginary parts of $\mathbf{x}$. A complex vector is also called a *bivector*. The real vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ span a plane unless they are proportional. If they are, $\mathbf{x}$ is a real vector multiplied by a complex factor. Since an eigenvector is unique up to a constant multiplier, without loss in generality we may take $\mathbf{x}$ to be real whenever $\mathbf{e}_1$ and $\mathbf{e}_2$ are proportional. A complex vector is *genuinely complex* if its real and imaginary parts are not proportional. We assume in this section that $\mathbf{x}$ is genuinely complex so that $\mathbf{e}_1$ and $\mathbf{e}_2$ span a plane.

The vectors $\mathbf{e}_1$ and $\mathbf{e}_2$ are not unique. The non-uniqueness is more than a real constant multiplier. Let

$$\mathbf{x}^\circ = \mathbf{e}_1^\circ + i\mathbf{e}_2^\circ$$

be an eigenvector. Then $\mathbf{x}^\circ$ multiplied by a complex factor $\rho e^{-i\psi}$ where ρ and ψ are real and ρ is positive is also an eigenvector. Let $\rho = 1$ and

$$\mathbf{x} = e^{-i\psi}\mathbf{x}^\circ = \mathbf{e}_1 + i\mathbf{e}_2, \tag{1.8-1}$$

where

$$\begin{aligned} \mathbf{e}_1 &= \mathbf{e}_1^\circ \cos\psi + \mathbf{e}_2^\circ \sin\psi, \\ \mathbf{e}_2 &= -\mathbf{e}_1^\circ \sin\psi + \mathbf{e}_2^\circ \cos\psi. \end{aligned} \tag{1.8-2}$$

It is seen that $\mathbf{e}_1$ and $\mathbf{e}_2$ are not simply $\mathbf{e}_1^\circ$ and $\mathbf{e}_2^\circ$ multiplied by a real factor. Since $\mathbf{e}_1$ and $\mathbf{e}_2$ are both linear combinations of $\mathbf{e}_1^\circ$ and $\mathbf{e}_2^\circ$, $\mathbf{e}_1$ and $\mathbf{e}_2$ lie on the same plane spanned by $\mathbf{e}_1^\circ$ and $\mathbf{e}_2^\circ$. If $\rho \neq 1$, the real and imaginary parts of $\mathbf{x}$ still lie on the same plane. Therefore the plane is called the *eigenplane*. Regardless of the choice of the complex multiplier, the real and imaginary parts of a complex eigenvector lie on an eigenplane. It is clear that $\mathbf{x}$ and its complex conjugate $\bar{\mathbf{x}}$ span the same eigenplane.

As ψ varies, (1.8-2) show that $\mathbf{e}_1$ and $\mathbf{e}_2$ trace an ellipse, Fig. 1.1. The rotation of the vector $\mathbf{e}_1$ as ψ increases from zero is from $\mathbf{e}_1^o$ to $\mathbf{e}_2^o$. A pair of diameters in an ellipse are said to be *conjugates* if all chords parallel to one diameter are bisected by the other diameter. Therefore the tangent to the ellipse at the extremity of one diameter is parallel to the other diameter. It can be shown that $\mathbf{e}_1$ and $\mathbf{e}_2$ form a pair of *conjugate radii*. One could choose ψ such that $\mathbf{e}_1$ and $\mathbf{e}_2$ are orthogonal, and hence are the principal radii of the ellipse.

Let

$$\xi_o = |\mathbf{e}_1^o|, \quad \eta_o = |\mathbf{e}_2^o|, \quad \delta_o = \mathbf{e}_1^o \cdot \mathbf{e}_2^o, \tag{1.8-3}$$

where ξ_o and η_o are, respectively, the lengths of the vectors $\mathbf{e}_1^o$ and $\mathbf{e}_2^o$. We then have from (1.8-2)

$$\xi^2 = \xi_o^2 \cos^2\psi + \eta_o^2 \sin^2\psi + 2\delta_o \sin\psi\cos\psi,$$
$$\eta^2 = \xi_o^2 \sin^2\psi + \eta_o^2 \cos^2\psi - 2\delta_o \sin\psi\cos\psi.$$

Hence

$$\xi^2 + \eta^2 = \xi_o^2 + \eta_o^2 \tag{1.8-4}$$

which is *invariant* with ψ. Thus the sum of the square of the lengths of the conjugate radii is invariant with ψ.

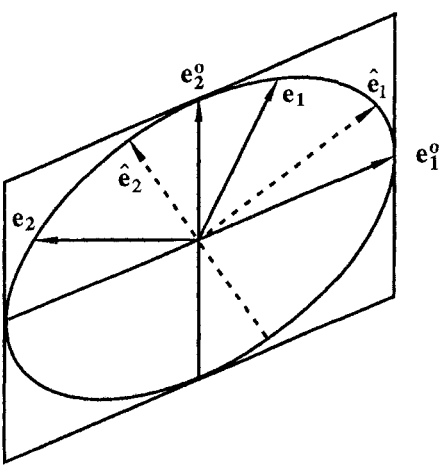

Fig. 1.1. The real and imaginary parts of a complex eigenvector $\mathbf{x}^o = \mathbf{e}_1^o + i\,\mathbf{e}_2^o$ span an eigenplane. The real and imaginary parts of $\mathbf{x} = e^{-i\psi}\mathbf{x}^o = \mathbf{e}_1 + i\,\mathbf{e}_2$, as ψ varies, describe an ellipse. They are a pair of conjugate radii of the ellipse.

Let $\hat{\mathbf{e}}_1$ and $\hat{\mathbf{e}}_2$ be the principal radii of the ellipse and

$$|\hat{\mathbf{e}}_1| \geq |\hat{\mathbf{e}}_2|. \tag{1.8-5}$$

From (1.8-2)

$$\mathbf{e}_1 \cdot \mathbf{e}_2 = -\tfrac{1}{2}(\xi_o^2 - \eta_o^2)\sin 2\psi + \delta_o \cos 2\psi.$$

If $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$ with $|\mathbf{e}_1| \geq |\mathbf{e}_2|$ at $\psi = \hat{\psi}$, $\mathbf{e}_1$ and $\mathbf{e}_2$ are the principal radii of the ellipse. The major axis is located at

$$\tan 2\hat{\psi} = \frac{2\delta_o}{\xi_o^2 - \eta_o^2}.$$

The condition (1.8-5) allows us to determine $\hat{\psi}$ uniquely if we limit

$$-\frac{\pi}{2} < \hat{\psi} \leq \frac{\pi}{2}.$$

It is not difficult to show that $\mathbf{e}_1^o$ and $\mathbf{e}_2^o$ are the principal axes of the ellipse if and only if $\mathbf{e}_1^o \cdot \mathbf{e}_2^o = 0$. We then have $\hat{\psi} = 0$, $\pi/2$, or arbitrary depending on whether $|\mathbf{e}_1^o|$ is larger, smaller, or equal to $|\mathbf{e}_2^o|$

Geometrical interpretations of the orthogonality of two bivectors and other subjects related to the algebra of bivectors can be found in Gibbs (1881), Synge (1966), Stone (1963), and Boulanger and Hayes (1991).

1.9 Square Roots of a Matrix

What are the *square roots* of a square matrix? In other words, what square matrices **B** of order n satisfy

$$\mathbf{B}^2 = \mathbf{A} \tag{1.9-1}$$

when a square matrix **A** is given? The matrix **B** is a square root of **A** and is denoted by $\sqrt{\mathbf{A}}$ or $\mathbf{A}^{1/2}$. Equation (1.9-1) is written as

$$\sqrt{\mathbf{A}}\sqrt{\mathbf{A}} = \mathbf{A} = \mathbf{A}^{1/2}\mathbf{A}^{1/2}. \tag{1.9-2}$$

When **A** is an identity matrix we have

$$\sqrt{\mathbf{A}}\sqrt{\mathbf{A}} = \mathbf{I}. \tag{1.9-3}$$

The obvious solutions to (1.9-3) are

$$\sqrt{\mathbf{A}} = \pm\mathbf{I} \tag{1.9-4}$$

but these two are not the only solutions if $n > 1$. For $n = 2$,

$$\sqrt{\mathbf{A}} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \begin{bmatrix} \sqrt{2} & -1 \\ 1 & -\sqrt{2} \end{bmatrix}, \quad \begin{bmatrix} i & 1 \\ 2 & -i \end{bmatrix}, \tag{1.9-5}$$

all satisfy (1.9-3). They are special solutions of

$$\sqrt{\mathbf{A}} = \pm\begin{bmatrix} \sqrt{1-ab} & a \\ b & -\sqrt{1-ab} \end{bmatrix} \tag{1.9-6}$$

in which a and b are arbitrary. Thus there are infinitely many solutions to (1.9-3). As shown in (1.9-5), square roots of a real symmetric matrix need not be real or symmetric. An example of a nontrivial real $\sqrt{\mathbf{A}}$ for (1.9-3) for $n = 6$ can be seen in (6.5-19) in Section 6.5.

How about the solutions to the following equation

$$\sqrt{\mathbf{A}}\sqrt{\mathbf{A}} = -\mathbf{I} \ ? \tag{1.9-7}$$

Again, the obvious solutions are

$$\sqrt{\mathbf{A}} = \pm i\mathbf{I}. \tag{1.9-8}$$

Less obvious solutions are

$$\sqrt{\mathbf{A}} = \pm\begin{bmatrix} \sqrt{ab-1} & a \\ -b & -\sqrt{ab-1} \end{bmatrix}. \tag{1.9-9}$$

In this case, not only $\sqrt{\mathbf{A}}$ can be real or complex, symmetric or non-symmetric, it can be skew-symmetric. Examples of a nontrivial real $\sqrt{\mathbf{A}}$ for (1.9-7) for n=6 can be found in (6.5-4), (8.8-41), and (8.8-43).

It can be shown from (1.9-2) that if μ is an eigenvalue of $\sqrt{\mathbf{A}}$, μ^2 is an eigenvalue of $\mathbf{A}$. Conversely, if λ is an eigenvalue of $\mathbf{A}$, the eigenvalue of $\sqrt{\mathbf{A}}$ can be $\sqrt{\lambda}$ or $-\sqrt{\lambda}$. Therefore there are two possible choices of eigenvalues for $\sqrt{\mathbf{A}}$ for each λ. When $\mathbf{A}$ is a 3×3 matrix with the eigenvalues λ_1, λ_2, and λ_3, the followings are eight possible sets of eigenvalues for $\sqrt{\mathbf{A}}$,

$$\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \pm\sqrt{\lambda_3}\right), \tag{1.9-10a}$$

$$\left(\sqrt{\lambda_1}, -\sqrt{\lambda_2}, \pm\sqrt{\lambda_3}\right), \tag{1.9-10b}$$

$$\left(-\sqrt{\lambda_1}, \sqrt{\lambda_2}, \pm\sqrt{\lambda_3}\right), \tag{1.9-10c}$$

$$\left(-\sqrt{\lambda_1}, -\sqrt{\lambda_2}, \pm\sqrt{\lambda_3}\right). \tag{1.9-10d}$$

If $\mathbf{x}$ and $\mathbf{y}$ are, respectively, the right and left eigenvectors of $\sqrt{\mathbf{A}}$, they are also the right and left eigenvectors of $\mathbf{A}$. The converse however does not necessarily hold. In (1.9-4) any vector is an eigenvector of $\mathbf{A}=\mathbf{I}$ but not necessarily an eigenvector of $\sqrt{\mathbf{A}}$ shown in (1.9-5) or (1.9-6). When $\mathbf{A}$ is simple, i.e., when the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ are distinct, the converse holds. Employing the diagonalization (1.3-5),

$$\mathbf{A} = \mathbf{X}\langle\lambda_*\rangle\mathbf{Y}^T = \mathbf{X}\langle\lambda_*\rangle^{1/2}\langle\lambda_*\rangle^{1/2}\mathbf{Y}^T$$

$$= \mathbf{X}\langle\lambda_*\rangle^{1/2}\mathbf{Y}^T\mathbf{X}\langle\lambda_*\rangle^{1/2}\mathbf{Y}^T = \left\{\mathbf{X}\langle\lambda_*\rangle^{1/2}\mathbf{Y}^T\right\}^2$$

suggests that

$$\sqrt{\mathbf{A}} = \mathbf{X}\langle\lambda_*\rangle^{1/2}\mathbf{Y}^T \tag{1.9-11}$$

is a solution. The vectors $\mathbf{x}$ and $\mathbf{y}$ are unique except an arbitrary scalar multiplier. If $\mathbf{x}$ is multiplied by k, $\mathbf{y}$ is multiplied by k^{-1} so that $\mathbf{x}\cdot\mathbf{y}=1$. Since $\langle\lambda_*\rangle^{1/2}$ is a diagonal matrix, the multiplier k in a column of $\mathbf{X}$ is cancelled by the multiplier k^{-1} in the corresponding row of $\mathbf{Y}^T$. Therefore the expression (1.9-11) is independent of the multipliers on the eigenvectors and is unique when a particular set of $\langle\lambda_*\rangle^{1/2}$ is chosen. If $\mathbf{A}$ is a 3×3 matrix there are eight choices of $\langle\lambda_*\rangle^{1/2}$ whose diagonal elements are given in (1.9-10).

> **Theorem 1.9-1** The square root $\sqrt{\mathbf{A}}$ of a square matrix $\mathbf{A}$ is unique when $\mathbf{A}$ is simple and the eigenvalues of $\sqrt{\mathbf{A}}$ are specified.

What if $\mathbf{A}$ is nonsemisimple? We will discuss a particular case in which $\mathbf{A}$ is a 3×3 matrix with $\lambda_1=\lambda_2$ which form a chain of length two. We may consider $\mathbf{A}$ a limiting case of a simple matrix with $\lambda_1\to\lambda_2$, $\mathbf{x}_1\to\mathbf{x}_2$, and $\mathbf{y}_1\to\mathbf{y}_2$ (Ting and Hwu 1988). By Theorem 1.9-1 there are eight solutions for $\sqrt{\mathbf{A}}$ when $\mathbf{A}$ is simple. The eigenvalues of $\sqrt{\mathbf{A}}$ are given in (1.9-10). As $\lambda_1\to\lambda_2$, only (1.9-10a) and (1.9-10d) have repeated eigenvalues while (1.9-10b) and (1.9-10c) have distinct eigenvalues. In this book we will be interested in the eigenvalues given by (1.9-10a) and (1.9-10d) with $\lambda_1=\lambda_2$. For these four cases of interest the solutions for $\sqrt{\mathbf{A}}$ can be obtained as follows. From (1.4-9) $\mathbf{A}$ can be written as

$$\mathbf{A} = \mathbf{X}\mathbf{J}\mathbf{Y}^T = \mathbf{X}\mathbf{J}^{1/2}\mathbf{J}^{1/2}\mathbf{Y}^T$$

$$= \mathbf{X}\mathbf{J}^{1/2}\mathbf{Y}^T\mathbf{X}\mathbf{J}^{1/2}\mathbf{Y}^T = \left\{\mathbf{X}\mathbf{J}^{1/2}\mathbf{Y}^T\right\}^2$$

and hence

$$\sqrt{\mathbf{A}} = \mathbf{X}\mathbf{J}^{1/2}\mathbf{Y}^T. \tag{1.9-12}$$

In the above

$$\mathbf{J} = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix},$$

and, setting $\eta = \frac{1}{2}\lambda_1^{-1/2}$,

$$\mathbf{J}^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & \eta & 0 \\ 0 & \sqrt{\lambda_1} & 0 \\ 0 & 0 & \pm\sqrt{\lambda_3} \end{bmatrix} \text{ or } \begin{bmatrix} -\sqrt{\lambda_1} & -\eta & 0 \\ 0 & -\sqrt{\lambda_1} & 0 \\ 0 & 0 & \pm\sqrt{\lambda_3} \end{bmatrix}. \tag{1.9-13}$$

With (1.9-13), (1.9-12) provides the desired four solutions for $\sqrt{\mathbf{A}}$.

It should be pointed out that (1.9-12) is not a diagonalization of $\sqrt{\mathbf{A}}$ because $\mathbf{J}^{1/2}$ as given by (1.9-13) are not Jordan canonical forms. However, $\mathbf{J}^{1/2}$ of (1.9-13) can be written as

$$\mathbf{J}^{1/2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \pm 2\sqrt{\lambda_1} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \pm\sqrt{\lambda_1} & 1 & 0 \\ 0 & \pm\sqrt{\lambda_1} & 0 \\ 0 & 0 & (\pm)\sqrt{\lambda_3} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \pm\eta & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Let

$$\mathbf{X}^* = \mathbf{X}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \pm 2\sqrt{\lambda_1} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{Y}^* = \mathbf{Y}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \pm\eta & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1.9-14}$$

It is easily shown that

$$\mathbf{X}^*(\mathbf{Y}^*)^T = \mathbf{X}\mathbf{Y}^T = \mathbf{I}, \quad (\mathbf{Y}^*)^T\mathbf{X}^* = \mathbf{Y}^T\mathbf{X} = \mathbf{I}.$$

Equation (1.9-12) can now be written in a diagonalized form

$$\sqrt{\mathbf{A}} = \mathbf{X}^* \begin{bmatrix} \pm\sqrt{\lambda_1} & 1 & 0 \\ 0 & \pm\sqrt{\lambda_1} & 0 \\ 0 & 0 & (\pm)\sqrt{\lambda_3} \end{bmatrix} (\mathbf{Y}^*)^T. \tag{1.9-15}$$

From $(1.9\text{-}14)_1$, the columns of $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ and $(\mathbf{x}_1, \pm 2\sqrt{\lambda_1}\mathbf{x}_2, \mathbf{x}_3)$ are, respectively, the right eigenvectors of $\mathbf{A}$ and $\sqrt{\mathbf{A}}$. The vectors $\mathbf{x}_2$ and $\pm 2\sqrt{\lambda_1}\mathbf{x}_2$ are the generalized right eigenvectors of $\mathbf{A}$ and $\sqrt{\mathbf{A}}$, respectively. Equation $(1.9\text{-}14)_2$ tells us that $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3)$ and $(\mathbf{y}_1, \pm\eta\mathbf{y}_2, \mathbf{y}_3)$ are, respectively, the left eigenvectors of $\mathbf{A}$ and $\sqrt{\mathbf{A}}$. The

vector $\mathbf{y}_1$ is the generalized left eigenvector for both $\mathbf{A}$ and $\sqrt{\mathbf{A}}$. The four solutions given by (1.9-12) and (1.9-13) are nonsemisimple.