Chapter 2

Matrix Algebra

2.1 Basic Definitions

Definition 2.1. Let $a_{ij} \in \mathcal{F}$, $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, where $\mathcal{F}$ is a suitable space, such as the one-dimensional Euclidean or complex space. Then, the ordered rectangular array

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

is said to be a matrix of dimension $m \times n$.

Remark 2.1. Note that the first subscript locates the **row** in which the typical element lies, whereas the second subscript locates the **column**. For example, a_{ks} denotes the element lying in the k th row and s th column of the matrix A. When writing a matrix, we usually write its typical element as well as its dimension. Thus,

$$A = (a_{ij}), \qquad i = 1, 2, \ldots, m, \; j = 1, 2, \ldots, n,$$

denotes a matrix whose typical element is a_{ij} and which has m rows and n columns.

Convention 2.1. Occasionally, we have reason to refer to the columns or rows of the matrix individually. If A is a matrix we shall denote its j th column by $a_{\cdot j}$, i.e.

$$a_{\cdot j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix},$$

P.J. Dhrymes, *Mathematics for Econometrics*,
DOI 10.1007/978-1-4614-8145-4_2,

and its i th row by

$$a_{i\cdot} = (a_{i1}, a_{i2}, \ldots, a_{in}).$$

Definition 2.2. Let A be a matrix as in Definition 2.1. Its transpose, denoted by A', is defined to be the $n \times m$ matrix

$$A' = [a_{ji}], \qquad j = 1, 2, \ldots, n, \ i = 1, 2, \ldots, m,$$

i.e. it is obtained by interchanging rows and columns.

Definition 2.3. Let A be as in Definition 2.1. If $m = n$, A is said to be a **square matrix**.

Definition 2.4. If A is a square matrix, it is said to be **symmetric** if and only if

$$A' = A.$$

If A is a square matrix with, say, n rows and n columns, it is said to be a **diagonal matrix** if and only if

$$a_{ij} = 0, \qquad i \neq j.$$

In this case, it is denoted by

$$A = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn}).$$

Remark 2.2. If A is square matrix, then, evidently, it is not necessary to refer to the number of its rows and columns separately. If it has, say n rows and n columns, we say that A is of **dimension** (or **order**) n.

Definition 2.5. Let A be a square matrix of order n. It is said to be an **upper triangular matrix** if and only if

$$a_{ij} = 0, \qquad i > j.$$

It is said to be a **lower triangular matrix** if and only if

$$a_{ij} = 0, \qquad i < j.$$

Remark 2.3. As the terms imply, for a **lower triangular matrix** all elements above the main diagonal must be zero, while for an **upper triangular matrix** all elements below the main diagonal must be zero.

Definition 2.6. The identity matrix of order n, denoted by I_n ,[1] is a diagonal matrix all of whose non-null elements are unity.

[1] In most of the literature, the subscript is typically omitted. In this volume we shall include it more often than not for the greater clarity it brings to the discussion.

Definition 2.7. The **null** matrix of dimension $m \times n$ is a matrix all of whose elements are null (zeros).

Definition 2.8. Let A be a square matrix of order n. It is said to be an **idempotent** matrix if and only if

$$AA = A.$$

Usually, but not necessarily, idempotent matrices encountered in econometrics are also **symmetric**.

2.2 Basic Operations

Let A, B be two $m \times n$ matrices with elements in $\mathcal{F}$, and let c be a scalar in $\mathcal{F}$. Then, we have:

i. Scalar multiplication:

$$cA = [ca_{ij}].$$

ii. Matrix addition:

$$A + B = [a_{ij} + b_{ij}].$$

Remark 2.4. Note that while scalar multiplication is defined for every matrix, matrix addition for A and B **is not defined unless both have the same dimensions**.

Let A be $m \times n$ and B be $q \times r$, both with elements in $\mathcal{F}$; then, we have:

iii. Matrix multiplication:

$$AB = \left[\sum_{s=1}^{n} a_{is} b_{sj}\right] \quad \text{provided } n = q;$$

$$BA = \left[\sum_{k=1}^{r} b_{ik} a_{kj}\right] \quad \text{provided } r = m.$$

Remark 2.5. Notice that matrix multiplication is not defined for any arbitrary two matrices A, B. They must satisfy certain conditions of **dimensional conformability**. Notice further that if the product

$$AB$$

is defined, the product

$$BA$$

need not be defined, and if it is, it is not **generally** true that

$$AB = BA.$$

Remark 2.6. If two matrices are such that a given operation between them is defined, we say that they are **conformable** with respect to that operation. Thus, for example, if A is $m \times n$ and B is $n \times r$ we say that A and B are **conformable** with respect to the operation of right multiplication, i.e. multiplying A on the right by B. If A is $m \times n$ and B is $q \times m$ we shall say that A and B are **conformable** with respect to the operation of left multiplication, i.e. multiplying A on the left by B. Or if A and B are both $m \times n$ we shall say that A and B are **conformable** with respect to matrix addition. Because being precise is rather cumbersome, we often merely say that two matrices are conformable, and we let the context define precisely the sense in which conformability is to be understood.

An immediate consequence of the preceding definitions is

Proposition 2.1. Let A be $m \times n$, and B be $n \times r$. The jth column of

$$C = AB$$

is given by

$$c_{\cdot j} = \sum_{s=1}^{n} a_{\cdot s} b_{sj}, \qquad j = 1, 2, \ldots, r.$$

Proof: Obvious from the definition of matrix multiplication.

Proposition 2.2. Let A be $m \times n$, B be $n \times r$. The ith row of

$$C = AB$$

is given by

$$c_{i\cdot} = \sum_{q=1}^{n} a_{iq} b_{q\cdot}, \qquad i = 1, 2, \ldots, m.$$

Proof: Obvious from the definition of matrix multiplication.

Proposition 2.3. Let A, B be $m \times n$, and $n \times r$, respectively. Then,

$$C' = B'A',$$

where

$$C = AB.$$

Proof: The typical element of C is given by

$$c_{ij} = \sum_{s=1}^{n} a_{is} b_{sj}.$$

By definition, the typical (i, j) element of C', say c'_{ij}, is given by

$$c'_{ij} = c_{ji} = \sum_{s=1}^{n} a_{js} b_{si}.$$

But

$$a_{js} = a'_{sj}, \qquad b_{si} = b'_{is},$$

i.e. a_{js} is the (s, j) element of A', say a'_{sj}, and b_{si} is the (i, s) element of B', say b'_{is}. Consequently,

$$c'_{ij} = c_{ji} = \sum_{s=1}^{n} a_{js} b_{si} = \sum_{s=1}^{n} b'_{is} a'_{sj},$$

which shows that the (i, j) element of C' is the (i, j) element of $B'A'$.

q.e.d.

2.3 Rank and Inverse of a Matrix

Definition 2.9. Let A be $m \times n$. The **column rank** of A is the maximum number of linearly independent columns it contains. The **row rank** of A is the maximum number of linearly independent rows it contains.

Remark 2.7. It may be shown—but not here—that the row rank of A is **equal** to its column rank. Hence, the concept of rank is unambiguous, and we denote by

$$r(A)$$

the rank of A. Thus, if we are told that A is $m \times n$ we can immediately conclude that

$$r(A) \leq \min(m, n).$$

Definition 2.10. Let A be $m \times n$, $m \leq n$. We say that A is of **full rank** if and only if

$$r(A) = m.$$

Definition 2.11. Let A be a square matrix of order m. We say that A is nonsingular if and only if

$$r(A) = m.$$

Remark 2.8. An example of a nonsingular matrix is the diagonal matrix

$$A = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{mm})$$

for which

$$a_{ii} \neq 0, \qquad i = 1, 2, \ldots, m.$$

We are now in a position to define a matrix operation that corresponds to division for scalars. For example, if $c \in \mathcal{F}$ and $c \neq 0$, we know that for any $a \in \mathcal{F}$

$$\frac{a}{c}$$

means the operation of defining

$$\frac{1}{c}$$

(the "inverse" of c) and multiplying that by a. The "inverse" of a scalar, say c, is another scalar, say b, such that

$$bc = cb = 1.$$

We have a similar operation for square matrices.

2.3.1 Matrix Inversion

Let A be a square matrix of order m. Its inverse, say B, is another square matrix of order m such that B, if it exists, is defined by the property

$$AB = BA = I_m$$

Definition 2.12. Let A be a square matrix of order m. If its inverse exists it is denoted by A^{-1}, and the matrix A is said to be **invertible**.

Remark 2.9. The terms **invertible**, **nonsingular**, and **of full rank** are synonymous for square matrices. This is made clear below.

Proposition 2.4. Let A be a square matrix of order m. Then A is invertible if and only if

$$r(A) = m.$$

Proof: Necessity: Suppose A is invertible; then there exists a square matrix B (of order m) such that

$$AB = I_m. \tag{2.1}$$

Let $c \neq 0$ be any m-element vector and note that Eq. (2.1) implies

$$ABc = c.$$

Since $c \neq 0$ we must have that

$$Ad = c, \qquad d = Bc \neq 0.$$

But this means that if c is any m-dimensional vector it can be expressed as a linear combination of the columns of A, which in turn means that the columns of A span the vector space V_m consisting of all m-dimensional vectors with elements in $\mathcal{F}$. Because the dimension of this space is m, it follows that the (m) columns of A are linearly independent; hence, its rank is m.

Sufficiency: Conversely, suppose that

$$r(A) = m.$$

Then, its columns form a basis for V_m. The unit vectors (see Chap. 1) $\{e_{\cdot i}: i = 1, 2, \dots, m\}$ all belong to V_m. Thus, we can write

$$e_{\cdot i} = Ab_{\cdot i} = \sum_{s=1}^{m} a_{\cdot s} b_{si}, \qquad i = 1, 2, \dots, m.$$

The matrix

$$B = [b_{si}]$$

has the property[2]

$$AB = I_m.$$

q.e.d.

Corollary 2.1. Let A be a square matrix of order m. If A is invertible then the following is true for its inverse B: B is of rank m and thus B also is invertible; the inverse of B is A.

Proof: Obvious from the definition of the inverse and the proposition.

It is useful here to introduce the following definition

Definition 2.13. Let A be $m \times n$. The **column space** of A, denoted by $C(A)$, is the set of m-dimensional (column) vectors

$$C(A) = \{\xi : \xi = Ax\},$$

where x is n-dimensional with elements in $\mathcal{F}$. Similarly, the **row space** of A, $R(A)$, is the set of n-dimensional(row) vectors

$$R(A) = \{\zeta : \zeta = yA\},$$

where y is a row vector of dimension m with elements in $\mathcal{F}$.

[2] Strictly speaking, we should also provide an argument based on the rows of A and on $BA = I_m$, but this is repetitious and is omitted for the sake of simplicity.

Remark 2.10. It is clear that the column space of A is a vector space and that it is **spanned** by the columns of A. Moreover, the dimension of this vector space is simply the rank of A, i.e. $\dim C(A) = \mathrm{r}\,(A)$. Similarly, the row space of A is a vector space spanned by its rows, and the dimension of this space is also equal to the rank of A because the row rank of A is equal to its column rank.

Definition 2.14. Let A be $m \times n$. The (column) **null space** of A, denoted by $N(A)$, is the set

$$N(A) = \{x: \quad Ax = 0\}.$$

Remark 2.11. A similar definition can be made for the (row) null space of A.

Definition 2.15. Let A be $m \times n$, and consider its null space $N(A)$. This is a vector space; its dimension is termed the **nullity** of A and is denoted by

$$n(A).$$

We now have an important relation between the column space and column null space of any matrix.

Proposition 2.5. Let A be $p \times q$. Then,

$$r(A) + n(A) = q.$$

Proof: Suppose the nullity of A is $n(A) = n \leq q$, and let $\{\xi_i : \; i = 1, 2, \ldots, n\}$ be a basis for $N(A)$. Note that each ξ_i is a q-dimensional (column) vector with elements in $\mathcal{F}$. We can extend this to a basis for V_q, the vector space containing all q-dimensional vectors with elements in $\mathcal{F}$; thus, let

$$\{\xi_1, \xi_2, \ldots, \xi_n, \zeta_1, \zeta_2, \ldots, \zeta_{q-n}\}$$

be such a basis. If x is any q-dimensional vector, we can write, uniquely,

$$x = \sum_{i=1}^{n} c_i \xi_i + \sum_{j=1}^{q-n} f_j \zeta_j .$$

Now, define

$$y = Ax \in C(A)$$

and note that

$$y = \sum_{i=1}^{n} c_i A\xi_i + \sum_{j=1}^{q-n} f_j A\zeta_j = \sum_{j=1}^{q-n} f_j (A\zeta_j). \tag{2.2}$$

This is so since

$$A\xi_i = 0, \qquad i = 1, 2, \ldots, n,$$

owing to the fact that the ξ 's are a basis for the null space of A.

But Eq. (2.2) means that the vectors

$$\{A\zeta_j : j = 1, 2, \ldots, q-n\}$$

span $C(A)$, since x and (hence) y are arbitrary. We show that these vectors are linearly independent, and hence a basis for $C(A)$. Suppose not. Then, there exist scalars, g_j, $j = 1, 2, \ldots, q-n$, not all of which are zero, such that

$$0 = \sum_{j=1}^{q-n} (A\zeta_j) g_j = A \left(\sum_{j=1}^{q-n} \zeta_j g_j \right). \tag{2.3}$$

Equation (2.3) implies that

$$\zeta = \sum_{j=1}^{q-n} \zeta_j g_j \tag{2.4}$$

lies in the null space of A, because it states $A\zeta = 0$. As such, $\zeta \in V_q$ and has a unique representation in terms of the basis of that vector space, say

$$\zeta = \sum_{i=1}^{n} d_i \xi_i + \sum_{j=1}^{q-n} k_j \zeta_j. \tag{2.5}$$

Moreover, since $\zeta \in N(A)$, we know that in Eq. (2.5)

$$k_j = 0, \qquad j = 1, 2, \ldots, q-n.$$

But Eqs. (2.5) and (2.4) give two dissimilar representations of ζ in terms of a single basis for V_q, which is a contradiction, unless

$$g_j = 0, \qquad j = 1, 2, \ldots, q-n,$$
$$d_i = 0, \qquad i = 1, 2, \ldots, n.$$

This shows that Eq. (2.3) can be satisfied only by null g_j, $j = 1, 2, \ldots, q-n$; hence, the set $\{A\zeta_j : \ j = 1, 2, \ldots, q-n\}$ is linearly independent and, consequently, a basis for $C(A)$. Therefore, since the dimension of $C(A) = \mathrm{r}(A)$, by Remark 2.10, we have

$$\dim[C(A)] = r(A) = q - n.$$

q.e.d.

Another useful result is the following.

Proposition 2.6. Let A be $p \times q$, let B be a nonsingular matrix of order q, and put $D = AB$. Then

$$r(D) = r(A).$$

Proof: We shall show that $C(A) = C(D)$, which, by the discussion in the proof of Proposition 2.5, is equivalent to the claim of the proposition.

Suppose $y \in C(A)$. Then, there exists a vector $x \in V_q$ such that $y = Ax$. Since B is nonsingular, define the vector $\xi = B^{-1}x$. We note $D\xi = ABB^{-1}x = y$, which shows that

$$C(A) \subset C(D). \tag{2.6}$$

Conversely, suppose $z \in C(D)$. This means there exists a vector $\xi \in V_q$ such that $z = D\xi$. Define the vector $x = B\xi$ and note that

$$Ax = AB\xi = D\xi = z;$$

this means that $z \in C(A)$, which shows

$$C(D) \subset C(A). \tag{2.7}$$

But Eqs. (2.6) and (2.7) together imply $C(A) = C(D)$.

q.e.d.

Finally, we have

Proposition 2.7. Let A be $p \times q$ and B $q \times r$, and put

$$D = AB.$$

Then

$$r(D) \leq \min[r(A), r(B)].$$

Proof: Since $D = AB$, we note that if $x \in N(B)$ then $x \in N(D)$; hence, we conclude

$$N(B) \subset N(D),$$

and thus that

$$n(B) \leq n(D). \tag{2.8}$$

But from

$$r(D) + n(D) = r,$$
$$r(B) + n(B) = r,$$

we find, in view of Eq. (2.8),

$$r(D) \leq r(B). \tag{2.9}$$

Next, suppose that $y \in C(D)$. This means that there exists a vector, say, $x \in V_r$, such that $y = Dx$ or $y = ABx = A(Bx)$, so that $y \in C(A)$. But this means that

$$C(D) \subset C(A),$$

or that

$$r(D) \leq r(A). \tag{2.10}$$

Together Eqs. (2.9) and (2.10) imply

$$r(D) \leq \min[r(A), r(B)].$$

q.e.d.

Remark 2.12. The preceding results can be stated in the following useful form: multiplying two (and therefore any finite number of) matrices results in a matrix whose rank cannot exceed the rank of the lowest ranked factor. The product of nonsingular matrices is nonsingular. Multiplying a matrix by a nonsingular matrix does not change its rank.

2.4 Hermite Forms and Rank Factorization

We begin with a few elementary aspects of matrix operations.

Definition 2.16. Let A be $m \times n$; any one of the following operations is said to be an **elementary transformation** of A :

i. Interchanging two rows (or columns);

ii. Multiplying the elements of a row (or column) by a (nonzero) scalar c;

iii. Multiplying the elements of a row (or column) by a (nonzero) scalar c and adding the result to another row (or column).

The operations above are said to be **elementary row (or column) operations**.

Remark 2.13. The matrix performing operation i is the matrix obtained from the identity matrix by interchanging the two rows (or columns) in question.

The matrix performing operation ii is obtained from the identity matrix by multiplying the corresponding row (or column) by the scalar c.

Finally, the matrix performing operation iii is obtained from the identity matrix as follows: if it is desired to add c times the k th row to the i th row of a given matrix A, simply insert the scalar c in the (i,k) position of the appropriate identity matrix and use the resulting matrix to multiply A on the left.

Such matrices are termed **elementary matrices**. An elementary **row** operation is performed on A by multiplying A on the left by the corresponding elementary matrix, E, i.e. EA.

An elementary **column** operation is performed (*mutatis mutandis*) by AE.

Example 2.1. Let

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

and suppose we want to interchange the position of the first and third rows (columns). Define

$$E_1 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

Then,

$$E_1 A = \begin{bmatrix} a_{31} & a_{32} & a_{33} \\ a_{21} & a_{22} & a_{23} \\ a_{11} & a_{12} & a_{13} \end{bmatrix}; \qquad AE_1 = \begin{bmatrix} a_{13} & a_{12} & a_{11} \\ a_{23} & a_{22} & a_{21} \\ a_{33} & a_{32} & a_{31} \end{bmatrix}.$$

Suppose we wish to multiply the second row (column) of A by the scalar c. Define

$$E_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

Then,

$$E_2 A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ ca_{21} & ca_{22} & ca_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \qquad AE_2 = \begin{bmatrix} a_{11} & ca_{12} & a_{13} \\ a_{21} & ca_{22} & a_{23} \\ a_{31} & ca_{32} & a_{33} \end{bmatrix}.$$

Finally, suppose we wish to add c times the first row (column) to the third row (column). Define

$$E_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ c & 0 & 1 \end{bmatrix}, \quad \text{and note that}$$

$$E_3A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ ca_{11}+a_{31} & ca_{12}+a_{32} & ca_{13}+a_{33} \end{bmatrix}$$

$$AE_3 = \begin{bmatrix} a_{11}+ca_{13} & a_{12} & a_{13} \\ a_{21}+ca_{23} & a_{22} & a_{23} \\ a_{31}+ca_{33} & a_{32} & a_{33} \end{bmatrix}.$$

The result below follows immediately.

Proposition 2.8. Every elementary transformation matrix is nonsingular, and its inverse is a matrix of the same type.

Proof: For matrices of type E_1 it is clear that $E_1E_1 = I$. The inverse of a matrix of type E_2 is of the same form but with c replaced by $1/c$. Similarly, the inverse of a matrix of type E_3 is of the same form but with c replaced by $-c$.

q.e.d.

Definition 2.17. An $m \times n$ matrix C is said to be an (**upper**) **echelon matrix** if:

i. It can be partitioned:
$$C = \begin{pmatrix} C_1 \\ 0 \end{pmatrix},$$
where C_1 is $r \times n$ $(r \leq n)$ and there is no row in C_1 consisting entirely of zeros;

ii. The first nonzero element appearing in each row of C_1 is unity and, if the first nonzero element in row i is c_{ij} then all other elements in column j are zero, i.e. $c_{ij} = 0$ for $j > i$;

iii. When the first nonzero element in the kth row of C_1 is c_{kj_k}, then $j_1 < j_2 < j_3 < \cdots < j_k$.

An immediate consequence of the definition is

Proposition 2.9. Let A be $m \times n$; there exists a nonsingular $(m \times m)$ matrix B such that
$$BA = C$$
and C is an (upper) echelon matrix.

Proof: Consider the first column of A, and suppose it contains a nonzero element (if not, consider the second column, etc.). Without loss of generality, we suppose this to be a_{11} (if not, simply interchange rows so that it does become the first element). Multiply the first row by $1/a_{11}$. This is accomplished through multiplication on the left by a matrix of type E_2. Next, multiply the first row of the resulting matrix by $-a_{s1}$ and add to the sth row. This is accomplished through multiplication on the left by a matrix of type E_3. Continuing in this fashion, we make all elements in the first column zero except the first, which is unity. Repeat this for the second column, third column, and, in general, for all other columns of A. In the end some rows may consist entirely of zeros. If they do not all occur at the end (of the rows of the matrix) interchange rows so that all zero rows occur at the end. This can be done through multiplication on the left by a matrix of type E_1. The resulting matrix is, thus, in upper echelon form, and has been obtained through multiplication on the left by a number of elementary matrices. Because the latter are nonsingular, we have

$$BA = C,$$

where B is nonsingular and C is in upper echelon form.

q.e.d.

Proposition 2.10. Let A be $m \times n$, and suppose it can be reduced to an upper echelon matrix

$$BA = \begin{pmatrix} C_1 \\ 0 \end{pmatrix} = C$$

such that C_1 is $r \times n$. Then,

$$\mathrm{r}(A) = r.$$

Proof: By construction, the rows of C_1 are linearly independent; thus,

$$\mathrm{r}(C_1) = \mathrm{r}(C) = r \quad \text{and} \quad \mathrm{r}(C) \le \mathrm{r}(A).$$

We also have $A = B^{-1}C$. Hence

$$\mathrm{r}(A) \le \mathrm{r}(C),$$

which shows

$$\mathrm{r}(A) = \mathrm{r}(C) = r.$$

q.e.d.

Definition 2.18. An $n \times n$ matrix H^* is said to be in (**upper**) **Hermite form** if and only if:

i. H^* is (upper) triangular;

ii. The elements along the main diagonal of H^* are either zero or one;

iii. If a main diagonal element of H^* is zero, all elements in the **row** in which the null diagonal element occurs are zero;

iv. If a main diagonal element of H^* is unity then all other elements in the **column** in which the unit element occurs are zero.

Definition 2.19. An $n \times m$ matrix H is said to be in (**upper**) **Hermite canonical** form if and only if

$$H = \begin{bmatrix} I & H_1 \\ 0 & 0 \end{bmatrix}.$$

Proposition 2.11. Every matrix in Hermite form can be put in Hermite canonical form by elementary row and column operations.

Proof: Let H^* be a matrix in Hermite form; by interchanging rows we can put all zero rows at the end so that for some nonsingular matrix B_1 we have

$$B_1 H^* = \begin{pmatrix} H_1^* \\ 0 \end{pmatrix},$$

where the first nonzero element in each row of H_1^* is unity and H_1^* contains no zero rows. By interchanging columns, we can place the (unit) first nonzero elements of the rows of H_1^* along the main diagonal, so that there exists a nonsingular matrix B_2 for which

$$B_1 H^* B_2 = \begin{bmatrix} I & H_1 \\ 0 & 0 \end{bmatrix}.$$

q.e.d.

Proposition 2.12. The rank of a matrix H^*, in Hermite form, is equal to the dimension of the identity block in its Hermite canonical form.

Proof: Obvious.

Proposition 2.13. Every (square) matrix H^* in Hermite form is idempotent, although it is obviously not necessarily symmetric.

Proof: We have to show that

$$H^* H^* = H^*.$$

Because H^* is **upper triangular**, we know that H^*H^* is also **upper triangular**, and thus we need only determine its (i,j) element for $i \le j$. Now, the (i,j) element of H^*H^* is

$$(H^*H^*)_{ij} = \sum_{k=1}^{n} h^*_{ik} h^*_{kj} = \sum_{k=i}^{j} h^*_{ik} h^*_{kj}.$$

If $h^*_{ii} = 0$ then $h^*_{ij} = 0$ for all j; hence $(H^*H^*)_{ij} = 0$ for all j. If $h^*_{ii} = 1$ then

$$(H^*H^*)_{ij} = h^*_{ij} + h^*_{i,i+1} h^*_{i+1,j} + \cdots + h^*_{ij} h^*_{jj}.$$

Now, $h^*_{i+1,i+1}$ is either zero or one; if zero then $h^*_{i+1,j} = 0$ for all j, and hence the second term on the right side in the equation above is zero. If $h^*_{i+1,i+1} = 1$ then $h^*_{i,i+1} = 0$, so that again the second term is null. Similarly, if $h^*_{i+2,i+2} = 0$, then the third term is null; if $h^*_{i+2,i+2} = 1$ then $h^*_{i,i+2} = 0$, so that again the third term on the right side of the equation defining $(H^*H^*)_{ij}$ is zero. Finally, if $h^*_{jj} = 1$ then $h^*_{ij} = 0$, and if $h^*_{jj} = 0$ then again $h^*_{ij} = 0$. Consequently, it is always the case that

$$(H^*H^*)_{ij} = h^*_{ij}$$

and thus

$$H^*H^* = H^*.$$

q.e.d.

2.4.1 Rank Factorization

Proposition 2.14. Let A be $n \times n$ of rank $r \le n$. There exist nonsingular matrices Q_1, Q_2 such that

$$Q_1^{-1} A Q_2^{-1} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Proof: By Proposition 2.9, there exists a nonsingular matrix Q_1 such that

$$Q_1^{-1} A = \begin{pmatrix} A_1^* \\ 0 \end{pmatrix},$$

i.e. $Q_1^{-1}A$ is an (upper) echelon matrix. By transposition of columns, we obtain

$$\begin{pmatrix} A_1^* \\ 0 \end{pmatrix} B_1 = \begin{bmatrix} I_r & A_1 \\ 0 & 0 \end{bmatrix},$$

which in upper Hermite canonical form.

By elementary column operations, we can eliminate A_1, i.e.

$$\begin{bmatrix} I_r & A_1 \\ 0 & 0 \end{bmatrix} B_2 = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Take $Q_2^{-1} = B_1 B_2$, and note that we have

$$Q_1^{-1} A Q_2^{-1} = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}.$$

q.e.d.

Proposition 2.15 (Rank Factorization)**.** Let A be $m \times n$ $(m \leq n)$ of rank $r \leq m$. There exists an $m \times r$ matrix C_1 of rank r and an $r \times n$ matrix C_2 of rank r such that

$$A = C_1 C_2.$$

Proof: Let

$$A_0 = \begin{pmatrix} A \\ 0 \end{pmatrix},$$

where A_0 is $n \times n$ of rank r. By Proposition 2.14, there exist nonsingular matrices Q_1, Q_2 such that

$$A_0 = Q_1 \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} Q_2.$$

Partition

$$Q_1 = \begin{bmatrix} C_1 & C_{11} \\ C_{21} & C_{22} \end{bmatrix}, \qquad Q_2 = \begin{pmatrix} C_2 \\ C^* \end{pmatrix}$$

so that C_1 is $m \times r$ and C_2 is $r \times n$ (of rank r). Thus,

$$A_0 = \begin{pmatrix} C_1 C_2 \\ C_{21} C_2 \end{pmatrix};$$

hence,

$$A = C_1 C_2.$$

Since

$$r = r(A) \leq \min[r(C_1), r(C_2)] = \min[r(C_1), r],$$

we must have that

$$r(C_1) = r.$$

q.e.d.

Remark 2.14. Proposition 2.15 is the so-called **rank factorization theorem**.

2.5 Trace and Determinants

Associated with square matrices are two important scalar functions, the **trace** and the **determinant**.

Definition 2.20. Let A be a square matrix of order m. Its trace is denoted by $\mathrm{tr}(A)$ and is defined by

$$\mathrm{tr}(A) = \sum_{i=1}^{m} a_{ii}.$$

An immediate consequence of the definition is

Proposition 2.16. Let A, B be two square matrices of order m. Then,

$$\begin{aligned} \mathrm{tr}(A+B) &= \mathrm{tr}(A) + \mathrm{tr}(B), \\ \mathrm{tr}(AB) &= \mathrm{tr}(BA). \end{aligned}$$

Proof: By definition, the typical element of $A + B$ is $a_{ij} + b_{ij}$. Hence,

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

Similarly, the typical element of AB is

$$\sum_{k=1}^{m} a_{ik} b_{kj}.$$

Hence,

$$\mathrm{tr}(AB) = \sum_{i=1}^{m} \sum_{k=1}^{m} a_{ik} b_{ki}.$$

The typical element of BA is

$$\sum_{i=1}^{m} b_{ki} a_{ij}.$$

Thus,

$$\mathrm{tr}(BA) = \sum_{k=1}^{m} \sum_{i=1}^{m} b_{ki} a_{ik} = \sum_{i=1}^{m} \sum_{k=1}^{m} a_{ik} b_{ki},$$

which shows

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

q.e.d.

Definition 2.21. Let A be a square matrix of order m; its **determinant**, denoted by $|A|$ or by $\det A$, is given by

$$|A| = \sum (-1)^s a_{1j_1} a_{2j_2} \cdots a_{mj_m},$$

where $j_1, j_2, \dots, j_m$ is a permutation of the numbers $1, 2, \dots, m$, and s is zero or one depending on whether the number of transpositions required to restore $j_1, j_2, \dots, j_m$ to the natural sequence $1, 2, 3, \dots, m$ is even or odd; the sum is taken over all possible such permutations.

Example 2.2. Consider the matrix

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

According to the definition, its determinant is given by

$$\begin{aligned} |A| &= (-1)^{s_1} a_{11}a_{22}a_{33} + (-1)^{s_2} a_{11}a_{23}a_{32} \\ &\quad + (-1)^{s_3} a_{12}a_{21}a_{33} + (-1)^{s_4} a_{12}a_{23}a_{31} \\ &\quad + (-1)^{s_5} a_{13}a_{21}a_{32} + (-1)^{s_6} a_{13}a_{22}a_{31}. \end{aligned}$$

To determine s_1, we note that the second subscripts in the corresponding term are in natural order; hence $s_1 = 0$. For the second term, we note that one transposition restores the second subscripts to the natural order; hence $s_2 = 1$. For the third term, $s_3 = 1$. For the fourth term two transpositions are required; hence, $s_4 = 0$. For the fifth term two transpositions are required; hence, $s_5 = 0$. For the sixth term one transposition is required; hence, $s_6 = 1$.

Remark 2.15. It should be noted that although Definition 2.21 is stated with the **rows** in natural order, a completely equivalent definition is one in which the **columns** are in natural order. Thus, for example, we could just as well have defined

$$|A| = \sum (-1)^d a_{i_1 1} a_{i_2 2} \cdots a_{i_m m}$$

where d is zero or one, according as the number of transpositions required to restore $i_1, i_2, \dots, i_m$ to the natural order $1, 2, 3, \dots, m$ is even or odd.

Example 2.3. Consider the matrix A of Example 2.2 and obtain the determinant in accordance with Remark 2.15. Thus

$$\begin{aligned} |A| &= (-1)^{d_1} a_{11}a_{22}a_{33} + (-1)^{d_2} a_{11}a_{32}a_{23} \\ &= +(-1)^{d_3} a_{21}a_{12}a_{33} + (-1)^{d_4} a_{21}a_{32}a_{13} \\ &\quad (-1)^{d_5} a_{31}a_{12}a_{23} + (-1)^{d_6} a_{31}a_{22}a_{13}. \end{aligned}$$

It is easily determined that $d_1 = 0,\ d_2 = 1,\ d_3 = 1,\ d_4 = 0,\ d_5 = 0,\ d_6 = 1$. Noting, in comparison with Example 2.2, that $s_1 = d_1,\ s_2 = d_2,\ s_3 = d_3,\ s_4 = d_5,\ s_5 = d_4,\ s_6 = d_6$, we see that we have exactly the same terms.

An immediate consequence of the definition is the following proposition.

Proposition 2.17. Let A be a square matrix of order m. Then

$$|A'| = |A|.$$

Proof: Obvious from Definition 2.21 and Remark 2.15.

Proposition 2.18. Let A be a square matrix of order m, and consider the matrix B that is obtained by interchanging the k th and r th rows of A $(k \leq r)$. Then,

$$|B| = -|A|.$$

Proof: By definition,

$$|B| = \sum (-1)^s b_{1j_1} b_{2j_2} \cdots b_{mj_m},$$

$$|A| = \sum (-1)^s a_{1j_1} a_{2j_2} \cdots a_{mj_m}. \tag{2.11}$$

However, each term in $|B|$ (except possibly for sign) is exactly the same as in $|A|$ but for the interchange of the k th and r th rows. Thus, for example, we can write

$$|B| = \sum (-1)^s a_{1j_1} \cdots a_{k-1 j_{k-1}} a_{rj_k} a_{k+1 j_{k+1}}$$

$$\cdots a_{r-1 j_{r-1}} a_{kr_r} a_{r+1 j_{r+1}} \cdots a_{mj_m}. \tag{2.12}$$

Now, if we restore to their natural order the first subscripts in Eq. (2.12), we will have an expression like the one for $|A|$ in Eq. (2.11), except that in Eq. (2.12) we would require an odd number of additional transpositions to restore the second subscripts to the natural order $1, 2, \ldots, m$. Hence, the sign of each term in Eq. (2.12) is exactly the opposite of the corresponding term in Eq. (2.11). Consequently,

$$|B| = -|A|.$$

q.e.d.

Proposition 2.19. Let A be a square matrix of order m, and suppose it has two identical rows. Then

$$|A| = 0.$$

Proof: Let B be the matrix obtained by interchanging the two identical rows. Then by Proposition 2.18,

$$|B| = -|A|. \tag{2.13}$$

Since these two rows are identical $B = A$, and thus

$$|B| = |A|. \tag{2.14}$$

But Eqs. (2.13) and (2.14) imply

$$|A| = 0.$$

q.e.d.

Proposition 2.20. Let A be a square matrix of order m, and suppose all elements in its r th row are zero. Then,

$$|A| = 0.$$

Proof: By the definition of a determinant, we have

$$|A| = \sum (-1)^s a_{1j_1} a_{2j_2} a_{3j_3} \cdots a_{mj_m},$$

and it is clear that every term above contains an element from the i th row, say a_{ij_i}. Hence, all terms vanish and thus

$$|A| = 0.$$

q.e.d.

Remark 2.16. It is clear that, in any of the propositions regarding determinants, we may substitute "column" for "row" without disturbing the conclusion. This is clearly demonstrated by Remark 2.15 and the example following. Thus, while most of the propositions are framed in terms of rows, an equivalent result would hold in terms of columns.

Proposition 2.21. Let A be a square matrix of order m. Let B be the matrix obtained when we multiply the i th row by a scalar k. Then

$$|B| = k|A|.$$

Proof: By Definition 2.21,

$$\begin{aligned} |B| &= \sum (-1)^s b_{1j} b_{2j_1} \cdots b_{mj_m} \\ &= k \sum (-1)^s a_{1j_m} a_{2j_m} \cdots a_{mj_m} = k|A|. \end{aligned}$$

This is so because

$$\begin{aligned} b_{sj_s} &= a_{sj_s} \quad \text{for } s \neq i, \\ &= k a_{sj_s} \quad \text{for } s = i. \end{aligned}$$

q.e.d.

Proposition 2.22. Let A be a square matrix of order m. Let B be the matrix obtained when to the r th row of A we add k times its s th row. Then,

$$|B| = |A|.$$

Proof: By Definition 2.21,

$$\begin{aligned} |B| &= \sum (-1)^s b_{1j_1} b_{2j_2} \cdots b_{mj_m} \\ &= \sum (-1)^s a_{1j_1} \cdots a_{r-1j_{r-1}} (a_{rj_r} + k a_{sj_r}) \cdots a_{mj_m} \\ &= \sum (-1)^s a_{1j_1} \cdots a_{r-1j_{r-1}} a_{rj_r} \cdots a_{mj_n} + k \sum (-1)^s a_{1j_1} \\ &\quad \cdots a_{r-1j_{r-1}} a_{sj_r} \cdots a_{mj_m}. \end{aligned}$$

The first term on the rightmost member of the equation above gives $|A|$, and the second term represents k times the determinant of a matrix having two identical rows. By Proposition 2.19, that determinant is zero. Hence,

$$|B| = |A|.$$

q.e.d.

Remark 2.17. It is evident, by a simple extension of the argument above, that if we add a linear combination of any number of the remaining rows (or columns) to the r th row (or column) of A, we do not affect the determinant of A.

Remark 2.18. While Definition 2.21 is intuitively illuminating and, indeed, leads rather easily to the derivation of certain important properties of the determinant, it is not particularly convenient for computational purposes. We give below a number of useful alternatives for evaluating determinants.

Definition 2.22. Let A be a square matrix of order m and let B_{ij} be the matrix obtained by deleting from A its i th row and j th column. The quantity

$$A_{ij} = (-1)^{i+j} |B_{ij}|$$

is said to be the **cofactor** of the element a_{ij} of A. The matrix B_{ij} is said to be an $(m-1)$**-order minor** of A.

Proposition 2.23. (Expansion by cofactors) Let A be a square matrix of order m. Then,

$$|A| = \sum_{j=1}^{m} a_{ij} A_{ij}, \qquad |A| = \sum_{i=1}^{m} a_{ij} A_{ij}.$$

Proof: For definiteness, we shall prove this for a specific value of i (for the expansion by cofactors in a given row). By the definition of a determinant, $|A| = \sum (-1)^s a_{1j_1} a_{2j_2} \cdots a_{mj_m}$. This can also be written more suggestively as follows:

$$\begin{aligned} |A| &= a_{11} \sum (-1)^s a_{2j_2} \cdots a_{mj_m} + a_{12} \sum (-1)^s a_{2j_2} \cdots a_{mj_m} \\ &\quad + \cdots + a_{1m} \sum (-1)^s a_{2j_2} \cdots a_{mj_m} \\ &= a_{11} f_{11} + a_{12} f_{12} + a_{1m} f_{1m}, \end{aligned} \tag{2.15}$$

where, for example, $f_{1r} = \sum (-1)^s a_{2j_2} \cdots a_{mj_m}$, and the numbers $j_2, j_3, \ldots, j_m$ represent some arrangement (permutation) of the integers $1, 2, \ldots, m$, excluding the integer $r \leq m$. But it is clear that, except (possibly) for sign, f_{1r} is simply the determinant of an $(m-1)$-order minor of A obtained by deleting its first row and r th column. In that determinant, s would be zero or one depending on whether the number of transpositions required to restore $j_2, j_3, \ldots, j_m$ to the natural order $1, 2, \ldots, r-1, r+1, \ldots, m$ is even or odd. In Eq. (2.15), however, the corresponding s would be zero or one depending on whether the number of transpositions required to restore $r,\ j_2, j_3, \ldots, j_m$ to the natural order $1, 2, \ldots, r-1,\ r,\ r+1, \ldots, m$ is even or odd. But for $r > 1$, this would be exactly $r-1$ more than before, and would be exactly the same if $r = 1$. Thus,

$$f_{11} = |B_{11}|, \ \ f_{12} = (-1)|B_{12}|, \ \ f_{13} = (-1)^2 |B_{13}|, \ \ \ldots, \ \ f_{1m} = (-1)^{m-1} |B_{1m}|.$$

Noting that $A_{1j} = (-1)^{j+1} |B_{1j}|$, and $(-1)^{j+1} = (-1)^{j-1}$, we conclude that $f_{1j} = A_{1j}$. A similar argument can be made for the expansion along any row—and not merely the first—as well as expansion along any column.

q.e.d.

Remark 2.19. Expansion by cofactors is a very useful way of evaluating a determinant; it is also the one most commonly used in actual computations. For certain instances, however, another method—the Laplace expansion—is preferable. Its proof, however, is cumbersome and will be omitted.

Definition 2.23. Let A be a square matrix of order m. Let P be an n-order $(n < m)$ minor formed by the rows $i_1, i_2, \ldots, i_n$ and the columns $j_1, j_2, \ldots, j_n$ of A, and let Q be the $(m-n)$-order minor formed by taking the remaining rows and columns. Then Q is said to be the **complementary minor of** P (and conversely P is said to be the complementary minor of Q.) Moreover,

$$M = [(-1)^{\sum_{i=1}^{n}(i_r+j_r)}]|Q|$$

is said to be the **complementary cofactor of** P.

Proposition 2.24 (Laplace expansion)**.** Let A be a square matrix of order m. Let $P(i_1, i_2, \ldots, i_n \mid j_1, j_2, \ldots, j_n)$ be an n-order minor of A formed by rows $i_1, i_2, \ldots, i_n$ and columns $j_1, j_2, \ldots, j_n$, $n < m$. Let M be its associated complementary cofactor. Then

$$|A| = \sum_{j_1<j_2<\cdots<j_n} |P(i_1, i_2, \ldots, i_n \mid j_1, j_2, \ldots, j_n)|M,$$

where the sum is taken over all possible choices of n columns of P, the number of which is in fact

$$\binom{m}{n};$$

similarly,

$$|A| = \sum_{i_1<i_2<\cdots<i_n} |P(i_1, i_2, \ldots, i_n \mid j_1, j_2, \ldots, j_n)|M,$$

the sum now chosen over all $\binom{m}{n}$ ways in which n of the rows of P may be chosen.

Proof: The proof, while conceptually simple, is rather cumbersome and not particularly instructive. The interested reader is referred to Hadley (1961).

Remark 2.20. The first representation in Proposition 2.24 refers to an expansion by n columns, the second to an expansion by n rows. It is simple to see that this method is a generalization of the method of expansion by cofactors. The usefulness of the Laplace expansion lies chiefly in the evaluation of determinants of partitioned matrices, a fact that will become apparent in later discussion.

In dealing with determinants, it is useful to establish rules for evaluating such quantities for sums or products of matrices. We have

Proposition 2.25. Let A, B be two square matrices of order m. Then, in general,

$$|A + B| \neq |A| + |B|,$$

in the sense that any of the following three relations is possible:

$$|A+B| = |A| + |B|,$$
$$|A+B| > |A| + |B|,$$
$$|A+B| < |A| + |B|.$$

Proof: We establish the validity of the proposition by a number of examples. Thus, for

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

we have $|A| = 6$, $|B| = -\frac{2}{3}$, $|A+B| = 5\frac{1}{3}$, and we see that

$$|A+B| = |A| + |B|.$$

For the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

we find $|A| = 1$, $|B| = 1$, $|A+B| = 4$. Thus, $|A+B| > |A| + |B|$.

For the matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix},$$

we find $|A| = 1$, $|B| = 1$, $|A+B| = 0$. Thus, $|A+B| < |A| + |B|$.

q.e.d.

Proposition 2.26. Let A, B be two square matrices of order m. Then,

$$|AB| = |A|\,|B|.$$

Proof: Define the $2m \times 2m$ matrix

$$C = \begin{bmatrix} A & 0 \\ -I & B \end{bmatrix}.$$

Multiply the last m rows of C on the left by A (i.e. take m linear combinations of such rows) and add them to the first rows. The resulting matrix is

$$C^* = \begin{bmatrix} 0 & AB \\ -I & B \end{bmatrix}.$$

By Proposition 2.22 and Remark 2.17, we have

$$|C| = |C^*|. \tag{2.16}$$

Expand $|C^*|$ by the method of Proposition 2.24 and note that, using m-order minors involving the last m rows, their associated complementary cofactors will vanish (since they involve the determinant of a matrix containing a zero column), except for the one corresponding to $-I$. The complementary cofactor for that minor is

$$[(-1)^{\sum_{i=1}^{m}(i+m+i)}]|AB| = (-1)^{m^2+m^2+m}|AB|.$$

Moreover,

$$|-I| = (-1)^m.$$

Hence,

$$|C^*| = (-1)^{2m^2+2m}|AB| = |AB|. \tag{2.17}$$

Similarly, expand $|C|$ by the same method using m-order minors involving the first m rows. Notice now that all m-order minors involving the first m rows of C have a zero determinant except for the one corresponding to A, whose determinant is, evidently, $|A|$. Its associated complementary cofactor is

$$[(-1)^{\sum_{i=1}^{m}(i+i)}]|B| = (-1)^{2[m(m+1)/2]}|B| = |B|.$$

Hence, we have $|C| = |A|\,|B|$. But this result, together with Eqs. (2.16) and (2.17), implies $|AB| = |A|\,|B|$.

q.e.d.

Corollary 2.2. Let A be an invertible matrix of order m. Then,

$$|A^{-1}| = \frac{1}{|A|}.$$

Proof: By definition, $|AA^{-1}| = |I| = 1$—the last equality following immediately from the fundamental definition of the determinant. Since $|AA^{-1}| = |A|\,|A^{-1}|$, we have $|A^{-1}| = |A|^{-1}$.

q.e.d.

We conclude this section by introducing

Definition 2.24. Let A be a square matrix of order m. Let A_{ij} be the cofactor of the (i,j) element of A, a_{ij}, and define

$$B = (A_{ij}), \qquad i,j = 1,2,\ldots,m.$$

The **adjoint** of A, denoted by adj A, is defined by

$$\text{adj } A = B'.$$

2.6 Computation of the Inverse

We defined earlier the inverse of a matrix, say A, to be another matrix, say B, having the properties $AB = BA = I$.

Although this describes the essential property of the inverse, it does not provide a useful way to determine the elements of B. In the preceding section, we have laid the foundation for providing a practicable method for determining the elements of the inverse of a given matrix. To this end we have the following proposition.

Proposition 2.27. Let A be an invertible square matrix of order m. Its inverse, denoted by A^{-1}, is given by

$$A^{-1} = \frac{\text{adj } A}{|A|}.$$

Proof: In the standard notation for the inverse, denote the (i, j) element of A^{-1} by a^{ij}; then, the proposition asserts that

$$a^{ij} = \frac{A_{ji}}{|A|},$$

where A_{ji} is the cofactor of the element in the j th row and i th column of A. Let us now verify the validity of the assertion by determining the typical element of AA^{-1}. It is given by

$$\sum_{k=1}^{m} a_{ik} a^{kj} = \frac{1}{|A|} \sum_{k=1}^{m} a_{ik} A_{jk}. \tag{2.18}$$

Now for $i = j$ we have the expansion by cofactors along the i th row of A. Hence all diagonal elements of AA^{-1} are unity. For $i \neq j$, we may evaluate the quantity in Eq. (2.18) as follows. Strike out the j th row of A and replace it by the i th row. The resulting matrix has two identical rows and as such its determinant is zero. Now, expand by cofactors along the j th row. The cofactors of such elements are plainly A_{jk} because the other rows of A have not been disturbed. Thus, expanding by cofactors along the j th row we conclude $\sum_{k=1}^{m} a_{ik} A_{jk} = 0$. This is so because above we have a representation of the determinant of a matrix with two identical rows. Thus, $AA^{-1} = I$. Similarly, consider the typical element of $A^{-1}A$, namely

$$\sum_{k=1}^{m} a^{ik} a_{kj} = \frac{1}{|A|} \sum_{k=1}^{m} a_{kj} A_{ki}. \tag{2.19}$$

Again for $j = i$ we have, in the right-hand summation, the determinant of A evaluated by an expansion along the i th column. Hence, all diagonal elements of $A^{-1}A$ are unity. For $i \neq j$, consider the matrix obtained when we strike out the ith column of A and replace it by its jth column. The resulting matrix has two identical columns and hence its determinant is zero. Evaluating its determinant by expansion along the ith column, we note that the cofactors are given by A_{ki} because the other columns of A have not been disturbed. But then we have

$$\sum_{k=1}^{m} a_{kj}A_{ki} = 0, \qquad i \neq j,$$

and thus we conclude that $A^{-1}A = I_m$.

q.e.d.

Proposition 2.28. Let A, B be two invertible matrices of order m. Then, $(AB)^{-1} = B^{-1}A^{-1}$.

Proof: We verify

$$\begin{aligned}(AB)(B^{-1}A^{-1}) &= A(BB^{-1})A^{-1} = AA^{-1} = I_m,\\ (B^{-1}A^{-1})(AB) &= B^{-1}A^{-1}AB = B^{-1}B = I_m.\end{aligned}$$

q.e.d.

Remark 2.21. For any two conformable and invertible matrices, A, B, we have

$$(A+B)^{-1} \neq A^{-1} + B^{-1},$$

and, indeed, $A+B$ need not be invertible. For example, suppose $B = -A$. Then, even though A^{-1}, B^{-1} exist, $A+B = 0$, which is evidently not invertible since its determinant is zero.

2.7 Partitioned Matrices

Frequently, we find it convenient to deal with partitioned matrices. In this section, we derive certain useful results that will facilitate operations with such matrices. Let A be $m \times n$ and write

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

where A_{11} is $m_1 \times n_1$, A_{22} is $m_2 \times n_2$, A_{12} is $m_1 \times n_2$, and A_{21} is $m_2 \times n_1$, where $(m_1 + m_2 = m$, and $n_1 + n_2 = n)$. The above is said to be a **partition** of the matrix A.

Now, let B be $m \times n$, and partition it conformably with A, i.e. put

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

where B_{11} is $m_1 \times n_1$, B_{22} is $m_2 \times n_2$, and so on.

Addition of (conformably) partitioned matrices is defined by

$$A + B = \begin{bmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{bmatrix}.$$

If A is $m \times n$, C is $n \times q$, and A is partitioned as above, let

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix},$$

where C_{11} is $n_1 \times q_1$, C_{22} is $n_2 \times q_2$, and so on.

Multiplication of two (conformably) partitioned matrices is defined by

$$AC = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$$

$$= \begin{bmatrix} A_{11}C_{11} + A_{12}C_{21} & A_{11}C_{12} + A_{12}C_{22} \\ A_{21}C_{11} + A_{22}C_{21} & A_{21}C_{12} + A_{22}C_{22} \end{bmatrix}.$$

In general, and for either matrix addition or matrix multiplication, readers will not commit an error if, upon (conformably) partitioning two matrices, they proceed to regard the partition blocks as ordinary (scalar) elements and apply the usual rules except for division. Thus, for example, consider

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1s} \\ A_{21} & A_{22} & \cdots & A_{2s} \\ \vdots & \vdots & & \vdots \\ A_{s1} & A_{s2} & \cdots & A_{ss} \end{bmatrix},$$

where A_{ij} is $m_i \times n_j$, the matrix A is $m \times n$, and

$$\sum_{i=1}^{s} m_i = m, \qquad \sum_{j=1}^{s} n_j = n.$$

Similarly, consider

$$B = \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1s} \\ B_{21} & B_{22} & \cdots & B_{2s} \\ \vdots & \vdots & & \vdots \\ B_{s1} & B_{s2} & \cdots & B_{ss} \end{bmatrix},$$

where B_{ij} is $m_i \times n_j$ as above. Then A and B are conformably partitioned with respect to matrix addition and their sum is simply

$$A + B = \begin{bmatrix} A_{11} + B_{11} & \cdots & A_{1s} + B_{1s} \\ A_{21} + B_{21} & \cdots & A_{2s} + B_{2s} \\ \vdots & & \vdots \\ A_{s1} + B_{s1} & \cdots & A_{ss} + B_{ss} \end{bmatrix},$$

If, instead of being $m \times n$, B is $n \times q$ and its partition blocks B_{ij} are $n_i \times q_j$ matrices such that

$$\sum_{i=1}^{s} n_i = n, \qquad \sum_{j=1}^{s} q_j = q,$$

then A and B are conformably partitioned with respect to multiplication, and their product is given by

$$AB = \begin{bmatrix} \sum_{r=1}^{s} A_{1r}B_{r1} & \cdots & \sum_{r=1}^{s} A_{1r}B_{rs} \\ \sum_{r=1}^{s} A_{2r}B_{r1} & \cdots & \sum_{r=1}^{s} A_{2r}B_{rs} \\ \vdots & & \vdots \\ \sum_{r=1}^{s} A_{sr}B_{r1} & \cdots & \sum_{r=1}^{s} A_{sr}B_{rs} \end{bmatrix},$$

where the (i, j) block of AB, namely

$$\sum_{r=1}^{s} A_{ir}B_{rj},$$

is a matrix of dimension $m_i \times q_j$.

For inverses and determinants of partitioned matrices, we may prove certain useful results.

Proposition 2.29. Let A be a square matrix of order m. Partition

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and let A_{ij} be $m_i \times m_j$, $i, j = 1, 2$, $m_1 + m_2 = m$. Also, let

$$A_{21} = 0.$$

Then

$$|A| = |A_{11}|\,|A_{22}|.$$

Proof: This follows immediately from the Laplace expansion by noting that if we expand along the last m_2 rows the only $m_2 \times m_2$ minor with non-vanishing determinant is A_{22}. Its complementary cofactor is

$$[(-1)^{\sum_{i=m_1+1}^{m_1+m_2}(i+i)}]|A_{11}| = |A_{11}|.$$

Consequently,

$$|A| = |A_{11}|\,|A_{22}|.$$

q.e.d.

Corollary 2.3. If, instead, we had assumed

$$A_{12} = 0,$$

then

$$|A| = |A_{11}|\,|A_{22}|.$$

Proof: Obvious from the preceding.

Definition 2.25. A matrix of the form

$$A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

(as in Proposition 2.29) is said to be an **upper block triangular** matrix. A matrix of the form

$$A = \begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix}$$

is said to be a **lower block triangular** matrix.

Definition 2.26. Let A be as in Proposition 2.29, but suppose

$$A_{12} = 0, \qquad A_{21} = 0,$$

i.e. A is of the form

$$A = \begin{bmatrix} A_{11} & 0 \\ 0 & A_{22} \end{bmatrix}.$$

Then, A is said to be a **block diagonal** matrix and is denoted by

$$A = \mathrm{diag}(A_{11}, A_{22}).$$

Corollary 2.4. Let A be a block diagonal matrix as above. Then,

$$|A| = |A_{11}|\,|A_{22}|.$$

Proof: Obvious.

Remark 2.22. Note that, in the definition of block triangular (or block diagonal) matrices, the blocks A_{11}, A_{22} need not be triangular (or diagonal) matrices.

Proposition 2.30. Let A be a partitioned square matrix of order m,

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

where the A_{ii} are nonsingular square matrices of order m_i, $i = 1, 2, ,$ $m_1 + m_2 = m$. Then,

$$|A| = |A_{22}|\,|A_{11} - A_{12}A_{22}^{-1}A_{21}|, \quad \text{and} \quad |A| = |A_{11}|\,|A_{22} - A_{21}A_{11}^{-1}A_{12}|.$$

Proof: Consider the matrix

$$A_* = \begin{bmatrix} I_{m_1} & -A_{12}A_{22}^{-1} \\ 0 & I_{m_2} \end{bmatrix} A = \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ A_{21} & A_{22} \end{bmatrix}.$$

By Proposition 2.29 and Corollary 2.3,

$$\det \begin{bmatrix} I_{m_1} & -A_{12}A_{22}^{-1} \\ 0 & I_{m_2} \end{bmatrix} = 1;$$

thus, we conclude $|A_*| = |A|$. Again by Proposition 2.29, the determinant of A_* may be evaluated as

$$|A_*| = |A_{22}|\,|A_{11} - A_{12}A_{22}^{-1}A_{21}|.$$

Hence, we conclude

$$|A| = |A_{22}|\,|A_{11} - A_{12}A_{22}^{-1}A_{21}|.$$

Similarly, consider

$$A^* = \begin{bmatrix} I_{m_1} & 0 \\ -A_{21}A_{11}^{-1} & I_{m_2} \end{bmatrix} A = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} - A_{21}A_{11}^{-1}A_{12} \end{bmatrix}$$

and thus conclude

$$|A| = |A_{11}|\,|A_{22} - A_{21}A_{11}^{-1}A_{12}|.$$

q.e.d.

Next, we turn to the determination of inverses of partitioned matrices. We have

Proposition 2.31. Let A be a square nonsingular matrix of order m, and partition

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

such that the A_{ii}, $i = 1, 2$, are nonsingular matrices of order m_i, $i = 1, 2$, respectively $(m_1 + m_2 = m)$. Then,

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

where

$$B_{11} = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \quad B_{12} = -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1},$$

$$B_{21} = -A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}, \quad B_{22} = (A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}.$$

Proof: By definition of the inverse B, we have

$$AB = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} & A_{11}B_{12} + A_{12}B_{22} \\ A_{21}B_{11} + A_{22}B_{21} & A_{21}B_{12} + A_{22}B_{22} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix},$$

which implies

$$\begin{aligned} A_{11}B_{11} + A_{12}B_{21} &= I_{m_1}, & A_{11}B_{12} + A_{12}B_{22} &= 0, \\ A_{21}B_{11} + A_{22}B_{21} &= 0, & A_{21}B_{12} + A_{22}B_{22} &= I_{m_2}. \end{aligned}$$

Solving these equations by substitution, we have the proposition.

q.e.d.

The result above may be utilized to obtain the inverse of certain types of matrices that occur frequently in econometrics.

Proposition 2.32. Let A be $m \times n$, B be $n \times m$, and suppose $I_m + AB$, $I_n + BA$ are nonsingular matrices. Then,

$$(I_m + AB)^{-1} = I_m - A(I_n + BA)^{-1}B.$$

Proof: Observe that

$$\begin{bmatrix} I_n & -B \\ 0 & I_m \end{bmatrix} \begin{bmatrix} I_n & B \\ -A & I_m \end{bmatrix} = \begin{bmatrix} I_n + BA & 0 \\ -A & I_m \end{bmatrix}.$$

Consequently,

$$\begin{bmatrix} (I_n + BA)^{-1} & 0 \\ A(I_n + BA)^{-1} & I_m \end{bmatrix} = \begin{bmatrix} (I_n + BA)^{-1} & -(I_n + BA)^{-1}B \\ A(I_n + BA)^{-1} & (I_m + AB)^{-1} \end{bmatrix} \begin{bmatrix} I_n & B \\ 0 & I_m \end{bmatrix},$$

which implies, in particular, $(I_m + AB)^{-1} + A(I_n + BA)^{-1}B = I_m$, which further implies $(I_m + AB)^{-1} = I_m - A(I_n + BA)^{-1}B$.

q.e.d.

Corollary 2.5. Let C, D be square nonsingular matrices of order m and n, respectively; let X be $m \times n$, Y be $n \times m$, and suppose the nonsingularity conditions of Proposition 2.32 hold. Then,

$$(C + XDY)^{-1} = C^{-1} - C^{-1}X(D^{-1} + YC^{-1}X)^{-1}YC^{-1}.$$

Proof: Note that

$$(C + XDY)^{-1} = [C(I_m + C^{-1}XDY)]^{-1} = (I_m + C^{-1}XDY)^{-1}C^{-1}.$$

Let $C^{-1}X$, DY be, respectively, the matrices A, B of Proposition 2.32. Then,

$$\begin{aligned}(I_m + C^{-1}XDY)^{-1} &= I_m - C^{-1}X(I_n + DYC^{-1}X)^{-1}DY \\ &= I_m - C^{-1}X(D^{-1} + YC^{-1}X)^{-1}Y.\end{aligned}$$

Consequently,

$$\begin{aligned}(C + XDY)^{-1} &= (I_m + C^{-1}XDY)^{-1}C^{-1} \\ &= C^{-1} - C^{-1}X(D^{-1} + YC^{-1}X)^{-1}YC^{-1}.\end{aligned}$$

q.e.d.

Remark 2.23. A certain special case occurs sufficiently frequently in econometrics to deserve special notice. Precisely, let $n = 1$ so that D is now a **scalar**, say d. Let x, y be two m-element column vectors, so that with

$$n = 1, \quad X = x, \quad Y = y', \quad D = d,$$

the result of the corollary becomes

$$[C + dxy']^{-1} = C^{-1} - \alpha C^{-1}xy'C^{-1}, \quad \alpha = \frac{d}{1 + dy'C^{-1}x}.$$

The determinant of such matrices may also be expressed in relatively simple form. In order to do so we present a very useful result on determinants, which we will rederive later as a by-product of more general considerations.

Proposition 2.33. Let A be $m \times n$, and B be $n \times m$; then,

$$|I_m + AB| = |I_n + BA|.$$

Proof: Note that

$$\begin{bmatrix} I_m & A \\ -B & I_n \end{bmatrix} \begin{bmatrix} I_m & 0 \\ B & I_n \end{bmatrix} = \begin{bmatrix} I_m + AB & A \\ 0 & I_n \end{bmatrix},$$

$$\begin{bmatrix} I_m & 0 \\ B & I_n \end{bmatrix} \begin{bmatrix} I_m & A \\ -B & I_n \end{bmatrix} = \begin{bmatrix} I_m & A \\ 0 & I_n + BA \end{bmatrix}.$$

Using Proposition 2.29, and the results above, we conclude

$$|I_n + BA| = \det \begin{bmatrix} I_m & A \\ -B & I_n \end{bmatrix} = |I_m + AB|.$$

q.e.d.

We immediately have

Corollary 2.6. Let C, D, X, and Y be as in Corollary 2.5. Then,

$$|C + XDY| = |C|\,|D|\,|D^{-1} + YC^{-1}X|.$$

Proof: Since $C + XDY = C(I_m + C^{-1}XDY)$, we find, by Proposition 2.26, $|C + XDY| = |C|\,|I_m + C^{-1}XDY|$. By Proposition 2.33,

$$|I_m + C^{-1}XDY| = |I_m + DYC^{-1}X| = |D|\,|D^{-1} + YC^{-1}X|$$

and consequently $|C + XDY| = |C|\,|D|\,|D^{-1} + YC^{-1}X|$.

q.e.d.

Remark 2.24. Again, the special case when $n = 1$ and thus D is a **scalar**, say, d, deserves special mention. Thus, let x, y be m-element column vectors and

$$D = d, \quad X = x, \quad Y = y'.$$

The result of Corollary 2.6, for the special case $n = 1$, is rendered as

$$|C + dxy'| = |C|(1 + dy'C^{-1}x).$$

Remark 2.25. In view of Proposition 2.33, we see that in the statement of Proposition 2.32, it is not necessary to explicitly assume that both

$$I_m + AB \quad \text{and} \quad I_n + BA$$

are nonsingular, because if one is, the other must be as well.

In the discussion above we examined the question of obtaining the inverse of a partitioned matrix

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

where A_{11} and A_{22} were assumed to be nonsingular.

In many instances, the situation arises that such matrices are singular, while A is nonsingular. In this context Proposition 2.31 is inapplicable; instead, the following proposition is applicable.

Proposition 2.31a. Let A be a nonsingular matrix, and partition it as

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Let A_{11}, A_{22} be singular, and suppose the inverses

$$\begin{aligned} V_{11} &= (A_{11} + A_{12}A_{21})^{-1}, \\ V_{22} &= (A_{22} - A_{21}V_{11}A_{12} - A_{21}V_{11}A_{12}A_{22})^{-1} \end{aligned}$$

exist. Then,

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

where

$$B_{11} = V_{11} + V_{11}A_{12}(I + A_{22})V_{22}A_{21}V_{11},$$

$$B_{12} = V_{11}A_{12} + V_{11}A_{12}(I + A_{22})V_{22}A_{21}V_{11}A_{12} - V_{11}A_{12}(I + A_{22})V_{22}$$

$$B_{21} = -V_{22}A_{21}V_{11},$$

$$B_{22} = V_{22} - V_{22}A_{21}V_{11}A_{12}.$$

Proof: Obvious by direct verification.

Remark 2.26. One arrives at the result of Proposition 2.31a by multiplying A, on the left, first by the matrix

$$C_1 = \begin{bmatrix} I & A_{12} \\ 0 & I \end{bmatrix}$$

and then, also on the left, by

$$C_2 = \begin{bmatrix} I & 0 \\ -A_{21}V_{11} & I \end{bmatrix}$$

to obtain the block triangular matrix

$$\begin{bmatrix} V_{11}^{-1} & A_{12}(I + A_{22}) \\ 0 & V_{22}^{-1} \end{bmatrix}.$$

Inverting this matrix and multiplying on the **right** by C_2C_1 gives the desired result.

Some simplification of the representation of the inverse may be possible, but it hardly seems worthwhile without additional assumptions on the nature of the matrix A.

In the case of symmetric matrices, we have the following very useful result.

Corollary 2.7. Let A be as in Proposition 2.31a, and suppose, in addition, that:

i. A is symmetric;

ii. The matrices $(I + A_{22})$, $I - A_{21}V_{11}A_{12}$ are nonsingular.

Then,

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

where

$$B_{11} = V_{11} + V_{11}A_{12}(I + A_{22})V_{22}A_{21}V_{11},$$

$$B_{12} = -V_{11}A_{12}V_{22}',$$

$$B_{21} = -V_{22}A_{21}V_{11},$$

$$B_{22} = V_{22} - V_{22}A_{21}V_{11}A_{12}.$$

Proof: We note that since A is symmetric $A_{21}' = A_{12}$. Moreover,

$$(I + A_{22})V_{22} = (V_{22}^{-1}(I + A_{22})^{-1})^{-1},$$

$$A_{22}(I + A_{22})^{-1} = I - (I + A_{22})^{-1}.$$

Hence,

$$(I + A_{22})V_{22} = (I - A_{21}V_{11}A_{12} - (I + A_{22})^{-1})^{-1},$$

which shows first that

$$(I + A_{22})V_{22} = V_{22}'(I + A_{22})$$

i.e. that the matrix above is symmetric, and second, since V_{11} is evidently a symmetric matrix, that B_{11} is also symmetric, as required.

We note further the identity

$$\begin{aligned}(I - A_{21}V_{11}A_{12})V_{22}' &= (V_{22}'^{-1}(I - A_{21}V_{11}A_{12})^{-1})^{-1} \\ &= (I + A_{22} - (I - A_{21}V_{11}A_{12})^{-1})^{-1},\end{aligned}$$

which shows that the matrix

$$B_{22} = V_{22} - V_{22}A_{21}V_{11}A_{12}$$

is symmetric, as is also required. To complete the proof, we need only show that, in this case, the representation of B_{12} in Proposition 2.31a reduces to that given in this corollary.

Regrouping the elements in B_{12} of Proposition 2.31a, we find

$$B_{12} = V_{11}A_{12} - V_{11}A_{12}(I + A_{22})V_{22}(I - A_{21}V_{11}A_{12})$$

and, using the identity above, we conclude

$$(I + A_{22})V_{22}(I - A_{21}V_{11}A_{12}) = (I - (I - A_{21}V_{11}A_{12})^{-1}(I + A_{22})^{-1})^{-1}.$$

Finally, using the identity in Corollary 2.5, with

$$C = I, \quad D = I, \quad X = -(I - A_{21}V_{11}A_{12})^{-1}, \quad Y = (I + A_{22})^{-1},$$

we conclude that

$$(I + A_{22})V_{22}(I - A_{21}V_{11}A_{12}) = I + V_{22}'.$$

Consequently,

$$B_{12} = -V_{11}A_{12}V_{22}'.$$

q.e.d.

Remark 2.27. In some econometric problems, namely those involving consumer expenditure systems, we need to minimize a certain function subject to a set of constraints, involving a matrix of restrictions, R, which is of full **row** rank.

In order to solve the system, and thus obtain estimators for the parameters and the relevant Lagrange multipliers, we would need to invert the matrix

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

with $A_{12} = R'$, $A_{21} = R$, $A_{22} = 0$ and singular (positive semidefinite) A_{11}.

Since R is of full row rank, and A is nonsingular, it follows that the matrices

$$V_{11}^{-1} = A_{11} + R'R, \qquad RV_{11}R'$$

are nonsingular. Moreover, in this case

$$I + A_{22} = I, \qquad V_{22} = -(RV_{11}R')^{-1}$$

and, consequently, we have that the required inverse is

$$\begin{bmatrix} V_{11} - V_{11}R'(RV_{11}R')^{-1}RV_{11} & V_{11}R'(RV_{11}R')^{-1} \\ \\ (RV_{11}R')^{-1}RV_{11} & I - (RV_{11}R')^{-1} \end{bmatrix}.$$

2.8 Kronecker Products of Matrices

Definition 2.27. Let A be $m \times n$, and B be $p \times q$. The **Kronecker product** of the two matrices, denoted by

$$A \otimes B,$$

is defined by

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \cdots & a_{1n}B \\ a_{21}B & a_{22}B & \cdots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \cdots & a_{mm}B \end{bmatrix}.$$

Often, it is written more compactly as

$$A \otimes B = [a_{ij}B]$$

and is a matrix of dimension $(mp) \times (nq)$.

One operates with Kronecker products as follows.

Matrix addition. Let A_1, A_2 be matrices of dimension $m \times n$ and B_1, B_2, be matrices of dimension $p \times q$, and put

$$D_i = (A_i \otimes B_1), \qquad i = 1, 2.$$

Then,

$$D_1 + D_2 = (A_1 + A_2) \otimes B_1.$$

Similarly, if

$$E_i = (A_1 \otimes B_i), \qquad i = 1, 2$$

then

$$E_1 + E_2 = A_1 \otimes (B_1 + B_2).$$

Scalar multiplication. Let

$$C_i = (A_i \otimes B_i), \qquad i = 1, 2,$$

with the A_i, B_i as previously defined, and let α be a scalar. Then,

$$\alpha C_i = (\alpha A_i \otimes B_i) = (A_i \otimes \alpha B_i).$$

Matrix multiplication. Let C_i, $i = 1, 2$, be two Kronecker product matrices,

$$C_i = A_i \otimes B_i, \qquad i = 1, 2,$$

and suppose A_1 is $m \times n$, A_2 is $n \times r$, B_1 is $p \times q$, and B_2 is $q \times s$. Then,

$$C_1 C_2 = A_1 A_2 \otimes B_1 B_2.$$

Matrix inversion. Let C be a Kronecker product

$$C = A \otimes B$$

and suppose A, B are invertible matrices of order m and n, respectively. Then,

$$C^{-1} = A^{-1} \otimes B^{-1}.$$

All of the above can be verified directly either from the rules for operating with partitioned matrices or from other appropriate definitions.

We have

Proposition 2.34. Let A, B be square matrices of orders m, and n, respectively. Then,

$$\mathrm{tr}(A \otimes B) = \mathrm{tr}(A)\mathrm{tr}(B).$$

Proof: Since $A \otimes B$ is a **block** matrix, its trace is the sum of the traces of the diagonal blocks. Hence,

$$\mathrm{tr}(A \otimes B) = \sum_{i=1}^{m} (\mathrm{tr} a_{ii} B) = \left(\sum_{i=1}^{m} a_{ii} \right) \mathrm{tr}(B) = \mathrm{tr}(A)\mathrm{tr}(B).$$

q.e.d.

Proposition 2.35. Let A, and B be nonsingular matrices of orders m, and n, respectively. Then,

$$|A \otimes B| = |A|^n |B|^m.$$

Proof: Denote by A_{i+1} the matrix obtained when we suppress the first i rows and columns of A. Similarly, denote by $a^i_{i\cdot}$ the ith row of A after its first i elements are suppressed, and by $a^{(i)}_{\cdot i}$ the ith column of A after its first i elements have been suppressed.

Now, partition

$$A = \begin{bmatrix} a_{11} & a_{1\cdot}^{(1)} \\ a_{\cdot 1}^{(1)} & A_2 \end{bmatrix}$$

and write the Kronecker product in the partitioned form

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{1\cdot} \otimes B \\ a_{\cdot 1} \otimes B & A_2 \otimes B \end{bmatrix}.$$

Apply Proposition 2.30 to the partitioned matrix above to obtain

$$|A \otimes B| = |a_{11}B - (a_{1\cdot} \otimes B)(A_2 \otimes B)^{-1}(a_{\cdot 1} \otimes B)|\,|A_2 \otimes B|.$$

Because A is nonsingular, we assume that A_2 is also nonsingular,[3] as well as A_3, $A_4, \ldots,$ etc. Thus, we may evaluate

$$(a_{1\cdot} \otimes B)(A_2 \otimes B)^{-1}(a_{\cdot 1} \otimes B) = a_{1\cdot}A_2^{-1}a_{\cdot 1} \otimes B,$$

where, evidently, $a_{1\cdot}^{(1)}A_2^{-1}a_{\cdot 1}^{(1)}$ is a scalar. Consequently,

$$\begin{aligned} |a_{11}B - \left(a_{1\cdot}^{(1)} \otimes B\right)(A_2 \otimes B)^{-1}\left(a_{\cdot 1}^{(1)} \otimes B\right)| &= |\left(a_{11} - a_{1\cdot}^{(1)}A_2^{-1}a_{\cdot 1}^{(1)}\right) \otimes B| \\ &= (a_{11} - a_{1\cdot}^{(1)}A_2^{-1}a_{\cdot 1}^{(1)})^n|B|. \end{aligned}$$

Applying Proposition 2.30 to the partition of A above, we note that

$$|A| = \left(a_{11} - a_{1\cdot}^{(1)}A_2^{-1}a_{\cdot 1}^{(1)}\right)|A_2|,$$

and so we find

$$|A \otimes B| = |A|^n|A_2|^{-n}|B|\,|A_2 \otimes B|.$$

Applying the same procedure, we also find

$$|A_2 \otimes B| = |A_2|^n|A_3|^{-n}|B|\,|A_3 \otimes B|,$$

and thus

$$|A \otimes B| = |A|^n|A_3|^{-n}|B|^2|A_3 \otimes B|.$$

Continuing in this fashion $m-1$ times, we have

$$|A \otimes B| = |A|^n|A_m|^{-n}|B|^{m-1}|A_{m-1} \otimes B|.$$

But

$$A_m = a_{mm}$$

[3] This involves some loss of generality but makes a proof by elementary methods possible. The results stated in the proposition are valid without these restrictive assumptions.

and

$$|A_m \otimes B| = |a_{mm} B| = a_{mm}^n |B|.$$

Since

$$|A_m| = a_{mm},$$

we conclude

$$|A \otimes B| = |A|^n |B|^m.$$

q.e.d.

2.9 Characteristic Roots and Vectors

Definition 2.28. Let A be a square matrix of order m; let λ, x be, respectively, a scalar and an m-element non-null vector. If

$$Ax = \lambda x,$$

λ is said to be a **characteristic root** of A and x its associated **characteristic vector**.

Remark 2.28. Characteristic vectors are evidently not unique. If x is a characteristic vector and c a non-null scalar, cx is also a characteristic vector. We render characteristic vectors unique by imposing the requirement, or convention, that their length be unity, i.e. that $x^{\prime} x = 1$.

Proposition 2.36. Let A be a square matrix of order m and let Q be an invertible matrix of order m. Then,

$$B = Q^{-1} A Q$$

has the same characteristic roots as A, and if x is a characteristic vector of A then $Q^{-1}x$ is a characteristic vector of B.

Proof: Let (λ, x) be any pair of characteristic root and associated characteristic vector of A. They satisfy

$$Ax = \lambda x.$$

Pre-multiply by Q^{-1} to obtain

$$Q^{-1} A x = \lambda Q^{-1} x.$$

But we may also write

$$Q^{-1} A = Q^{-1} A Q Q^{-1}.$$

Thus, we obtain

$$Q^{-1}AQ(Q^{-1}x) = \lambda(Q^{-1}x),$$

which shows that the pair $(\lambda, Q^{-1}x)$ is a characteristic root and associated characteristic vector of B.

q.e.d.

Remark 2.29. If A and Q are as above, and

$$B = Q^{-1}AQ,$$

B and A are said to be **similar** matrices.

Moreover, it is clear that if there exists a matrix P such that

$$P^{-1}AP = D,$$

where D is a diagonal matrix, then P must be the matrix of the characteristic vectors of A, and D the matrix of its characteristic roots, provided the columns of P have unit length. This is so because the equation above implies

$$AP = PD$$

and the columns of this relation read

$$Ap_{\cdot i} = d_i p_{\cdot i}, \qquad i = 1, 2, \ldots, m,$$

thus defining the pair $(d_i, p_{\cdot i})$ as a characteristic root and its associated characteristic vector.

We now investigate the conditions under which a matrix A is similar to a diagonal matrix.

Proposition 2.37. Let A be a square matrix of order m, and suppose

$$r_i, \qquad i = 1, 2, \ldots, n,\ n \leq m,$$

are the **distinct** characteristic roots of A. If

$$\{x_{\cdot i} : i = 1, 2, \ldots, n\}$$

is the set of associated characteristic vectors, then it is a linearly independent set.

Proof: Put

$$X = (x_{\cdot 1}, x_{\cdot 2}, \ldots, x_{\cdot n})$$

and note that X is $m \times n$, and $n \leq m$. Suppose the columns of X are not linearly independent. Then, there exists a non-null vector

$$b = (b_1, b_2, \ldots, b_n)'$$

such that

$$Xb = 0. \tag{2.20}$$

Let

$$R = \operatorname{diag}(r_1, r_2, \ldots, r_n)$$

and note that, since

$$AX = XR,$$

multiplying Eq. (2.20) (on the left) by A we have

$$0 = AXb = XRb.$$

Repeating this j times, we find

$$XR^j b = 0, \qquad j = 1, 2, \ldots, n-1, \tag{2.21}$$

where

$$R^j = \operatorname{diag}(r_1^j, r_2^j, \ldots, r_n^j).$$

Consider now the matrix whose j th column is $XR^j b$, with the understanding that for $j = 0$ we have Eq. (2.20). In view of Eq. (2.21) this is the null matrix. But note also that

$$0 = (Xb, XRb, \ldots, XR^{n-1}b) = XBV, \tag{2.22}$$

where

$$B = \operatorname{diag}(b_1, b_2, \ldots, b_n)$$

and V is the so-called **Vandermonde** matrix

$$V = \begin{bmatrix} 1 & r_1 & r_1^2 & \cdots & r_1^{n-1} \\ 1 & r_2 & r_2^2 & \cdots & r_2^{n+1} \\ \vdots & \vdots & \vdots & & \\ 1 & r_n & r_n^2 & \cdots & r_n^{n-1} \end{bmatrix}.$$

It may be shown (see Proposition 2.38) that if the r_i are distinct, V is nonsingular. Hence, from Eq. (2.22) we conclude

$$XB = 0.$$

But this means

$$b_i x_{\cdot i} = 0, \qquad i = 1, 2, \ldots, n.$$

Thus, unless

$$b_i = 0, \qquad i = 1, 2, \ldots, n,$$

we must have, for some i, say i_0,

$$x_{\cdot i_0} = 0.$$

This is a contradiction and shows that Eq. (2.20) cannot hold for non-null b; hence, the characteristic vectors corresponding to distinct characteristic roots are linearly independent.

q.e.d.

Proposition 2.38. Let

$$V = \begin{bmatrix} 1 & r_1 & r_1^2 & \cdots & r_1^{n-1} \\ 1 & r_2 & r_2^2 & \cdots & r_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & r_n & r_n^2 & \cdots & r_n^{n-1} \end{bmatrix}$$

and suppose the r_i, $i = 1, 2, \ldots, n$, are distinct. Then

$$|V| \neq 0.$$

Proof: Expand $|V|$ by cofactors along the first row to obtain

$$|V| = a_0 + a_1 r_1 + a_2 r_1^2 + \cdots + a_{n-1} r_1^{n-1},$$

where a_i is the cofactor of r_1^i. This shows $|V|$ to be a polynomial of degree $n-1$ in r_1; it is immediately evident that r_2, $r_3, \ldots, r_n$ are its roots since if for r_1 we substitute r_i, $i \geq 2$, we have the determinant of a matrix with two identical rows. From the fundamental theorem of algebra, we can thus write

$$|V| = a_{n-1} \prod_{j=2}^{n} (r_1 - r_j).$$

But

$$a_{n-1} = (-1)^{n+1} |V_1|,$$

where V_1 is the matrix obtained by striking out the first row and n th column of V. Hence, we can also write

$$|V| = |V_1| \prod_{j=2}^{n} (r_j - r_1).$$

But V_1 is of exactly the same form as V except that it is of dimension $n-1$ and does not contain r_1.

Applying a similar procedure to V_1, we find

$$|V_1| = |V_2| \prod_{j=3}^{n} (r_j - r_2),$$

where V_2 is evidently the matrix obtained when we strike out the first and second rows as well as columns n and $n-1$ of V. Continuing in this fashion we find

$$|V| = \sum_{i=1}^{n-1} \prod_{j_i=i+1}^{n} (r_{j_i} - r_i).$$

Since $r_{j_i} \neq r_i$ it is evident that

$$|V| \neq 0.$$

q.e.d.

An immediate consequence of Proposition 2.37 is

Proposition 2.39. Let A be a square matrix of order m, and suppose all its roots are distinct. Then, A is similar to a diagonal matrix.

Proof: Let $(\lambda_i, x_{\cdot i})$, $i = 1, 2, \ldots, m$, be the characteristic roots and associated characteristic vectors of A. Let

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m), \qquad X = (x_{\cdot 1}, x_{\cdot 2}, \ldots, x_{\cdot m}),$$

and note that the relationship

$$Ax_{\cdot i} = \lambda_i x_{\cdot i}, \qquad i = 1, 2, \ldots, m,$$

between A and its characteristic roots and vectors may be written compactly as

$$AX = X\Lambda.$$

By Proposition 2.37, X is nonsingular; hence,

$$X^{-1}AX = \Lambda.$$

q.e.d.

The usefulness of this proposition is enhanced by the following approximation result.

Proposition 2.40. Let A be a square matrix of order m. Then, there exists a square matrix of order m, say B, such that B has distinct roots and

$$\sum_{i,j=1}^{m} |a_{ij} - b_{ij}| < \varepsilon,$$

where ε is any arbitrary preassigned positive quantity however small.

Proof: The proof of this result lies entirely outside the scope of this volume. The interested reader is referred to Bellman (1960), pp. 199 ff.

In the preceding, we have established a number of properties regarding the characteristic roots and their associated characteristic vectors without explaining how such quantities may be obtained. It is thus useful to deal with these aspects of the problem before we proceed.

By the definition of characteristic roots and vectors of a square matrix A, we have

$$Ax = \lambda x,$$

or more revealingly

$$(\lambda I - A)x = 0, \tag{2.23}$$

where λ is a characteristic root and x the associated characteristic vector. We recall that, for a characteristic vector, we require

$$x \neq 0, \quad x^{'}x = 1. \tag{2.24}$$

Clearly, Eq. (2.23) together with Eq. (2.24) implies that the columns of $\lambda I - A$ are linearly dependent. Hence, we can find all λ's for which Eqs. (2.23) and (2.24) are satisfied for appropriate x's, by finding the λ's for which

$$|\lambda I - A| = 0. \tag{2.25}$$

Definition 2.29. Let A be a square matrix of order m. The relation in Eq. (2.25) regarded as an equation in λ is said to be the **characteristic equation** of the matrix A.

From the basic definition of a determinant, we easily see that Eq. (2.25) represents a polynomial of degree m in λ. This is so because in evaluating a determinant we take the sum of all possible products involving the choice of one element from each row and column. In this case, the largest power of λ occurs in the term involving the choice of the diagonal elements of $\lambda I - A$. This term is

$$(\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{mm}),$$

and we easily see that the highest power of λ occurring in the characteristic equation is λ^m, and its coefficient is unity. Moreover, collecting terms involving λ^j, $j = 0, 1, 2, \ldots, m$, we can write the characeristic equation as

$$0 = |\lambda I - A| = \lambda^m + b_{m-1}\lambda^{m-1} + b_{m-2}\lambda^{m-2} + \cdots + b_0. \tag{2.26}$$

It is also clear from this discussion that

$$b_0 = |-A| = (-1)^m |A|. \tag{2.27}$$

The fundamental theorem of algebra assures us that, over the field of complex numbers, the polynomial of degree m in Eq. (2.26) has m roots. These may be numbered, say, in order of decreasing magnitude: λ_1, λ_2, $\lambda_3, \ldots, \lambda_m$. The characteristic equation of Eq. (2.26) can also be written as

$$0 = |\lambda I - A| = \prod_{i=1}^{m} (\lambda - \lambda_i). \tag{2.28}$$

The roots of the characteristic equation of A, as exhibited Eq. (2.26) or Eq. (2.28) are said to be the **characteristic roots** of A. If λ_i is one of the characteristic roots of A, the columns of $\lambda_i I - A$ are linearly dependent; it follows, therefore, that there exists at least one non-null vector, say $x_{\cdot i}$, such that

$$(\lambda_i I - A) x_{\cdot i} = 0.$$

But this means that the pair $(\lambda_i, x_{\cdot i})$ represents a characteristic root and its associated characteristic vector, provided the latter is normalized so that its length is unity.

Thus, obtaining the characteristic roots of a matrix involves solving a polynomial equation of degree m; obtaining the characteristic vectors involves solving a system of m linear equations. An immediate consequence of the preceding discussion is

Proposition 2.41. Let A be a square matrix of order m. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be its characteristic roots. Then,

$$|A| = \prod_{i=1}^{m} \lambda_i.$$

Proof: If in Eq. (2.28) we compute the constant term of the polynomial in the right member, we find

$$\prod_{i=1}^{m} (-\lambda_i) = (-1)^m \prod_{i=1}^{m} \lambda_i.$$

From Eq. (2.27), we see that

$$b_0 = |-A| = (-1)^m |A|.$$

Because Eqs. (2.28) and (2.26) are two representations of the same polynomial, we conclude

$$|A| = \prod_{i=1}^{m} \lambda_i.$$

q.e.d.

Remark 2.30. The preceding proposition implies that if A is a singular matrix, then at least one of its roots is zero. It also makes clear the terminology **distinct** and **repeated** characteristic roots. In particular, let $s < m$ and suppose Eq. (2.28) turns out to be of the form

$$|\lambda I - A| = \prod_{j=1}^{s} (\lambda - \lambda_{(j)})^{m_j},$$

where

$$\sum_{j=1}^{s} m_j = m, \qquad \lambda_{(j)} \neq \lambda_{(i)} \quad \text{for } i \neq j.$$

Then, we say that A has s **distinct** roots, viz. the roots $\lambda_{(1)}$, $\lambda_{(2)}, \ldots, \lambda_{(s)}$, and that the root $\lambda_{(i)}$ is **repeated** m_i times, since the factor corresponding to it in the factorization of the characteristic equation is raised to the m_i power.

Remark 2.31. It may further be shown, but will not be shown here, that if A is a square matrix of order m and rank $r \leq m$ then it has r nonzero roots and $m - r$ zero roots, i.e. the zero root is repeated $m - r$ times or, alternatively, its characteristic equation is of the form

$$|\lambda I - A| = \lambda^{m-r} f(\lambda), \tag{2.29}$$

where

$$f(\lambda_i) = 0, \qquad i = 1, 2, \ldots, r,$$

and

$$\lambda_i \neq 0, \qquad i = 1, 2, \ldots, r.$$

From the method for obtaining characteristic roots, we easily deduce

Proposition 2.42. Let A be a square matrix of order m and let

$$\lambda_i, \qquad i = 1, 2, \ldots, m,$$

be its characteristic roots. Then,

i. The characteristic roots of A' are exactly those of A, and

ii. If A is nonsingular, the characteristic roots of A^{-1} are given by

$$\mu_i = \frac{1}{\lambda_i}, \qquad i = 1, 2, \ldots, m.$$

Proof: The characteristic roots of A are simply the solution of $|\lambda I - A| = 0$. The characteristic roots of A' are obtained by solving $|vI - A'| = 0$. Since $vI - A' = (vI - A)'$, Proposition 2.17 implies that the determinant of $(vI - A)'$ is exactly the same as the determinant of $vI - A$. Hence, if by v_i, $i = 1, 2, \ldots, m$, we denote the characteristic roots of A', we conclude

$$v_i = \lambda_i, \qquad i = 1, 2, \ldots, m,$$

which proves part i.

For part ii,

$$|\mu I - A^{-1}| = 0$$

is the characteristic equation for A^{-1}, and moreover

$$\mu I - A^{-1} = A^{-1}(\mu A - I) = -\mu A^{-1}\left(\frac{1}{\mu}I - A\right).$$

Thus,

$$|\mu I - A^{-1}| = (-1)^m \mu^m |A^{-1}|\,|\lambda I - A|, \qquad \lambda = \frac{1}{\mu},$$

and we see that since $\mu = 0$ is not a root,

$$|\mu I - A^{-1}| = 0$$

if and only if

$$|\lambda I - A| = 0,$$

where

$$\lambda = \frac{1}{\mu}.$$

Hence, if μ_i are the roots of A^{-1}, we must have

$$\mu_i = \frac{1}{\lambda_i}, \qquad i = 1, 2, \ldots, m.$$

q.e.d.

Another important result that may be derived by using the characteristic equation is

Proposition 2.43. Let A, B be two square matrices of order m. Then, the characteristic roots of AB are exactly the characteristic roots of BA.

Proof: The characteristic roots of AB and BA are, respectively, the solutions of

$$|\lambda I - AB| = 0, \qquad |\lambda I - BA| = 0.$$

We show that

$$|\lambda I - AB| = |\lambda I - BA|,$$

thus providing the desired result. For some square matrix C of order m, consider

$$\psi(t) = |\lambda I + tC|,$$

where t is an indeterminate. Quite clearly, $\psi(t)$ is a polynomial of degree m. As such it may be represented by a Taylor series expansion about $t = 0$. If the expansion contains $m + 1$ terms, the resulting representation will be exact. Expanding, we find

$$\psi(t) = \psi(0) + \psi'(0)t + \tfrac{1}{2}\psi''(0)t^2 + \cdots + \frac{1}{m!}\psi^{(m)}(0)t^m.$$

By the usual rules for differentiating determinants (see Sect. 4.3), we easily find that

$$\psi(0) = \lambda^m, \qquad \psi'(0) = \lambda^{m-1} \operatorname{tr} C$$

and, in general,

$$\frac{1}{j!}\psi^{(j)}(0) = \lambda^{m-j} h_j(C),$$

where $h_j(C)$ depends only on $\operatorname{tr} C$, $\operatorname{tr} C^2, \ldots, \operatorname{tr} C^j$. Evaluating ψ at $t = -1$, gives the characteristic equation for C as

$$0 = \psi(-1) = |\lambda I - C| = \lambda^m - \lambda^{m-1} \operatorname{tr} C + \lambda^{m-2} h_2(C) \\ -\lambda^{m-3} h_3(C) + \cdots + (-1)^m h_m(C).$$

Let

$$C_1 = AB, \qquad C_2 = BA$$

and note that

$$\operatorname{tr} C_1 = \operatorname{tr} C_2, \qquad \operatorname{tr} C_1^2 = \operatorname{tr} C_2^2,$$

and, in general,

$$\operatorname{tr} C_1^j = \operatorname{tr} C_2^j.$$

This is so because

$$C_1^j = (AB)(AB)\cdots(AB),$$

$$C_2^j = (BA)(BA)\cdots(BA) = \overbrace{B(AB)(AB)\cdots(AB)}^{j-1 \text{ terms}} A = BC_1^{j-1}A.$$

Thus,

$$\text{tr } C_2^j = \text{tr } BC_1^{j-1}A = \text{tr } C_1^{j-1}AB = \text{tr } C_1^j.$$

Consequently, we see that

$$h_j(C_1) = h_j(C_2)$$

and, moreover,

$$|\lambda I - AB| = |\lambda I - BA|.$$

q.e.d.

Corollary 2.8. Let A and B be, respectively, $m \times n$ and $n \times m$ matrices, where $m \leq n$. Then, the characteristic roots of BA, an $n \times n$ matrix, consist of $n - m$ zeros and the m characteristic roots of AB, an $m \times m$ matrix.

Proof: Define the matrices

$$A_* = \begin{pmatrix} A \\ 0 \end{pmatrix}, \qquad B_* = (B, 0)$$

such that A_* and B_* are $n \times n$ matrices. By Proposition 2.43, the characteristic roots of A_*B_* are exactly those of B_*A_*. But

$$A_*B_* = \begin{bmatrix} AB & 0 \\ 0 & 0 \end{bmatrix}, \qquad B_*A_* = BA.$$

Thus,

$$\lambda I - A_*B_* = \begin{bmatrix} \lambda I - AB & 0 \\ 0 & \lambda I \end{bmatrix}$$

and, consequently, $|\lambda I - BA| = |\lambda I - B_*A_*| = |\lambda I - A_*B_*| = \lambda^{n-m}|\lambda I - AB|$.

q.e.d.

Corollary 2.9. Let A be a square matrix of order m, and let λ_i, $i = 1, 2, \ldots, m$, be its characteristic roots. Then,

$$\text{tr } A = \sum_{i=1}^{m} \lambda_i.$$

Proof: From the proof of Proposition 2.43, we have that

$$|\lambda I - A| = \lambda^m - \lambda^{m-1}\,\mathrm{tr}\,A + \lambda^{m-2}h_2(A) + \cdots + (-1)^m h_m(A).$$

From the factorization of polynomials, we have

$$|\lambda I - A| = \prod_{i=1}^{m}(\lambda - \lambda_i) = \lambda^m - \lambda^{m-1}\left(\sum_{i=1}^{m}\lambda_i\right) + \cdots + (-1)^m \prod_{i=1}^{m}\lambda_i.$$

Equating the coefficients for λ^{m-1}, we find $\mathrm{tr}\,A = \sum_{i=1}^{m}\lambda_i$.

q.e.d.

Proposition 2.44. Let A be a square matrix of order m. Then A is diagonalizable, i.e. it is similar to a diagonal matrix, if and only if for each characteristic root λ of A the multiplicity of λ is equal to the nullity of $\lambda I - A$.

Proof: Suppose A is diagonalizable. Then, we can write

$$Q^{-1}AQ = \Lambda, \qquad \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m).$$

Now, suppose the **distinct** roots are $\lambda_{(i)}$, $i = 1, 2, \ldots, s$, $s \le m$. Let the multiplicity of $\lambda_{(i)}$ be m_i, where

$$\sum_{i=1}^{s} m_i = m.$$

It is clear that $\lambda_{(i)}I - \Lambda$ has m_i zeros on its diagonal and hence is of rank

$$r[\lambda_{(i)}I - \Lambda] = m - m_i = \sum_{j \ne i} m_j.$$

Since

$$\lambda_{(i)}I - A = \lambda_{(i)}I - Q\Lambda Q^{-1} = Q(\lambda_{(i)}I - \Lambda)Q^{-1},$$

it follows that

$$r(\lambda_{(i)}I - A) = r(\lambda_{(i)}I - \Lambda) = \sum_{j \ne i} m_j.$$

But $\lambda_{(i)}I - A$ is an $m \times m$ matrix, and by Proposition 2.5 its nullity obeys

$$n[\lambda_{(i)}I - A] = m - r[\lambda_{(i)}I - A] = m_i,$$

which is the multiplicity of $\lambda_{(i)}$.

Conversely, suppose that the nullity of $\lambda_{(i)}I-A$ is m_i and $\sum_{i=1}^{s} m_i = m$. Choose the basis

$$\xi_{\cdot 1}, \xi_{\cdot 2}, \ldots, \xi_{\cdot m_1}$$

for the null space of $\lambda_{(1)}I - A$,

$$\xi_{\cdot m_1+1}, \ldots, \xi_{\cdot m_1+m_2}$$

for the null space of $\lambda_{(2)}I - A$, and so on until the null space of $\lambda_{(s)}I - A$.

Thus, we have m, m-element, vectors

$$\xi_{\cdot 1}, \xi_{\cdot 2}, \ldots, \xi_{\cdot m},$$

and each (appropriate) subset of m_i vectors, $i = 1, 2, \ldots, s$, is linearly independent. We claim that the entire set of m vectors is linearly independent. Suppose not. Then, we can find a set of scalars a_i, not all of which are zero, such that

$$\sum_{k=1}^{m} \xi_{\cdot k} a_k = 0.$$

We can also write the equation above as

$$\sum_{i=1}^{s} \zeta_{\cdot i} = 0, \quad \zeta_{\cdot i} = \sum_{j=m_1+\cdots+m_{i-1}+1}^{m_1+\cdots+m_i} \xi_{\cdot j} a_j, \qquad i = 1, 2, \ldots, s, \tag{2.30}$$

it being understood that $m_0 = 0$. Because of the way in which we have chosen the $\xi_{\cdot k}$, $k = 1, 2, \ldots, m$, the second equation of Eq. (2.30) implies that the $\zeta_{\cdot i}$ obey

$$(\lambda_{(i)}I - A)\zeta_{\cdot i} = 0,$$

i.e. that they are characteristic vectors of A corresponding to the distinct roots $\lambda_{(i)}$, $i = 1, 2, \ldots, s$. The first equation in Eq. (2.30) then implies that the $\zeta_{\cdot i}$ are linearly dependent. By Proposition 2.37, this is a contradiction. Hence,

$$a_k = 0, \qquad k = 1, 2, \ldots, m,$$

and the $\xi_{\cdot i}$, $i = 1, 2, \ldots, m$ are a linearly independent set. Let

$$X = (\xi_{\cdot 1}, \xi_{\cdot 2}, \ldots, \xi_{\cdot m})$$

and arrange the (distinct) roots

$$|\lambda_{(1)}| > |\lambda_{(2)}| > \cdots > |\lambda_{(s)}|.$$

Putting

$$\Lambda = \operatorname{diag}\bigl(\lambda_{(1)}I_{m_1}, \lambda_{(2)}I_{m_2}, \ldots, \lambda_{(s)}I_{m_s}\bigr),$$

we must have

$$AX = X\Lambda.$$

Because X is nonsingular, we conclude

$$X^{-1}AX = \Lambda.$$

q.e.d.

2.9.1 Kronecker Product Matrices

Although Kronecker product matrices were examined in an earlier section, and their characteristic roots and vectors may be determined by the preceding discussion, it is very useful here to explain and make explicit their connection to the corresponding entities of their constituent matrices.

Proposition 2.45. Let $D = A \otimes B$, where A is $m \times m$ and B is $n \times n$, with characteristic roots and vectors, respectively,

$$\{(\lambda_i,\ x_{\cdot i}):\ i = 1, 2, \ldots, m\},\quad \{(\mu_j, y_{\cdot j}):\ j = 1, 2, \ldots, n\}.$$

The following statements are true:

i. The characteristic roots and associated characteristic vectors of D are given by
$$\{(\nu_{ij}, z_{\cdot ij}):\ \nu_{ij} = \lambda_i \mu_j,\ z_{\cdot ij} = x_{\cdot i} \otimes y_{\cdot j}\},$$
for $i = 1, 2, \ldots, m,\ j = 1, 2, \ldots n$.

ii. $\mathrm{r}\,(D) = \mathrm{r}\,(A)\,\mathrm{r}\,(B)$.

Proof: Since by hypothesis

$$Ax_{\cdot i} = \lambda_i x_{\cdot i},\quad By_{\cdot j} = \mu_j y_{\cdot j},\quad i = 1, 2, \ldots, m,\ j = 1, 2, \ldots, n,$$

we have

$$Dz_{ij} = Ax_{\cdot i} \otimes By_{\cdot j} = \lambda_i x_{\cdot i} \otimes \mu_j y_{\cdot j} = (\lambda_i \otimes \mu_j)(x_{\cdot i} \otimes y_{\cdot j}) = \nu_{ij} z_{ij},$$

which proves part i.

To prove part ii, we shall also assume that A, B are diagonalizable[4] and let Q_1, Q_2 be nonsingular matrices of order m, n, respectively, such that

$$Q_1^{-1}AQ_1 = \Lambda,\quad Q_2^{-1}BQ_2 = M,$$

[4]This is an assumption that simplifies the proof considerably. Strictly speaking, it is not necessary. The cost it imposes on the generality of the result is, in any event, miniscule in view of Proposition 2.40.

where Λ and M are the diagonal matrices containing the characteristic roots of A and B, respectively. Consequently,

$$(Q_1^{-1} \otimes Q_2^{-1})D(Q_1 \otimes Q_2) = \Lambda \otimes M.$$

Since, evidently, $\mathrm{r}(A) = \mathrm{r}(\Lambda)$, $\mathrm{r}(B) = \mathrm{r}(M)$, i.e. they are equal to the nonzero characteristic roots of A and B, respectively and, moreover, the number of nonzero roots of D is equal to the product of the nonzero roots of A and the nonzero roots of B, we conclude that

$$\mathrm{r}(D) = \mathrm{r}(A)\mathrm{r}(B).$$

q.e.d.

2.10 Orthogonal Matrices

Although the term orthogonal was informally defined in Chap. 1, we repeat the formal definition for completeness.

Definition 2.30. Let a, b be two m-element vectors. They are said to be (mutually) **orthogonal** if and only if

$$a'b = 0. \tag{2.31}$$

They are said to be **orthonormal** if Eq. (2.31) holds and, in addition, $a'a = 1$, $b'b = 1$.

Definition 2.31. Let Q be a square matrix of order m. It is said to be **orthogonal** if and only if its columns are orthonormal.

An immediate consequence of the definition is the proposition below.

Proposition 2.46. Let Q be an orthogonal matrix of order m. Then, it is nonsingular.

Proof: It will suffice to shall show that its columns are linearly independent. Suppose there exist scalars c_i, $i = 1, 2, \ldots, m$, such that

$$\sum_{i=1}^{m} c_i q_{\cdot i} = 0, \tag{2.32}$$

the $q_{\cdot i}$ being the (orthonormal) columns of Q. Pre-multiply Eq. (2.32) by $q'_{\cdot j}$, $j = 1, 2, 3, \ldots, m$, **and note that we obtain**

$$c_j q'_{\cdot j} q_{\cdot j} = 0, \quad j = 1, 2, 3, \ldots, m.$$

But since

$$q'_{\cdot j} q_{\cdot j} = 1$$

we conclude that Eq. (2.32) implies

$$c_j = 0, \; j = 1, 2, \ldots, m.$$

q.e.d.

A further consequence is

Proposition 2.47. Let Q be an orthogonal matrix of order m. Then, $Q' = Q^{-1}$.

Proof: By the definition of an orthogonal matrix,

$$Q'Q = I_m.$$

By Proposition 2.46, its inverse exists. Multiplying on the right by Q^{-1} we find

$$Q' = Q^{-1}.$$

q.e.d.

Proposition 2.48. Let Q be an orthogonal matrix of order m. The following statements are true:

i. $|Q| = 1$ or $|Q| = -1$;

ii. If $\lambda_i, \; i = 1, 2, \ldots, m,$ are the characteristic roots of Q, $\lambda_i = \pm 1, \; i = 1, 2, \ldots, m.$

Proof: The validity of i follows immediately from $Q'Q = I_m$, which implies

$$|Q|^2 = 1, \qquad |Q| = \pm 1.$$

For ii, we note that, by Proposition 2.42, the characteristic roots of Q' are exactly those of Q, and the characteristic roots of Q^{-1} are $1/\lambda_i, \; i = 1, 2, \ldots, m,$ where the λ_i are the characteristic roots of Q. Because for an orthogonal matrix $Q' = Q^{-1}$, we conclude $\lambda_i = 1/\lambda_i$, which implies

$$\lambda_i = \pm 1.$$

q.e.d.

It is interesting that given a set of linearly independent vectors we can transform them into an orthonormal set. This procedure, known as Gram-Schmidt orthogonalization, is explained below.

Proposition 2.49 (Gram-Schmidt orthogonalization)**.** If $\xi_{\cdot i},\ \ i=1,2,\ldots,m,$ is a set of m linearly independent, m-element **column** vectors, they can be transformed into a set of orthonormal vectors.

Proof: First, we transform the $\xi_{\cdot i}$ into an orthogonal set, and then divide each resulting vector by its modulus to produce the desired orthonormal set. To this end, define

$$\begin{aligned}
y_{\cdot 1} &= \xi_{\cdot 1}\\
y_{\cdot 2} &= a_{12}\xi_{\cdot 1}+\xi_{\cdot 2}\\
y_{\cdot 3} &= a_{13}\xi_{\cdot 1}+a_{23}\xi_{\cdot 2}+\xi_{\cdot 3}\\
&\vdots\\
y_{\cdot m} &= a_{1m}\xi_{\cdot 1}+a_{2m}\xi_{\cdot 2}\cdots+a_{m-1,m}\xi_{\cdot m-1}+\xi_{\cdot m}.
\end{aligned}$$

The condition for defining the a_{ij} is that

$$y_{\cdot i}'y_{\cdot j}=0, \qquad i=1,2,\ldots,j-1. \tag{2.33}$$

But since $y_{\cdot i}$ depends only on $\xi_{\cdot 1},\xi_{\cdot 2},\ldots,\xi_{\cdot i},$ a condition equivalent to Eq. (2.33) is

$$\xi_{\cdot i}'y_{\cdot j}=0, \qquad i=1,2,\ldots,j-1.$$

To make the notation compact, put

$$X_j=(\xi_{\cdot 1},\xi_{\cdot 2},\ldots,\xi_{\cdot j-1}), \qquad a_{\cdot j}=(a_{1j},a_{2j},\ldots,a_{j-1,j})',$$

and note that the y's may be written compactly as

$$\begin{aligned}
y_{\cdot 1} &= \xi_{\cdot 1}\\
y_{\cdot j} &= X_j a_{\cdot j}+\xi_{\cdot j}, \qquad j=2,\ldots,m.
\end{aligned}$$

We wish the $y_{\cdot j}$ to satisfy the condition

$$X_j'y_{\cdot j}=X_j'X_ja_{\cdot j}+X_j'\xi_{\cdot j}=0. \tag{2.34}$$

The matrix $X_j'X_j$ is nonsingular because the columns of X_j are linearly independent.[5] Hence,

$$a_{\cdot j}=-(X_j'X_j)^{-1}X_j'\xi_{\cdot j}, \qquad j=2,3,\ldots,m, \tag{2.35}$$

[5] A simple proof of this is as follows. Suppose there exists a non-null vector c such that $X_j'X_jc=0$. But $c'X_jX_jc=0,$ implies $X_jc=0,$ which is a contradiction.

and we can define the desired orthogonal set by $y_{\cdot 1} = \xi_{\cdot 1}$
$y_{\cdot i} = \xi_{\cdot i} - X_i(X_i'X_i)^{-1}X_i'\xi_{\cdot i}, \quad i \geq 2$. Put

$$\zeta_{\cdot i} = \frac{y_{\cdot i}}{(y_{\cdot i}'y_{\cdot i})^{1/2}}, \qquad i = 1, 2, \ldots, m,$$

and note that $\zeta_{\cdot i}'\zeta_{\cdot i} = 1, \ i = 1, 2, \ldots, m$. The set

$$\{\zeta_{\cdot i}: i = 1, 2, \ldots, m\}$$

is the desired orthonormal set.

q.e.d.

A simple consequence is

Proposition 2.50. Let a be an m-element non-null **column** vector with unit length (modulus). Then, there exists an orthogonal matrix with a as the first column.

Proof: Given a, there certainly exist m-element vectors $\xi_{\cdot 2}, \ \xi_{\cdot 3}, \ldots, \xi_{\cdot m}$ such that the set

$$\{a, \xi_{\cdot 2}, \ldots, \xi_{\cdot m}\}$$

is linearly independent. The desired matrix is then obtained by applying Gram-Schmidt orthogonalization to this set.

q.e.d.

Remark 2.32. Evidently, Propositions 2.49 and 2.50 are applicable to row vectors.

2.11 Symmetric Matrices

In this section, we shall establish certain useful properties of symmetric matrices.

Proposition 2.51. Let S be a symmetric matrix of order m whose elements are real. Then, its characteristic roots are also real.

Proof: Let λ be any characteristic root of S and let z be its associated characteristic vector. Put

$$\lambda = \lambda_1 + i\lambda_2, \qquad z = x + iy,$$

so that we allow that $\lambda, \ z$ may be complex. Since they form a pair of characteristic root and its associated characteristic vector, they satisfy

$$Sz = \lambda z. \tag{2.36}$$

Pre-multiply by $\bar{z}'$, $\bar{z}$ being the complex conjugate of z. We find

$$\bar{z}'Sz = \lambda\bar{z}'z. \tag{2.37}$$

We note that since z is a characteristic vector

$$\bar{z}'z = x'x + y'y > 0.$$

In Eq. (2.36) take the complex conjugate to obtain

$$S\bar{z} = \bar{\lambda}\,\bar{z}, \tag{2.38}$$

because the elements of S are real. Pre-multiply Eq. (2.38) by z' to find

$$z'S\bar{z} = \bar{\lambda}z'\bar{z}. \tag{2.39}$$

Since $\bar{z}'Sz$ is a scalar (a 1×1 "matrix"),

$$(z'S\bar{z}) = (z'S\bar{z})' = \bar{z}'Sz$$

and, moreover, $\bar{z}'z = z'\bar{z}$. Subtracting Eq. (2.39) from Eq. (2.37), we find $0 = (\lambda - \bar{\lambda})\bar{z}'z$. Since $\bar{z}'z > 0$, we conclude $\lambda = \bar{\lambda}$. But

$$\lambda = \lambda_1 + i\lambda_2, \qquad \bar{\lambda} = \lambda_1 - i\lambda_2,$$

which implies $\lambda_2 = -\lambda_2$ or $\lambda_2 = 0$. Hence, $\lambda = \lambda_1$, and the characteristic root is real.

q.e.d.

Another important property is

Proposition 2.52. Let S be a symmetric matrix of order m. Let its distinct roots be $\lambda_{(i)}$, $i = 1, 2, \ldots, s$, $s \le m$, and let the multiplicity of $\lambda_{(i)}$ be m_i, $\sum_{i=1}^{s} m_i = m$. Then, corresponding to the root $\lambda_{(i)}$ there exist m_i **linearly independent** orthonormal characteristic vectors.

Proof: Since $\lambda_{(i)}$ is a characteristic root of S, let $q_{\cdot 1}$ be its associated characteristic vector (of unit length). By Proposition 2.49, there exist vectors

$$p_{\cdot j}^{(i)}, \qquad j = 2, 3, \ldots, m,$$

such that

$$Q_1 = \left(q_{\cdot 1}, p_{\cdot 2}^{(1)}, p_{\cdot 3}^{(1)}, \ldots, p_{\cdot m}^{(1)}\right)$$

is an orthogonal matrix. Consider

$$S_1 = Q_1'SQ_1 = \begin{bmatrix} \lambda_{(i)} & 0 \\ 0 & A_1 \end{bmatrix},$$

where A_1 is a matrix whose i, j element is

$$p_{\cdot i}^{(1)'} S p_{\cdot j}^{(1)}, \qquad i, j = 2, 3, \ldots, m.$$

But S and S_1 have exactly the same roots. Hence, if $m_i \geq 2$,

$$\begin{aligned} |\lambda I - S| = |\lambda I - S_1| &= \begin{vmatrix} \lambda - \lambda_{(i)} & 0 \\ 0 & \lambda I_{m-1} - A_1 \end{vmatrix} \\ &= (\lambda - \lambda_{(i)}) |\lambda I_{m-1} - A_1| = 0 \end{aligned}$$

implies that $\lambda_{(i)}$ is also a root of

$$|\lambda I_{m-1} - A_1| = 0.$$

Hence, the nullity of $\lambda_{(i)} I - S$ is at least two, i.e.,

$$n(\lambda_{(i)} I - S) \geq 2,$$

and we can thus find another vector, say, $q_{\cdot 2}$, satisfying

$$(\lambda_{(i)} I - S) q_{\cdot 2} = 0$$

and such that $q_{\cdot 1}, q_{\cdot 2}$ are linearly independent and of unit length, and such that the matrix

$$Q_2 = \left[q_{\cdot 1}, q_{\cdot 2}, p_{\cdot 3}^{(2)}, p_{\cdot 4}^{(2)}, \ldots, p_{\cdot m}^{(2)} \right]$$

is orthogonal.

Define

$$S_2 = Q_2' S Q_2$$

and note that S_2 has exactly the same roots as S. Note further that

$$\begin{aligned} |\lambda I - S| = |\lambda I - S_2| &= \begin{vmatrix} \lambda - \lambda_{(i)} & 0 & \\ 0 & \lambda - \lambda_{(i)} & 0 \\ & 0 & \lambda I_{m-2} - A_2 \end{vmatrix} \\ &= (\lambda - \lambda_{(i)})^2 |\lambda I_{m-2} - A_2| = 0 \end{aligned}$$

and $m_i > 2$ implies

$$|\lambda_{(i)} I_{m-2} - A_2| = 0.$$

Hence, $n(\lambda_{(i)} I - S) \geq 3$ and consequently we can choose another characteristic vector, $q_{\cdot 3}$, of unit length orthogonal to $q_{\cdot 1}, q_{\cdot 2}$ and such that

$$Q_3 = \left[q_{\cdot 1}, q_{\cdot 2}, q_{\cdot 3}, p_{\cdot 4}^{(3)}, p_{\cdot 5}^{(3)}, \ldots, p_{\cdot m}^{(3)} \right]$$

is an orthogonal matrix.

Continuing in this fashion, we can choose m_i orthonormal vectors

$$q_{\cdot 1}, q_{\cdot 2}, \ldots, q_{\cdot m_i}$$

corresponding to $\lambda_{(i)}$ whose multiplicity is m_i. It is clear that we cannot choose more than m_i such vectors since, after the choice of $q_{\cdot m_i}$, we will be dealing with

$$\begin{aligned} |\lambda I - S| = |\lambda I - S_{m_i}| &= \begin{vmatrix} (\lambda - \lambda_{(i)}) I_{m_i} & 0 \\ 0 & \lambda I_{m_i^*} - A_{m_i} \end{vmatrix} \\ &= (\lambda - \lambda_{(i)})^{m_i} |\lambda I_{m_i^*} - A_{m_i}| = 0, \end{aligned}$$

where $m_i^* = m - m_i$. It is evident that

$$|\lambda I - S| = (\lambda - \lambda_{(i)})^{m_i} |\lambda I_{m_i^*} - A_{m_i}| = 0$$

implies

$$|\lambda_{(i)} I_{m_i^*} - A_{m_i}| \neq 0, \tag{2.40}$$

for, if not, the multiplicity of $\lambda_{(i)}$ would exceed m_i. In turn Eq. (2.40) means that

$$r(\lambda_{(i)} I - S) = m - m_i$$

and thus

$$n(\lambda_{(i)} I - S) = m_i.$$

Because we have chosen m_i linearly independent characteristic vectors corresponding to $\lambda_{(i)}$, they form a basis for the null space of $\lambda_{(i)} I - S$ and thus a larger number of such vectors would form a linearly dependent set.

q.e.d.

Corollary 2.10. If S is as in Proposition 2.51, the multiplicity of the root $\lambda_{(i)}$ is equal to the nullity of

$$\lambda_{(i)} I - S.$$

Proof: Obvious from the proof of the proposition above.

q.e.d.

An important consequence of the preceding is

Proposition 2.53. Let S be a symmetric matrix of order m. Then, the characteristic vectors of S can be chosen to be an orthonormal set, i.e. there exists an orthogonal matrix Q such that

$$Q' S Q = \Lambda,$$

or equivalently S is orthogonally similar to a diagonal matrix.

Proof: Let the distinct characteristic roots of S be $\lambda_{(i)}$, $i = 1, 2, \ldots, s$, $s \leq m$, where $\lambda_{(i)}$ is of multiplicity m_i, and $\sum_{i=1}^{s} m_i = m$. By Corollary 2.10, the nullity of $\lambda_{(i)}I - S$ is equal to the multiplicity m_i of the root $\lambda_{(i)}$. By Proposition 2.52, there exist m_i orthonormal characteristic vectors corresponding to $\lambda_{(i)}$. By Proposition 2.37, characteristic vectors corresponding to distinct characteristic roots are linearly independent. Hence, the matrix

$$Q = (q_{\cdot 1}, q_{\cdot 2}, \ldots, q_{\cdot m}),$$

where the first m_1 columns are the characteristic vectors corresponding to $\lambda_{(1)}$, the next m_2 columns are those corresponding to $\lambda_{(2)}$, and so on, is an orthogonal matrix. Define

$$\Lambda = \mathrm{diag}(\lambda_{(1)} I_{m_1}, \lambda_{(2)} I_{m_2}, \ldots, \lambda_{(s)} I_{m_s})$$

and note that we have

$$SQ = Q\Lambda.$$

Consequently, $Q'SQ = \Lambda$.

q.e.d.

Proposition 2.54 (Simultaneous diagonalization)**.** Let A, B be two symmetric matrices of order m; Then, there exists an orthogonal matrix Q such that

$$Q'AQ = D_1, \qquad Q'BQ = D_2,$$

where the D_i, $i = 1, 2$, are diagonal matrices if and only if

$$AB = BA.$$

Proof: Sufficiency: This part is trivial since if such an orthogonal matrix exists,

$$\begin{aligned} Q'AQQ'BQ &= D_1 D_2, \\ Q'BQQ'AQ &= D_2 D_1. \end{aligned}$$

But the two equations above imply

$$\begin{aligned} AB &= QD_1 D_2 Q', \\ BA &= QD_2 D_1 Q', \end{aligned}$$

which shows that

$$AB = BA \tag{2.41}$$

because the diagonal matrices D_1, D_2 commute.

Necessity: Suppose Eq. (2.41) holds. Since A is symmetric, let Λ be the diagonal matrix containing its (real) characteristic roots and let Q_1 be the matrix of associated characteristic vectors. Thus,

$$Q_1' A Q_1 = \Lambda.$$

Define

$$C = Q_1' B Q_1$$

and note that

$$\begin{aligned}\Lambda C &= Q_1' A Q_1 Q_1' B Q_1\\ &= Q_1' A B Q_1\\ &= Q_1' B A Q_1 = Q_1' B Q_1 Q_1' A Q_1 = C\Lambda.\end{aligned} \tag{2.42}$$

If all the roots of A are distinct, we immediately conclude from Eq. (2.42) that

$$C = Q_1' B Q_1$$

is a diagonal matrix. Thus, taking

$$D_1 = \Lambda, \qquad D_2 = C,$$

the proof is completed. If not, let

$$\lambda_{(i)}, \qquad i = 1, 2, \ldots, s,$$

be the distinct roots of A and let $\lambda_{(i)}$ be of multiplicity m_i, where $\sum_{i=1}^{s} m_i = m$. We may write

$$\Lambda = \begin{bmatrix} \lambda_{(1)} I_{m_1} & & 0 & \cdots & 0 \\ & & & \ddots & \cdots \\ 0 & & \lambda_{(2)} I_{m_2} & & \\ \vdots & \ddots & & 0 & \\ 0 & \cdots & & 0 & \lambda_{(s)} I_{m_s} \end{bmatrix}.$$

Partition C conformably with Λ, i.e.

$$C = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1s} \\ C_{21} & C_{22} & \cdots & C_{2s} \\ \vdots & \vdots & & \vdots \\ C_{s1} & C_{s2} & \cdots & C_{ss} \end{bmatrix},$$

so that C_{ij} is a matrix of dimension $m_i \times m_j$. From Eq. (2.42), we thus conclude

$$\lambda_{(i)} C_{ij} = \lambda_{(j)} C_{ij}. \tag{2.43}$$

But for $i \neq j$ we have

$$\lambda_{(i)} \neq \lambda_{(j)},$$

and Eq. (2.43) implies

$$C_{ij} = 0, \qquad i \neq j. \tag{2.44}$$

Thus, C is the block diagonal matrix

$$C = \mathrm{diag}(C_{11}, C_{22}, \ldots, C_{ss}).$$

Clearly, the C_{ii}, $i = 1, 2, \ldots, s$, are symmetric matrices. Thus, there exist orthogonal matrices, say,

$$Q_i^*, \qquad i = 1, 2, 3, \ldots, s,$$

that diagonalize them, i.e.

$$Q_i^{*'} C_{ii} Q_i^* = D_i^*, \qquad i = 1, 2, \ldots, m,$$

the D_i^* being diagonal matrices. Define

$$Q_2 = \mathrm{diag}(Q_1^*, Q_2^*, \ldots, Q_s^*)$$

and note that Q_2 is an orthogonal matrix such that

$$D_2 = Q_2' C Q_2 = Q_2' Q_1' B Q_1 Q_2$$

with

$$D_2 = \mathrm{diag}(D_1^*, D_2^*, \ldots, D_s^*).$$

Evidently, D_2 is a diagonal matrix. Define $Q = Q_1 Q_2$ and note:

i. $Q'Q = Q_2' Q_1' Q_1 Q_2 = Q_2' Q_2 = I_m$, so that Q is an orthogonal matrix;

ii. $Q'AQ = Q_2' \Lambda Q_2 = \Lambda$, which follows from the construction of Q_2;

iii. $Q'BQ = D_2$.

Taking $D_1 = \Lambda$, we see that

$$Q'AQ = D_1, \qquad Q'BQ = D_2.$$

q.e.d.

Corollary 2.11. Let A, B be two symmetric matrices of order m such that

$$AB = 0.$$

Then, there exists an orthogonal matrix Q such that $Q'AQ = D_1$, $Q'BQ = D_2$, and, moreover, $D_1 D_2 = 0$.

Proof: Since A, B are symmetric and $AB = 0$, we see that $0 = (AB)' = B'A' = BA = AB$.

By Proposition 2.54, there exists an orthogonal matrix Q such that

$$Q'AQ = D_1, \qquad Q'BQ = D_2.$$

Moreover,

$$D_1 D_2 = Q'AQQ'BQ = Q'ABQ = 0.$$

q.e.d.

We close this section by stating an interesting result that connects the rank of a matrix to the number of the latter's nonzero characteristic roots. This result holds for all matrices, as implied by Proposition 2.14; however, the discussion and proof are greatly simplified in the case of symmetric matrices. Thus, we have

Corollary 2.12. Let S be as in Proposition 2.53; then,

$$r(S) = r \leq m$$

if and only if the number of nonzero characteristic roots of S is r.

Proof: From Proposition 2.53, there exists an orthogonal matrix Q such that

$$S = Q\Lambda Q', \quad \Lambda = Q'SQ, \quad \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m),$$

the λ_i, $i = 1, 2, \ldots, m$, being the characteristic roots of S. But the first two relations above imply

$$r(S) = r(\Lambda).$$

Now, suppose

$$r(S) = r.$$

Then, $(m - r)$ of the diagonal elements of Λ must be zero; hence, only r of the characteristic roots of S are nonzero. Conversely, if only r of the characteristic roots of S are nonzero, then

$$r(\Lambda) = r$$

and, consequently,

$$r(S) = r.$$

q.e.d.

2.12 Idempotent Matrices

We recall from Definition 2.8 that a square matrix A is said to be idempotent if and only if

$$AA = A.$$

An easy consequence of the definition is

Proposition 2.55. Let A be a square matrix of order m; suppose further that A is idempotent. Then, its characteristic roots are either zero or one.

Proof: Let λ, x be a pair consisting of a characteristic root and its associated (normalized) characteristic vector. Thus,

$$Ax = \lambda x. \tag{2.45}$$

Pre-multiplying by A, we find

$$Ax = AAx = \lambda Ax = \lambda^2 x. \tag{2.46}$$

But Eqs. (2.45) and (2.46) imply, after pre-multiplication by x',

$$\lambda = \lambda^2.$$

This condition is satisfied only by

$$\lambda = 0, \quad \text{or} \quad \lambda = 1.$$

q.e.d.

Remark 2.33. In idempotent matrices, we have a non-obvious example of a matrix with repeated roots. If A is a **symmetric** idempotent matrix it is **diagonalizable**, i.e. it is similar to a diagonal matrix. The result above is to be understood in the context of Propositions 2.37, 2.39, 2.40, and 2.44.

Example 2.4. An example of a (non-symmetric) matrix whose characteristic vectors are not linearly independent is

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad |\lambda I_2 - A| = (\lambda - 1)^2,$$

which has a repeated root, namely $\lambda = 1$, of multiplicity 2. The nullity of the matrix $\lambda I - A$ for $\lambda = 1$ is defined by the dimension of the (null) space of this matrix, i.e. by the dimension of the collection of vectors

$$\left(x : \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} x = 0 \right).$$

This (null) space is generated (spanned) by the vector $(1,0)^{'}$ and is thus of dimension 1; since the repeated root is of multiplicity 2, the characteristic vectors of A cannot be linearly **independent**. In fact, the equations for the characteristic vectors associated with the unit root (of multiplicity 2) are given by

$$x_1 + x_2 = x_1, \quad x_2 = x_2,$$

which implies that $x_2 = 0$, and x_1 is arbitrary. Thus, the characteristic vectors corresponding to the repeated root 1 are $(1,0)^{'}$ and $(c,0)^{'}$, where c is an arbitrary constant, and the matrix of characteristic vectors is **singular**. Consequently, A is non-diagonalizable.

Proposition 2.56. Let A be an idempotent matrix of order m and rank r. Then,

$$\text{tr } A = r(A).$$

Proof: From Corollary 2.9, we have

$$\text{tr } A = \sum_{i=1}^{m} \lambda_i .$$

By Proposition 2.55,

$$\lambda_i = 0 \quad \text{or} \quad \lambda_i = 1.$$

Hence,

$$\text{tr } A = \text{number of nonzero roots}$$

or

$$\text{tr } A = r(A).$$

q.e.d.

2.13 Semi-definite and Definite Matrices

Definition 2.32. Let A be a square matrix of order m and let x be an m-element vector. Then, A is said to be **positive semi-definite** if and only if for all vectors x

$$x'Ax \geq 0.$$

The matrix A is said to be **positive definite** if and only if for non-null x

$$x'Ax > 0.$$

Definition 2.33. Let A be a square matrix of order m. Then, A is said to be **negative (semi)definite** if and only if $-A$ is positive (semi)definite.

Remark 2.34. It is clear that we need only study the properties of positive (semi)definite matrices, since the properties of negative (semi)definite matrices can easily be derived from the latter.

Remark 2.35. A definite or semi-definite matrix B need not be symmetric. However, because the defining property of such matrices involves the **quadratic form** $x'Bx$, we see that if we put

$$A = \tfrac{1}{2}(B + B')$$

we have $x'Ax = x'Bx$, with A symmetric. Thus, whatever properties may be ascribed to B, by virtue of the fact that for any x, say

$$x'Bx \geq 0,$$

can also be ascribed to A. Thus, we sacrifice no generality if we always take definite or semi-definite matrices to be symmetric. In subsequent discussion, it should be understood that if we say that A is positive (semi) definite we also mean that A is symmetric as well.

Certain properties follow immediately from the definition of definite and semi-definite matrices.

Proposition 2.57. Let A be a square matrix of order m. If A is positive definite, it is also positive semi-definite. The converse, however, is not true.

Proof: The first part is obvious from the definition since if x is any m-element vector and A is positive definite, $x'Ax \geq 0$, so that A is also positive semi-definite.

That the converse is not true is established by an example. Take

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

For any vector $x = (x_1, x_2)'$, $x'Ax = (x_1 + x_2)^2 \geq 0$, so that A is positive semi-definite. For the choice of $x_1 = -x_2$, $x_2 \neq 0$, we have $x'Ax = 0$, which shows that A is not positive definite.

q.e.d.

Proposition 2.58. Let A be a square matrix of order m. Then,

i. If A is positive definite,

$$a_{ii} > 0, \qquad i = 1, 2, \ldots, m$$

ii. If A is only positive semi-definite,

$$a_{ii} \geq 0, \qquad i = 1, 2, \ldots, m.$$

Proof: Let $e_{\cdot i}$ be the m-element unit vector (all of whose elements are zero except the ith, which is unity). If A is positive definite, since $e_{\cdot i}$ is not the null vector, we must have

$$e'_{\cdot i}Ae_{\cdot i} > 0, \qquad i = 1, 2, \ldots, m.$$

But

$$e'_{\cdot i}Ae_{\cdot i} = a_{ii}, \qquad i = 1, 2, \ldots, m.$$

If A is positive semi-definite but not positive definite, then, repeating the argument above we find

$$a_{ii} = e'_{\cdot i}Ae_{\cdot i} \geq 0, \qquad i = 1, 2, \ldots, m.$$

q.e.d.

Another interesting property is the following.

Proposition 2.59 (Triangular decomposition theorem)**.** Let A be a positive definite matrix of order m. Then, there exists a lower triangular matrix T such that

$$A = TT'.$$

Proof: Let

$$T = \begin{bmatrix} t_{11} & 0 & & \cdots & 0 \\ t_{21} & t_{22} & & \ddots & \vdots \\ t_{31} & t_{32} & t_{33} & & \\ & & & \ddots & 0 \\ t_{m1} & t_{m2} & t_{m3} & \cdots & t_{mm} \end{bmatrix}.$$

Setting

$$A = TT'$$

we obtain the equations (by equating the (i, j) elements of A and TT')

$$t_{11}^2 = a_{11}, \quad t_{11}t_{21} = a_{12}, \quad t_{11}t_{31} = a_{13}, \quad \ldots, \quad t_{11}t_{m1} = a_{1m}$$
$$t_{21}t_{11} = a_{21}, \quad t_{21}^2 + t_{22}^2 = a_{22}, \quad t_{21}t_{31} + t_{22}t_{32} = a_{23}, \quad \ldots,$$
$$t_{21}t_{m1} + t_{22}t_{m2} = a_{2m}$$
$$\vdots$$
$$t_{m1}t_{11} = a_{m1}, \quad t_{m1}t_{21} + t_{m2}t_{22} = a_{m2}, \quad \ldots, \quad \sum_{i=1}^{m} t_{mi}^2 = a_{mm}.$$

In solving the equations as they are arranged, line by line, we see that we are dealing with a recursive system. From the first line, we have

$$t_{11} = \pm\sqrt{a_{11}}, \quad t_{21} = \frac{a_{12}}{t_{11}}, \quad t_{31} = \frac{a_{13}}{t_{11}}, \quad \ldots, \quad t_{m1} = \frac{a_{1m}}{t_{11}}.$$

From the second line, we have

$$t_{21} = \frac{a_{21}}{t_{11}}, \qquad t_{22} = \pm\left(\frac{a_{22}a_{11} - a_{21}^2}{a_{11}}\right)^{1/2},$$

and in general

$$t_{i2} = \frac{a_{2i} - t_{21}t_{i1}}{t_{22}}, \qquad i = 3, 4, \ldots, m.$$

Similarly, in the third line, we find

$$t_{33} = \pm\left(a_{33} - \frac{a_{31}^2}{t_{31}^2} - \frac{(a_{23} - t_{21}t_{31})^2}{t_{22}^2}\right)^{1/2},$$

$$t_{i3} = \frac{a_{3i} - t_{31}t_{i1} - t_{32}t_{i2}}{t_{33}}, \qquad i = 4, 5, \ldots, m,$$

and so on.

q.e.d.

Remark 2.36. Evidently, the lower triangular matrix above is not unique. In particular, we see that for t_{11} we have the choice

$$t_{11} = \sqrt{a_{11}} \quad \text{or} \quad t_{11} = -\sqrt{a_{11}}.$$

Similarly, for t_{22} we have the choice

$$t_{22} = \left(\frac{a_{22}a_{11} - a_{21}^2}{a_{11}}\right)^{1/2} \quad \text{or} \quad t_{22} = -\left(\frac{a_{22}a_{11} - a_{21}^2}{a_{11}}\right)^{1/2},$$

and so on. The matrix T can be rendered unique if we specify, say, that all diagonal elements must be positive.

Notice further that the same argument as in Proposition 2.59 can establish the existence of a unique **upper triangular** matrix T^* such that $A = T^*T^{*'}$.

In the literature of econometrics, the triangular decomposition of Proposition 2.59 is occasionally referred to as the **Choleski decomposition**.

The properties of characteristic roots of (semi)definite matrices are established in

Proposition 2.60. Let A be a symmetric matrix of order m and let λ_i, $i = 1, 2, \ldots, m$, be its (real) characteristic roots. If A is positive definite,

$$\lambda_i > 0, \qquad i = 1, 2, \ldots, m.$$

If it is only positive semi-definite,

$$\lambda_i \geq 0, \qquad i = 1, 2, \ldots, m.$$

Proof: Let $x_{\cdot i}$ be the normalized characteristic vector corresponding to the root λ_i of A. If A is positive definite,

$$x'_{\cdot i} A x_{\cdot i} = \lambda_i > 0, \qquad i = 1, 2, \ldots, m.$$

If A is merely positive semi-definite, we can only assert

$$x'_{\cdot i} A x_{\cdot i} = \lambda_i \geq 0, \qquad i = 1, 2, \ldots, m.$$

q.e.d.

By now, the reader should have surmised that positive definite matrices are nonsingular, and positive semi-definite matrices (which are not also positive definite) are singular matrices. This is formalized in

Proposition 2.61. Let A be a symmetric matrix of order m. If A is positive definite then

$$r(A) = m.$$

If A is merely positive semi-definite, i.e. it is not also positive definite, then

$$r(A) < m.$$

Proof: Since A is symmetric, let Λ denote the diagonal matrix of its (real) characteristic roots and Q the associated (orthogonal) matrix of characteristic vectors. We have

$$AQ = Q\Lambda.$$

By Propositions 2.51 and 2.44, Q is nonsingular so that A is similar to the diagonal matrix Λ, which is evidently of rank m. Thus r(A) $=$ m and its inverse exists and is given by

$$A^{-1} = Q\Lambda^{-1}Q',$$

which also shows that the characteristic roots of A^{-1} are the diagonal elements of Λ^{-1}, as exhibited below

$$\Lambda^{-1} = \operatorname{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \ldots, \frac{1}{\lambda_m}\right).$$

This establishes the first part of the proposition.

For the second part suppose A is **only** positive semi-definite. From Proposition 2.60, we merely know that $\lambda_i \geq 0$, $i = 1, 2, \ldots, m$. We now establish that at least one root must be zero, thus completing the proof of the proposition.

We have the representation

$$Q'AQ = \Lambda.$$

Consequently, for any vector y,

$$y'Q'AQy = \sum_{i=1}^{m} \lambda_i y_i^2.$$

Now, if x is any non-null vector by the semi-definiteness of A we have

$$0 \leq x'Ax = x'QQ'AQQ'x = x'Q\Lambda Q'x = \sum_{i=1}^{m} \lambda_i y_i^2, \tag{2.47}$$

where now we have put

$$y = Q'x.$$

Since x is non-null, y is also non-null.

If none of the λ_i is zero, Eq. (2.47) implies that for any non-null x $x'Ax > 0$, thus showing A to be positive definite. Consequently, at least one of the λ_i, $i = 1, 2, \ldots, m$, must be zero, and there must exist at least one non-null x such that $x'Ax = \sum_{i=1}^{m} \lambda_i y_i^2 = 0$. But this shows that $r(A) < m$.

q.e.d.

Remark 2.37. Roughly speaking, positive definite and semi-definite matrices correspond to positive and nonnegative numbers in the usual number system. The reader's intuitive comprehension would be aided if he thinks of them as a sort of matrix generalization of positive and nonnegative real numbers. Just as a nonnegative number can always be written as the square of some other number, the same holds *mutatis mutandis* for definite and semi-definite matrices. In fact, this conceptualization leads to the concept of the square root for such matrices.

Proposition 2.62. Let A be a symmetric matrix of order m. Then A is positive definite if and only if there exists a matrix S of dimension $n \times m$ and rank m $(n \geq m)$ such that

$$A = S'S.$$

It is positive semi-definite if and only if

$$r(S) < m.$$

Proof: Sufficiency: If A is positive (semi)definite then, as in the proof of Proposition 2.61, we have the representation

$$A = Q\Lambda Q'.$$

Taking

$$S = \Lambda^{1/2} Q',$$

we have $A = S'S$. If A is positive definite, Λ is nonsingular, and thus

$$\mathrm{r}(S) = m.$$

If A is merely positive semi-definite, $\mathrm{r}\,(\Lambda) < m$, and hence

$$\mathrm{r}(S) < m.$$

Necessity: Suppose $A = S'S$, where S is $n \times m$ $(n \geq m)$ of rank m. Let x be any non-null vector and note $x'Ax = x'S'Sx$. The right member of the equation above is a sum of squares and thus is zero if and only if

$$Sx = 0. \tag{2.48}$$

If A is positive definite, Eq. (2.48) can be satisfied only with null x. Hence the rank of S is m.

Evidently, for any x, $x'Ax = x'S'Sx \geq 0$, and if A is positive semi-definite (but not positive definite) there exists at least one non-null x such that $x'Ax = 0$; hence, for that x, $Sx = 0$ thus S is of rank less than m.

q.e.d.

An obvious consequence of the previous discussion is

Corollary 2.13. If A is a positive definite matrix, $|A| > 0$, $\mathrm{tr}(A) > 0$.

Proof: Let λ_i, $i = 1, 2, \ldots, m$, be the characteristic roots of A. Since

$$|A| = \prod_{i=1}^{m} \lambda_i, \qquad \mathrm{tr}(A) = \sum_{i=1}^{m} \lambda_i,$$

the result follows immediately from Proposition 2.60.

q.e.d.

Corollary 2.14. Let A be a positive semi-definite, but not a positive definite, matrix. Then $|A| = 0$, $\mathrm{tr}(A) \geq 0$; $\mathrm{tr}(A) = 0$ if and only if A is the null matrix.

Proof: From the representation

$$|A| = \prod_{i=1}^{m} \lambda_i$$

we conclude that $|A| = 0$ by Proposition 2.61.

For the second part, we note that

$$A = Q\Lambda Q' \tag{2.49}$$

and

$$\mathrm{tr}(A) = 0$$

if and only if

$$\mathrm{tr}(\Lambda) = 0.$$

But

$$\mathrm{tr}(\Lambda) = \sum_{i=1}^{m} \lambda_i = 0, \qquad \lambda_i \geq 0 \;\; \text{all } i$$

implies

$$\lambda_i = 0, \qquad i = 1, 2, \ldots, m.$$

If this holds, then Eq. (2.49) implies

$$A = 0.$$

Consequently, if A is not a null matrix,

$$\mathrm{tr}(A) > 0.$$

q.e.d.

Corollary 2.15 (Square root of a positive definite matrix)**.** Let A be a positive definite matrix of order m. Then, there exists a nonsingular matrix W such that

$$A = W'W.$$

Proof: Obvious from Propositions 2.59, 2.61, and 2.62. A particular choice of W may be $W = Q\Lambda^{1/2}Q'$, which is the usual expression for defining the **square root of a positive definite matrix** A .

q.e.d.

In previous discussions, when considering characteristic roots and characteristic vectors, we did so in the context of the characteristic equation

$$|\lambda I - A| = 0.$$

Often it is more convenient to broaden the definition of characteristic roots and vectors as follows.

Definition 2.34. Let A, B be two matrices of order m, where B is nonsingular. The **characteristic roots of A in the metric of B**, and their associated characteristic vectors, are connected by the relation

$$Ax = \lambda Bx,$$

where λ is a characteristic root and x is the associated (non-null) characteristic vector.

Remark 2.38. It is evident that the **characteristic roots of A in the metric of B** are found by solving the polynomial equation

$$|\lambda B - A| = 0.$$

It is also clear that this is a simple generalization of the ordinary definition of characteristic roots where the role of B is played by the identity matrix.

Definition 2.34 is quite useful in dealing with differences of positive (semi)definite matrices, and particularly in determining whether such differences are positive (semi)definite or not. This is intimately connected with the question of relative efficiency in comparing two estimators. We have

Proposition 2.63. Let B be a positive definite matrix and let A be positive (semi)definite, both of order m. Then, the characteristic roots of A in the metric of B, say λ_i, obey

$$\lambda_i > 0, \qquad i = 1, 2, \ldots, m,$$

if A is positive definite, and

$$\lambda_i \geq 0, \qquad i = 1, 2, \ldots, m,$$

if A is positive semi-definite.

Proof: Consider

$$|\lambda B - A| = 0.$$

Since B is positive definite, by Corollary 2.15 there exists a nonsingular matrix P such that

$$B = P^{'-1} P^{-1}.$$

Consequently, by Proposition 2.26,

$$0 = |\lambda B - A| = |\lambda P^{'-1} P^{-1} - A| = |\lambda I - P'AP|\,|P|^{-2}.$$

Thus, the **characteristic roots of A in the metric of B** are simply the usual characteristic roots of $P'AP$, i.e. the solution of

$$|\lambda I - P'AP| = 0.$$

If A is positive definite, $P'AP$ is also positive definite; if A is only positive semi-definite, $P'AP$ is only positive semi-definite. Hence, in the former case

$$\lambda_i > 0, \qquad i = 1, 2, \ldots, m,$$

whereas in the latter case

$$\lambda_i \geq 0, \qquad i = 1, 2, \ldots, m.$$

q.e.d.

A very useful result in this context is

Proposition 2.64 (Simultaneous decomposition)**.** Let B be a positive definite matrix and A positive (semi)definite, both of order m. Let

$$\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$$

be the diagonal matrix of the characteristic roots of A in the metric of B. Then, there exists a nonsingular matrix W such that

$$B = W'W, \qquad A = W'\Lambda W.$$

Proof: From Proposition 2.63, we have that the λ_i are also the (ordinary) characteristic roots of $P'AP$, where P is such that

$$B = P'^{-1}P^{-1}.$$

Let Q be the (orthogonal) matrix of (ordinary) characteristic vectors of $P'AP$. Thus, we have

$$P'APQ = Q\Lambda. \tag{2.50}$$

From Eq. (2.50), we easily establish

$$A = P'^{-1}Q\Lambda Q'P^{-1}.$$

Putting $W = Q'P^{-1}$, we have

$$A = W'\Lambda W, \quad B = W'W.$$

q.e.d.

From the preceding two propositions flow a number of useful results regarding differences of positive (semi)definite matrices. Thus,

Proposition 2.65. Let B be a positive definite matrix and A be positive (semi)definite. Then $B-A$ is positive (semi)definite if and only if

$$\lambda_i < 1 \qquad (\lambda_i \leq 1),$$

respectively, where the λ_i are the characteristic roots of A in the metric of B, $i=1,2,\ldots,m$.

Proof: From Proposition 2.64, there exists a nonsingular matrix W such that

$$B = W'W, \qquad A = W'\Lambda W.$$

Hence,

$$B - A = W'(I-\Lambda)W.$$

Let x be any m-element vector, and note

$$x'(B-A)x = y'(I-\Lambda)y = \sum_{i=1}^{m}(1-\lambda_i)y_i^2, \tag{2.51}$$

where $y = Wx$. If, for arbitrary non-null x, $x'(B-A)x > 0$, we must have $1-\lambda_i > 0$, or

$$\lambda_i < 1, \qquad i=1,2,\ldots,m, \tag{2.52}$$

thus concluding the if part of the proof; conversely, if Eq. (2.52) holds, then y is **non-null**, for **arbitrary non-null** x; consequently, it follows from Eq. (2.51) that $B-A$ is positive definite, thus concluding the only if part for positive definite matrices.

If, on the other hand, $B-A$ is only positive semi-definite, then for at least one index i we must have

$$\lambda_i = 1,$$

and conversely.

q.e.d.

Another useful result easily obtained from the simultaneous decomposition of matrices is given in

Proposition 2.66. Let A, B be two positive definite matrices, both of order m. If $B-A$ is positive definite, $A^{-1}-B^{-1}$ is also positive definite. If $B-A$ is positive semidefinite, so is $A^{-1}-B^{-1}$.

Proof: We may write $B = W'W$, $A = W'\Lambda W$; by Proposition 2.65, the diagonal elements of Λ (i.e., the roots of A in the metric of B) are less than unity. Hence, $B^{-1} = W^{-1}W'^{-1}$, $A^{-1} = W^{-1}\Lambda^{-1}W'^{-1}$. Thus,

$$A^{-1} - B^{-1} = W^{-1}(\Lambda^{-1} - I_m)W'^{-1}.$$

The diagonal elements of $\Lambda^{-1} - I$ are given by

$$\frac{1}{\lambda_i} - 1 > 0, \qquad i = 1, 2, \ldots, m,$$

and thus

$$A^{-1} - B^{-1}$$

is positive definite by Proposition 2.61.

If $B - A$ is only positive semi-definite, then for at least one of the roots we have $\lambda_{i_0} = 1$. Hence,

$$A^{-1} - B^{-1} = W^{-1}(\Lambda^{-1} - I_m)W^{'-1} \geq 0,$$

owing to the fact that at least one of the diagonal elements of $\Lambda^{-1} - I_m$ is zero.

q.e.d.

Finally, we have

Proposition 2.67. Let B be positive definite, and A be positive (semi)definite. If $B - A$ is positive (semi)definite then

$$|B| > |A|, \quad (|B| \geq |A|), \quad \mathrm{tr}(B) > \mathrm{tr}(A), \quad (\mathrm{tr}(B) \geq \mathrm{tr}(A)).$$

Proof: As in Proposition 2.66, we can write $B = W'W, \quad A = W'\Lambda W,$ and by Proposition 2.65 we know that the diagonal elements of Λ, viz., the λ_i, obey $\lambda_i < 1,$ for all i. Consequently, by Proposition 2.26,

$$|B| = |W|^2, \qquad |A| = |W|^2|\Lambda|.$$

Moreover, by Proposition 2.65, if $B - A$ is positive definite, $|\lambda_i| < 1,$ for all $i,$ and hence $|B| > |A|$.

Evidently, the inequality above is automatically satisfied if A itself is merely positive semi-definite. Moreover,

$$\mathrm{tr}(B) - \mathrm{tr}(A) = \mathrm{tr}(B - A) > 0.$$

On the other hand, if $B - A$ is merely positive semi-definite then we can only assert $|\Lambda| \leq 1,$ and hence we conclude that

$$|B| \geq |A|, \qquad \mathrm{tr}(B) \geq \mathrm{tr}(A).$$

q.e.d.

Corollary 2.16. In Proposition 2.67, the strict inequalities hold unless

$$B = A.$$

Proof: Since

$$\frac{|A|}{|B|} = |\Lambda| = \prod_{i=1}^{m} \lambda_i,$$

we see that $|A| = |B|$ implies $\lambda_i = 1, \quad i = 1, 2, \ldots, m$. Hence, $B = A$. Moreover, from the proof of Proposition 2.65,

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

implies

$$0 = \mathrm{tr}(B - A) = \mathrm{tr}[W'(I - \Lambda)W].$$

But this means

$$W'(I - \Lambda)W = 0,$$

which in turn implies

$$\Lambda = I_m,$$

and consequently

$$B = A.$$

q.e.d.

We now present the very useful **singular value decomposition** theorem for an arbitrary matrix A.

In Sect. 2.5 we examined the **rank factorization** theorem and showed that if A is $m \times n$ of rank $r \leq n \leq m$, then there exist matrices C_1, C_2, respectively, of dimension $m \times r$, $r \times n$ and both of rank r, such that

$$A = C_1 C_2.$$

The matrices C_1, C_2 are, of course, non-unique. The construction given in that section proceeds from first principles and essentially utilizes elementary row and column operations. Although conceptually simple and straightforward, that construction is not particularly useful for applied work. In view of the ready availability of computer software for obtaining the characteristic roots and vectors of symmetric matrices, the following result is perhaps more convenient.

Proposition 2.68 (Singular value decomposition theorem)**.** Let A be $m \times n$ of rank r, $r \leq n \leq m$. Then, there exist matrices B_1, B_2 and a diagonal matrix, D, with positive diagonal elements, such that

$$A = B_1 D B_2.$$

Proof: Consider the matrices AA', $A'A$; both are of rank r and of dimension $m \times m$, $n \times n$, respectively.

By Proposition 2.52, we have the representation

$$AA' = Q\Lambda Q',$$

where Q is the (orthogonal) matrix of characteristic vectors and Λ is the (diagonal) matrix of the corresponding characteristic roots. Similarly,

$$A'A = RMR'$$

where again R, M are, respectively, the matrices of characteristic vectors and corresponding characteristic roots of $A'A$.

By Corollary 2.11, we conclude that since AA' and $A'A$ are both of rank r, only r of their characteristic roots are positive, the remaining being zero. Hence, we can write

$$\Lambda = \begin{bmatrix} \Lambda_r & 0 \\ 0 & 0 \end{bmatrix}, \qquad M = \begin{bmatrix} M_r & 0 \\ 0 & 0 \end{bmatrix}.$$

Partition Q, R conformably with Λ and M, respectively, i.e.

$$Q = (Q_r, Q_*), \qquad R = (R_r, R_*)$$

such that Q_r is $m \times r$, R_r is $n \times r$ and correspond, respectively, to the non-zero characteristic roots of AA' and $A'A$.

Take

$$B_1 = Q_r,\ B_2 = R_r', \qquad \Lambda_r = D = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_r),$$

where λ_i, $i = 1, 2, \ldots, r$ are the positive characteristic roots of AA' and hence, by Corollary 2.8, those of $A'A$ as well.

Now, define

$$S = Q_r \Lambda_r^{1/2} R_r' = B_1 D B_2.$$

We show that $S = A$, thus completing the proof. We easily verify that

$$S'S = A'A, \qquad SS' = AA'.$$

From the first relation above we conclude that for an arbitrary orthogonal matrix, say P_1,

$$S = P_1 A$$

whereas from the second we conclude that for an arbitrary orthogonal matrix P_2 we must have

$$S = AP_2.$$

The preceding, however, implies that for arbitrary orthogonal matrices P_1, P_2 the matrix A satisfies

$$AA' = P_1 AA' P_1', \qquad A'A = P_2' A' A P_2,$$

which in turn implies that

$$P_1 = I_m, \qquad P_2 = I_n.$$

Thus,

$$A = S = Q_r \Lambda_r^{1/2} R_r' = B_1 D B_2.$$

q.e.d.

We close this chapter by formally defining the "square root" of a positive (semi)definite matrix.

Definition 2.35. Let A be a positive (semi)definite matrix, Λ the diagonal matrix of its (nonnegative) characteristic roots, and Q the matrix of its associated characteristic vectors. The **square root** of A, denoted by $A^{1/2}$, is defined by

$$A^{1/2} = Q\Lambda^{1/2}Q', \quad \Lambda = \mathrm{diag}(\lambda_1^{1/2}, \lambda_2^{1/2}, \ldots, \lambda_m^{1/2})$$

Springer
http://www.springer.com/978-1-4614-8144-7

Mathematics for Econometrics
Dhrymes, P.J.
2013, XVII, 419 p. 1 illus., Softcover
ISBN: 978-1-4614-8144-7