APPENDIX A

Matrix Theory

A.1. DEFINITION OF A MATRIX AND OPERATIONS ON MATRICES

In this appendix we summarize some of the well-known definitions and theorems of matrix algebra. A number of results that are not always contained in books on matrix algebra are proved here.

An $m \times n$ matrix $\boldsymbol{A}$ is a rectangular array of real numbers

(1)
$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix},$$

which may be abbreviated (a_{ij}), $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$. Capital boldface letters will be used to denote matrices whose elements are the corresponding lowercase letters with appropriate subscripts. The sum of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same numbers of rows and columns, respectively, is defined by

(2)
$$\boldsymbol{A} + \boldsymbol{B} = (a_{ij}) + (b_{ij}) = (a_{ij} + b_{ij}).$$

The product of a matrix by a real number λ is defined by

(3)
$$\lambda \boldsymbol{A} = \boldsymbol{A}\lambda = (\lambda a_{ij}).$$

An Introduction to Multivariate Statistical Analysis, Third Edition. By T. W. Anderson
ISBN 0-471-36091-0

These operations have the algebraic properties

$$A + B = B + A, \tag{4}$$

$$(A + B) + C = A + (B + C), \tag{5}$$

$$A + (-1)A = (0), \tag{6}$$

$$(\lambda + \mu)A = \lambda A + \mu A, \tag{7}$$

$$\lambda(A + B) = \lambda A + \lambda B, \tag{8}$$

$$\lambda(\mu A) = (\lambda\mu)A. \tag{9}$$

The matrix (0) with all elements 0 is denoted as $\mathbf{0}$. The operation $A + (-1B)$ is denoted as $A - B$.

If A has the same number of columns as B has rows, that is, $A = (a_{ij})$, $i = 1, \ldots, l$, $j = 1, \ldots, m$, $B = (b_{jk})$, $j = 1, \ldots, m$, $k = 1, \ldots, n$, then A and B can be multiplied according to the rule

$$AB = (a_{ij})(b_{jk}) = \left(\sum_{j=1}^{m} a_{ij} b_{jk} \right), \qquad i = 1, \ldots, l, \quad k = 1, \ldots, n; \tag{10}$$

that is, AB is a matrix with l rows and n columns, the element in the ith row and kth column being $\sum_{j=1}^{m} a_{ij} b_{jk}$. The matrix product has the properties

$$(AB)C = A(BC), \tag{11}$$

$$A(B + C) = AB + AC, \tag{12}$$

$$(A + B)C = AC + BC. \tag{13}$$

The relationships (11)–(13) hold provided one side is meaningful (i.e., the numbers of rows and columns are such that the operations can be performed); it follows then that the other side is also meaningful. Because of (11) we can write

$$(AB)C = A(BC) = ABC. \tag{14}$$

The product BA may be meaningless even if AB is meaningful, and even when both are meaningful they are not necessarily equal.

The *transpos*' of the $l \times m$ matrix $A = (a_{ij})$ is defined to be the $m \times l$ matrix A' which has in the jth row and ith column the element that A has in the ith row and jth column. The operation of transposition has the properties

$$(A')' = A, \tag{15}$$

$$(A+B)' = A' + B', \tag{16}$$

$$(AB)' = B'A', \tag{17}$$

again with the restriction (which is understood throughout this book) that at least one side is meaningful.

A vector x with m components can be treated as a matrix with m rows and one column. Therefore, the above operations hold for vectors.

We shall now be concerned with square matrices of the same size, which can be added and multiplied at will. The number of rows and columns will be taken to be p. A is called *symmetric* if $A = A'$. A particular matrix of considerable interest is the *identity* matrix

$$I = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix} = (\delta_{ij}), \tag{18}$$

where δ_{ij}, the Kronecker delta, is defined by

$$\begin{aligned} \delta_{ij} &= 1, & i &= j, \\ &= 0, & i &\neq j. \end{aligned} \tag{19}$$

The identity matrix satisfies

$$IA = AI = A. \tag{20}$$

We shall write the identity as I_p when we wish to emphasize that it is of order p. Associated with any square matrix A is the determinant $|A|$, defined by

$$|A| = \sum (-1)^{f(j_1, \dots, j_p)} \prod_{i=1}^{p} a_{ij_i}, \tag{21}$$

where the summation is taken over all permutations $(j_1, \dots, j_p)$ of the set of integers $(1, \dots, p)$, and $f(j_1, \dots, j_p)$ is the number of transpositions required to change $(1, \dots, p)$ into $(j_1, \dots, j_p)$. A transposition consists of interchanging two numbers, and it can be shown that, although one can transform $(1, \dots, p)$ into $(j_1, \dots, j_p)$ by transpositions in many different ways, the number of transpositions required is always even or always odd, so that $(-1)^{f(j_1,\ldots,j_p)}$ is consistently defined. Then

(22) $$|AB| = |A| \cdot |B|.$$

Also

(23) $$|A| = |A'|.$$

A *submatrix* of A is a rectangular array obtained from A by deleting rows and columns. A *minor* is the determinant of a square submatrix of A. The minor of an element a_{ij} is the determinant of the submatrix of a square matrix A obtained by deleting the ith row and jth column. The *cofactor* of a_{ij}, say A_{ij}, is $(-1)^{i+j}$ times the minor of a_{ij}. It follows from (21) that

(24) $$|A| = \sum_{j=1}^{p} a_{ij} A_{ij} = \sum_{j=1}^{p} a_{jk} A_{jk}.$$

If $|A| \neq 0$, there exists a unique matrix B such that $AB = I$. Then B is called the *inverse* of A and is denoted by A^{-1}. Let a^{hk} be the element of A^{-1} in the hth row and kth column. Then

(25) $$a^{hk} = \frac{A_{kh}}{|A|}.$$

The operation of taking the inverse satisfies

(26) $$(AC)^{-1} = C^{-1}A^{-1},$$

since

(27) $$(AC)(C^{-1}A^{-1}) = A(CC^{-1})A^{-1} = AIA^{-1} = AA^{-1} = I.$$

Also $I^{-1} = I$ and $A^{-1}A = I$. Furthermore, since the transposition of (27) gives $(A^{-1})'A' = I$, we have $(A^{-1})' = (A')^{-1}$.

A matrix whose determinant is not zero is called *nonsingular*. If $|A| \neq 0$, then the only solution to

(28) $$Az = \mathbf{0}$$

is the trivial one $z = \mathbf{0}$ [by multiplication of (28) on the left by A^{-1}]. If $|A| = 0$, there is at least one nontrivial solution (that is, $z \neq \mathbf{0}$). Thus an equivalent definition of A being nonsingular is that (28) have only the trivial solution.

A set of vectors $z_1, \ldots, z_r$ is said to be *linearly independent* if there exists no set of scalars $c_1, \ldots, c_r$, not all zero, such that $\sum_{i=1}^{r} c_i z_i = \mathbf{0}$. A $q \times p$

matrix D is said to be of *rank* r if the maximum number of linearly independent columns is r. Then every minor of order $r+1$ must be zero (from the remarks in the preceding paragraph applied to the relevant square matrix of order $r+1$), and at least one minor of order r must be nonzero. Conversely, if there is at least one minor of order r that is nonzero, there is at least one set of r columns (or rows) which is linearly independent. If all minors of order $r+1$ are zero, there cannot be any set of $r+1$ columns (or rows) that are linearly independent, for such linear independence would imply a nonzero minor of order $r+1$, but this contradicts the assumption. Thus rank r is equivalently defined by the maximum number of linearly independent rows, by the maximum number of linearly independent columns, or by the maximum order of nonzero minors.

We now consider the quadratic form

$$x'Ax = \sum_{i,j=1}^{p} a_{ij}x_i x_j, \tag{29}$$

where $x' = (x_1, \ldots, x_p)$ and $A = (a_{ij})$ is a symmetric matrix. This matrix A and the quadratic form are called *positive semidefinite* if $x'Ax \geq 0$ for all x. If $x'Ax > 0$ for all $x \neq 0$, then A and the quadratic form are called *positive definite*. In this book *positive definite* implies the matrix is symmetric.

Theorem A.1.1. *If C with p rows and columns is positive definite, and if B with p rows and q columns, $q \leq p$, is of rank q, then $B'CB$ is positive definite.*

Proof. Given a vector $y \neq 0$, let $x = By$. Since B is of rank q, $By = x \neq 0$. Then

$$\begin{aligned} y'(B'CB)y &= (By)'C(By) \\ &= x'Cx > 0. \end{aligned} \tag{30}$$

The proof is completed by observing that $B'CB$ is symmetric. As a converse, we observe that $B'CB$ is positive definite only if B is of rank q, for otherwise there exists $y \neq 0$ such that $By = 0$. ■

Corollary A.1.1. *If C is positive definite and B is nonsingular, then $B'CB$ is positive definite.*

Corollary A.1.2. *If C is positive definite, then C^{-1} is positive definite.*

Proof. C must be nonsingular; for if $Cx = 0$ for $x \neq 0$, then $x'Cx = 0$ for this x, but that is contrary to the assumption that C is positive definite. Let

B in Theorem A.1.1 be C^{-1}. Then $B'CB = (C^{-1})'CC^{-1} = (C^{-1})'$. Transposing $CC^{-1} = I$, we have $(C^{-1})'C' = (C^{-1})'C = I$. Thus $C^{-1} = (C^{-1})'$. ■

Corollary A.1.3. *The $q \times q$ matrix formed by deleting $p - q$ rows of a positive definite matrix C and the corresponding $p - q$ columns of C is positive definite.*

Proof. This follows from Theorem A.1.1 by forming B by taking the $p \times p$ identity matrix and deleting the columns corresponding to those deleted from C. ■

The *trace* of a square matrix A is defined as $\operatorname{tr} A = \sum_{i=1}^{p} a_{ii}$. The following properties are verified directly:

$$\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B, \tag{31}$$

$$\operatorname{tr} AB = \operatorname{tr} BA. \tag{32}$$

A square matrix A is said to be *diagonal* if $a_{ij} = 0$, $i \neq j$. Then $|A| = \prod_{i=1}^{p} a_{ii}$, for in (24) $|A| = a_{11}A_{11}$, and in turn A_{11} is evaluated similarly.

A square matrix A is said to be *triangular* if $a_{ij} = 0$ for $i > j$ or alternatively for $i < j$. If $a_{ij} = 0$ for $i > j$, the matrix is *upper* triangular, and, if $a_{ij} = 0$ for $i < j$, it is *lower* triangular. The product of two upper triangular matrices A, B is upper triangular, for the i, jth term $(i > j)$ of AB is $\sum_{k=1}^{p} a_{ik}b_{kj} = 0$ since $a_{ik} = 0$ for $k < i$ and $b_{kj} = 0$ for $k > j$. Similarly, the product of two lower triangular matrices is lower triangular. The determinant of a triangular matrix is the product of the diagonal elements. The inverse of a nonsingular triangular matrix is triangular in the same way.

Theorem A.1.2. *If A is nonsingular, there exists a nonsingular lower triangular matrix F such that $FA = A^*$ is nonsingular upper triangular.*

Proof. Let $A = A_1$. Define recursively $A_g = (a_{ij}^{(g)}) = F_{g-1}A_{g-1}$, $g = 2, \ldots, p$, where $F_{g-1} = (f_{ij}^{(g-1)})$ has elements

$$f_{jj}^{(g-1)} = 1, \qquad j = 1, \ldots, p, \tag{33}$$

$$f_{i,g-1}^{(g-1)} = -\frac{a_{i,g-1}^{(g-1)}}{a_{g-1,g-1}^{(g-1)}}, \qquad i = g, \ldots, p, \tag{34}$$

$$f_{ij}^{(g-1)} = 0, \qquad \text{otherwise}. \tag{35}$$

Then

$$(36)\quad a_{ij}^{(g)} = 0, \qquad i = j+1,\ldots,p, \quad j = 1,\ldots,g-1,$$

$$(37)\quad a_{ij}^{(g)} = a_{ij}^{(g-1)}, \qquad i = 1,\ldots,g-1, \quad j = 1,\ldots,p,$$

$$(38)\quad a_{ij}^{(g)} = a_{ij}^{(g-1)} + f_{i,g-1}^{(g-1)} a_{g-1,j}^{(g-1)} = a_{ij}^{(g-1)} - \frac{a_{i,g-1}^{(g-1)} a_{g-1,j}^{(g-1)}}{a_{g-1,g-1}^{(g-1)}}, \qquad i, j = g,\ldots,p.$$

Note that $F = F_{p-1},\ldots,F_1$ is lower triangular and the elements of A_g in the first $g-1$ columns below the diagonal are 0; in particular $A^* = FA$ is upper triangular. From $|A| \neq 0$ and $|F_{g-1}| = 1$, we have $|A_{g-1}| \neq 0$. Hence $a_{11}^{(1)},\ldots,a_{g-2,g-2}^{(g-2)}$ are different from 0 and the last $p-g$ columns of A_{g-1} can be numbered so $a_{g-1,g-1}^{(g-1)} \neq 0$; then $f_{i,g-1}^{(g-1)}$ is well defined. ■

The equation $FA = A^*$ can be solved to obtain $A = LR$, where $R = A^*$ is upper triangular and $L = F^{-1}$ is lower triangular and has 1's on the main diagonal (because F is lower triangular and has 1's on the main diagonal). This is known as the *LR decomposition.*

Corollary A.1.4. *If A is positive definite, there exists a lower triangular nonsingular matrix F such that FAF' is diagonal and positive definite.*

Proof. From Theorem A.1.2, there exists a lower triangular nonsingular matrix F such that FA is upper triangular and nonsingular. Then FAF' is upper triangular and symmetric; hence it is diagonal. ■

Corollary A.1.5. *The determinant of a positive definite matrix A is positive.*

Proof. From the construction of FAF',

$$(39)\qquad FAF' = \begin{bmatrix} a_{11}^{(1)} & 0 & 0 & \cdots & 0 \\ 0 & a_{22}^{(2)} & 0 & \cdots & 0 \\ 0 & 0 & a_{33}^{(3)} & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_{pp}^{(p)} \end{bmatrix}$$

is positive definite, and hence $a_{gg}^{(g)} > 0$, $g = 1,\ldots,p$, and $0 < |FAF'| = |F| \cdot |A| \cdot |F| = |A|$. ■

Corollary A.1.6. *If A is positive definite, there exists a lower triangular matrix G such that $GAG' = I$.*

Proof. Let $FAF' = D^2$, and let D be the diagonal matrix whose diagonal elements are the positive square roots of the diagonal elements of D^2. Then $C = D^{-1}F$ serves the purpose. ■

Corollary A.1.7 (Cholsky Decomposition). *If A is positive definite, there exists a unique lower triangular matrix T ($t_{ij} = 0$, $i < j$) with positive diagonal elements such that $A = TT'$.*

Proof. From Corollary A.1.6, $A = G^{-1}(G')^{-1}$, where G is lower triangular. Then $T = G^{-1}$ is lower triangular. ■

In effect this theorem was proved in Section 7.2 for $A = VV'$.

A.2. CHARACTERISTIC ROOTS AND VECTORS

The *characteristic roots* of a square matrix B are defined as the roots of the *characteristic equation*

$$|B - \lambda I| = 0. \tag{1}$$

Alternative terms are *latent roots* and *eigenvalues*. For example, with

$$B = \begin{pmatrix} 5 & 2 \\ 2 & 5 \end{pmatrix},$$

we have

$$|B - \lambda I| = \begin{vmatrix} 5-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} = 25 - 4 - 10\lambda + \lambda^2 = \lambda^2 - 10\lambda + 21. \tag{2}$$

The degree of the polynomial equation (1) is the order of the matrix B and the constant term is $|B|$.

A matrix C is said to be *orthogonal* if $C'C = I$; it follows that $CC' = I$. Let the vectors $x' = (x_1, \dots, x_p)$ and $y' = (y_1, \dots, y_p)$ represent two points in a p-dimensional Euclidean space. The distance squared between them is $D(x, y) = (x - y)'(x - y)$. The transformation $z = Cx$ can be thought of as a change of coordinate axes in the p-dimensional space. If C is orthogonal, the transformation is distance-preserving, for

$$(3)\qquad D(Cx,Cy) = (Cy - Cx)'(Cy - Cx) = (y-x)'C'C(y-x) = (y-x)'(y-x) = D(x,y).$$

Since the angles of a triangle are determined by the lengths of its sides, the transformation $z = Cx$ also preserves angles. It consists of a rotation together with a possible reflection of one or more axes. We shall denote $\sqrt{x'x}$ by $\|x\|$.

Theorem A.2.1. *Given any symmetric matrix B, there exists an orthogonal matrix C such that*

$$(4)\qquad C'BC = D = \begin{pmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_p \end{pmatrix}.$$

If B is positive semidefinite, then $d_i \geq 0$, $i = 1,\ldots,p$; if B is positive definite, then $d_i > 0$, $i = 1,\ldots,p$.

The proof is given in the discussion of principal components in Section 11.2 for the case of B positive semidefinite and holds for B symmetric. The characteristic equation (1) under transformation by C becomes

$$(5)\qquad \begin{aligned} 0 &= |C'| \cdot |B - \lambda I| \cdot |C| = |C'(B - \lambda I)C| \\ &= |C'BC - \lambda I| = |D - \lambda I| \\ &= \begin{vmatrix} d_1 - \lambda & 0 & \cdots & 0 \\ 0 & d_2 - \lambda & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & d_p - \lambda \end{vmatrix} = \prod_{i=1}^{p} (d_i - \lambda). \end{aligned}$$

Thus the characteristic roots of B are the diagonal elements of the transformed matrix D.

If λ_i is a characteristic root of B, then a vector x_i not identically $\mathbf{0}$ satisfying

$$(6)\qquad (B = \lambda_i I)x_i = \mathbf{0}$$

is called a *characteristic vector* (or eigenvector) of the matrix B corresponding to the characteristic root λ_i. Any scalar multiple of x_i is also a characteristic vector. When B is symmetric, $x_i'(B - \lambda_i I) = \mathbf{0}$. If the roots are distinct, $x_j'Bx_i = 0$ and $x_j'x_i = 0$, $i \neq j$. Let $c_i = (1/\|x_i\|)x_i$ be the ith normalized characteristic vector, and let $C = (c_1, \ldots, c_p)$. Then $C'C = I$ and $BC = CD$. These lead to (4). If a characteristic root has multiplicity m, then a set of m corresponding characteristic vectors can be replaced by m linearly independent linear combinations of them. The vectors can be chosen to satisfy (6) and $x_j'x_i = 0$ and $x_j'Bx_i = 0$, $i \neq j$.

A characteristic vector lies in the direction of the principal axis (see Chapter 11). The characteristic roots of B are proportional to the squares of the reciprocals of the lengths of the principal axes of the ellipsoid

$$x'Bx = 1 \tag{7}$$

since this becomes under the rotation $y = Cx$

$$1 = y'Dy = \sum_{i=1}^{p} d_i y_i^2. \tag{8}$$

For a pair of matrices A (nonsingular) and B we shall also consider equations of the form

$$|B - \lambda A| = 0. \tag{9}$$

The roots of such equations are of interest because of their invariance under certain transformations. In fact, for nonsingular C, the roots of

$$|C'BC - \lambda(C'AC)| = 0 \tag{10}$$

are the same as those of (9) since

$$|C'BC - \lambda C'AC| = |C'(B - \lambda A)C| = |C'| \cdot |B - \lambda A| \cdot |C| \tag{11}$$

and $|C'| = |C| \neq 0$.

By Corollary A.1.6 we have that if A is positive definite there is a matrix E such that $E'AE = I$. Let $E'BE = B^*$. From Theorem A.2.1 we deduce that there exists an orthogonal matrix C such that $C'B^*C = D$, where D is diagonal. Defining EC as F, we have the following theorem:

Theorem A.2.2. *Given B positive semidefinite and A positive definite, there exists a nonsingular matrix F such that*

$$F'BF = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}, \tag{12}$$

$$F'AF = I, \tag{13}$$

where $\lambda_1 \geq \cdots \geq \lambda_p$ (≥ 0) are the roots of (9). *If B is positive definite*, $\lambda_i > 0$, $i = 1, \ldots, p$.

Corresponding to each root λ_i there is a vector x_i satisfying

$$(B - \lambda_i A)x_i = \mathbf{0} \tag{14}$$

and $x_i'Ax_i = 1$. If the roots are distinct $x_j'Bx_i = 0$ and $x_j'Ax_i = 0$, $i \neq j$. Then $F = (x_1, \dots, x_p)$. If a root has multiplicity m, then a set of m linearly independent x_i's can be replaced by m linearly independent combinations of them. The vectors can be chosen to satisfy (14) and $x_j'Bx_i = 0$ and $x_j'Ax_i = 0$, $i \neq j$.

Theorem A.2.3 (The Singular Value Decomposition). *Given an $n \times p$ matrix X, $n \geq p$, there exists an $n \times n$ orthogonal matrix P, a $p \times p$ orthogonal matrix Q, and an $n \times p$ matrix D consisting of a $p \times p$ diagonal positive semidefinite matrix and an $(n-p) \times p$ zero matrix such that*

$$X = PDQ. \tag{15}$$

Proof. From Theorem A.2.1, there exists a $p \times p$ orthogonal matrix Q and a diagonal matrix E such that

$$QX'XQ' = \begin{pmatrix} E_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \tag{16}$$

where E_1 is diagonal and positive definite. Let $XQ' = Y = (Y_1 \ Y_2)$, where the number of columns of Y_1 is the order of E_1. Then $Y_2'Y_2 = \mathbf{0}$, and hence $Y_2 = \mathbf{0}$. Let $P_1 = Y_1 E_1^{-\frac{1}{2}}$. Then $P_1'P_1 = I$. An $n \times n$ orthogonal matrix $P = (P_1 \ P_2)$ satisfying the theorem is obtained by adjoining P_2 to make P orthogonal. Then the upper left-hand corner of D is $E_1^{\frac{1}{2}}$, and the rest of D consists of zeros. ■

Theorem A.2.4. *Let A be positive definite and B be positive semidefinite. Then*

$$\lambda_p \leq \frac{x'Bx}{x'x} \leq \lambda_1, \tag{17}$$

where λ_1 and λ_p are the largest and smallest roots of (1), *and*

$$\lambda_p \leq \frac{x'Bx}{x'Ax} \leq \lambda_1, \tag{18}$$

where λ_1 and λ_p are the largest and smallest roots of (9).

Proof. The inequalities (17) were essentially proved in Section 11.2, and can also be derived from (4). The inequalities (18) follow from Theorem A.2.2. ■

A square matrix A is *idempotent* if $A^2 = A$. If λ satisfies $|A - \lambda I| = 0$, there exists a vector $x \neq 0$ such that $\lambda x = Ax = A^2 x$. However, $A^2 x = A(Ax) = A\lambda x = \lambda^2 x$. Thus $\lambda^2 = \lambda$, and λ is either 0 or 1. The multiplicity of $\lambda = 1$ is the rank of A. If A is $p \times p$, then $I_p - A$ is idempotent of rank $p -$ (rank A), and A and $I_p - A$ are orthogonal. If A is symmetric, there is an orthogonal matrix O such that

$$OAO' = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \qquad O(I-A)O' = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}. \tag{19}$$

A.3. PARTITIONED VECTORS AND MATRICES

Consider the matrix A defined by (1) of Section A.1. Let

$$\begin{aligned} A_{11} &= (a_{ij}), & i &= 1,\dots,p, \quad i = 1,\dots,q, \\ A_{12} &= (a_{ij}), & i &= 1,\dots,p, \quad j = q+1,\dots,n, \\ A_{21} &= (a_{ij}), & i &= p+1,\dots,m, \quad j = 1,\dots,q, \\ A_{22} &= (a_{ij}), & i &= p+1,\dots,m, \quad j = q+1,\dots,n. \end{aligned} \tag{1}$$

Then we can write

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}. \tag{2}$$

We say that A has been *partitioned* into submatrices A_{ij}. Let B $(m \times n)$ be partitioned similarly into submatrices B_{ij}, $i, j = 1, 2$. Then

$$A + B = \begin{pmatrix} A_{11} + B_{11} & A_{12} + B_{12} \\ A_{21} + B_{21} & A_{22} + B_{22} \end{pmatrix}. \tag{3}$$

Now partition C $(n \times r)$ as

$$C = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}, \tag{4}$$

where C_{11} and C_{12} have q rows and C_{11} and C_{21} have s columns. Then

$$\begin{aligned} AC &= \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix} \\ &= \begin{pmatrix} A_{11}C_{11} + A_{12}C_{21} & A_{11}C_{12} + A_{12}C_{22} \\ A_{21}C_{11} + A_{22}C_{21} & A_{21}C_{12} + A_{22}C_{22} \end{pmatrix}. \end{aligned} \tag{5}$$

To verify this, consider an element in the first p rows and first s columns of AC. The i, jth element is

$$\sum_{k=1}^{n} a_{ik}c_{kj}, \qquad i \leq p, \quad j \leq s. \tag{6}$$

This sum can be written

$$\sum_{k=1}^{q} a_{ik}c_{kj} + \sum_{k=q+1}^{n} a_{ik}c_{kj}. \tag{7}$$

The first sum is the i, jth element of $A_{11}C_{11}$, the second sum is the i, jth element of $A_{12}C_{21}$, and therefore the entire sum (6) is the i, jth element of $A_{11}C_{11} + A_{12}C_{21}$. In a similar fashion we can verify that the other submatrices of AC can be written as in (5).

We note in passing that if A is partitioned as in (2), then the transpose of A can be written

$$A' = \begin{pmatrix} A'_{11} & A'_{21} \\ A'_{12} & A'_{22} \end{pmatrix}. \tag{8}$$

If $A_{12} = \mathbf{0}$ and $A_{21} = \mathbf{0}$, then for A positive definite and A_{11} square,

$$A^{-1} = \begin{pmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & A_{22}^{-1} \end{pmatrix}. \tag{9}$$

The matrix on the right exists because A_{11} and A_{22} are nonsingular. That the right-hand matrix is the inverse of A is verified by multiplication:

$$\begin{pmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & A_{22} \end{pmatrix} \begin{pmatrix} A_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & A_{22}^{-1} \end{pmatrix} = \begin{pmatrix} I & \mathbf{0} \\ \mathbf{0} & I \end{pmatrix}, \tag{10}$$

which is a partitioned form of I_p.

We also note that

$$\begin{vmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & A_{22} \end{vmatrix} = \begin{vmatrix} A_{11} & \mathbf{0} \\ \mathbf{0} & I \end{vmatrix} \cdot \begin{vmatrix} I & \mathbf{0} \\ \mathbf{0} & A_{22} \end{vmatrix} = |A_{11}| \cdot |A_{22}|. \tag{11}$$

The evaluation of the first determinant in the middle is made by expanding according to minors of the last row; the only nonzero element in the sum is the last, which is 1 times a determinant of the same form with I of order one less. The procedure is repeated until $|A_{11}|$ is the minor. Similarly,

$$\begin{vmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{vmatrix} = \begin{vmatrix} I & 0 \\ 0 & A_{22} \end{vmatrix} \cdot \begin{vmatrix} A_{11} & A_{12} \\ 0 & I \end{vmatrix} = |A_{11}| \cdot |A_{22}|. \tag{12}$$

A useful fact is that if A_1 of q rows and p columns is of rank q, there exists a matrix A_2 of $p-q$ rows and p columns such that

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} \tag{13}$$

is nonsingular. This statement is verified by numbering the columns of A so that A_{11} consisting of the first q columns of A_1 is nonsingular (at least one $q \times q$ minor of A_1 is different from zero) and then taking A_2 as $(\mathbf{0} \;\; I)$; then

$$|A| = \begin{vmatrix} A_{11} & A_{12} \\ 0 & I \end{vmatrix} = |A_{11}|, \tag{14}$$

which is not equal to zero.

Theorem A.3.1. *Let the square matrix A be partitioned as in* (2) *so that A_{22} is square. If A_{22} is nonsingular, let*

$$B = \begin{pmatrix} I & -A_{12}A_{22}^{-1} \\ 0 & I \end{pmatrix}, \qquad C = \begin{pmatrix} I & 0 \\ -A_{22}^{-1}A_{21} & I \end{pmatrix}. \tag{15}$$

Then

$$BA = \begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ A_{21} & A_{22} \end{pmatrix}, \qquad AC = \begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & A_{12} \\ 0 & A_{22} \end{pmatrix}, \tag{16}$$

$$BAC = \begin{pmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} & 0 \\ 0 & A_{22} \end{pmatrix}. \tag{17}$$

If A is symmetric, $C = B'$.

Theorem A.3.2. *Let the square matrix A be partitioned as in* (2) *so that A_{22} is square. If A_{22} is nonsingular,*

$$|A| = |A_{11} - A_{12}A_{22}^{-1}A_{21}| \cdot |A_{22}|. \tag{18}$$

Proof. Equation (18) follows from (16) because $|\boldsymbol{B}| = 1$. ■

Corollary A.3.1. *For C nonsingular*

$$\begin{vmatrix} C & y \\ y' & 1 \end{vmatrix} = |C - yy'| = \begin{vmatrix} 1 & y' \\ y & C \end{vmatrix} = |C|(1 - y'C^{-1}y). \tag{19}$$

Theorem A.3.3. *Let the nonsingular matrix A be partitioned as in* (2) *so that A_{22} is square. If A_{22} is nonsingular, let $A_{11\cdot 2} = A_{11} - A_{12}A_{22}^{-1}A_{21}$. Then*

$$A^{-1} = \begin{pmatrix} A_{11\cdot 2}^{-1} & -A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} \\ -A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1} & A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} + A_{22}^{-1} \end{pmatrix}. \tag{20}$$

Proof. From Theorem A.3.1,

$$\boldsymbol{A} = \boldsymbol{B}^{-1}\begin{pmatrix} A_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & A_{22} \end{pmatrix}\boldsymbol{C}^{-1}. \tag{21}$$

Hence

$$\begin{aligned} A^{-1} &= C\begin{pmatrix} A_{11\cdot 2} & \mathbf{0} \\ \mathbf{0} & A_{22} \end{pmatrix}^{-1} B \\ &= \begin{pmatrix} I & \mathbf{0} \\ -A_{22}^{-1}A_{21} & I \end{pmatrix}\begin{pmatrix} A_{11\cdot 2}^{-1} & \mathbf{0} \\ \mathbf{0} & A_{22}^{-1} \end{pmatrix}\begin{pmatrix} I & -A_{12}A_{22} \\ \mathbf{0} & I \end{pmatrix}. \end{aligned} \tag{22}$$

Multiplication gives the desired result. ■

Corollary A.3.2. *If $x' = (x^{(1)\prime}\ x^{(2)\prime})$, then*

$$x'A^{-1}x = \left(x^{(1)} - A_{12}A_{22}^{-1}x^{(2)}\right)' A_{11\cdot 2}^{-1}\left(x^{(1)} - A_{12}A_{22}^{-1}x^{(2)}\right) + x^{(2)\prime}A_{22}^{-1}x^{(2)}. \tag{23}$$

Proof. From the theorem

$$\begin{aligned} x'A^{-1}x &= x^{(1)\prime}A_{11\cdot 2}^{-1}x^{(1)} - x^{(1)\prime}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1}x^{(2)} \\ &\quad - x^{(2)\prime}A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}x^{(1)} + x^{(2)\prime}\left(A_{22}^{-1}A_{21}A_{11\cdot 2}^{-1}A_{12}A_{22}^{-1} + A_{22}^{-1}\right)x^{(2)}, \end{aligned} \tag{24}$$

which is equal to the right-hand side of (23). ■

Theorem A.3.4. *Let the nonsingular matrix A be partitioned as in* (2) *so that A_{22} is square. If A_{22} is nonsingular,*

$$(25) \quad \left(A_{22} - A_{21}A_{11}^{-1}A_{12}\right)^{-1} = A_{22}^{-1}A_{21}\left(A_{11} - A_{12}A_{22}^{-1}A_{21}\right)^{-1}A_{12}A_{22}^{-1} + A_{22}^{-1}.$$

Proof. The lower right-hand corner of A^{-1} is the right-hand side of (25) by Theorem A.3.3 and is also the left-hand side of (25) by interchange of 1 and 2. ■

Theorem A.3.5. *Let U be $p \times m$. The conditions for $I_p - UU'$, $I_m - U'U$, and*

$$(26) \quad \begin{pmatrix} I_p & U \\ U' & I_m \end{pmatrix}$$

to be positive definite are the same.

Proof. We have

$$(27) \quad (v' \quad w')\begin{pmatrix} I_p & U \\ U' & I_m \end{pmatrix}\begin{pmatrix} v \\ w \end{pmatrix} = v'v + v'Uw + w'U'v + w'w$$

$$= v'(I_m - UU')v + (U'v + w)'(U'v + w).$$

The second term on the right-hand side is nonnegative; the first term is positive for all $v \neq 0$ if and only if $I_m - U'U$ is positive definite. Reversing the roles of v and w shows that (26) is positive definite if and only if $I_p - UU'$ is positive definite. ■

A.4. SOME MISCELLANEOUS RESULTS

Theorem A.4.1. *Let C be $p \times p$, positive semidefinite, and of rank r $(\leq p)$. Then there is a nonsingular matrix A such that*

$$(1) \quad ACA' = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Proof. Since C is of rank r, there is a $(p - r) \times p$ matrix A_2 such that

$$(2) \quad A_2C = 0.$$

Choose B $(r \times p)$ such that

$$(3) \quad \begin{pmatrix} B \\ A_2 \end{pmatrix}$$

is nonsingular. Then

$$(4)\qquad \begin{pmatrix} B \\ A_2 \end{pmatrix} C(B' \quad A_2') = \begin{pmatrix} BC \\ 0 \end{pmatrix} (B' \quad A_2') = \begin{pmatrix} BCB' & 0 \\ 0 & 0 \end{pmatrix}.$$

This matrix is of rank r, and therefore BCB' is nonsingular. By Corollary A.1.6 there is a nonsingular matrix D such that $D(BCB')D' = I_r$. Then

$$(5)\qquad A = \begin{pmatrix} DB \\ A_2 \end{pmatrix} = \begin{pmatrix} D & 0 \\ 0 & I \end{pmatrix} \begin{pmatrix} B \\ A_2 \end{pmatrix}$$

is a nonsingular matrix such that (1) holds. ∎

Lemma A.4.1. *If E is $p \times p$, symmetric, and nonsingular, there is a nonsingular matrix F such that*

$$(6)\qquad FEF' = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix},$$

where the order of I is the number of positive characteristic roots of E and the order of $-I$ is the number of negative characteristic roots of E.

Proof. From Theorem A.2.1 we know there is an orthogonal matrix G such that

$$(7)\qquad GEG' = \begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & h_p \end{pmatrix},$$

where $h_1 \geq \cdots \geq h_q > 0 > h_{q+1} \geq \cdots \geq h_p$ are the characteristic roots of E. Let

$$(8)\qquad K = \begin{pmatrix} 1/\sqrt{h_1} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1/\sqrt{h_q} & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1/\sqrt{-h_{q+1}} & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1/\sqrt{-h_p} \end{pmatrix}.$$

Then

$$KGEG'K' = (KG)E(KG)' = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}. \quad \blacksquare \tag{9}$$

Corollary A.4.1. *Let C be $p \times p$, symmetric, and of rank r $(\leq p)$. Then there is a nonsingular matrix A such that*

$$ACA' = \begin{pmatrix} I & 0 & 0 \\ 0 & -I & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{10}$$

where the order of I is the number of positive characteristic roots of C and the order of $-I$ is the number of negative characteristic roots, the sum of the orders being r.

Proof. The proof is the same as that of Theorem A.4.1 except that Lemma A.4.1 is used instead of Corollary A.1.6. ■

Lemma A.4.2. *Let A be $n \times m$ $(n > m)$ such that*

$$A'A = I_m. \tag{11}$$

There exists an $n \times (n-m)$ matrix B such that $(A \ B)$ is orthogonal.

Proof. Since A is of rank m, there exists an $n \times (n-m)$ matrix C such that $(A \ C)$ is nonsingular. Take D as $C - AA'C$; then $D'A = 0$. Let E $[(n-m) \times (n-m)]$ be such that $E'D'DE = I$. Then B can be taken as DE. ■

Lemma A.4.3. *Let x be a vector of n components. Then there exists an orthogonal matrix O such that*

$$Ox = \begin{pmatrix} c \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{12}$$

where $c = \sqrt{x'x}$.

Proof. Let the first row of O be $(1/c)x'$. The other rows may be chosen in any way to make the matrix orthogonal. ■

Lemma A.4.4. *Let $B = (b_{ij})$ be a $p \times p$ matrix. Then*

$$\frac{\partial |B|}{\partial b_{ij}} = B_{ij}, \qquad i, j = 1, \dots, p. \tag{13}$$

Proof. The expansion of $|B|$ by elements of the ith row is

$$|B| = \sum_{h=1}^{p} b_{ih} B_{ih}. \tag{14}$$

Since B_{ih} does not contain b_{ij}, the lemma follows. ■

Lemma A.4.5. *Let $b_{ij} = \beta_{ij}(c_1, \dots, c_n)$ be the i, jth element of a $p \times p$ matrix B. Then for $g = 1, \dots, n$,*

$$\frac{\partial |B|}{\partial c_g} = \sum_{i,h=1}^{p} \frac{\partial |B|}{\partial b_{ih}} \cdot \frac{\partial \beta_{ih}(c_1, \dots, c_n)}{\partial c_g} = \sum_{i,h=1}^{p} B_{ih} \frac{\partial \beta_{ih}(c_1, \dots, c_n)}{\partial c_g}. \tag{15}$$

Theorem A.4.2. *If $A = A'$,*

$$\frac{\partial |A|}{\partial a_{ii}} = A_{ii}, \tag{16}$$

$$\frac{\partial |A|}{\partial a_{ij}} = 2A_{ij}, \qquad i \neq j. \tag{17}$$

Proof. Equation (16) follows from the expansion of $|A|$ according to elements of the ith row. To prove (17) let $b_{ij} = b_{ji} = a_{ij}$, $i, j = 1, \dots, p$, $i \leq j$. Then by Lemma A.4.5,

$$\frac{\partial |B|}{\partial a_{ij}} = B_{ij} + B_{ji}, \tag{18}$$

Since $|A| = |B|$ and $B_{ij} = B_{ji} = A_{ij} = A_{ji}$, (17) follows. ■

Theorem A.4.3.

$$\frac{\partial}{\partial x}(x'Ax) = 2Ax, \tag{19}$$

where $\partial / \partial x$ denotes taking partial derivatives with respect to each component of x and arranging the partial derivatives in a column.

Proof. Let h be a column vector of as many components as x. Then

$$\begin{aligned}(x+h)'A(x+h) &= x'Ax + h'Ax + x'Ah + h'Ah \\ &= x'Ax + 2h'Ax + h'Ah.\end{aligned} \tag{20}$$

The partial derivative vector is the vector multiplying h' in the second term on the right. ■

Definition A.4.1. *Let $A = (a_{ij})$ be a $p \times m$ matrix and $B = (b_{\alpha\beta})$ be a $q \times n$ matrix. The $pq \times mn$ matrix with $a_{ij} b_{\alpha\beta}$ as the element in the i, αth row and the*

j, βth column is called the Kronecker *or* direct product *of A and B and is denoted by $A \otimes B$; that is,*

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \cdots & a_{1m}B \\ a_{21}B & a_{22}B & \cdots & a_{2m}B \\ \vdots & \vdots & & \vdots \\ a_{p1}B & a_{p2}B & \cdots & a_{pm}B \end{pmatrix}. \tag{21}$$

Some properties are the following when the orders of matrices permit the indicated operations:

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD), \tag{22}$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}. \tag{23}$$

Theorem A.4.4. *Let the ith characteristic root of A $(p \times p)$ be λ_i and the corresponding characteristic vector be $x_i = (x_{1i}, \dots, x_{pi})'$, and let the αth root of B $(q \times q)$ be ν_α and the corresponding characteristic vector be y_α, $\alpha = 1, \dots, q$. Then the i, αth root of $A \otimes B$ is $\lambda_i \nu_\alpha$, and the corresponding characteristic vector is $x_i \otimes y_\alpha = (x_{1i}y_\alpha', \dots, x_{pi}y_\alpha')'$, $i = 1, \dots, p$, $\alpha = 1, \dots, q$.*

Proof.

$$(A \otimes B)(x_i \otimes y_\alpha) = \begin{pmatrix} a_{11}B & \cdots & a_{1p}B \\ \vdots & & \vdots \\ a_{p1}B & \cdots & a_{pp}B \end{pmatrix} \begin{pmatrix} x_{1i}y_\alpha \\ \vdots \\ x_{pi}y_\alpha \end{pmatrix} \tag{24}$$

$$= \begin{pmatrix} \sum_j a_{1j}x_{ji}By_\alpha \\ \vdots \\ \sum_j a_{pj}x_{ji}By_\alpha \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_i x_{1i}By_\alpha \\ \vdots \\ \lambda_i x_{pi}By_\alpha \end{pmatrix} = \lambda_i \nu_\alpha \begin{pmatrix} x_{1i}y_\alpha \\ \vdots \\ x_{pi}y_\alpha \end{pmatrix}. \qquad \blacksquare$$

Theorem A.4.5

$$|A \otimes B| = |A|^q |B|^p. \tag{25}$$

Proof. The determinant of any matrix is the product of its roots; therefore

$$|A \otimes B| = \prod_{i=1}^{p} \prod_{\alpha=1}^{q} \lambda_i \nu_\alpha = \left(\prod_{i=1}^{p} \lambda_i \right)^q \left(\prod_{\alpha=1}^{q} \nu_\alpha \right)^p. \tag{26}$$ ■

Definition A.4.2. *If the* $p \times m$ *matrix* $A = (a_1, \dots, a_m)$, *then* $\operatorname{vec} A = (a_1', \dots, a_m')'$.

Some properties of the vec operator [e.g., Magnus (1988)] are

$$\operatorname{vec} ABC = (C' \otimes A) \operatorname{vec} B, \tag{27}$$

$$\operatorname{vec} xy' = y \otimes x. \tag{28}$$

Theorem A.4.6. *The Jacobian of the transformation* $E = Y^{-1}$ *(from* E *to* Y*) is* $|Y|^{-2p}$, *where* p *is the order of* E *and* Y.

Proof. From $EY = I$, we have

$$\left(\frac{\partial}{\partial \theta} E \right) Y + E \left(\frac{\partial}{\partial \theta} Y \right) = 0, \tag{29}$$

where

$$\left(\frac{\partial}{\partial \theta} E \right) = \begin{pmatrix} \frac{\partial e_{11}}{\partial \theta} & \cdots & \frac{\partial e_{1p}}{\partial \theta} \\ \vdots & & \vdots \\ \frac{\partial e_{p1}}{\partial \theta} & \cdots & \frac{\partial e_{pp}}{\partial \theta} \end{pmatrix}. \tag{30}$$

Then

$$\left(\frac{\partial}{\partial \theta} E \right) = -E \left(\frac{\partial}{\partial \theta} Y \right) E = -Y^{-1} \left(\frac{\partial}{\partial \theta} Y \right) Y^{-1}. \tag{31}$$

If $\theta = y_{\alpha\beta}$, then

$$\left(\frac{\partial}{\partial y_{\alpha\beta}} E \right) = -E \varepsilon_{\alpha\beta} E = -\mathbf{e}_{\cdot\alpha} \mathbf{e}_{\beta\cdot}, \tag{32}$$

where $\varepsilon_{\alpha\beta}$ is a $p \times p$ matrix with all elements 0 except the element in the αth row and βth column, which is 1; and $e_{\cdot\alpha}$ is the αth column of E and $e_{\beta\cdot}$ is its βth row. Thus $\partial e_{ij}/\partial y_{\alpha\beta} = -e_{i\alpha}e_{\beta j}$. Then the Jacobian is the determinant of a $p^2 \times p^2$ matrix

$$(33) \quad \operatorname{mod}\left|\frac{\partial e_{ij}}{\partial y_{\alpha\beta}}\right| = |e_{i\alpha}e_{\beta j}| = |E \otimes E'| = |E|^p|E'|^p = |E|^{2p} = |Y|^{-2p}. \qquad \blacksquare$$

Theorem A.4.7. *Let A and B be symmetric matrices with characteristic roots $a_1 \geq a_2 \geq \cdots \geq a_p$ and $b_1 \geq b_2 \geq \cdots \geq b_p$, respectively, and let H be a $p \times p$ orthogonal matrix. Then*

$$(34) \qquad \max_H \operatorname{tr} HAH'B = \sum_{j=1}^{p} a_j b_j, \qquad \min_H HA'H'B = \sum_{j=1}^{p} a_j b_{p+1-j}.$$

Proof. Let $A = H_a D_a H_a'$ and $B = H_b D_b H_b'$, where H_a and H_b are orthogonal and D_a and D_b are diagonal with diagonal elements $a_1, \ldots, a_p$ and $b_1, \ldots, b_p$ respectively. Then

$$(35) \qquad \begin{aligned} \max_{H^*} \operatorname{tr} H^*AH^{*\prime}B &= \max_{H^*} \operatorname{tr} H^*H_aD_aH_a'H^{*\prime}H_bD_bH_b' \\ &= \max_{H^*} \operatorname{tr} H_b'H^*H_aD_a(H_b'H^*H_a)'D_b \\ &= \max_H \operatorname{tr} HD_aH'D_b, \end{aligned}$$

where $H = H_b'H^*H_a$. We have

$$(36) \quad \begin{aligned} \operatorname{tr} HD_aH'D_b &= \sum_{i=1}^{p} (HD_aH')_{ii} b_i \\ &= \sum_{i=1}^{p-1} \sum_{j=1}^{i} (HD_aH')_{jj}(b_i - b_{i+1}) + b_p \sum_{j=1}^{p} (HD_aH')_{jj} \\ &\leq \sum_{i=1}^{p-1} \sum_{j=1}^{i} a_j(b_i - b_{i+1}) + b_p \sum_{j=1}^{p} a_j \\ &= \sum_{i=1}^{p} a_i b_i \end{aligned}$$

by Lemma A.4.6 below. The minimum in (34) is treated as the negative of the maximum with B replaced by $-B$ [von Neumann (1937)]. $\blacksquare$

Lemma A.4.6. *Let $P = (p_{ij})$ be a doubly stochastic matrix ($p_{ij} \geq 0$, $\sum_{i=1}^{p} p_{ij} = 1$, $\sum_{j=1}^{p} p_{ij} = 1$). Let $y_1 \geq y_2 \geq \cdots \geq y_p$. Then*

$$\sum_{i=1}^{k} y_i \geq \sum_{i=1}^{k} \sum_{j=1}^{n} p_{ij} y_j, \qquad k = 1, \ldots, p. \tag{37}$$

Proof.

$$\sum_{i=1}^{k} \sum_{j=1}^{p} p_{ij} y_j = \sum_{j=1}^{p} g_j y_j, \tag{38}$$

where $g_j = \sum_{i=1}^{k} p_{ij}$, $j = 1, \ldots, p$ ($0 \leq g_j \leq 1$, $\sum_{j=1}^{p} g_j = k$). Then

$$\begin{aligned} \sum_{j=1}^{p} g_j y_j - \sum_{j=1}^{k} y_j &= -\sum_{j=1}^{k} y_j + y_k\left(k - \sum_{j=1}^{p} g_j\right) + \sum_{j=1}^{p} g_j y_j \\ &= \sum_{j=1}^{k} (y_j - y_k)(g_j - 1) + \sum_{j=k+1}^{p} (y_j - y_k) g_j \\ &\leq 0. \end{aligned} \tag{39}$$ ■

Corollary A.4.2. *Let A be a symmetric matrix with characteristic roots $a_1 \geq a_2 \geq \cdots a_p$. Then*

$$\max_{R'R = I_k} \operatorname{tr} R'AR = \sum_{i=1}^{k} a_i. \tag{40}$$

Proof. In Theorem A.4.7 let

$$B = \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix}. \tag{41}$$ ■

Theorem A.4.8.

$$|I + xC| = 1 + x \operatorname{tr} C + O(x^2). \tag{42}$$

Proof. The determinant (42) is a polynomial in x of degree p; the coefficient of the linear term is the first derivative of the determinant evaluated at $x = 0$. In Lemma A.4.5 let $n = 1$, $c_1 = x$, $\beta_{ih}(x) = \delta_{ih} + x c_{ih}$, where $\delta_{ii} = 1$ and $\delta_{ih} = 0$, $i \neq h$. Then $d\beta_{ih}(x)/dx = c_{ih}$, $B_{ii} = 1$ for $x = 0$, and $B_{ih} = 0$ for $x = 0$, $i \neq h$. Thus

$$\left.\frac{d|B(x)|}{dx}\right|_{x=0} = \sum_{i=1}^{p} c_{ii}. \tag{43}$$ ■

A.5. GRAM–SCHMIDT ORTHOGONALIZATION AND THE SOLUTION OF LINEAR EQUATIONS

A.5.1. Gram–Schmidt Orthogonalization

The derivation of the Wishart density in Section 7.2 included the Gram–Schmidt orthogonalization of a set of vectors; we shall review that development here. Consider the p linearly independent n-dimensional vectors $v_1, \ldots, v_p$ $(p \le n)$. Define $w_1 = v_1$,

$$(1) \qquad w_i = v_i - \sum_{j=1}^{i-1} \frac{v_i' w_j}{\|w_j\|^2} w_j, \qquad i = 2, \ldots, p.$$

Then $w_i \neq \mathbf{0}$, $i = 1, \ldots, p$, because $v_1, \ldots, v_p$ are linearly independent, and $w_i' w_j = 0$, $i \neq j$, as was proved by induction in Section 7.2. Let $u_i = (1/\|w_i\|) w_i$, $i = 1, \ldots, p$. Then $u_1, \ldots, u_p$ are *orthonormal*; that is, they are orthogonal and of unit length. Let $U = (u_1, \ldots, u_p)$. Then $U'U = I$. Define $t_{ii} = \|w_i\|$ (> 0),

$$(2) \qquad t_{ij} = \frac{v_i' w_j}{\|w_j\|} = v_i' u_j, \qquad j = 1, \ldots, i-1, \quad i = 2, \ldots, p,$$

and $t_{ij} = 0$, $j = i+1, \ldots, p$, $i = 1, \ldots, p-1$. Then $T = (t_{ij})$ is a lower triangular matrix. We can write (1) as

$$(3) \qquad v_i = \|w_i\| u_i + \sum_{j=1}^{i-1} (v_i' u_j) u_j = \sum_{j=1}^{i} t_{ij} u_j, \qquad i = 1, \ldots, p,$$

that is,

$$(4) \qquad V = (v_1, \ldots, v_p) = UT'.$$

Then

$$(5) \qquad A = V'V = TU'UT' = TT'$$

as shown in Section 7.2. Note that if V is square, we have decomposed an arbitrary nonsingular matrix into the product of an orthogonal matrix and an upper triangular matrix with positive diagonal elements; this is sometimes known as the *QR decomposition*. The matrices U and T in (4) are unique.

These operations can be done in a different order. Let $V = (v_1^{(0)}, \ldots, v_p^{(0)})$. For $k = 1, \ldots, p-1$ define recursively

$$(6) \qquad t_{kk} = \|v_k^{(k-1)}\|, \qquad u_k = \frac{1}{\|v_k^{(k-1)}\|} v_k^{(k-1)} = \frac{1}{t_{kk}} v_k^{(k-1)},$$

$$(7) \qquad t_{jk} = v_j^{(k-1)\prime} u_k, \qquad j = k+1, \ldots, p,$$

$$(8) \qquad v_j^{(k)} = v_j^{(k-1)} - t_{jk} u_k, \qquad j = k+1, \ldots, p.$$

Finally $t_{pp} = \|v_p^{(p-1)}\|$ and $u_p = (1/t_{pp})v_p^{(p-1)}$. The same orthonormal vectors $u_1, \ldots, u_p$ and the same triangular matrix (t_{ij}) are given by the two procedures.

The numbering of the columns of V is arbitrary. For numerical stability it is usually best at any given stage to select the largest of $\|v_j^{(k-1)}\|$ to call t_{kk}.

Instead of constructing w_i as orthogonal to $w_1, \ldots, w_{i-1}$, we can equivalently construct it as orthogonal to $v_1, \ldots, v_{i-1}$. Let $w_1 = v_1$, and define

$$w_i = v_i + \sum_{j=1}^{i-1} f_{ij} v_j \tag{9}$$

such that

$$\begin{aligned} 0 = v_h' w_i &= v_h' v_i + \sum_{j=1}^{i-1} f_{ij} v_h' v_j \\ &= a_{hi} + \sum_{j=1}^{i-1} a_{hj} f_{ij}, \qquad h = 1, \ldots, i-1. \end{aligned} \tag{10}$$

Let $F = (f_{ij})$, where $f_{ii} = 1$ and $f_{ij} = 0$, $i < j$. Then

$$W = (w_1, \ldots, w_p) = VF'. \tag{11}$$

Let D_t be the diagonal matrix with $\|w_j\| = t_{jj}$ as the jth diagonal element. Then $U = WD_t^{-1} = VF'D_t^{-1}$. Comparison with $V = UT'$ shows that $F = DT^{-1}$. Since $A = TT'$, we see that $FA = DT'$ is upper triangular. Hence F is the matrix defined in Theorem A.1.2.

There are other methods of accomplishing the QR decomposition that may be computationally more efficient or more stable. A *Householder matrix* has the form $H = I_n - 2\alpha\alpha'$, where $\alpha'\alpha = 1$, and is orthogonal and symmetric. Such a matrix H_1 (i.e., a vector α) can be selected so that the first column of $H_1 V$ has 0's in all positions except the first, which is positive. The next matrix has the form

$$H_2 = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I_{n-1} \end{pmatrix} - 2\begin{pmatrix} 0 \\ \alpha \end{pmatrix}(0 \quad \alpha') = \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & I_{n-1} - \alpha\alpha' \end{pmatrix}. \tag{12}$$

The $(n-1)$-component vector α is chosen so that the second column of $H_1 V$ has all 0's except the first two components, the second being positive. This process is continued until

$$H_{p-1} \cdots H_2 H_1 V = \begin{bmatrix} T' \\ \mathbf{0} \end{bmatrix}, \tag{13}$$

where T' is upper triangular and $\mathbf{0}$ is $(n-p)\times p$. Let

$$H' = H_1 \cdots H_{p-1} = (H^{(1)} \quad H^{(2)}), \tag{14}$$

where $H^{(1)}$ has p columns. Then from (13) we obtain $V = H^{(1)}T'$. Since the decomposition is unique, $H^{(1)} = U$.

Another procedure uses *Givens matrices*. A Givens matrix G_{ij} is I except for the elements $g_{ii} = \cos\theta = g_{jj}$ and $g_{ij} = \sin\theta = -g_{ji}$, $i \neq j$. It is orthogonal. Multiplication of V on the left by such a matrix leaves all rows unchanged except the ith and jth; θ can be chosen so that the i,jth element of $G_{ij}V$ is 0. Givens matrices $G_{21},\dots,G_{n1}$ can be chosen in turn so $G_{n1}\cdots G_{21}V$ has all 0's in the first column except the first element, which is positive. Next $G_{32},\dots,G_{n2}$ can be selected in turn so that when they are applied the resulting matrix has 0's in the second column except for the first two elements. Let

$$G' = G'_{21}\cdots G'_{n1}G'_{32}\cdots G'_{n,p-1} = (G^{(1)} \quad G^{(2)}). \tag{15}$$

Then we obtain

$$V = G'\begin{bmatrix} T' \\ \mathbf{0} \end{bmatrix} = G^{(1)}T', \tag{16}$$

and $G^{(1)} = U$.

A.5.2. Solution of Linear Equations

In the computation of regression coefficients and other statistics, we need to solve linear equations

$$Ax = y, \tag{17}$$

where A is $p\times p$ and positive definite. One method of solution is Gaussian elimination of variables, or pivotal condensation. In the proof of Theorem A.1.2 we constructed a lower triangular matrix F with diagonal elements 1 such that $FA = A^*$ is upper triangular. If $Fy = y^*$, then the equation is

$$A^*x = y^*. \tag{18}$$

In coordinates this is

$$\sum_{j=1}^{p} a_{ij}^* x_j = y_i^*. \tag{19}$$

Let $a_{ij}^{**} = a_{ij}^{*}/a_{ii}^{*}$, $y_i^{**} = y_i^{*}/a_{ii}^{*}$, $j = i, i+1, \ldots, p$, $i = 1, \ldots, p$. Then

$$x_i = y_i^{**} - \sum_{j=i+1}^{p} a_{ij}^{**} x_j; \tag{20}$$

these equations are to be solved successively for $x_p, x_{p-1}, \ldots, x_1$. The calculation of $FA = A^*$ is known as the *forward* solution, and the solution of (18) as the *backward* solution.

Since $FAF' = A^*F' = D^2$ diagonal, (20) is $A^{**}x = y^{**}$, where $A^{**} = D^{-2}A^*$ and $y^{**} = D^{-2}y^*$. Solving this equation gives

$$x = A^{**-1}y^{**} = F'y^{**}. \tag{21}$$

The computation is

$$x = F_1' \cdots F_{p-1}' D^{-2} F_{p-1} \cdots F_1 y. \tag{22}$$

The multiplier of y in (22) indicates a sequence of row operations which yields A^{-1}.

The operations of the forward solution transform A to the upper triangular matrix A^*. As seen in Section A.5.1, the triangularization of a matrix can be done by a sequence of Householder transformations or by a sequence of Givens transformations.

From $FA = A^*$, we obtain

$$|A| = \prod_{i=1}^{p} a_{ii}^{(i)}, \tag{23}$$

which is the product of the diagonal elements of A^*, resulting from the forward solution. We also have

$$\begin{aligned} y'A^{-1}y &= (Fy)'D^{-2}(Fy) = y^{*\prime}D^{-2}y^* \\ &= y^{*\prime}y^{**}. \end{aligned} \tag{24}$$

The forward solution gives a computation for the quadratic form which occurs in T^2 and other statistics.

For more on matrix computations consult Golub and Von Loan (1989).