Appendix B

Review of Matrix Concepts

B.1 INTRODUCTION

The following matrix notation, definitions, and theorems are used extensively in this book. Much of this material is based on Graybill (1961).

- A matrix $\mathbf{A}$ will have elements denoted by a_{ij}, where i refers to the row and j to the column.
- $\mathbf{A}^T$ will denote the transpose of $\mathbf{A}$.
- $\mathbf{A}^{-1}$ will denote the inverse of $\mathbf{A}$.
- $|\mathbf{A}|$ will denote the determinant of $\mathbf{A}$.
- The dimension of a matrix is the number of its rows by the number of its columns.
- An $n \times m$ matrix $\mathbf{A}$ will have n rows and m columns.
- If $m = 1$, the matrix will be called an $n \times 1$ vector.
- Given the matrices $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$, the product $\mathbf{AB} = \mathbf{C} = (c_{ij})$ is defined as the matrix $\mathbf{C}$ with the pqth element equal to

$$\sum_{s=1}^{m} a_{ps}\, b_{sq} \tag{B.1.1}$$

where m is the column dimension of $\mathbf{A}$ and the row dimension of $\mathbf{B}$.

- Given

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

and

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

then

$$\begin{aligned} \mathbf{AB} &= \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix} \end{aligned} \tag{B.1.2}$$

provided the elements of $\mathbf{A}$ and $\mathbf{B}$ are conformable.

- For $\mathbf{AB}$ to be defined, the number of columns in $\mathbf{A}$ must equal the number of rows in $\mathbf{B}$.

- For $\mathbf{A} + \mathbf{B}$ to be defined, $\mathbf{A}$ and $\mathbf{B}$ must have the same dimension.

- The transpose of $\mathbf{A}^T$ equals $\mathbf{A}$; that is, $(\mathbf{A}^T)^T = \mathbf{A}$.

- The inverse of $\mathbf{A}^{-1}$ is $\mathbf{A}$; that is, $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$.

- The transpose and inverse symbols may be permuted; that is, $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$.

- $(\mathbf{AB})^T = \mathbf{B}^T\,\mathbf{A}^T$.

- $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\,\mathbf{A}^{-1}$ if $\mathbf{A}$ and $\mathbf{B}$ are each nonsingular.

- A scalar commutes with every matrix; that is, $k\,\mathbf{A} = \mathbf{A}\,k$.

- For any matrix $\mathbf{A}$, we have $\mathbf{IA} = \mathbf{AI} = \mathbf{A}$.

- All diagonal matrices of the same dimension are commutative.

- If $\mathbf{D}_1$ and $\mathbf{D}_2$ are diagonal matrices, then the product is diagonal.

- If $\mathbf{X}$ and $\mathbf{Y}$ are vectors and if $\mathbf{A}$ is a nonsingular matrix and if the equation $\mathbf{Y} = \mathbf{AX}$ holds, then $\mathbf{X} = \mathbf{A}^{-1}\,\mathbf{Y}$.

B.2 RANK

- The *rank* of a matrix is the dimension of its largest square nonsingular submatrix; that is, one whose determinant is nonzero.
- The rank of the product $\mathbf{AB}$ of the two matrices $\mathbf{A}$ and $\mathbf{B}$ is less than or equal to the rank of $\mathbf{A}$ and is less than or equal to the rank of $\mathbf{B}$.
- The rank of the sum of $\mathbf{A} + \mathbf{B}$ is less than or equal to the rank of $\mathbf{A}$ plus the rank of $\mathbf{B}$.
- If $\mathbf{A}$ is an $n \times n$ matrix and if $|\mathbf{A}| = 0$, then the rank of $\mathbf{A}$ is less than n.
- If the rank of $\mathbf{A}$ is less than n, then all the rows of $\mathbf{A}$ are not independent; likewise, all the columns of $\mathbf{A}$ are not independent ($\mathbf{A}$ is $n \times n$).
- If the rank of $\mathbf{A}$ is $m \leq n$, then the number of linearly independent rows is m; also, the number of linearly independent columns is m ($\mathbf{A}$ is $n \times n$).
- If $\mathbf{A}^T\mathbf{A} = 0$, then $\mathbf{A} = 0$.
- The rank of a matrix is unaltered by multiplication by a nonsingular matrix; that is, if $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are matrices such that $\mathbf{AB}$ and $\mathbf{BC}$ exist and if $\mathbf{A}$ and $\mathbf{C}$ are nonsingular, then $\rho(\mathbf{AB}) = \rho(\mathbf{BC}) = \rho(\mathbf{B})$. $\rho(\mathbf{B})$ = rank of $\mathbf{B}$.
- If the product $\mathbf{AB}$ of two square matrices is 0, then $\mathbf{A} = 0, \mathbf{B} = 0$, or $\mathbf{A}$ and $\mathbf{B}$ are both singular.
- If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices of rank r and s, respectively, then the rank of $\mathbf{AB}$ is greater than or equal to $r + s - n$.
- The rank of $\mathbf{AA}^T$ equals the rank of $\mathbf{A}^T\mathbf{A}$, equals the rank of $\mathbf{A}$, equals the rank of $\mathbf{A}^T$.

B.3 QUADRATIC FORMS

- The rank of the quadratic form $\mathbf{Y}^T\mathbf{AY}$ is defined as the rank of the matrix $\mathbf{A}$ where $\mathbf{Y}$ is a vector and $\mathbf{Y} \neq 0$.
- The quadratic form $\mathbf{Y}^T\mathbf{AY}$ is said to be *positive definite* if and only if $\mathbf{Y}^T\mathbf{AY} > 0$ for all vectors $\mathbf{Y}$ where $\mathbf{Y} \neq 0$.
- A quadratic form $\mathbf{Y}^T\mathbf{AY}$ is said to be *positive semidefinite* if and only if $\mathbf{Y}^T\mathbf{AY} \geq 0$ for all $\mathbf{Y}$, and $\mathbf{Y}^T\mathbf{AY} = 0$ for some vector $\mathbf{Y} \neq 0$.

- A quadratic form $\mathbf{Y}^T\mathbf{AY}$ that may be either positive definite or positive semidefinite is called *nonnegative definite.*

- The matrix $\mathbf{A}$ of a quadratic form $\mathbf{Y}^T\mathbf{AY}$ is said to be positive definite (semidefinite) when the quadratic form is positive definite (semidefinite).

- If $\mathbf{P}$ is a nonsingular matrix and if $\mathbf{A}$ is positive definite (semidefinite), then $\mathbf{P}^T\mathbf{AP}$ is positive definite (semidefinite).

- A necessary and sufficient condition for the symmetric matrix $\mathbf{A}$ to be positive definite is that there exist a nonsingular matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{PP}^T$.

- A necessary and sufficient condition that the matrix $\mathbf{A}$ be positive definite, where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

is that the following inequalities hold:

$$a_{11} > 0, \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \ldots \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} > 0.$$

- If $\mathbf{A}$ is an $m \times n$ matrix of rank $n < m$, then $\mathbf{A}^T\mathbf{A}$ is positive definite and $\mathbf{AA}^T$ is positive semidefinite.

- If $\mathbf{A}$ is an $m \times n$ matrix of rank $k < n$ and $k < m$, then $\mathbf{A}^T\mathbf{A}$ and $\mathbf{AA}^T$ are each positive semidefinite.

- A matrix that may be either positive definite or positive semidefinite is said to be nonnegative definite.

- If $\mathbf{A}$ and $\mathbf{B}$ are symmetric conformable matrices, $\mathbf{A}$ is said to be greater than $\mathbf{B}$ if $\mathbf{A} - \mathbf{B}$ is nonnegative definite.

B.4 DETERMINANTS

- For each square matrix $\mathbf{A}$, there is a uniquely defined scalar called the *determinant* of $\mathbf{A}$ and denoted by $|\mathbf{A}|$.

- The determinant of a diagonal matrix is equal to the product of the diagonal elements.
- If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ matrices, then $|\mathbf{AB}| = |\mathbf{BA}| = |\mathbf{A}|\,|\mathbf{B}|$.
- If $\mathbf{A}$ is singular if and only if $|\mathbf{A}| = 0$.
- If $\mathbf{C}$ is an $n \times n$ matrix such that $\mathbf{C}^T\mathbf{C} = \mathbf{I}$, then $\mathbf{C}$ is said to be an *orthogonal* matrix, and $\mathbf{C}^T = \mathbf{C}^{-1}$.
- If $\mathbf{C}$ is an orthogonal matrix, then $|\mathbf{C}| = +1$ or $|\mathbf{C}| = -1$.
- If $\mathbf{C}$ is an orthogonal matrix, then $|\mathbf{C}^T\mathbf{AC}| = |\mathbf{A}|$.
- The determinant of a positive definite matrix is positive.
- The determinant of a triangular matrix is equal to the product of the diagonal elements.
- The determinant of a matrix is equal to the product of its eigenvalues.
- $|\mathbf{A}| = |\mathbf{A}^T|$
- $|\mathbf{A}^{-1}| = 1 \,/\, |\mathbf{A}|$, if $|\mathbf{A}| \neq 0$.
- If $\mathbf{A}$ is a square matrix such that

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

  where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square matrices, and if $\mathbf{A}_{12} = 0$ or $\mathbf{A}_{21} = 0$, then $|\mathbf{A}| = |\mathbf{A}_{11}|\,|\mathbf{A}_{22}|$.
- If $\mathbf{A}_1$ and $\mathbf{A}_2$ are symmetric and $\mathbf{A}_2$ is positive definite and if $\mathbf{A}_1 - \mathbf{A}_2$ is positive semidefinite (or positive definite), then $|\mathbf{A}_1| \geq |\mathbf{A}_2|$.

B.5 MATRIX TRACE

- The *trace* of a matrix $\mathbf{A}$, which will be written $\mathrm{tr}\,(\mathbf{A})$, is equal to the sum of the diagonal elements of $\mathbf{A}$; that is,

$$tr\,(\mathbf{A}) = \sum_{i=1}^{n} a_{ii}. \tag{B.5.1}$$

- tr(AB) = tr(BA).

- tr(ABC) = tr(CAB) = tr(BCA); that is, the trace of the product of matrices is invariant under any cyclic permutation of the matrices.
- Note that the trace is defined only for a square matrix.
- If $\mathbf{C}$ is an orthogonal matrix, $\mathrm{tr}\,(\mathbf{C}^T\mathbf{AC}) = \mathrm{tr}\,(\mathbf{A})$.

B.6 EIGENVALUES AND EIGENVECTORS

- A *characteristic root* (*eigenvalue*) of a $p \times p$ matrix $\mathbf{A}$ is a scalar λ such that $\mathbf{AX} = \lambda\mathbf{X}$ for some vector $\mathbf{X} \neq 0$.
- The vector $\mathbf{X}$ is called the *characteristic vector* (*eigenvector*) of the matrix $\mathbf{A}$.
- The characteristic root of a matrix $\mathbf{A}$ can be defined as a scalar λ such that $|\mathbf{A} - \lambda\,\mathbf{I}| = 0$.
- $|\mathbf{A} - \lambda\,\mathbf{I}|$ is a pth degree polynomial in λ.
- This polynomial is called the *characteristic polynomial*, and its roots are the characteristic roots of the matrix $\mathbf{A}$.
- The number of nonzero characteristic roots of a matrix $\mathbf{A}$ is equal to the rank of $\mathbf{A}$.
- The characteristic roots of $\mathbf{A}$ are identical with the characteristic roots of $\mathbf{CAC}^{-1}$. If $\mathbf{C}$ is an orthogonal matrix, it follows that $\mathbf{A}$ and $\mathbf{CAC}^T$ have identical characteristic roots; that is, $\mathbf{C}^T = \mathbf{C}^{-1}$.
- The characteristic roots of a symmetric matrix are real; that is, if $\mathbf{A} = \mathbf{A}^T$, the characteristic polynomial of $|\mathbf{A} = \lambda\,\mathbf{I}| = 0$ has all real roots.
- The characteristic roots of a positive definite matrix $\mathbf{A}$ are positive; the characteristic roots of a positive semidefinite matrix are nonnegative.

B.7 THE DERIVATIVES OF MATRICES AND VECTORS

- Let $\mathbf{X}$ be an $n \times 1$ vector and let Z be a scalar that is a function of $\mathbf{X}$. The derivative of Z with respect to the vector $\mathbf{X}$, which will be written $\partial Z \,/\, \partial\mathbf{X}$, will mean the $1 \times n$ row vector*

$$\mathbf{C} \equiv \left[\frac{\partial Z}{\partial x_1} \; \frac{\partial Z}{\partial x_2} \cdots \frac{\partial Z}{\partial x_n} \right]. \tag{B.7.1}$$

*Generally this partial derivative would be defined as a column vector. However, it is defined as a row vector here because we have defined $\widetilde{H} = \frac{\partial G(\mathbf{X})}{\partial \mathbf{X}}$ as a row vector in the text.

- If $\mathbf{X}$, $\mathbf{C}$, and Z are as defined previously, then

$$\partial Z \,/\, \partial \mathbf{X} \;=\; \mathbf{C}. \tag{B.7.2}$$

- If $\mathbf{A}$ and $\mathbf{B}$ are $m \times 1$ vectors, which are a function of the $n \times 1$ vector $\mathbf{X}$, and we define

$$\frac{\partial(\mathbf{A}^T\mathbf{B})}{\partial \mathbf{X}}$$

to be a row vector as in Eq. (B.7.1), then

$$\partial(\mathbf{A}^T\mathbf{B}) \,/\, \partial \mathbf{X} \;=\; \mathbf{B}^T \frac{\partial \mathbf{A}}{\partial \mathbf{X}} + \mathbf{A}^T \frac{\partial \mathbf{B}}{\partial \mathbf{X}} \tag{B.7.3}$$

where

$$\frac{\partial \mathbf{A}}{\partial \mathbf{X}}$$

is an $m \times n$ matrix whose ij element is

$$\frac{\partial A_i}{\partial X_j}$$

and

$$\frac{\partial(\mathbf{A}^T\mathbf{B})}{\partial \mathbf{X}}$$

is a $1 \times n$ row vector.†

- If $\mathbf{A}$ is an $m \times 1$ vector that is a function of the $n \times 1$ vector $\mathbf{X}$, and W is an $m \times m$ symmetric matrix such that

$$Z = \mathbf{A}^T W \mathbf{A} = \mathbf{A}^T W^{1/2} W^{1/2} \mathbf{A}.$$

Let $\mathbf{B} \equiv W^{1/2}\mathbf{A}$, then

$$Z = \mathbf{B}^T \mathbf{B}.$$

From Eq. (B.7.3)

$$\frac{\partial Z}{\partial \mathbf{X}} = 2\mathbf{B}^T \frac{\partial \mathbf{B}}{\partial \mathbf{X}} \tag{B.7.4}$$

where

$$\frac{\partial \mathbf{B}}{\partial \mathbf{X}} = W^{1/2} \frac{\partial \mathbf{A}}{\partial \mathbf{X}}.$$

---

†If $\frac{\partial Z}{\partial \mathbf{X}}$ is defined to be a column vector, $\frac{\partial(A^T B)}{\partial \mathbf{X}}$ would be given by the transpose of Eq. (B.7.3).

- Let $\mathbf{A}$ be a $p \times 1$ vector, $\mathbf{B}$ be a $q \times 1$ vector, and C be a $p \times q$ matrix whose ij^{th} element equals c_{ij}. Let

$$Z = \mathbf{A}^T C \mathbf{B} = \sum_{m=1}^{q} \sum_{n=1}^{p} a_n c_{nm} b_m. \tag{B.7.5}$$

Then $\partial Z / \partial C = \mathbf{A}\mathbf{B}^T$.

Proof: $\partial Z / \partial C$ will be a $p \times q$ matrix whose ij^{th} element is $\partial Z / \partial c_{ij}$.

Assuming that C is not symmetric and that the elements of C are independent,

$$\frac{\partial Z}{\partial c_{ij}} = \frac{\partial \left(\sum_{m=1}^{q} \sum_{n=1}^{p} a_n c_{nm} b_m \right)}{\partial c_{ij}} = a_i \, b_j. \tag{B.7.6}$$

Thus the ij^{th} element of $\partial Z / \partial C$ is $a_i b_j$. Therefore, it follows that

$$\frac{\partial Z}{\partial C} = \mathbf{A}\mathbf{B}^T.$$

- The derivative of a matrix product with respect to a scalar is given by

$$\frac{d}{dt}\{\mathbf{A}(t)\mathbf{B}(t)\} = \frac{d\mathbf{A}(t)}{dt}\mathbf{B}(t) + \mathbf{A}(t)\frac{d\mathbf{B}(t)}{dt}. \tag{B.7.7}$$

See Graybill (1961) for additional discussion of the derivatives of matrices and vectors.

B.8 MAXIMA AND MINIMA

- If $y = f(x_1, x_2, ..., x_n)$ is a function of n variables and if all partial derivatives $\partial y / \partial x_i$ are continuous, then y attains its maxima and minima only at the points where

$$\frac{\partial y}{\partial x_1} = \frac{\partial y}{\partial x_2} = ... = \frac{\partial y}{\partial x_n} = 0. \tag{B.8.1}$$

- If $f(x_1, x_2, ..., x_n)$ is such that all the first and second partial derivatives are continuous, then at the point where

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = ... = \frac{\partial f}{\partial x_n} = 0 \tag{B.8.2}$$

the function has

- a minimum, if the matrix $\mathbf{K}$, where the ijth element of $\mathbf{K}$ is $\partial^2 f \,/\, \partial x_i \partial x_j$, is positive definite.
- a maximum, if the matrix $-\mathbf{K}$ is positive definite.

In these two theorems on maxima and minima, remember that the x_i are independent variables.

- If the x_i are not independent, that is, there are constraints relating them, we use the method of Lagrange multipliers. Suppose that we have a function $f(x_1, x_2, \ldots, x_n)$ we wish to maximize (or minimize) subject to the constraint that $h(x_1, x_2, \ldots, x_n) = 0$. The equation $h = 0$ describes a surface in space and the problem is one of maximizing $f(x_1, x_2, \ldots, x_n)$ as $x_1, x_2, \ldots, x_n$ vary on the curve of intersection of the two surfaces. At a maximum point the derivative of f must be zero along the intersection curve; that is, the directional derivative along the tangent must be zero. The directional derivative is the component of the vector ∇f along the tangent. Hence, ∇f must lie in a plane normal to the intersection curve at this point. This plane must also contain ∇h; that is, ∇f and ∇h are coplanar at this point. Hence, there must exist a scalar λ such that

$$\nabla f + \lambda \nabla h = 0 \tag{B.8.3}$$

at the maximum point. If we define

$$F \equiv f + \lambda h$$

then Eq. (B.8.3) is equivalent to $\nabla F = 0$. Hence,

$$\frac{\partial F}{\partial x_1} = \frac{\partial F}{\partial x_2} = \cdots = \frac{\partial F}{\partial x_n} = 0.$$

These n equations together with $h = 0$ provide us with $n + 1$ equations and $n + 1$ unknowns $(x_1, x_2, \ldots, x_n, \lambda)$. We have assumed that all first partial derivatives are continuous and that $\partial h / \partial x_i \neq 0$ for all i at the point.

- If there are additional constraints we introduce additional Lagrange multipliers in Eq. (B.8.3); for example,

$$\nabla f + \lambda_1 \nabla h_1 + \lambda_2 \nabla h_2 + \cdots \lambda_k \nabla h_k = 0. \tag{B.8.4}$$

B.9 USEFUL MATRIX INVERSION THEOREMS

Theorem 1: Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ positive definite (PD) matrices. If $\mathbf{A}^{-1} + \mathbf{B}^{-1}$ is PD, then $\mathbf{A} + \mathbf{B}$ is PD and

$$(\mathbf{A} + \mathbf{B})^{-1} = \mathbf{B}^{-1} (\mathbf{A}^{-1} + \mathbf{B}^{-1})^{-1} \mathbf{A}^{-1}$$

$$= \mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{B}^{-1}. \tag{B.9.1}$$

Proof: From the identity

$$(\mathbf{A}+\mathbf{B})^{-1} = [\mathbf{A}(\mathbf{A}^{-1}+\mathbf{B}^{-1})\mathbf{B}]^{-1} = \mathbf{B}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1}$$

or

$$(\mathbf{A}+\mathbf{B})^{-1} = [\mathbf{B}(\mathbf{B}^{-1}+\mathbf{A}^{-1})\mathbf{A}]^{-1} = \mathbf{A}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{B}^{-1}.$$

Theorem 2: Let $\mathbf{A}$ and $\mathbf{B}$ be $n \times n$ PD matrices. If $\mathbf{A}+\mathbf{B}$ is PD, then $\mathbf{I}+\mathbf{A}\mathbf{B}^{-1}$ and $\mathbf{I}+\mathbf{B}\mathbf{A}^{-1}$ are PD and

$$\begin{aligned}(\mathbf{A}+\mathbf{B})^{-1} &= \mathbf{B}^{-1} - \mathbf{B}^{-1}(\mathbf{I}+\mathbf{A}\mathbf{B}^{-1})^{-1}\mathbf{A}\mathbf{B}^{-1}\\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{I}+\mathbf{B}\mathbf{A}^{-1})^{-1}\mathbf{B}\mathbf{A}^{-1}.\end{aligned} \tag{B.9.2}$$

Proof: From the identity

$$\mathbf{A}^{-1} = (\mathbf{A}^{-1}+\mathbf{B}^{-1}) - \mathbf{B}^{-1}$$

premultiply by $\mathbf{B}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}$ and use Theorem 1

$$\begin{aligned}\mathbf{B}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{A}^{-1} &= \mathbf{B}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})\\ &\quad -\mathbf{B}^{-1}(\mathbf{A}^{-1}+\mathbf{B}^{-1})^{-1}\mathbf{B}^{-1}\\ &= \mathbf{B}^{-1} - \mathbf{B}^{-1}[\mathbf{A}^{-1}(\mathbf{I}+\mathbf{A}\mathbf{B}^{-1})]^{-1}\mathbf{B}^{-1}\\ &= \mathbf{B}^{-1} - \mathbf{B}^{-1}(\mathbf{I}+\mathbf{A}\mathbf{B}^{-1})^{-1}\mathbf{A}\mathbf{B}^{-1}.\end{aligned}$$

The left-hand side of this equation is $(\mathbf{A}+\mathbf{B})^{-1}$ (from Theorem 1). Hence,

$$(\mathbf{A}+\mathbf{B})^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1}(\mathbf{I}+\mathbf{A}\mathbf{B}^{-1})^{-1}\mathbf{A}\mathbf{B}^{-1}.$$

Theorem 3: If $\mathbf{A}$ and $\mathbf{B}$ are PD matrices of order n and m, respectively, and if $\mathbf{C}$ is of order $n \times m$, then

$$(\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}+\mathbf{B}^{-1})^{-1}\mathbf{C}^T\mathbf{A}^{-1} = \mathbf{B}\mathbf{C}^T(\mathbf{A}+\mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1} \tag{B.9.3}$$

provided the inverse exists.

Proof: From the identity

$$\mathbf{C}^T(\mathbf{A}^{-1}\mathbf{C}\mathbf{B}\mathbf{C}^T+\mathbf{I})(\mathbf{I}+\mathbf{A}^{-1}\mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1} \equiv \mathbf{C}^T$$

we have

$$(\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}\mathbf{B}\mathbf{C}^T + \mathbf{C}^T)(\mathbf{A}^{-1}(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^T))^{-1} = \mathbf{C}^T$$

or

$$(\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C} + \mathbf{B}^{-1})\mathbf{B}\mathbf{C}^T(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1}\mathbf{A} = \mathbf{C}^T.$$

Now premultiply by $(\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C}+\mathbf{B}^{-1})^{-1}$ and postmultiply by $\mathbf{A}^{-1}$, which yields

$$\mathbf{B}\mathbf{C}^T(\mathbf{A} + \mathbf{C}\mathbf{B}\mathbf{C}^T)^{-1} = (\mathbf{C}^T\mathbf{A}^{-1}\mathbf{C} + \mathbf{B}^{-1})^{-1}\mathbf{C}^T\mathbf{A}^{-1}.$$

Theorem 4: The Schur Identity or insideout rule. If $\mathbf{A}$ is a PD matrix of order n, and if $\mathbf{B}$ and $\mathbf{C}$ are any conformable matrices such that $\mathbf{BC}$ is order n, then

$$(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \tag{B.9.4}$$

Proof: Define

$$\mathbf{X} = (\mathbf{A} + \mathbf{BC})^{-1}.$$

Then

$$\begin{aligned} (\mathbf{A} + \mathbf{BC})\,\mathbf{X} &= \mathbf{I} \qquad \text{(B.9.5)} \\ \mathbf{AX} + \mathbf{BCX} &= \mathbf{I}. \end{aligned}$$

Solve Eq. (B.9.5) for $\mathbf{CX}$. First multiply by $\mathbf{A}^{-1}$ to yield

$$\mathbf{X} + \mathbf{A}^{-1}\mathbf{BCX} = \mathbf{A}^{-1}. \tag{B.9.6}$$

Premultiply Eq. (B.9.6) by $\mathbf{C}$

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

Then

$$\mathbf{CX} = (\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}. \tag{B.9.7}$$

Substitute Eq. (B.9.7) into Eq. (B.9.6) to yield

$$\mathbf{X} = (\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}\,(\mathbf{I} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1}.$$

B.10 REFERENCE

Graybill, F. A., *An Introduction to Linear Statistical Models*, McGraw-Hill, New York, 1961.