APPENDIX A

MATRIX ALGEBRA

A.1 TERMINOLOGY

A **matrix** is a rectangular array of numbers, denoted

$$\mathbf{A} = [a_{ik}] = [\mathbf{A}]_{ik} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ & & \cdots & \\ a_{n1} & a_{n2} & \cdots & a_{nK} \end{bmatrix}. \quad \textbf{(A-1)}$$

The typical element is used to denote the matrix. A subscripted element of a matrix is always read as $a_{\text{row, column}}$. An example is given in Table A.1. In these data, the rows are identified with years and the columns with particular variables.

A vector is an ordered set of numbers arranged either in a row or a column. In view of the preceding, a **row vector** is also a matrix with one row, whereas a **column vector** is a matrix with one column. Thus, in Table A.1, the five variables observed for 1972 (including the date) constitute a row vector, whereas the time series of nine values for consumption is a column vector.

A matrix can also be viewed as a set of column vectors or as a set of row vectors.[1] The **dimensions** of a matrix are the numbers of rows and columns it contains. "**A** is an $n \times K$ matrix" (read "n by K") will always mean that **A** has n rows and K columns. If n equals K, then **A** is a **square matrix**. Several particular types of square matrices occur frequently in econometrics.

- A **symmetric matrix** is one in which $a_{ik} = a_{ki}$ for all i and k.
- A **diagonal matrix** is a square matrix whose only nonzero elements appear on the **main diagonal**, that is, moving from upper left to lower right.
- A **scalar matrix** is a diagonal matrix with the same value in all diagonal elements.
- An **identity matrix** is a scalar matrix with ones on the diagonal. This matrix is always denoted **I**. A subscript is sometimes included to indicate its size, or **order**. For example, $\mathbf{I}_4$ indicates a 4×4 identity matrix.
- A **triangular matrix** is one that has only zeros either above or below the main diagonal. If the zeros are above the diagonal, the matrix is **lower triangular**.

[1]Henceforth, we shall denote a matrix by a boldfaced capital letter, as is **A** in (A-1), and a vector as a boldfaced lowercase letter, as in **a**. Unless otherwise noted, a vector will always be assumed to be a *column vector*.

TABLE A.1 Matrix of Macroeconomic Data

	Column				
Row	*1 Year*	*2 Consumption (billions of dollars)*	*3 GNP (billions of dollars)*	*4 GNP Deflator*	*5 Discount Rate (N.Y Fed., avg.)*
1	1972	737.1	1185.9	1.0000	4.50
2	1973	812.0	1326.4	1.0575	6.44
3	1974	808.1	1434.2	1.1508	7.83
4	1975	976.4	1549.2	1.2579	6.25
5	1976	1084.3	1718.0	1.3234	5.50
6	1977	1204.4	1918.3	1.4005	5.46
7	1978	1346.5	2163.9	1.5042	7.46
8	1979	1507.2	2417.8	1.6342	10.28
9	1980	1667.2	2633.1	1.7864	11.77

Source: Data from the *Economic Report of the President* (Washington, D.C.: U.S. Government Printing Office, 1983).

A.2 ALGEBRAIC MANIPULATION OF MATRICES

A.2.1 EQUALITY OF MATRICES

Matrices (or vectors) **A** and **B** are equal if and only if they have the same dimensions and each element of **A** equals the corresponding element of **B**. That is,

$$\mathbf{A} = \mathbf{B} \quad \text{if and only if } a_{ik} = b_{ik} \quad \text{for all } i \text{ and } k. \tag{A-2}$$

A.2.2 TRANSPOSITION

The **transpose** of a matrix **A**, denoted $\mathbf{A}'$, is obtained by creating the matrix whose *kth* row is the *kth* column of the original matrix.[2] Thus, if $\mathbf{B} = \mathbf{A}'$, then each column of **A** will appear as the corresponding row of **B**. If **A** is $n \times K$, then $\mathbf{A}'$ is $K \times n$.

An equivalent definition of the transpose of a matrix is

$$\mathbf{B} = \mathbf{A}' \Leftrightarrow b_{ik} = a_{ki} \quad \text{for all } i \text{ and } k. \tag{A-3}$$

The definition of a symmetric matrix implies that

$$\text{if (and only if) } \mathbf{A} \text{ is symmetric, then } \mathbf{A} = \mathbf{A}'. \tag{A-4}$$

It also follows from the definition that for any **A**,

$$(\mathbf{A}')' = \mathbf{A}. \tag{A-5}$$

Finally, the transpose of a column vector, **a**, is a row vector:

$$\mathbf{a}' = [a_1 \quad a_2 \quad \cdots \quad a_n].$$

[2]Authors sometimes denote the transpose of a matrix with a superscript "T," as in $\mathbf{A}^{\mathrm{T}}$ = the transpose of **A**. We will use the prime notation throughout this book .

A.2.3 VECTORIZATION

In some derivations and analyses, it is occasionally useful to reconfigure a matrix into a vector (rarely the reverse). The matrix function Vec(**A**) takes the columns of an $n \times K$ matrix and rearranges them in a long $nK \times 1$ vector. Thus, $Vec\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = [1, 2, 2, 4]'$. A related operation is the half vectorization, which collects the lower triangle of a symmetric matrix in a column vector. For example, $Vech\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$.

A.2.4 MATRIX ADDITION

The operations of addition and subtraction are extended to matrices by defining

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = [a_{ik} + b_{ik}]. \tag{A-6}$$

$$\mathbf{A} - \mathbf{B} = [a_{ik} - b_{ik}]. \tag{A-7}$$

Matrices cannot be added unless they have the same dimensions, in which case they are said to be **conformable for addition**. A **zero matrix** or **null matrix** is one whose elements are all zero. In the addition of matrices, the zero matrix plays the same role as the scalar 0 in scalar addition; that is,

$$\mathbf{A} + \mathbf{0} = \mathbf{A}. \tag{A-8}$$

It follows from (A-6) that matrix addition is commutative,

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}. \tag{A-9}$$

and associative,

$$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}), \tag{A-10}$$

and that

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'. \tag{A-11}$$

A.2.5 VECTOR MULTIPLICATION

Matrices are multiplied by using the **inner product**. The inner product, or **dot product**, of two vectors, **a** and **b**, is a scalar and is written

$$\mathbf{a}'\mathbf{b} = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \Sigma_{j=1}^{n} a_jb_j. \tag{A-12}$$

Note that the inner product is written as the transpose of vector **a** times vector **b**, a row vector times a column vector. In (A-12), each term a_jb_j equals b_ja_j; hence

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a}. \tag{A-13}$$

A.2.6 A NOTATION FOR ROWS AND COLUMNS OF A MATRIX

We need a notation for the ith row of a matrix. Throughout this book, an untransposed vector will always be a column vector. However, we will often require a notation for the column vector that is the transpose of a row of a matrix. This has the potential to create some ambiguity, but the following convention based on the subscripts will suffice for our work throughout this text:

- $\mathbf{a}_k$, or $\mathbf{a}_l$ or $\mathbf{a}_m$ will denote *column k, l,* or *m* of the matrix **A**,
- $\mathbf{a}_i$, or $\mathbf{a}_j$ or $\mathbf{a}_t$ or $\mathbf{a}_s$ will denote the column vector formed by the transpose of row i, j, t, or s of matrix **A**. Thus, $\mathbf{a}_i'$ is row i of **A**. **(A-14)**

For example, from the data in Table A.1 it might be convenient to speak of $\mathbf{x}_i$, where $i = 1972$ as the 5×1 vector containing the five variables measured for the year 1972, that is, the transpose of the 1972 row of the matrix. In our applications, the common association of subscripts "i" and "j" with individual i or j, and "t" and "s" with time periods t and s will be natural.

A.2.7 MATRIX MULTIPLICATION AND SCALAR MULTIPLICATION

For an $n \times K$ matrix **A** and a $K \times M$ matrix **B**, the product matrix, $\mathbf{C} = \mathbf{AB}$, is an $n \times M$ matrix whose ikth element is the inner product of row i of **A** and column k of **B**. Thus, the product matrix **C** is

$$\mathbf{C} = \mathbf{AB} \Rightarrow c_{ik} = \mathbf{a}_i'\mathbf{b}_k. \tag{A-15}$$

[Note our use of (A-14) in (A-15).] To multiply two matrices, the number of columns in the first must be the same as the number of rows in the second, in which case they are **conformable for multiplication**.[3] Multiplication of matrices is generally not commutative. In some cases, **AB** may exist, but **BA** may be undefined or, if it does exist, may have different dimensions. In general, however, even if **AB** and **BA** do have the same dimensions, they will not be equal. In view of this, we define **premultiplication** and **postmultiplication** of matrices. In the product **AB**, **B** is *premultiplied* by **A**, whereas **A** is *postmultiplied* by **B**.

Scalar multiplication of a matrix is the operation of multiplying every element of the matrix by a given scalar. For scalar c and matrix **A**,

$$c\mathbf{A} = [ca_{ik}]. \tag{A-16}$$

If two matrices **A** and **B** have the same number of rows and columns, then we can compute the **direct product** (also called the **Hadamard product** or the Schur product or the entrywise product), which is a new matrix (or vector) whose ij element is the product of the corresponding elements of **A** and **B**. The usual symbol for this operation is "$\circ$." Thus,

$$\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \circ \begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} 1a & 2b \\ 2b & 3c \end{bmatrix} \text{ and } \begin{pmatrix} 3 \\ 5 \end{pmatrix} \circ \begin{pmatrix} 2 \\ 4 \end{pmatrix} = \begin{pmatrix} 6 \\ 20 \end{pmatrix}.$$

The product of a matrix and a vector is written

$$\mathbf{c} = \mathbf{Ab}.$$

[3]A simple way to check the conformability of two matrices for multiplication is to write down the dimensions of the operation, for example, $(n \times K)$ times $(K \times M)$. The inner dimensions must be equal; the result has dimensions equal to the outer values.

The number of elements in **b** must equal the number of columns in **A**; the result is a vector with number of elements equal to the number of rows in **A**. For example,

$$\begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 6 & 1 \\ 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix}.$$

We can interpret this in two ways. First, it is a compact way of writing the three equations

$$\begin{aligned} 5 &= 4a + 2b + 1c, \\ 4 &= 2a + 6b + 1c, \\ 1 &= 1a + 1b + 0c. \end{aligned}$$

Second, by writing the set of equations as

$$\begin{bmatrix} 5 \\ 4 \\ 1 \end{bmatrix} = a\begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix} + b\begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix} + c\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix},$$

we see that the right-hand side is a **linear combination** of the columns of the matrix where the coefficients are the elements of the vector. For the general case,

$$\mathbf{c} = \mathbf{A}\mathbf{b} = b_1\mathbf{a}_1 + b_2\mathbf{a}_2 + \cdots + b_K\mathbf{a}_K. \tag{A-17}$$

In the calculation of a matrix product $\mathbf{C} = \mathbf{AB}$, each column of $\mathbf{C}$ is a linear combination of the columns of $\mathbf{A}$, where the coefficients are the elements in the corresponding column of $\mathbf{B}$. That is,

$$\mathbf{C} = \mathbf{AB} \Leftrightarrow \mathbf{c}_k = \mathbf{A}\mathbf{b}_k. \tag{A-18}$$

Let $\mathbf{e}_k$ be a column vector that has zeros everywhere except for a one in the *kth* position. Then $\mathbf{A}\mathbf{e}_k$ is a linear combination of the columns of $\mathbf{A}$ in which the coefficient on every column but the kth is zero, whereas that on the kth is one. The result is

$$\mathbf{a}_k = \mathbf{A}\mathbf{e}_k. \tag{A-19}$$

Combining this result with (A-17) produces

$$(\mathbf{a}_1 \quad \mathbf{a}_2 \quad \cdots \quad \mathbf{a}_n) = \mathbf{A}(\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_n) = \mathbf{AI} = \mathbf{A}. \tag{A-20}$$

In matrix multiplication, the identity matrix is analogous to the scalar 1. For any matrix or vector $\mathbf{A}$, $\mathbf{AI} = \mathbf{A}$. In addition, $\mathbf{IA} = \mathbf{A}$, although if $\mathbf{A}$ is not a square matrix, the two identity matrices are of different orders.

A conformable matrix of zeros produces the expected result: $\mathbf{A0} = \mathbf{0}$.

Some general rules for matrix multiplication are as follows:

- **Associative law:** $(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$. **(A-21)**
- **Distributive law:** $\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC}$. **(A-22)**
- **Transpose of a product:** $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$. **(A-23)**
- **Transpose of an extended product:** $(\mathbf{ABC})' = \mathbf{C}'\mathbf{B}'\mathbf{A}'$. **(A-24)**

A.2.8 SUMS OF VALUES

Denote by **i** a vector that contains a column of ones. Then,

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n = \mathbf{i}'\mathbf{x}. \tag{A-25}$$

If all elements in $\mathbf{x}$ are equal to the same constant a, then $\mathbf{x} = a\mathbf{i}$ and

$$\sum_{i=1}^{n} x_i = \mathbf{i}'(a\mathbf{i}) = a(\mathbf{i}'\mathbf{i}) = na. \qquad \textbf{(A-26)}$$

For any constant a and vector $\mathbf{x}$,

$$\sum_{i=1}^{n} ax_i = a\sum_{i=1}^{n} x_i = a\mathbf{i}'\mathbf{x}. \qquad \textbf{(A-27)}$$

If $a = 1/n$, then we obtain the arithmetic mean,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\mathbf{i}'\mathbf{x}, \qquad \textbf{(A-28)}$$

from which it follows that

$$\sum_{i=1}^{n} x_i = \mathbf{i}'\mathbf{x} = n\bar{x}.$$

The sum of squares of the elements in a vector $\mathbf{x}$ is

$$\sum_{i=1}^{n} x_i^2 = \mathbf{x}'\mathbf{x}; \qquad \textbf{(A-29)}$$

while the sum of the products of the n elements in vectors $\mathbf{x}$ and $\mathbf{y}$ is

$$\sum_{i=1}^{n} x_i y_i = \mathbf{x}'\mathbf{y}. \qquad \textbf{(A-30)}$$

By the definition of matrix multiplication,

$$[\mathbf{X}'\mathbf{X}]_{kl} = [\mathbf{x}_k'\mathbf{x}_l] \qquad \textbf{(A-31)}$$

is the inner product of the kth and lth columns of $\mathbf{X}$. For example, for the data set given in Table A.1, if we define $\mathbf{X}$ as the 9×3 matrix containing (year, consumption, GNP), then

$$\begin{aligned}[\mathbf{X}'\mathbf{X}]_{23} &= \sum_{t=1972}^{1980} \text{consumption}_t \, \text{GNP}_t = 737.1(1185.9) + \cdots + 1667.2(2633.1) \\ &= 19{,}743{,}711.34.\end{aligned}$$

If $\mathbf{X}$ is $n \times K$, then [again using (A-14)]

$$\mathbf{X}'\mathbf{X} = \sum_{i=1}^{n} \mathbf{x}_i\mathbf{x}_i'.$$

This form shows that the $K \times K$ matrix $\mathbf{X}'\mathbf{X}$ is the sum of n $K \times K$ matrices, each formed from a single row (year) of $\mathbf{X}$. For the example given earlier, this sum is of nine 3×3 matrices, each formed from one row (year) of the original data matrix.

A.2.9 A USEFUL IDEMPOTENT MATRIX

A fundamental matrix in statistics is the "centering matrix" that is used to transform data to deviations from their mean. First,

$$\mathbf{i}\bar{x} = \mathbf{i}\frac{1}{n}\mathbf{i}'\mathbf{x} = \begin{bmatrix} \bar{x} \\ \bar{x} \\ \vdots \\ \bar{x} \end{bmatrix} = \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x}. \qquad \textbf{(A-32)}$$

The matrix $(1/n)\mathbf{i}\mathbf{i}'$ is an $n \times n$ matrix with every element equal to $1/n$. The set of values in deviations form is

$$\begin{bmatrix} x_1 - \bar{x} \\ x_2 - \bar{x} \\ \cdots \\ x_n - \bar{x} \end{bmatrix} = [\mathbf{x} - \mathbf{i}\bar{x}] = \left[\mathbf{x} - \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x}\right]. \tag{A-33}$$

Because $\mathbf{x} = \mathbf{I}\mathbf{x}$,

$$\left[\mathbf{x} - \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x}\right] = \left[\mathbf{I}\mathbf{x} - \frac{1}{n}\mathbf{i}\mathbf{i}'\mathbf{x}\right] = \left[\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}'\right]\mathbf{x} = \mathbf{M}^0\mathbf{x}. \tag{A-34}$$

Henceforth, the symbol $\mathbf{M}^0$ will be used only for this matrix. Its diagonal elements are all $(1 - 1/n)$, and its off-diagonal elements are $-1/n$. The matrix $\mathbf{M}^0$ is primarily useful in computing sums of squared deviations. Some computations are simplified by the result

$$\mathbf{M}^0\mathbf{i} = \left[\mathbf{I} - \frac{1}{n}\mathbf{i}\mathbf{i}'\right]\mathbf{i} = \mathbf{i} - \frac{1}{n}\mathbf{i}(\mathbf{i}'\mathbf{i}) = \mathbf{0},$$

which implies that $\mathbf{i}'\mathbf{M}^0 = \mathbf{0}'$. The sum of deviations about the mean is then

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \mathbf{i}'[\mathbf{M}^0\mathbf{x}] = \mathbf{0}'\mathbf{x} = 0. \tag{A-35}$$

For a single variable $\mathbf{x}$, the sum of squared deviations about the mean is

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \left(\sum_{i=1}^{n}x_i^2\right) - n\bar{x}^2. \tag{A-36}$$

In matrix terms,

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = (\mathbf{x} - \bar{x}\mathbf{i})'(\mathbf{x} - \bar{x}\mathbf{i}) = (\mathbf{M}^0\mathbf{x})'(\mathbf{M}^0\mathbf{x}) = \mathbf{x}'\mathbf{M}^{0\prime}\mathbf{M}^0\mathbf{x}.$$

Two properties of $\mathbf{M}^0$ are useful at this point. First, because all off-diagonal elements of $\mathbf{M}^0$ equal $-1/n$, $\mathbf{M}^0$ is symmetric. Second, as can easily be verified by multiplication, $\mathbf{M}^0$ is equal to its square; $\mathbf{M}^0\mathbf{M}^0 = \mathbf{M}^0$.

DEFINITION A.1 Idempotent Matrix
*An **idempotent** matrix,* $\mathbf{M}$, *is one that is equal to its square, that is,* $\mathbf{M}^2 = \mathbf{M}\mathbf{M} = \mathbf{M}$. *If* $\mathbf{M}$ *is a symmetric idempotent matrix (all of the idempotent matrices we shall encounter are symmetric), then* $\mathbf{M}'\mathbf{M} = \mathbf{M}$ as well.

Thus, $\mathbf{M}^0$ is a symmetric idempotent matrix. Combining results, we obtain

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \mathbf{x}'\mathbf{M}^0\mathbf{x}. \tag{A-37}$$

Consider constructing a matrix of sums of squares and cross products in deviations from the column means. For two vectors $\mathbf{x}$ and $\mathbf{y}$,

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = (\mathbf{M}^0\mathbf{x})'(\mathbf{M}^0\mathbf{y}), \quad \textbf{(A-38)}$$

so

$$\begin{bmatrix} \sum_{i=1}^{n}(x_i - \bar{x})^2 & \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) \\ \sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x}) & \sum_{i=1}^{n}(y_i - \bar{y})^2 \end{bmatrix} = \begin{bmatrix} \mathbf{x}'\mathbf{M}^0\mathbf{x} & \mathbf{x}'\mathbf{M}^0\mathbf{y} \\ \mathbf{y}'\mathbf{M}^0\mathbf{x} & \mathbf{y}'\mathbf{M}^0\mathbf{y} \end{bmatrix}. \quad \textbf{(A-39)}$$

If we put the two column vectors $\mathbf{x}$ and $\mathbf{y}$ in an $n \times 2$ matrix $\mathbf{Z} = [\mathbf{x}, \mathbf{y}]$, then $\mathbf{M}^0\mathbf{Z}$ is the $n \times 2$ matrix in which the two columns of data are in mean deviation form. Then

$$(\mathbf{M}^0\mathbf{Z})'(\mathbf{M}^0\mathbf{Z}) = \mathbf{Z}'\mathbf{M}^0\mathbf{M}^0\mathbf{Z} = \mathbf{Z}'\mathbf{M}^0 Z.$$

A.3 GEOMETRY OF MATRICES

A.3.1 VECTOR SPACES

The K elements of a column vector

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \\ \cdots \\ a_K \end{bmatrix}$$

can be viewed as the coordinates of a point in a K-dimensional space, as shown in Figure A.1 for two dimensions, or as the definition of the line segment connecting the origin and the point defined by $\mathbf{a}$.

Two basic arithmetic operations are defined for vectors, **scalar multiplication** and **addition**. A scalar multiple of a vector, $\mathbf{a}$, is another vector, say $\mathbf{a}^*$, whose coordinates are the scalar multiple of $\mathbf{a}$'s coordinates. Thus, in Figure A.1,

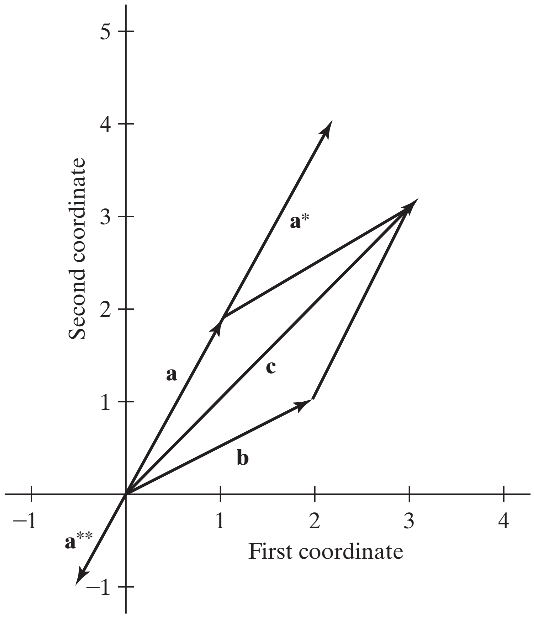
$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{a}^* = 2\mathbf{a} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}, \quad \mathbf{a}^{**} = -\frac{1}{2}\mathbf{a} = \begin{bmatrix} -\frac{1}{2} \\ -1 \end{bmatrix}.$$

The set of all possible scalar multiples of $\mathbf{a}$ is the line through the origin, $\mathbf{0}$ and $\mathbf{a}$. Any scalar multiple of $\mathbf{a}$ is a segment of this line. The sum of two vectors $\mathbf{a}$ and $\mathbf{b}$ is a third vector whose coordinates are the sums of the corresponding coordinates of $\mathbf{a}$ and $\mathbf{b}$. For example,

$$\mathbf{c} = \mathbf{a} + \mathbf{b} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}.$$

Geometrically, $\mathbf{c}$ is obtained by moving in the distance and direction defined by $\mathbf{b}$ from the tip of $\mathbf{a}$ or, because addition is commutative, from the tip of $\mathbf{b}$ in the distance and direction of $\mathbf{a}$. Note that scalar multiplication and addition of vectors are special cases of (A-16) and (A-6) for matrices.

FIGURE A.1 Vector Space.



The two-dimensional plane is the set of all vectors with two real-valued coordinates. We label this set $\mathbb{R}^2$ ("R two," not "R squared"). It has two important properties.

- $\mathbb{R}^2$ *is closed under scalar multiplication;* every scalar multiple of a vector in $\mathbb{R}^2$ is also in $\mathbb{R}^2$.
- $\mathbb{R}^2$ *is closed under addition;* the sum of any two vectors in the plane is always a vector in $\mathbb{R}^2$.

DEFINITION A.2 Vector Space
*A **vector space** is any set of vectors that is closed under scalar multiplication and addition.*

Another example is the set of all real numbers, that is, $\mathbb{R}^1$, that is, the set of vectors with one real element. In general, that set of K-element vectors all of whose elements are real numbers is a K-dimensional vector space, denoted $\mathbb{R}^K$. The preceding examples are drawn in $\mathbb{R}^2$.

A.3.2 LINEAR COMBINATIONS OF VECTORS AND BASIS VECTORS

In Figure A.2, $\mathbf{c} = \mathbf{a} + \mathbf{b}$ and $\mathbf{d} = \mathbf{a}^* + \mathbf{b}$. But since $\mathbf{a}^* = 2\mathbf{a}$, $\mathbf{d} = 2\mathbf{a} + \mathbf{b}$. Also, $\mathbf{e} = \mathbf{a} + 2\mathbf{b}$ and $\mathbf{f} = \mathbf{b} + (-\mathbf{a}) = \mathbf{b} - \mathbf{a}$. As this exercise suggests, any vector in $\mathbb{R}^2$ could be obtained as a **linear combination** of $\mathbf{a}$ and $\mathbf{b}$.

DEFINITION A.3 Basis Vectors
A set of vectors in a vector space is a ***basis*** *for that vector space if they are linearly independent and any vector in the vector space can be written as a linear combination of that set of vectors.*

As is suggested by Figure A.2, any pair of two-element vectors, including **a** and **b**, that point in different directions will form a basis for $\mathbb{R}^2$. Consider an arbitrary set of three vectors in $\mathbb{R}^2$, **a, b**, and **c**. If **a** and **b** are a basis, then we can find numbers α_1 and α_2 such that $\mathbf{c} = \alpha_1\mathbf{a} + \alpha_2\mathbf{b}$. Let

$$\mathbf{a} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix}.$$

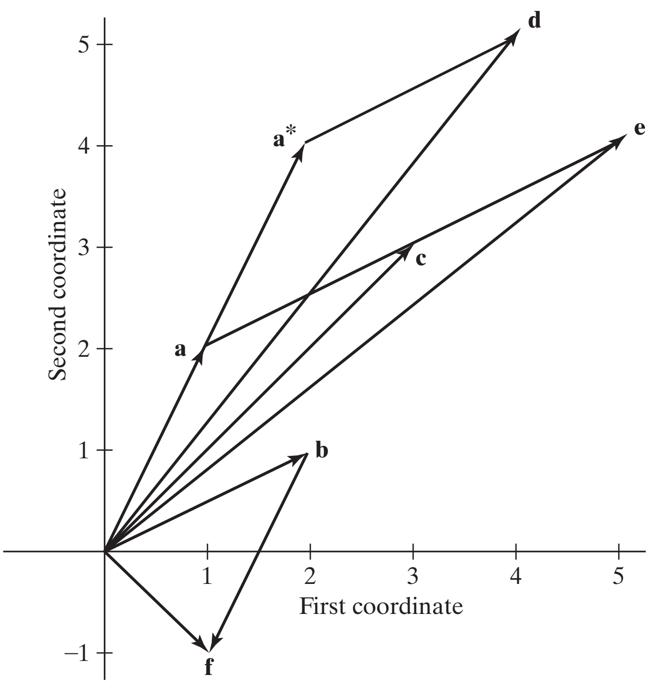
Then

$$\begin{aligned} c_1 &= \alpha_1 a_1 + \alpha_2 b_1, \\ c_2 &= \alpha_1 a_2 + \alpha_2 b_2. \end{aligned} \tag{A-40}$$

The solutions (α_1, α_2) to this pair of equations are

$$\alpha_1 = \frac{b_2c_1 - b_1c_2}{a_1b_2 - b_1a_2}, \quad \alpha_2 = \frac{a_1c_2 - a_2c_1}{a_1b_2 - b_1a_2}. \tag{A-41}$$

FIGURE A.2 Linear Combinations of Vectors.



This result gives a unique solution unless $(a_1b_2 - b_1a_2) = 0$. If $(a_1b_2 - b_1a_2) = 0$, then $a_1/a_2 = b_1/b_2$, which means that $\mathbf{b}$ is just a multiple of $\mathbf{a}$. This returns us to our original condition, that $\mathbf{a}$ and $\mathbf{b}$ must point in different directions. The implication is that if $\mathbf{a}$ and $\mathbf{b}$ are any pair of vectors for which the denominator in (A-41) is not zero, then any other vector $\mathbf{c}$ can be formed as a *unique* linear combination of $\mathbf{a}$ and $\mathbf{b}$. The basis of a vector space is not unique, since any set of vectors that satisfies the definition will do. But for any particular basis, only one linear combination of them will produce another particular vector in the vector space.

A.3.3 LINEAR DEPENDENCE

As the preceding should suggest, K vectors are required to form a basis for $\mathbb{R}^K$. Although the basis for a vector space is not unique, not every set of K vectors will suffice. In Figure A.2, $\mathbf{a}$ and $\mathbf{b}$ form a basis for $\mathbb{R}^2$, but $\mathbf{a}$ and $\mathbf{a}^*$ do not. The difference between these two pairs is that $\mathbf{a}$ and $\mathbf{b}$ are linearly *independent,* whereas $\mathbf{a}$ and $\mathbf{a}^*$ are linearly *dependent*.

> **DEFINITION A.4 Linear Dependence**
> *A set of $k \geq 2$ vectors is **linearly dependent** if at least one of the vectors in the set can be written as a linear combination of the others.*

Because $\mathbf{a}^*$ is a multiple of $\mathbf{a}$, $\mathbf{a}$ and $\mathbf{a}^*$ are linearly dependent. For another example, if

$$\mathbf{a} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ 3 \end{bmatrix}, \quad \text{and} \quad \mathbf{c} = \begin{bmatrix} 10 \\ 14 \end{bmatrix},$$

then

$$2\mathbf{a} + \mathbf{b} - \frac{1}{2}\mathbf{c} = \mathbf{0},$$

so $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are linearly dependent. Any of the three possible pairs of them, however, are linearly independent.

> **DEFINITION A.5 Linear Independence**
> *A set of vectors is **linearly independent** if and only if the only solution $(\alpha_1, \ldots, \alpha_K)$ to*
>
> $$\alpha_1\mathbf{a}_1 + \alpha_2\mathbf{a}_2 + \cdots + \alpha_K\mathbf{a}_K = \mathbf{0}$$
>
> is
>
> $$\alpha_1 = \alpha_2 = \cdots = \alpha_K = 0.$$

The preceding implies the following equivalent definition of a basis.

> **DEFINITION A.6 Basis for a Vector Space**
> *A basis for a vector space of K dimensions is any set of K linearly independent vectors in that vector space.*

Because any $(K + 1)$st vector can be written as a linear combination of the K basis vectors, it follows that any set of more than K vectors in $\mathbb{R}^K$ must be linearly dependent.

A.3.4 SUBSPACES

> **DEFINITION A.7 Spanning Vectors**
> *The set of all linear combinations of a set of vectors is the vector space that is* ***spanned*** *by those vectors.*

For example, by definition, the space spanned by a basis for $\mathbb{R}^K$ is $\mathbb{R}^K$. An implication of this is that if **a** and **b** are a basis for $\mathbb{R}^2$ and **c** is another vector in $\mathbb{R}^2$, the space spanned by [**a, b, c**] is, again, $\mathbb{R}^2$. Of course, **c** is superfluous. Nonetheless, any vector in $\mathbb{R}^2$ *can* be expressed as a linear combination of **a, b**, and **c**. (The linear combination will not be unique. Suppose, for example, that **a** and **c** are also a basis for $\mathbb{R}^2$.)

Consider the set of three coordinate vectors whose third element is zero. In particular,

$$\mathbf{a}' = [a_1 \quad a_2 \quad 0] \quad \text{and} \quad \mathbf{b}' = [b_1 \quad b_2 \quad 0].$$

Vectors **a** and **b** do not span the three-dimensional space $\mathbb{R}^3$. Every linear combination of **a** and **b** has a third coordinate equal to zero; thus, for instance, $\mathbf{c}' = [1 \quad 2 \quad 3]$ could not be written as a linear combination of **a** and **b**. If $(a_1b_2 - a_2b_1)$ is not equal to zero [see (A-41)]; however, then *any vector whose third element is zero can be expressed as a linear combination of* **a** *and* **b**. So, although **a** and **b** do not span $\mathbb{R}^3$, they do span something; they span the set of vectors in $\mathbb{R}^3$ whose third element is zero. This area is a plane (the "floor" of the box in a three-dimensional figure). This plane in $\mathbb{R}^3$ is a **subspace**, in this instance, a two-dimensional subspace. Note that *it is not* $\mathbb{R}^2$; it is the set of vectors in $\mathbb{R}^3$ whose third coordinate is 0. Any plane in $\mathbb{R}^3$ that contains the origin, $(0, 0, 0)$, regardless of how it is oriented, forms a two-dimensional subspace. Any two independent vectors that lie in that subspace will span it. But without a third vector that points in some other direction, we cannot span any more of $\mathbb{R}^3$ than this two-dimensional part of it. By the same logic, any line in $\mathbb{R}^3$ that passes through the origin is a one-dimensional subspace, in this case, the set of all vectors in $\mathbb{R}^3$ whose coordinates are multiples of those of the vector that define the line. A subspace is a vector space in all the respects in which we have defined it. We emphasize that it is *not* a vector space of lower dimension. For example, $\mathbb{R}^2$ is not a subspace of $\mathbb{R}^3$. The essential difference is the number of dimensions in the vectors. The vectors in $\mathbb{R}^3$ that form a two-dimensional subspace are still three-element vectors; they all just happen to lie in the same plane.

The space spanned by a set of vectors in $\mathbb{R}^K$ has at most K dimensions. If this space has fewer than K dimensions, it is a subspace, or **hyperplane.** But the important point in the preceding discussion is that *every set of vectors spans some space;* it may be the entire space in which the vectors reside, or it may be some subspace of it.

A.3.5 RANK OF A MATRIX

We view a matrix as a set of column vectors. The number of columns in the matrix equals the number of vectors in the set, and the number of rows equals the number of coordinates in each column vector. If the matrix contains K rows, its column space might have K dimensions. But,

DEFINITION A.8 Column Space
*The **column space** of a matrix is the vector space that is spanned by its column vectors.*

as we have seen, it might have fewer dimensions; the column vectors might be linearly dependent, or there might be fewer than K of them. Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 5 & 6 \\ 2 & 6 & 8 \\ 7 & 1 & 8 \end{bmatrix}.$$

It contains three vectors from $\mathbb{R}^3$, but the third is the sum of the first two, so the column space of this matrix cannot have three dimensions. Nor does it have only one, because the three columns are not all scalar multiples of one another. Hence, it has two, and the column space of this matrix is a two-dimensional subspace of $\mathbb{R}^3$. It follows that the column rank of a matrix is

DEFINITION A.9 Column Rank
*The **column rank** of a matrix is the dimension of the vector space that is spanned by its column vectors.*

equal to the largest number of linearly independent column vectors it contains. The column rank of $\mathbf{A}$ is 2. For another specific example, consider

$$\mathbf{B} = \begin{bmatrix} 1 & 2 & 3 \\ 5 & 1 & 5 \\ 6 & 4 & 5 \\ 3 & 1 & 4 \end{bmatrix}.$$

It can be shown (we shall see how later) that this matrix has a column rank equal to 3. Each column of $\mathbf{B}$ is a vector in $\mathbb{R}^4$, so the column space of $\mathbf{B}$ is a three-dimensional subspace of $\mathbb{R}^4$.

Consider, instead, the set of vectors obtained by using the *rows* of $\mathbf{B}$ instead of the columns. The new matrix would be

$$\mathbf{C} = \begin{bmatrix} 1 & 5 & 6 & 3 \\ 2 & 1 & 4 & 1 \\ 3 & 5 & 5 & 4 \end{bmatrix}.$$

This matrix is composed of four column vectors from $\mathbb{R}^3$. (Note that $\mathbf{C}$ is $\mathbf{B}'$.) The column space of $\mathbf{C}$ is at most $\mathbb{R}^3$, since four vectors in $\mathbb{R}^3$ must be linearly dependent. In fact, the column space of **C** *is* $\mathbb{R}^3$. Although this is not the same as the column space of **B**, it does have the same dimension. Thus, the column rank of **C** and the column rank of **B** are the same. But the columns of **C** are the rows of **B**. Thus, the column rank of **C** equals the **row rank** of **B**. That the column and row ranks of **B** are the same is not a coincidence. The general results (which are equivalent) are as follows:

THEOREM A.1 Equality of Row and Column Rank
The ***column rank*** *and* ***row rank*** *of a matrix are equal. By the definition of row rank and its counterpart for column rank, we obtain the corollary, the* ***row space*** *and* ***column space*** *of a matrix have the same dimension.* **(A-42)**

Theorem A.1 holds regardless of the actual row and column rank. If the column rank of a matrix happens to equal the number of columns it contains, then the matrix is said to have **full column rank**. **Full row rank** is defined likewise. Because the row and column ranks of a matrix are always equal, we can speak unambiguously of the **rank of a matrix**. For either the row rank or the column rank (and, at this point, we shall drop the distinction), it follows that

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}') \leq \min\,(\text{number of rows, number of columns}). \qquad \textbf{(A-43)}$$

In most contexts, we shall be interested in the columns of the matrices we manipulate. We shall use the term **full rank** to describe a matrix whose rank is equal to the number of columns it contains.

Of particular interest will be the distinction between full rank and **short rank matrices**. The distinction turns on the solutions to $\mathbf{Ax} = \mathbf{0}$. If a nonzero $\mathbf{x}$ for which $\mathbf{Ax} = \mathbf{0}$ exists, then **A** does not have full rank. Equivalently, if the nonzero $\mathbf{x}$ exists, then the columns of **A** are linearly dependent and at least one of them can be expressed as a linear combination of the others. For example, a nonzero set of solutions to

$$\begin{bmatrix} 1 & 3 & 10 \\ 2 & 3 & 14 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

is any multiple of $\mathbf{x}' = (2, 1, -\tfrac{1}{2})$.

In a product matrix $\mathbf{C} = \mathbf{AB}$, every column of **C** is a linear combination of the columns of **A**, so each column of **C** is in the column space of **A**. It is possible that the set of columns in **C** could span this space, but it is not possible for them to span a higher-dimensional space. At best, they could be a full set of linearly independent vectors in **A**'s column space. We conclude that the column rank of **C** could not be greater than that of **A**. Now, apply the same logic to the rows of **C**, which are all linear combinations of the rows of **B**. For the same reason that the column rank of **C** cannot exceed the column rank of **A**, the row rank of **C** cannot exceed the row rank of **B**. Row and column ranks are always equal, so we can conclude that

$$\text{rank}(\mathbf{AB}) \leq \min(\text{rank}(\mathbf{A}), \text{rank}(\mathbf{B})). \qquad \textbf{(A-44)}$$

A useful corollary to (A-44) is

If **A** is $M \times n$ and **B** is a square matrix of rank n, then $\text{rank}(\mathbf{AB}) = \text{rank}(\mathbf{A})$. **(A-45)**

Another application that plays a central role in the development of regression analysis is, for any matrix **A**,

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}'\mathbf{A}) = \text{rank}(\mathbf{A}\mathbf{A}'). \tag{A-46}$$

A.3.6 DETERMINANT OF A MATRIX

The determinant of a square matrix—determinants are not defined for nonsquare matrices—is a function of the elements of the matrix. There are various definitions, most of which are not useful for our work. Determinants figure into our results in several ways, however, that we can enumerate before we need formally to define the computations.

PROPOSITION
The determinant of a matrix is nonzero if and only if it has full rank.

Full rank and short rank matrices can be distinguished by whether or not their determinants are nonzero. There are some settings in which the value of the determinant is also of interest, so we now consider some algebraic results.

It is most convenient to begin with a diagonal matrix

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ & & & \cdots & \\ 0 & 0 & 0 & \cdots & d_K \end{bmatrix}.$$

The column vectors of **D** define a "box" in $\mathbb{R}^K$ whose sides are all at right angles to one another.[4] Its "volume," or determinant, is simply the product of the lengths of the sides, which we denote

$$|\mathbf{D}| = d_1 d_2 \ldots d_K = \prod_{k=1}^{K} d_k. \tag{A-47}$$

A special case is the identity matrix, which has, regardless of K, $|\mathbf{I}_K| = 1$. Multiplying **D** by a scalar c is equivalent to multiplying the length of each side of the box by c, which would multiply its volume by c^K. Thus,

$$|c\mathbf{D}| = c^K|\mathbf{D}|. \tag{A-48}$$

Continuing with this admittedly special case, we suppose that only one column of **D** is multiplied by c. In two dimensions, this would make the box wider but not higher, or vice versa. Hence, the "volume" (area) would also be multiplied by c. Now, suppose that each side of the box were multiplied by a different c, the first by c_1, the second by c_2, and so

[4]Each column vector defines a segment on one of the axes.

on. The volume would, by an obvious extension, now be $c_1 c_2 \dots c_K|\mathbf{D}|$. The matrix with columns defined by $[c_1\mathbf{d}_1\ c_2\mathbf{d}_2 \dots]$ is just $\mathbf{DC}$, where $\mathbf{C}$ is a diagonal matrix with c_i as its ith diagonal element. The computation just described is, therefore,

$$|\mathbf{DC}| = |\mathbf{D}| \cdot |\mathbf{C}|. \tag{A-49}$$

(The determinant of $\mathbf{C}$ is the product of the c_i's since $\mathbf{C}$, like $\mathbf{D}$, is a diagonal matrix.) In particular, note what happens to the whole thing if one of the c_i's is zero.

For 2×2 matrices, the computation of the determinant is

$$\begin{vmatrix} a & c \\ b & d \end{vmatrix} = ad - bc. \tag{A-50}$$

Notice that it is a function of all the elements of the matrix. This statement will be true, in general. For more than two dimensions, the determinant can be obtained by using an **expansion by cofactors**. Using *any* row, say, i, we obtain

$$|\mathbf{A}| = \sum_{k=1}^{K} a_{ik}(-1)^{i+k}|\mathbf{A}_{(ik)}|, \quad k = 1, \dots, K, \tag{A-51}$$

where $\mathbf{A}_{(ik)}$ is the matrix obtained from $\mathbf{A}$ by deleting row i and column k. The determinant of $\mathbf{A}_{(ik)}$ is called a **minor** of $\mathbf{A}$.[5] When the correct sign, $(-1)^{i+k}$, is added, it becomes a **cofactor**. This operation can be done using any column as well. For example, a 4×4 determinant becomes a sum of four 3×3s, whereas a 5×5 is a sum of five 4×4s, each of which is a sum of four 3×3s, and so on. Obviously, it is a good idea to base (A-51) on a row or column with many zeros in it, if possible. In practice, this rapidly becomes a heavy burden. It is unlikely, though, that you will ever calculate any determinants over 3×3 without a computer. A 3×3, however, might be computed on occasion; if so, the following shortcut known as Sarrus's rule will prove useful:

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

Although (A-48) and (A-49) were given for diagonal matrices, they hold for general matrices $\mathbf{C}$ and $\mathbf{D}$. One special case of (A-48) to note is that of $c = -1$. Multiplying a matrix by -1 does not necessarily change the sign of its determinant. It does so only if the order of the matrix is odd. By using the expansion by cofactors formula, an additional result can be shown:

$$|\mathbf{A}| = |\mathbf{A}'|. \tag{A-52}$$

A.3.7 A LEAST SQUARES PROBLEM

Given a vector $\mathbf{y}$ and a matrix $\mathbf{X}$, we are interested in expressing $\mathbf{y}$ as a linear combination of the columns of $\mathbf{X}$. There are two possibilities. If $\mathbf{y}$ lies in the column space of $\mathbf{X}$, then we shall be able to find a vector $\mathbf{b}$ such that

$$\mathbf{y} = \mathbf{Xb}. \tag{A-53}$$

[5] If i equals k, then the determinant is a principal minor.

FIGURE A.3 Least Squares Projections.

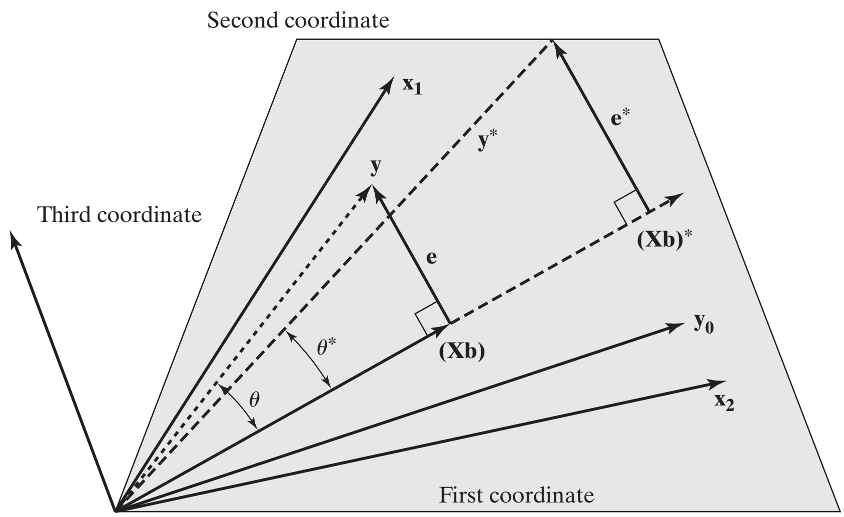


Figure A.3 illustrates such a case for three dimensions in which the two columns of $\mathbf{X}$ both have a third coordinate equal to zero. Only $\mathbf{y}$'s whose third coordinate is zero, such as $\mathbf{y}^0$ in the figure, can be expressed as $\mathbf{Xb}$ for some $\mathbf{b}$. For the general case, assuming that $\mathbf{y}$ is, indeed, in the column space of $\mathbf{X}$, we can find the coefficients $\mathbf{b}$ by solving the set of equations in (A-53). The solution is discussed in the next section.

Suppose, however, that $\mathbf{y}$ is not in the column space of $\mathbf{X}$. In the context of this example, suppose that $\mathbf{y}$'s third component is not zero. Then there is no $\mathbf{b}$ such that (A-53) holds. We can, however, write

$$\mathbf{y} = \mathbf{Xb} + \mathbf{e}, \tag{A-54}$$

where $\mathbf{e}$ is the difference between $\mathbf{y}$ and $\mathbf{Xb}$. By this construction, we find an $\mathbf{Xb}$ that is in the column space of $\mathbf{X}$, and $\mathbf{e}$ is the difference, or "residual." Figure A.3 shows two examples, $\mathbf{y}$ and $\mathbf{y}^*$. For the present, we consider only $\mathbf{y}$. We are interested in finding the $\mathbf{b}$ such that $\mathbf{y}$ is as close as possible to $\mathbf{Xb}$ in the sense that $\mathbf{e}$ is as short as possible.

DEFINITION A.10 Length of a Vector
The length, or ***norm****, of a vector* $\mathbf{e}$ *is given by the Pythagorean theorem:*

$$\|\mathbf{e}\| = \sqrt{\mathbf{e}'\mathbf{e}}. \tag{A-55}$$

The problem is to find the $\mathbf{b}$ for which

$$\|\mathbf{e}\| = \|\mathbf{y} - \mathbf{Xb}\|$$

is as small as possible. The solution is that $\mathbf{b}$ that makes $\mathbf{e}$ perpendicular, or *orthogonal,* to $\mathbf{Xb}$.

DEFINITION A.11 Orthogonal Vectors
Two nonzero vectors **a** *and* **b** *are* ***orthogonal****, written* $\mathbf{a} \perp \mathbf{b}$*, if and only if*

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = \mathbf{0}.$$

Returning once again to our fitting problem, we find that the **b** we seek is that for which

$$\mathbf{e} \perp \mathbf{Xb}.$$

Expanding this set of equations gives the requirement

$$\begin{aligned}(\mathbf{Xb})'\mathbf{e} &= \mathbf{0}\\ &= \mathbf{b}'\mathbf{X}'\mathbf{y} - \mathbf{b}'\mathbf{X}'\mathbf{Xb}\\ &= \mathbf{b}'[\mathbf{X}'\mathbf{y} - \mathbf{X}'\mathbf{Xb}],\end{aligned}$$

or, assuming **b** is not **0**, the set of equations

$$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{Xb}.$$

The means of solving such a set of equations is the subject of Section A.4.

In Figure A.3, the linear combination **Xb** is called the **projection** of **y** into the column space of **X**. The figure is drawn so that, although **y** and **y*** are different, they are similar in that the projection of **y** lies on top of that of **y***. The question we wish to pursue here is, Which vector, **y** or **y***, is closer to its projection in the column space of **X**? Superficially, it would appear that **y** is closer, because **e** is shorter than **e***. Yet **y*** is much more nearly parallel to its projection than **y**, so the only reason that its residual vector is longer is that **y*** is longer compared with **y**. A measure of comparison that would be unaffected by the length of the vectors is the angle between the vector and its projection (assuming that angle is not zero). By this measure, θ^* is smaller than θ, which would reverse the earlier conclusion.

THEOREM A.2 The Cosine Law
The angle θ *between two vectors* **a** *and* **b** *satisfies* $\cos\theta = \dfrac{\mathbf{a}'\mathbf{b}}{\|a\| \times \|\mathbf{b}\|}$.

The two vectors in the calculation would be **y** or **y*** and **Xb** or (**Xb**)*. A zero cosine implies that the vectors are orthogonal. If the cosine is one, then the angle is zero, which means that the vectors are the same. (They would be if **y** were in the column space of **X**.) By dividing by the lengths, we automatically compensate for the length of **y**. By this measure, we find in Figure A.3 that **y** * is closer to its projection, (**Xb**)* than **y** is to its projection, **Xb**.

A.4 SOLUTION OF A SYSTEM OF LINEAR EQUATIONS

Consider the set of n linear equations

$$\mathbf{Ax} = \mathbf{b}, \qquad \textbf{(A-56)}$$

in which the K elements of $\mathbf{x}$ constitute the unknowns. $\mathbf{A}$ is a known matrix of coefficients, and $\mathbf{b}$ is a specified vector of values. We are interested in knowing whether a solution exists; if so, then how to obtain it; and finally, if it does exist, then whether it is unique.

A.4.1 SYSTEMS OF LINEAR EQUATIONS

For most of our applications, we shall consider only square systems of equations, that is, those in which $\mathbf{A}$ is a square matrix. In what follows, therefore, we take n to equal K. Because the number of rows in $\mathbf{A}$ is the number of equations, whereas the number of columns in $\mathbf{A}$ is the number of variables, this case is the familiar one of "n equations in n unknowns."

There are two types of systems of equations.

> **DEFINITION A.12 Homogeneous Equation System**
> *A homogeneous system is of the form* $\mathbf{Ax} = \mathbf{0}$.

By definition, a nonzero solution to such a system will exist if and only if $\mathbf{A}$ does not have **full rank**. If so, then for at least one column of $\mathbf{A}$, we can write the preceding as

$$\mathbf{a}_k = -\sum_{m \neq k} \frac{x_m}{x_k}\mathbf{a}_m.$$

This means, as we know, that the columns of $\mathbf{A}$ are linearly dependent and that $|\mathbf{A}| = \mathbf{0}$.

> **DEFINITION A.13 Nonhomogeneous Equation System**
> *A nonhomogeneous system of equations is of the form* $\mathbf{Ax} = \mathbf{b}$, *where* $\mathbf{b}$ *is a nonzero vector.*

The vector $\mathbf{b}$ is chosen arbitrarily and is to be expressed as a linear combination of the columns of $\mathbf{A}$. Because $\mathbf{b}$ has K elements, this solution will exist only if the columns of $\mathbf{A}$ span the entire K-dimensional space, $\mathbb{R}^K$.[6] Equivalently, we shall require that the columns of $\mathbf{A}$ be linearly independent or that $|\mathbf{A}|$ not be equal to zero.

A.4.2 INVERSE MATRICES

To solve the system $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{x}$, something akin to division by a matrix is needed. Suppose that we could find a square matrix $\mathbf{B}$ such that $\mathbf{BA} = \mathbf{I}$. If the equation system is premultiplied by this $\mathbf{B}$, then the following would be obtained:

$$\mathbf{BAx} = \mathbf{Ix} = \mathbf{x} = \mathbf{Bb}. \qquad \textbf{(A-57)}$$

[6] If $\mathbf{A}$ does not have full rank, then the nonhomogeneous system will have solutions for *some* vectors $\mathbf{b}$, namely, any $\mathbf{b}$ in the column space of $\mathbf{A}$. But we are interested in the case in which there are solutions for *all* nonzero vectors $\mathbf{b}$, which requires $\mathbf{A}$ to have full rank.

If the matrix **B** exists, then it is the **inverse** of **A**, denoted

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

From the definition,

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

In addition, by premultiplying by **A**, postmultiplying by $\mathbf{A}^{-1}$, and then canceling terms, we find

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

as well.

If the inverse exists, then it must be unique. Suppose that it is not and that **C** is a different inverse of **A**. Then $\mathbf{CAB} = \mathbf{CAB}$, but $(\mathbf{CA})\mathbf{B} = \mathbf{IB} = \mathbf{B}$ and $\mathbf{C}(\mathbf{AB}) = \mathbf{C}$, which would be a contradiction if **C** did not equal **B**. Because, by (A-57), the solution is $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, the solution to the equation system is unique as well.

We now consider the calculation of the inverse matrix. For a 2×2 matrix, $\mathbf{AB} = \mathbf{I}$ implies that

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} = 1 \\ a_{11}b_{12} + a_{12}b_{22} = 0 \\ a_{21}b_{11} + a_{22}b_{21} = 0 \\ a_{21}b_{12} + a_{22}b_{22} = 1 \end{bmatrix}.$$

The solutions are

$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \frac{1}{a_{11}a_{22} - a_{12}a_{21}}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}. \qquad \textbf{(A-58)}$$

Notice the presence of the reciprocal of $|\mathbf{A}|$ in $\mathbf{A}^{-1}$. This result is not specific to the 2×2 case. We infer from it that if the determinant is zero, then the inverse does not exist.

DEFINITION A.14 Nonsingular Matrix
A matrix is nonsingular if and only if its inverse exists.

The simplest inverse matrix to compute is that of a diagonal matrix. If

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 & \cdots & 0 \\ 0 & d_2 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & d_K \end{bmatrix}, \quad \text{then} \quad \mathbf{D}^{-1} = \begin{bmatrix} 1/d_1 & 0 & 0 & \cdots & 0 \\ 0 & 1/d_2 & 0 & \cdots & 0 \\ & & \cdots & & \\ 0 & 0 & 0 & \cdots & 1/d_K \end{bmatrix},$$

which shows, incidentally, that $\mathbf{I}^{-1} = \mathbf{I}$.

We shall use a^{ik} to indicate the ikth element of $\mathbf{A}^{-1}$. The general formula for computing an inverse matrix is

$$a^{ik} = \frac{|\mathbf{C}_{ki}|}{|\mathbf{A}|}, \qquad \textbf{(A-59)}$$

where $|\mathbf{C}_{ki}|$ is the *ki*th cofactor of $\mathbf{A}$. [See (A-51).] It follows, therefore, that for $\mathbf{A}$ to be nonsingular, $|\mathbf{A}|$ must be nonzero. Notice the reversal of the subscripts

Some computational results involving inverses are

$$|\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|}, \tag{A-60}$$

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}, \tag{A-61}$$

$$(\mathbf{A}^{-1})' = (\mathbf{A}')^{-1}. \tag{A-62}$$

$$\text{If } \mathbf{A} \text{ is symmetric, then } \mathbf{A}^{-1} \text{ is symmetric.} \tag{A-63}$$

When both inverse matrices exist,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{A-64}$$

Note the condition preceding (A-64). It may be that $\mathbf{AB}$ is a square, nonsingular matrix when neither $\mathbf{A}$ nor $\mathbf{B}$ is even square. (Consider, e.g., $\mathbf{A}'\mathbf{A}$.) Extending (A-64), we have

$$(\mathbf{ABC})^{-1} = \mathbf{C}^{-1}(\mathbf{AB})^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}. \tag{A-65}$$

Recall that for a data matrix $\mathbf{X}$, $\mathbf{X}'\mathbf{X}$ is the sum of the *outer products* of the rows $\mathbf{X}$. Suppose that we have already computed $\mathbf{S} = (\mathbf{X}'\mathbf{X})^{-1}$ for a number of years of data, such as those given in Table A.1. The following result, which is called an **updating formula**, shows how to compute the new $\mathbf{S}$ that would result when a new row is added to $\mathbf{X}$: For symmetric, nonsingular matrix $\mathbf{A}$,

$$[\mathbf{A} \pm \mathbf{bb}']^{-1} = \mathbf{A}^{-1} \mp \left[\frac{1}{1 \pm \mathbf{b}'\mathbf{A}^{-1}\mathbf{b}}\right]\mathbf{A}^{-1}\,\mathbf{bb}'\mathbf{A}^{-1}. \tag{A-66}$$

Note the reversal of the sign in the inverse. Two more general forms of (A-66) that are occasionally useful are

$$[\mathbf{A} \pm \mathbf{bc}']^{-1} = \mathbf{A}^{-1} \mp \left[\frac{1}{1 \pm \mathbf{c}'\mathbf{A}^{-1}\mathbf{b}}\right]\mathbf{A}^{-1}\mathbf{bc}'\mathbf{A}^{-1}, \tag{A-66a}$$

$$[\mathbf{A} \pm \mathbf{BCB}']^{-1} = \mathbf{A}^{-1} \mp \mathbf{A}^{-1}\mathbf{B}[\mathbf{C}^{-1} \pm \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}]^{-1}\mathbf{B}'\mathbf{A}^{-1}. \tag{A-66b}$$

A.4.3 NONHOMOGENEOUS SYSTEMS OF EQUATIONS

For the nonhomogeneous system

$$\mathbf{Ax} = \mathbf{b},$$

if $\mathbf{A}$ is nonsingular, then the unique solution is

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

A.4.4 SOLVING THE LEAST SQUARES PROBLEM

We now have the tool needed to solve the least squares problem posed in Section A.3.7. We found the solution vector, $\mathbf{b}$ to be the solution to the nonhomogenous system

$\mathbf{X}'\mathbf{y} = \mathbf{X}'\mathbf{X}\mathbf{b}$. Let $\mathbf{a}$ equal the vector $\mathbf{X}'\mathbf{y}$ and let $\mathbf{A}$ equal the square matrix $\mathbf{X}'\mathbf{X}$. The equation system is then

$$\mathbf{A}\mathbf{b} = \mathbf{a}.$$

By the preceding results, if $\mathbf{A}$ is nonsingular, then

$$\mathbf{b} = \mathbf{A}^{-1}\mathbf{a} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{y})$$

assuming that the matrix to be inverted is nonsingular. We have reached the irreducible minimum. If the columns of $\mathbf{X}$ are linearly independent, that is, if $\mathbf{X}$ has full rank, then this is the solution to the least squares problem. If the columns of $\mathbf{X}$ are linearly dependent, then this system has no unique solution.

A.5 PARTITIONED MATRICES

In formulating the elements of a matrix, it is sometimes useful to group some of the elements in **submatrices**. Let

$$\mathbf{A} = \left[\begin{array}{cc|c} 1 & 4 & 5 \\ 2 & 9 & 3 \\ \hline 8 & 9 & 6 \end{array}\right] = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}.$$

$\mathbf{A}$ is a **partitioned matrix**. The subscripts of the submatrices are defined in the same fashion as those for the elements of a matrix. A common special case is the **block-diagonal matrix**:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix},$$

where $\mathbf{A}_{11}$ and $\mathbf{A}_{22}$ are square matrices.

A.5.1 ADDITION AND MULTIPLICATION OF PARTITIONED MATRICES

For conformably partitioned matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} + \mathbf{B}_{11} & \mathbf{A}_{12} + \mathbf{B}_{12} \\ \mathbf{A}_{21} + \mathbf{B}_{21} & \mathbf{A}_{22} + \mathbf{B}_{22} \end{bmatrix}, \tag{A-67}$$

and

$$\mathbf{A}\mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{B}_{11} & \mathbf{B}_{12} \\ \mathbf{B}_{21} & \mathbf{B}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{bmatrix}. \tag{A-68}$$

In all these, the matrices must be conformable for the operations involved. For addition, the dimensions of $\mathbf{A}_{ik}$ and $\mathbf{B}_{ik}$ must be the same. For multiplication, the number of columns in $\mathbf{A}_{ij}$ must equal the number of rows in $\mathbf{B}_{jl}$ for all pairs i and j. That is, all the necessary matrix products of the submatrices must be defined. Two cases frequently encountered are of the form

$$\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix}' \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} = [\mathbf{A}_1' \quad \mathbf{A}_2'] \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} = [\mathbf{A}_1'\mathbf{A}_1 + \mathbf{A}_2'\mathbf{A}_2], \tag{A-69}$$

and

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}' \begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{11}'\mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}'\mathbf{A}_{22} \end{bmatrix}. \quad \textbf{(A-70)}$$

A.5.2 DETERMINANTS OF PARTITIONED MATRICES

The determinant of a block-diagonal matrix is obtained analogously to that of a diagonal matrix:

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{11}| \times |\mathbf{A}_{22}|. \quad \textbf{(A-71)}$$

The determinant of a general 2×2 partitioned matrix is

$$\begin{vmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{vmatrix} = |\mathbf{A}_{22}| \times |\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}| = |\mathbf{A}_{11}| \times |\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}|. \quad \textbf{(A-72)}$$

A.5.3 INVERSES OF PARTITIONED MATRICES

The inverse of a block-diagonal matrix is

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{22}^{-1} \end{bmatrix}, \quad \textbf{(A-73)}$$

which can be verified by direct multiplication. For the general 2×2 partitioned matrix, one form of the **partitioned inverse** is

$$\begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}_{11}^{-1}(\mathbf{I} + \mathbf{A}_{12}\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1}) & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{F}_2 \\ -\mathbf{F}_2\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \mathbf{F}_2 \end{bmatrix}, \quad \textbf{(A-74)}$$

where

$$\mathbf{F}_2 = (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}.$$

The upper left block could also be written as

$$\mathbf{F}_1 = (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}.$$

A.5.4 DEVIATIONS FROM MEANS

Suppose that we begin with a column vector of n values $\mathbf{x}$ and let

$$\mathbf{A} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} = \begin{bmatrix} \mathbf{i}'\mathbf{i} & \mathbf{i}'\mathbf{x} \\ \mathbf{x}'\mathbf{i} & \mathbf{x}'\mathbf{x} \end{bmatrix}.$$

We are interested in the lower-right-hand element of $\mathbf{A}^{-1}$. Upon using the definition of $\mathbf{F}_2$ in (A-74), this is

$$\begin{aligned} \mathbf{F}_2 = [\mathbf{x}'\mathbf{x} - (\mathbf{x}'\mathbf{i})(\mathbf{i}'\mathbf{i})^{-1}(\mathbf{i}'\mathbf{x})]^{-1} &= \left\{\mathbf{x}'\left[\mathbf{I}\mathbf{x} - \mathbf{i}\left(\frac{1}{n}\right)\mathbf{i}'\mathbf{x}\right]\right\}^{-1} \\ &= \left\{\mathbf{x}'\left[\mathbf{I} - \left(\frac{1}{n}\right)\mathbf{i}\mathbf{i}'\right]\mathbf{x}\right\}^{-1} = (\mathbf{x}'\mathbf{M}^0\mathbf{x})^{-1}. \end{aligned}$$

Therefore, the lower-right-hand value in the inverse matrix is

$$(\mathbf{x}'\mathbf{M}^0\mathbf{x})^{-1} = \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = a^{22}.$$

Now, suppose that we replace **x** with **X**, a matrix with several columns. We seek the lower-right block of $(\mathbf{Z}'\mathbf{Z})^{-1}$, where $\mathbf{Z} = [\mathbf{i}, \mathbf{X}]$. The analogous result is

$$(\mathbf{Z}'\mathbf{Z})^{22} = [\mathbf{X}'\mathbf{X} - \mathbf{X}'\mathbf{i}(\mathbf{i}'\mathbf{i})^{-1}\mathbf{i}'\mathbf{X}]^{-1} = (\mathbf{X}'\mathbf{M}^0\mathbf{X})^{-1},$$

which implies that the $K \times K$ matrix in the lower-right corner of $(\mathbf{Z}'\mathbf{Z})^{-1}$ is the inverse of the $K \times K$ matrix whose jkth element is $\sum_{i=1}^{n}(x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)$. Thus, when a data matrix contains a column of ones, the elements of the inverse of the matrix of sums of squares and cross products will be computed from the original data in the form of deviations from the respective column means.

A.5.5 KRONECKER PRODUCTS

A calculation that helps to condense the notation when dealing with sets of regression models(see Chapter 10) is the **Kronecker product**. For general matrices **A** and **B**,

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1K}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2K}\mathbf{B} \\ & & \cdots & \\ a_{n1}\mathbf{B} & a_{n2}\mathbf{B} & \cdots & a_{nK}\mathbf{B} \end{bmatrix}. \quad \textbf{(A-75)}$$

Notice that there is no requirement for conformability in this operation. The Kronecker product can be computed for any pair of matrices. If **A** is $K \times L$ and **B** is $m \times n$, then $\mathbf{A} \otimes \mathbf{B}$ is $(Km) \times (Ln)$.

For the Kronecker product,

$$(\mathbf{A} \otimes \mathbf{B})^{-1} = (\mathbf{A}^{-1} \otimes \mathbf{B}^{-1}), \quad \textbf{(A-76)}$$

If **A** is $M \times M$ and **B** is $n \times n$, then

$$|\mathbf{A} \otimes \mathbf{B}| = |\mathbf{A}|^n|\mathbf{B}|^M,$$
$$(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}',$$
$$\text{trace}(\mathbf{A} \otimes \mathbf{B}) = \text{trace}(\mathbf{A})\,\text{trace}(\mathbf{B}).$$

(The trace of a matrix is defined in Section A.6.7.) For **A**, **B**, **C**, and **D** such that the products are defined,

$$(\mathbf{A} \otimes B)(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD}.$$

A.6 CHARACTERISTIC ROOTS AND VECTORS

A useful set of results for analyzing a square matrix **A** arises from the solutions to the set of equations

$$\mathbf{Ac} = \lambda\mathbf{c}. \quad \textbf{(A-77)}$$

The pairs of solutions $(\mathbf{c},\lambda)$ are the **characteristic vectors c** and **characteristic roots** λ. If $\mathbf{c}$ is any nonzero solution vector, then $k\mathbf{c}$ is also for any value of K. To remove the indeterminancy, $\mathbf{c}$ is **normalized** so that $\mathbf{c}'\mathbf{c} = 1$.

The solution then consists of λ and the $n - 1$ unknown elements in $\mathbf{c}$.

A.6.1 THE CHARACTERISTIC EQUATION

Solving (A-77) can, in principle, proceed as follows. First, (A-77) implies that

$$\mathbf{Ac} = \lambda\mathbf{Ic},$$

or that

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

This equation is a homogeneous system that has a nonzero solution only if the matrix $(\mathbf{A} - \lambda\mathbf{I})$ is singular or has a zero determinant. Therefore, if λ is a solution, then

$$|\mathbf{A} - \lambda\mathbf{I}| = 0. \tag{A-78}$$

This polynomial in λ is the **characteristic equation** of $\mathbf{A}$. For example, if

$$\mathbf{A} = \begin{bmatrix} 5 & 1 \\ 2 & 4 \end{bmatrix},$$

then

$$|\mathbf{A} - \lambda\mathbf{I}| = \begin{vmatrix} 5-\lambda & 1 \\ 2 & 4-\lambda \end{vmatrix} = (5-\lambda)(4-\lambda) - 2(1) = \lambda^2 - 9\lambda + 18.$$

The two solutions are $\lambda = 6$ and $\lambda = 3$.

In solving the characteristic equation, there is no guarantee that the characteristic roots will be real. In the preceding example, if the 2 in the lower-left-hand corner of the matrix were -2 instead, then the solution would be a pair of complex values. The same result can emerge in the general $n \times n$ case. The characteristic roots of a symmetric matrix such as $\mathbf{X}'\mathbf{X}$ are real, however.[7] This result will be convenient because most of our applications will involve the characteristic roots and vectors of symmetric matrices.

For an $n \times n$ matrix, the characteristic equation is an nth-order polynomial in λ. Its solutions may be n distinct values, as in the preceding example, or may contain repeated values of λ, and may contain some zeros as well.

A.6.2 CHARACTERISTIC VECTORS

With λ in hand, the characteristic vectors are derived from the original problem,

$$\mathbf{Ac} = \lambda\mathbf{c},$$

or

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{c} = \mathbf{0}. \tag{A-79}$$

Neither pair determines the values of c_1 and c_2. But this result was to be expected; it was the reason $\mathbf{c}'\mathbf{c} = 1$ was specified at the outset. The additional equation $\mathbf{c}'\mathbf{c} = 1$, however, produces complete solutions for the vectors.

[7]A proof may be found in Theil (1971).

A.6.3 GENERAL RESULTS FOR CHARACTERISTIC ROOTS AND VECTORS

A $K \times K$ symmetric matrix has K distinct characteristic vectors, $\mathbf{c}_1, \mathbf{c}_2, \ldots \mathbf{c}_K$. The corresponding characteristic roots, $\lambda_1, \lambda_2, \ldots, \lambda_K$, although real, need not be distinct. The characteristic vectors of a symmetric matrix are orthogonal,[8] which implies that for every $i \neq j$, $\mathbf{c}_i'\mathbf{c}_j = 0$.[9] It is convenient to collect the K-characteristic vectors in a $K \times K$ matrix whose ith column is the $\mathbf{c}_i$ corresponding to λ_i,

$$\mathbf{C} = [\mathbf{c}_1 \quad \mathbf{c}_2 \quad \cdots \quad c_K],$$

and the K-characteristic roots in the same order, in a diagonal matrix,

$$\boldsymbol{\Lambda} = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \lambda_K \end{bmatrix}.$$

Then, the full set of equations

$$\mathbf{A}\mathbf{c}_k = \lambda_k \mathbf{c}_k$$

is contained in

$$\mathbf{AC} = \mathbf{C}\boldsymbol{\Lambda}. \tag{A-80}$$

Because the vectors are orthogonal and $\mathbf{c}_i'\mathbf{c}_i = 1$, we have

$$\mathbf{C}'\mathbf{C} = \begin{bmatrix} \mathbf{c}_1'\mathbf{c}_1 & \mathbf{c}_1'\mathbf{c}_2 & \cdots & \mathbf{c}_1'\mathbf{c}_K \\ \mathbf{c}_2'\mathbf{c}_1 & \mathbf{c}_2'\mathbf{c}_2 & \cdots & \mathbf{c}_2'\mathbf{c}_K \\ & & \vdots & \\ \mathbf{c}_K'\mathbf{c}_1 & \mathbf{c}_K'\mathbf{c}_2 & \cdots & \mathbf{c}_K'\mathbf{c}_K \end{bmatrix} = \mathbf{I}. \tag{A-81}$$

Result (A-81) implies that

$$\mathbf{C}' = \mathbf{C}^{-1}. \tag{A-82}$$

Consequently,

$$\mathbf{CC}' = \mathbf{CC}^{-1} = \mathbf{I} \tag{A-83}$$

as well, so the rows as well as the columns of $\mathbf{C}$ are orthogonal.

A.6.4 DIAGONALIZATION AND SPECTRAL DECOMPOSITION OF A MATRIX

By premultiplying (A-80) by $\mathbf{C}'$ and using (A-81), we can extract the characteristic roots of $\mathbf{A}$.

[8]For proofs of these propositions, see Strang (1988–2014).

[9]This statement is not true if the matrix is not symmetric. For instance, it does not hold for the characteristic vectors computed in the first example. For nonsymmetric matrices, there is also a distinction between "right" characteristic vectors, $\mathbf{Ac} = \lambda c$, and "left" characteristic vectors, $\mathbf{d}'\mathbf{A} = \lambda\mathbf{d}'$, which may not be equal.

DEFINITION A.15 Diagonalization of a Matrix
The ***diagonalization*** *of a matrix* $\mathbf{A}$ *is*

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{C}'\mathbf{C}\boldsymbol{\Lambda} = \mathbf{I}\boldsymbol{\Lambda} = \boldsymbol{\Lambda}. \quad \textbf{(A-84)}$$

Alternatively, by *post* multiplying (A-80) by $\mathbf{C}'$ and using (A-83), we obtain a useful representation of $\mathbf{A}$.

DEFINITION A.16 Spectral Decomposition of a Matrix
The spectral decomposition of $\mathbf{A}$ *is*

$$\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}' = \sum_{k=1}^{K} \lambda_k \mathbf{c}_k \mathbf{c}_k'. \quad \textbf{(A-85)}$$

In this representation, the $K \times K$ matrix $\mathbf{A}$ is written as a sum of K rank one matrices. This sum is also called the **eigenvalue** (or, "own" value) decomposition of $\mathbf{A}$. In this connection, the term *signature* of the matrix is sometimes used to describe the characteristic roots and vectors. Yet another pair of terms for the parts of this decomposition are the **latent roots** and **latent vectors** of $\mathbf{A}$.

A.6.5 RANK OF A MATRIX

The diagonalization result enables us to obtain the rank of a matrix very easily. To do so, we can use the following result.

THEOREM A.3 Rank of a Product
For any matrix $\mathbf{A}$ *and nonsingular matrices* $\mathbf{B}$ *and* $\mathbf{C}$, *the rank of* $\mathbf{BAC}$ *is equal to the rank of* $\mathbf{A}$. ***Proof:*** *By (A-45),* $\text{rank}(\mathbf{BAC}) = \text{rank}[(\mathbf{BA})\mathbf{C}] = \text{rank}(\mathbf{BA})$. *By (A-43),* $\text{rank}(\mathbf{BA}) = \text{rank}(\mathbf{A}'\mathbf{B}')$, *and applying (A-45) again,* $\text{rank}(\mathbf{A}'\mathbf{B}') = \text{rank}(\mathbf{A}')$ *because* $\mathbf{B}'$ *is nonsingular if* $\mathbf{B}$ *is nonsingular [once again, by (A-43)]. Finally, applying (A-43) again to obtain* $\text{rank}(\mathbf{A}') = \text{rank}(\mathbf{A})$ *gives the result.*

Because $\mathbf{C}$ and $\mathbf{C}'$ are nonsingular, we can use them to apply this result to (A-84). By an obvious substitution,

$$\text{rank}(\mathbf{A}) = \text{rank}(\boldsymbol{\Lambda}). \quad \textbf{(A-86)}$$

Finding the rank of $\boldsymbol{\Lambda}$ is trivial. Because $\boldsymbol{\Lambda}$ is a diagonal matrix, its rank is just the number of nonzero values on its diagonal. By extending this result, we can prove the following theorems. (Proofs are brief and are left for the reader.)

THEOREM A.4 Rank of a Symmetric Matrix
The rank of a symmetric matrix is the number of nonzero characteristic roots it contains.

Note how this result enters the spectral decomposition given earlier. If any of the characteristic roots are zero, then the number of rank one matrices in the sum is reduced correspondingly. It would appear that this simple rule will not be useful if $\mathbf{A}$ is not square. But recall that

$$\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{A}'\mathbf{A}). \tag{A-87}$$

Because $\mathbf{A}'\mathbf{A}$ is always square, we can use it instead of $\mathbf{A}$. Indeed, we can use it even if $\mathbf{A}$ is not square, which leads to a fully general result.

THEOREM A.5 Rank of a Matrix
The rank of any matrix $\mathbf{A}$ equals the number of nonzero characteristic roots in $\mathbf{A}'\mathbf{A}$.

The row rank and column rank of a matrix are equal, so we should be able to apply Theorem A.5 to $\mathbf{AA}'$ as well. This process, however, requires an additional result.

THEOREM A.6 Roots of an Outer Product Matrix
The nonzero characteristic roots of $\mathbf{AA}'$ are the same as those of $\mathbf{A}'\mathbf{A}$.

The proof is left as an exercise. A useful special case the reader can examine is the characteristic roots of $\mathbf{aa}'$ and $\mathbf{a}'\mathbf{a}$, where $\mathbf{a}$ is an $n \times 1$ vector.

If a characteristic root of a matrix is zero, then we have $\mathbf{Ac} = \mathbf{0}$. Thus, if the matrix has a zero root, it must be singular. Otherwise, no nonzero $\mathbf{c}$ would exist. In general, therefore, a matrix is singular; that is, it does not have full rank if and only if it has at least one zero root.

A.6.6 CONDITION NUMBER OF A MATRIX

As the preceding might suggest, there is a discrete difference between full rank and short rank matrices. In analyzing data matrices such as the one in Section A.2, however, we shall often encounter cases in which a matrix is not quite short ranked, because it has all nonzero roots, but it is close. That is, by some measure, we can come very close to being able to write one column as a linear combination of the others. This case is important; we shall examine it at length in our discussion of multicollinearity in Section 4.9.1. Our definitions of rank and determinant will fail to indicate this possibility, but an alternative measure, the **condition number**, is designed for that purpose. Formally, the condition number for a square matrix $\mathbf{A}$ is

$$\gamma = \left[\frac{\text{maximum root}}{\text{minimum root}}\right]^{1/2}. \tag{A-88}$$

For nonsquare matrices $\mathbf{X}$, such as the data matrix in the example, we use $\mathbf{A} = \mathbf{X}'\mathbf{X}$. As a further refinement, because the characteristic roots are affected by the scaling of the columns of $\mathbf{X}$, we scale the columns to have length 1 by dividing each column by its norm [see (A-55)]. For the $\mathbf{X}$ in Section A.2, the largest characteristic root of $\mathbf{A}$ is 4.9255 and the smallest is 0.0001543. Therefore, the condition number is 178.67, which is extremely large. (Values greater than 20 are large.) That the smallest root is close to zero compared with the largest means that this matrix is nearly singular. Matrices with large condition numbers are difficult to invert accurately.

A.6.7 TRACE OF A MATRIX

The **trace** of a square $K \times K$ matrix is the sum of its diagonal elements:

$$\mathrm{tr}(\mathbf{A}) = \sum_{k=1}^{K} a_{kk}.$$

Some easily proven results are

$$\mathrm{tr}(c\mathbf{A}) = c(\mathrm{tr}(\mathbf{A})), \tag{A-89}$$

$$\mathrm{tr}(\mathbf{A}') = \mathrm{tr}(\mathbf{A}), \tag{A-90}$$

$$\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B}), \tag{A-91}$$

$$\mathrm{tr}(\mathbf{I}_K) = K. \tag{A-92}$$

$$\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}). \tag{A-93}$$

$$\mathbf{a}'\mathbf{a} = \mathrm{tr}(\mathbf{a}'\mathbf{a}) = \mathrm{tr}(\mathbf{aa}')$$

$$\mathrm{tr}(\mathbf{A}'\mathbf{A}) = \sum_{k=1}^{K} \mathbf{a}_k'\mathbf{a}_k = \sum_{i=1}^{K}\sum_{k=1}^{K} a_{ik}^2.$$

The permutation rule can be extended to any *cyclic* permutation in a product:

$$\mathrm{tr}(\mathbf{ABCD}) = \mathrm{tr}(\mathbf{BCDA}) = \mathrm{tr}(\mathbf{CDAB}) = \mathrm{tr}(\mathbf{DABC}). \tag{A-94}$$

By using (A-84), we obtain

$$\mathrm{tr}(\mathbf{C}'\mathbf{AC}) = \mathrm{tr}(\mathbf{ACC}') = \mathrm{tr}(\mathbf{AI}) = \mathrm{tr}(\mathbf{A}) = \mathrm{tr}(\mathbf{\Lambda}). \tag{A-95}$$

Because $\mathbf{\Lambda}$ is diagonal with the roots of $\mathbf{A}$ on its diagonal, the general result is the following.

THEOREM A.7 Trace of a Matrix

The trace of a matrix equals the sum of its characteristic roots. **(A-96)**

A.6.8 DETERMINANT OF A MATRIX

Recalling how tedious the calculation of a determinant promised to be, we find that the following is particularly useful. Because

$$\begin{aligned}\mathbf{C}'\mathbf{A}\mathbf{C} &= \boldsymbol{\Lambda},\\ |\mathbf{C}'\mathbf{A}\mathbf{C}| &= |\boldsymbol{\Lambda}|.\end{aligned} \tag{A-97}$$

Using a number of earlier results, we have, for orthogonal matrix **C**,

$$\begin{aligned}|\mathbf{C}'\mathbf{A}\mathbf{C}| &= |\mathbf{C}'|\cdot|\mathbf{A}|\cdot|\mathbf{C}| = |\mathbf{C}'|\cdot|\mathbf{C}|\cdot|\mathbf{A}| = |\mathbf{C}'\mathbf{C}|\cdot|\mathbf{A}| = |\mathbf{I}|\cdot|\mathbf{A}| = 1\cdot|\mathbf{A}|\\ &= |\mathbf{A}|\\ &= |\boldsymbol{\Lambda}|.\end{aligned} \tag{A-98}$$

Because $|\boldsymbol{\Lambda}|$ is just the product of its diagonal elements, the following is implied.

THEOREM A.8 Determinant of a Matrix

The determinant of a matrix equals the product of its characteristic roots. **(A-99)**

Notice that we get the expected result if any of these roots is zero. The determinant is the product of the roots, so it follows that a matrix is singular if and only if its determinant is zero and, in turn, if and only if it has at least one zero characteristic root.

A.6.9 POWERS OF A MATRIX

We often use expressions involving powers of matrices, such as $\mathbf{AA} = \mathbf{A}^2$. For positive integer powers, these expressions can be computed by repeated multiplication. But this does not show how to handle a problem such as finding a $\mathbf{B}$ such that $\mathbf{B}^2 = \mathbf{A}$, that is, the square root of a matrix. The characteristic roots and vectors provide a solution. Consider, first

$$\mathbf{AA} = \mathbf{A}^2 = (\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}')(\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}') = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}' = \mathbf{C}\boldsymbol{\Lambda}\mathbf{I}\boldsymbol{\Lambda}\mathbf{C}' = \mathbf{C}\boldsymbol{\Lambda}\boldsymbol{\Lambda}\mathbf{C}' = \mathbf{C}\boldsymbol{\Lambda}^2\mathbf{C}'. \tag{A-100}$$

Two results follow. Because $\boldsymbol{\Lambda}^2$ is a diagonal matrix whose nonzero elements are the squares of those in $\boldsymbol{\Lambda}$, the following is implied.

For any symmetric matrix, the characteristic roots of $\mathbf{A}^2$ *are the squares of those of* $\mathbf{A}$*, and the characteristic vectors are the same.* **(A-101)**

The proof is obtained by observing that the last result in (A-100) is the spectral decomposition of the matrix $\mathbf{B} = \mathbf{AA}$. Because $\mathbf{A}^3 = \mathbf{AA}^2$ and so on, (A-101) extends to any positive integer. By convention, for any $\mathbf{A}$, $\mathbf{A}^0 = \mathbf{I}$. Thus, for any symmetric matrix $\mathbf{A}$, $\mathbf{A}^K = \mathbf{C}\boldsymbol{\Lambda}^K\mathbf{C}'$, $K = 0, 1, \ldots$. Hence, the characteristic roots of $\mathbf{A}^K$ are λ^K, whereas the characteristic vectors are the same as those of $\mathbf{A}$. If $\mathbf{A}$ is nonsingular, so that all its roots λ_i are nonzero, then this proof can be extended to negative powers as well.

If $\mathbf{A}^{-1}$ exists, then

$$\mathbf{A}^{-1} = (\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}')^{-1} = (\mathbf{C}')^{-1}\boldsymbol{\Lambda}^{-1}\mathbf{C}^{-1} = \mathbf{C}\boldsymbol{\Lambda}^{-1}\mathbf{C}', \tag{A-102}$$

where we have used the earlier result, $\mathbf{C}' = \mathbf{C}^{-1}$. This gives an important result that is useful for analyzing inverse matrices.

THEOREM A.9 Characteristic Roots of an Inverse Matrix
If $\mathbf{A}^{-1}$ exists, then the characteristic roots of $\mathbf{A}^{-1}$ are the reciprocals of those of $\mathbf{A}$, and the characteristic vectors are the same.

By extending the notion of repeated multiplication, we now have a more general result.

THEOREM A.10 Characteristic Roots of a Matrix Power
For any nonsingular symmetric matrix $\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'$, $\mathbf{A}^K = \mathbf{C}\boldsymbol{\Lambda}^K\mathbf{C}'$, $K = \ldots, -2, -1, 0, 1, 2, \ldots.$

We now turn to the general problem of how to compute the square root of a matrix. In the scalar case, the value would have to be nonnegative. The matrix analog to this requirement is that all the characteristic roots are nonnegative. Consider, then, the candidate

$$\mathbf{A}^{1/2} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}\mathbf{C} = \mathbf{C}\begin{bmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ & & \cdots & \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{bmatrix}\mathbf{C}'. \tag{A-103}$$

This equation satisfies the requirement for a square root, because

$$\mathbf{A}^{1/2}\mathbf{A}^{1/2} = \mathbf{C}\boldsymbol{\Lambda}^{1/2}\mathbf{C}'\mathbf{C}\boldsymbol{\Lambda}^{1/2}\mathbf{C}' = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}' = \mathbf{A}. \tag{A-104}$$

If we continue in this fashion, we can define the nonnegative powers of a matrix more generally, still assuming that all the characteristic roots are nonnegative. For example, $\mathbf{A}^{1/3} = \mathbf{C}\boldsymbol{\Lambda}^{1/3}\mathbf{C}'$. If all the roots are strictly positive, we can go one step further and extend the result to any real power. For reasons that will be made clear in the next section, we say that a matrix with positive characteristic roots is **positive definite.** It is the matrix analog to a positive number.

DEFINITION A.17 Real Powers of a Positive Definite Matrix

*For a **positive definite** matrix $\mathbf{A}$, $\mathbf{A}^r = \mathbf{C}\boldsymbol{\Lambda}^r\mathbf{C}'$, for any real number, r.* **(A-105)**

The characteristic roots of $\mathbf{A}^r$ are the rth power of those of $\mathbf{A}$, and the characteristic vectors are the same.

If $\mathbf{A}$ is only **nonnegative definite**—that is, has roots that are either zero or positive—then (A-105) holds only for nonnegative r.

A.6.10 IDEMPOTENT MATRICES

Idempotent matrices are equal to their squares [see (A-37) to (A-39)]. In view of their importance in econometrics, we collect a few results related to idempotent matrices at this point. First, (A-101) implies that if λ is a characteristic root of an idempotent matrix, then $\lambda = \lambda^K$ for all nonnegative integers K. As such, if $\mathbf{A}$ is a symmetric idempotent matrix, then all its roots are one or zero. Assume that all the roots of $\mathbf{A}$ are one. Then $\mathbf{\Lambda} = \mathbf{I}$, and $\mathbf{A} = \mathbf{C\Lambda C}' = \mathbf{CIC}' = \mathbf{CC}' = \mathbf{I}$. If the roots are not all one, then one or more are zero. Consequently, we have the following results for symmetric idempotent matrices:[10]

- *The only full rank, symmetric idempotent matrix is the identity matrix* $\mathbf{I}$. **(A-106)**
- *All symmetric idempotent matrices except the identity matrix are singular*. **(A-107)**

The final result on idempotent matrices is obtained by observing that the count of the nonzero roots of $\mathbf{A}$ is also equal to their sum. By combining Theorems A.5 and A.7 with the result that for an idempotent matrix, the roots are all zero or one, we obtain this result:

- *The rank of a symmetric idempotent matrix is equal to its trace*. **(A-108)**

A.6.11 FACTORING A MATRIX: THE CHOLESKY DECOMPOSITION

In some applications, we shall require a matrix $\mathbf{P}$ such that

$$\mathbf{P}'\mathbf{P} = \mathbf{A}^{-1}.$$

One choice is

$$\mathbf{P} = \mathbf{\Lambda}^{-1/2}\mathbf{C}',$$

so that

$$\mathbf{P}'\mathbf{P} = (\mathbf{C}')'(\mathbf{\Lambda}^{-1/2})'\mathbf{\Lambda}^{-1/2}\mathbf{C}' = \mathbf{C\Lambda}^{-1}\mathbf{C}',$$

as desired.[11] Thus, the **spectral decomposition** of $\mathbf{A}$, $\mathbf{A} = \mathbf{C\Lambda C}'$ is a useful result for this kind of computation.

The **Cholesky factorization** of a symmetric positive definite matrix is an alternative representation that is useful in regression analysis. Any symmetric positive definite matrix $\mathbf{A}$ may be written as the product of a **lower triangular matrix L** and its transpose (which is an **upper triangular matrix**) $\mathbf{L}' = \mathbf{U}$. Thus, $\mathbf{A} = \mathbf{LU}$. This result is the Cholesky decomposition of $\mathbf{A}$. The square roots of the diagonal elements of $\mathbf{L}$, d_i, are the **Cholesky values** of $\mathbf{A}$. By arraying these in a diagonal matrix $\mathbf{D}$, we may also write $\mathbf{A} = \mathbf{LD}^{-1}\mathbf{D}^2\mathbf{D}^{-1}\mathbf{U} = \mathbf{L}^*\mathbf{D}^2\mathbf{U}^*$, which is similar to the spectral decomposition in (A-85). The usefulness of this formulation arises when the inverse of $\mathbf{A}$ is required. Once $\mathbf{L}$ is

[10]Not all idempotent matrices are symmetric. We shall not encounter any asymmetric ones in our work, however.

[11]We say that this is "one" choice because if $\mathbf{A}$ is symmetric, as it will be in all our applications, there are other candidates. The reader can easily verify that $\mathbf{C\Lambda}^{-1/2}\mathbf{C}' = \mathbf{A}^{-1/2}$ works as well.

computed, finding $\mathbf{A}^{-1} = \mathbf{U}^{-1}\mathbf{L}^{-1}$ is also straightforward as well as extremely fast and accurate. Most recently developed econometric software packages use this technique for inverting positive definite matrices.

A.6.12 SINGULAR VALUE DECOMPOSITION

A third type of decomposition of a matrix is useful for numerical analysis when the inverse is difficult to obtain because the columns of **A** are "nearly" collinear. Any $n \times K$ matrix **A** for which $n \geq K$ can be written in the form $\mathbf{A} = \mathbf{UWV}'$, where **U** is an orthogonal $n \times K$ matrix—that is, $\mathbf{U}'\mathbf{U} = \mathbf{I}_K$—**W** is a $K \times K$ diagonal matrix such that $w_i \geq 0$, and **V** is a $K \times K$ matrix such that $\mathbf{V}'\mathbf{V} = \mathbf{I}_K$. This result is called the **singular value decomposition** (SVD) of **A**, and w_i are the singular values of **A**.[12] (Note that if **A** is square, then the spectral decomposition is a singular value decomposition.) As with the Cholesky decomposition, the usefulness of the SVD arises in inversion, in this case, of $\mathbf{A}'\mathbf{A}$. By multiplying it out, we obtain that $(\mathbf{A}'\mathbf{A})^{-1}$ is simply $\mathbf{VW}^{-2}\mathbf{V}'$. Once the SVD of **A** is computed, the inversion is trivial. The other advantage of this format is its numerical stability, which is discussed at length in Press et al. (2007).

A.6.13 QR DECOMPOSITION

Press et al. (2007) recommend the SVD approach as the method of choice for solving least squares problems because of its accuracy and numerical stability. A commonly used alternative method similar to the SVD approach is the QR decomposition. Any $n \times K$ matrix, **X**, with $n \geq K$ can be written in the form $\mathbf{X} = \mathbf{QR}$ in which the columns of **Q** are orthonormal ($\mathbf{Q}'\mathbf{Q} = \mathbf{I}$) and **R** is an upper triangular matrix. Decomposing **X** in this fashion allows an extremely accurate solution to the least squares problem that does not involve inversion or direct solution of the normal equations. Press et al. suggest that this method may have problems with rounding errors in problems when **X** is nearly of short rank, but based on other published results, this concern seems relatively minor.[13]

A.6.14 THE GENERALIZED INVERSE OF A MATRIX

Inverse matrices are fundamental in econometrics. Although we shall not require them much in our treatment in this book, there are more general forms of inverse matrices than we have considered thus far. A **generalized inverse** of a matrix **A** is another matrix $\mathbf{A}^+$ that satisfies the following requirements:

1. $\mathbf{AA}^+\mathbf{A} = \mathbf{A}$.
2. $\mathbf{A}^+\mathbf{AA}^+ = \mathbf{A}^+$.
3. $\mathbf{A}^+\mathbf{A}$ is symmetric.
4. $\mathbf{AA}^+$ is symmetric.

[12]Discussion of the singular value decomposition (and listings of computer programs for the computations) may be found in Press et al. (1986).

[13]The National Institute of Standards and Technology (NIST) has published a suite of benchmark problems that test the accuracy of least squares computations (http://www.nist.gov/itl/div898/strd). Using these problems, which include some extremely difficult, ill-conditioned data sets, we found that the QR method would reproduce all the NIST certified solutions to 15 digits of accuracy, which suggests that the QR method should be satisfactory for all but the worst problems. NIST's benchmark for hard to solve least squares problems, the "Filipelli problem," is solved accurately to at least 9 digits with the QR method. Evidently, other methods of least squares solution fail to produce an accurate result.

A unique $\mathbf{A}^{+}$ can be found for any matrix, whether $\mathbf{A}$ is singular or not, or even if $\mathbf{A}$ is not square.[14] The unique matrix that satisfies all four requirements is called the **Moore–Penrose inverse** or **pseudoinverse** of $\mathbf{A}$. If $\mathbf{A}$ happens to be square and nonsingular, then the generalized inverse will be the familiar ordinary inverse. But if $\mathbf{A}^{-1}$ does not exist, then $\mathbf{A}^{+}$ can still be computed.

An important special case is the overdetermined system of equations

$$\mathbf{Ab} = \mathbf{y},$$

where $\mathbf{A}$ has n rows, $K < n$ columns, and column rank equal to $R \leq K$. Suppose that R equals K, so that $(\mathbf{A}'\mathbf{A})^{-1}$ exists. Then the Moore–Penrose inverse of $\mathbf{A}$ is

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

which can be verified by multiplication. A "solution" to the system of equations can be written

$$\mathbf{b} = \mathbf{A}^{+}\mathbf{y}.$$

This is the vector that minimizes the length of $\mathbf{Ab} - \mathbf{y}$. Recall this was the solution to the least squares problem obtained in Section A.4.4. If $\mathbf{y}$ lies in the column space of $\mathbf{A}$, this vector will be zero, but otherwise, it will not.

Now suppose that $\mathbf{A}$ does not have full rank. The previous solution cannot be computed. An alternative solution can be obtained, however. We continue to use the matrix $\mathbf{A}'\mathbf{A}$. In the spectral decomposition of Section A.6.4, if $\mathbf{A}$ has rank R, then there are R terms in the summation in (A-85). In (A-102), the spectral decomposition using the reciprocals of the characteristic roots is used to compute the inverse. To compute the Moore–Penrose inverse, we apply this calculation to $\mathbf{A}'\mathbf{A}$, using only the nonzero roots, then postmultiply the result by $\mathbf{A}'$. Let $\mathbf{C}_1$ be the R characteristic vectors corresponding to the nonzero roots, which we array in the diagonal matrix, $\mathbf{\Lambda}_1$. Then the Moore–Penrose inverse is

$$\mathbf{A}^{+} = \mathbf{C}_1\mathbf{\Lambda}_1^{-1}\mathbf{C}_1'\mathbf{A}',$$

which is very similar to the previous result.

If $\mathbf{A}$ is a symmetric matrix with rank $R \leq K$, the Moore–Penrose inverse is computed precisely as in the preceding equation without postmultiplying by $\mathbf{A}'$. Thus, for a symmetric matrix $\mathbf{A}$**,**

$$\mathbf{A}^{+} = \mathbf{C}_1\mathbf{\Lambda}_1^{-1}\mathbf{C}_1',$$

where $\mathbf{\Lambda}_1^{-1}$ is a diagonal matrix containing the reciprocals of the *nonzero* roots of $\mathbf{A}$.

A.7 QUADRATIC FORMS AND DEFINITE MATRICES

Many optimization problems involve double sums of the form

$$q = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij}. \qquad \textbf{(A-109)}$$

[14]A proof of uniqueness, with several other results, may be found in Theil (1983).

This **quadratic form** can be written

$$q = \mathbf{x}'\mathbf{A}\mathbf{x}$$

where $\mathbf{A}$ is a symmetric matrix. In general, q may be positive, negative, or zero; it depends on $\mathbf{A}$ and $\mathbf{x}$. There are some matrices, however, for which q will be positive regardless of $\mathbf{x}$, and others for which q will always be negative (or nonnegative or nonpositive). For a given matrix $\mathbf{A}$,

1. If $\mathbf{x}'\mathbf{A}\mathbf{x} > (<)\ 0$ for all nonzero $\mathbf{x}$, then $\mathbf{A}$ is **positive (negative) definite.**
2. If $\mathbf{x}'\mathbf{A}\mathbf{x} \geq (\leq)\ 0$ for all nonzero $\mathbf{x}$, then $\mathbf{A}$ is **nonnegative definite** or **positive semidefinite** (nonpositive definite).

It might seem that it would be impossible to check a matrix for definiteness, since $\mathbf{x}$ can be chosen arbitrarily. But we have already used the set of results necessary to do so. Recall that a symmetric matrix can be decomposed into

$$\mathbf{A} = \mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'.$$

Therefore, the quadratic form can be written as

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{x}'\mathbf{C}\boldsymbol{\Lambda}\mathbf{C}'\mathbf{x}.$$

Let $\mathbf{y} = \mathbf{C}'\mathbf{x}$. Then

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \mathbf{y}'\boldsymbol{\Lambda}\mathbf{y} = \sum_{i=1}^{n} \lambda_i y_i^2. \qquad \textbf{(A-110)}$$

If λ_i is positive for all i, then regardless of $\mathbf{y}$—that is, regardless of $\mathbf{x}$—q will be positive. This case was identified earlier as a positive definite matrix. Continuing this line of reasoning, we obtain the following theorem.

THEOREM A.11 Definite Matrices
Let $\mathbf{A}$ *be a symmetric matrix. If all the characteristic roots of* $\mathbf{A}$ *are positive (negative), then* $\mathbf{A}$ *is* **positive definite (negative definite)**. *If some of the roots are zero, then* $\mathbf{A}$ *is* **nonnegative (nonpositive) definite** *if the remainder are positive (negative). If* $\mathbf{A}$ *has both negative and positive roots, then* $\mathbf{A}$ *is* **indefinite.**

The preceding statements give, in each case, the "if" parts of the theorem. To establish the "only if" parts, assume that the condition on the roots does not hold. This must lead to a contradiction. For example, if some λ can be negative, then $\mathbf{y}'\boldsymbol{\Lambda}\mathbf{y}$ could be negative for some $\mathbf{y}$, so $\mathbf{A}$ cannot be positive definite.

A.7.1 NONNEGATIVE DEFINITE MATRICES

A case of particular interest is that of nonnegative definite matrices. Theorem A.11 implies a number of related results.

- If $\mathbf{A}$ is nonnegative definite, then $|\mathbf{A}| \geq 0$. **(A-111)**

Proof: The determinant is the product of the roots, which are nonnegative.

The converse, however, is not true. For example, a 2×2 matrix with two negative roots is clearly not positive definite, but it does have a positive determinant.

- If $\mathbf{A}$ is positive definite, so is $\mathbf{A}^{-1}$. **(A-112)**

  Proof: The roots are the reciprocals of those of $\mathbf{A}$, which are, therefore positive.

- The identity matrix $\mathbf{I}$ is positive definite. **(A-113)**

  Proof: $\mathbf{x}'\mathbf{I}\mathbf{x} = \mathbf{x}'\mathbf{x} > 0 \quad \textbf{if} \quad \mathbf{x} \neq \mathbf{0}$.

A very important result for regression analysis is

- If $\mathbf{A}$ is $n \times K$ with full column rank and $n > K$, then $\mathbf{A}'\mathbf{A}$ is positive definite and $\mathbf{A}\mathbf{A}'$ is nonnegative definite. **(A-114)**

  Proof: By assumption, $\mathbf{Ax} \neq \mathbf{0}$. So $\mathbf{x}'\mathbf{A}'\mathbf{A}\mathbf{x} = (\mathbf{Ax})'(\mathbf{Ax}) = \mathbf{y}'\mathbf{y} = \sum_j y_j^2 > 0$.

A similar proof establishes the nonnegative definiteness of $\mathbf{A}\mathbf{A}'$. The difference in the latter case is that because $\mathbf{A}$ has more rows than columns there is an $\mathbf{x}$ such that $\mathbf{A}'x = 0$. Thus, in the proof, we only have $\mathbf{y}'\mathbf{y} \geq 0$. The case in which $\mathbf{A}$ does not have full column rank is the same as that of $\mathbf{A}\mathbf{A}'$.

- If $\mathbf{A}$ is positive definite and $\mathbf{B}$ is a nonsingular matrix, then $\mathbf{B}'\mathbf{A}\mathbf{B}$ is positive definite. **(A-115)**

  Proof: $\mathbf{x}'\mathbf{B}'\mathbf{A}\mathbf{B}\mathbf{x} = \mathbf{y}'\mathbf{A}\mathbf{y} > 0$, where $\mathbf{y} = \mathbf{Bx}$. But $\mathbf{y}$ cannot be $\mathbf{0}$ because $\mathbf{B}$ is nonsingular.

Finally, note that for $\mathbf{A}$ to be negative definite, all $\mathbf{A}$'s characteristic roots must be negative. But, in this case, $|\mathbf{A}|$ is positive if $\mathbf{A}$ is of even order and negative if $\mathbf{A}$ is of odd order.

A.7.2 IDEMPOTENT QUADRATIC FORMS

Quadratic forms in idempotent matrices play an important role in the distributions of many test statistics. As such, we shall encounter them fairly often. Two central results are of interest.

- Every symmetric idempotent matrix is nonnegative definite. **(A-116)**

  Proof: All roots are one or zero; hence, the matrix is nonnegative definite by definition.

Combining this with some earlier results yields a result used in determining the sampling distribution of most of the standard test statistics.

- If $\mathbf{A}$ is symmetric and idempotent, $n \times n$ with rank J, then every quadratic form in $\mathbf{A}$ can be written

$$\mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{j=1}^{J} y_j^2 \qquad \textbf{(A-117)}$$

  Proof: This result is (A-110) with $\lambda =$ one or zero.

A.7.3 COMPARING MATRICES

Derivations in econometrics often focus on whether one matrix is "larger" than another. We now consider how to make such a comparison. As a starting point, the two matrices must have the same dimensions. A useful comparison is based on

$$d = \mathbf{x}'\mathbf{A}\mathbf{x} - \mathbf{x}'\mathbf{B}\mathbf{x} = \mathbf{x}'(\mathbf{A} - \mathbf{B})\mathbf{x}.$$

If d is always positive for any nonzero vector, $\mathbf{x}$, then by this criterion, we can say that $\mathbf{A}$ is larger than $\mathbf{B}$. The reverse would apply if d is always negative. It follows from the definition that

$$\text{if } d > 0 \text{ for all nonzero } \mathbf{x}, \text{ then } \mathbf{A} - \mathbf{B} \text{ is positive definite.} \tag{A-118}$$

If d is only greater than or equal to zero, then $\mathbf{A} - \mathbf{B}$ is nonnegative definite. The ordering is not complete. For some pairs of matrices, d could have either sign, depending on $\mathbf{x}$. In this case, there is no simple comparison.

A particular case of the general result which we will encounter frequently is.

$$\begin{aligned}&\text{If } \mathbf{A} \text{ is positive definite and } \mathbf{B} \text{ is nonnegative definite,}\\ &\text{then } \mathbf{A} + \mathbf{B} \geq \mathbf{A}.\end{aligned} \tag{A-119}$$

Consider, for example, the "updating formula" introduced in (A-66). This uses a matrix

$$\mathbf{A} = \mathbf{B}'\mathbf{B} + \mathbf{b}\mathbf{b}' \geq \mathbf{B}'\mathbf{B}.$$

Finally, in comparing matrices, it may be more convenient to compare their inverses. The result analogous to a familiar result for scalars is:

$$\text{If } \mathbf{A} > \mathbf{B}, \text{ then } \mathbf{B}^{-1} > \mathbf{A}^{-1}. \tag{A-120}$$

To establish this intuitive result, we would make use of the following, which is proved in Goldberger (1964, Chapter 2):

THEOREM A.12 Ordering for Positive Definite Matrices
If $\mathbf{A}$ *and* $\mathbf{B}$ *are two positive definite matrices with the same dimensions and if every characteristic root of* $\mathbf{A}$ *is larger than (at least as large as) the corresponding characteristic root of* $\mathbf{B}$ *when both sets of roots are ordered from largest to smallest, then* $\mathbf{A} - \mathbf{B}$ *is positive (nonnegative) definite.*

The roots of the inverse are the reciprocals of the roots of the original matrix, so the theorem can be applied to the inverse matrices.

A.8 CALCULUS AND MATRIX ALGEBRA[15]

A.8.1 DIFFERENTIATION AND THE TAYLOR SERIES

A variable y is a function of another variable x written

$$y = f(x), \quad y = g(x), \quad y = y(x),$$

[15]For a complete exposition, see Magnus and Neudecker (2007).

and so on, if each value of x is associated with a single value of y. In this relationship, y and x are sometimes labeled the **dependent variable** and the **independent variable**, respectively. Assuming that the function $f(x)$ is continuous and differentiable, we obtain the following derivatives:

$$f'(x) = \frac{dy}{dx}, f''(x) = \frac{d^2y}{dx^2},$$

and so on.

A frequent use of the derivatives of $f(x)$ is in the **Taylor series approximation**. A Taylor series is a polynomial approximation to $f(x)$. Letting x^0 be an arbitrarily chosen expansion point

$$f(x) \approx f(x^0) + \sum_{i=1}^{P} \frac{1}{i!}\frac{d^i f(x^0)}{d(x^0)^i}(x - x^0)^i. \tag{A-121}$$

The choice of P, the number of terms, is arbitrary; the more that are used, the more accurate the approximation will be. The approximation used most frequently in econometrics is the **linear approximation**,

$$f(x) \approx \alpha + \beta x, \tag{A-122}$$

where, by collecting terms in (A-121), $\alpha = [f(x^0) - f'(x^0)x^0]$ and $\beta = f'(x^0)$. The superscript "0" indicates that the function is evaluated at x^0. The **quadratic approximation** is

$$f(x) \approx \alpha + \beta x + \gamma x^2, \tag{A-123}$$

where $\alpha = [f^0 - f'^0 x^0 + \frac{1}{2}f''^0(x^0)^2]$, $\beta = [f'^0 - f''^0 x^0]$ and $\gamma = \frac{1}{2}f''^0$.

We can regard a function $y = f(x_1, x_2, \dots, x_n)$ as a **scalar-valued function** of a vector; that is, $y = f(\mathbf{x})$. The vector of partial derivatives, or **gradient vector**, or simply **gradient**, is

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \partial y/\partial x_1 \\ \partial y/\partial x_2 \\ \cdots \\ \partial y/\partial x_n \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \cdots \\ f_n \end{bmatrix}. \tag{A-124}$$

The vector $\mathbf{g}(\mathbf{x})$ or $\mathbf{g}$ is used to represent the gradient. Notice that it is a column vector. The shape of the derivative is determined by the denominator of the derivative.

A **second derivatives matrix** or **Hessian** is computed as

$$\mathbf{H} = \begin{bmatrix} \partial^2 y/\partial x_1 \partial x_1 & \partial^2 y/\partial x_1 \partial x_2 & \cdots & \partial^2 y/\partial x_1 \partial x_n \\ \partial^2 y/\partial x_2 \partial x_1 & \partial^2 y/\partial x_2 \partial x_2 & \cdots & \partial^2 y/\partial x_2 \partial x_n \\ \cdots & \cdots & \cdots & \cdots \\ \partial^2 y/\partial x_n \partial x_1 & \partial^2 y/\partial x_n \partial x_2 & \cdots & \partial^2 y/\partial x_n \partial x_n \end{bmatrix} = [f_{ij}]. \tag{A-125}$$

In general, $\mathbf{H}$ is a square, symmetric matrix. (The symmetry is obtained for continuous and continuously differentiable functions from Young's theorem.) Each column of $\mathbf{H}$ is the derivative of $\mathbf{g}$ with respect to the corresponding variable in $\mathbf{x}'$. Therefore,

$$\mathbf{H} = \left[\frac{\partial(\partial y/\partial \mathbf{x})}{\partial x_1}\frac{\partial(\partial y/\partial \mathbf{x})}{\partial x_2}\cdots\frac{\partial(\partial y/\partial \mathbf{x})}{\partial x_n}\right] = \frac{\partial(\partial y/\partial \mathbf{x})}{\partial(x_1 \quad x_2 \cdots x_n)} = \frac{\partial(\partial y/\partial \mathbf{x})}{\partial \mathbf{x}'} = \frac{\partial^2 y}{\partial \mathbf{x}\partial \mathbf{x}'}.$$

The first-order, or linear Taylor series approximation is

$$y \approx f(\mathbf{x}^0) + \sum_{i=1}^{n} f_i(\mathbf{x}^0)(x_i - x_i^0). \tag{A-126}$$

The right-hand side is

$$f(\mathbf{x}^0) + \left[\frac{\partial f(\mathbf{x}^0)}{\partial \mathbf{x}^0}\right]'(\mathbf{x} - \mathbf{x}^0) = [f(\mathbf{x}^0) - \mathbf{g}(\mathbf{x}^0)'\mathbf{x}^0] + \mathbf{g}(\mathbf{x}^0)'\mathbf{x} = [f^0 - \mathbf{g}^{0\prime}\mathbf{x}^0] + \mathbf{g}^{0\prime}\mathbf{x}.$$

This produces the linear approximation,

$$y \approx \alpha + \boldsymbol{\beta}'\mathbf{x}.$$

The second-order, or quadratic, approximation adds the second-order terms in the expansion,

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} f_{ij}^0(x_i - x_i^0)(x_j - x_j^0) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^0)'\mathbf{H}^0(\mathbf{x} - \mathbf{x}^0),$$

to the preceding one. Collecting terms in the same manner as in (A-126), we have

$$y \approx \alpha + \boldsymbol{\beta}'\mathbf{x} + \frac{1}{2}\mathbf{x}'\boldsymbol{\Gamma}\mathbf{x}, \tag{A-127}$$

where

$$\alpha = f^0 - \mathbf{g}^{0\prime}\mathbf{x}^0 + \frac{1}{2}\mathbf{x}^{0\prime}\mathbf{H}^0\mathbf{x}^0, \quad \boldsymbol{\beta} = \mathbf{g}^0 - \mathbf{H}^0\mathbf{x}^0 \quad \text{and} \quad \boldsymbol{\Gamma} = \mathbf{H}^0.$$

A linear function can be written

$$y = \mathbf{a}'\mathbf{x} = \mathbf{x}'\mathbf{a} = \sum_{i=1}^{n} a_i x_i,$$

so

$$\frac{\partial(\mathbf{a}'\mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}. \tag{A-128}$$

Note, in particular, that $\partial(\mathbf{a}'\mathbf{x})/\partial\mathbf{x} = \mathbf{a}$, not $\mathbf{a}'$. In a set of linear functions

$$\mathbf{y} = \mathbf{A}\mathbf{x},$$

each element y_i of $\mathbf{y}$ is

$$y_i = \mathbf{a}_i'\mathbf{x},$$

where $\mathbf{a}_i'$ is the ith row of $\mathbf{A}$ [see (A-14)]. Therefore,

$$\frac{\partial y_i}{\partial \mathbf{x}} = \mathbf{a}_i = \text{transpose of } i\text{th row of } \mathbf{A},$$

and

$$\begin{bmatrix} \partial y_1/\partial \mathbf{x}' \\ \partial y_2/\partial \mathbf{x}' \\ \cdots \\ \partial y_n/\partial \mathbf{x}' \end{bmatrix} = \begin{bmatrix} \mathbf{a}_1' \\ \mathbf{a}_2' \\ \cdots \\ \mathbf{a}_n' \end{bmatrix}.$$

Collecting all terms, we find that $\partial \mathbf{Ax}/\partial \mathbf{x}' = \mathbf{A}$, whereas the more familiar form will be

$$\frac{\partial \mathbf{x}'\mathbf{A}'}{\partial \mathbf{x}} = \mathbf{A}'. \tag{A-129}$$

A quadratic form is written

$$\mathbf{x}'\mathbf{Ax} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j a_{ij}. \tag{A-130}$$

For example,

$$\mathbf{A} = \begin{bmatrix} 1 & 3 \\ 3 & 4 \end{bmatrix},$$

so that

$$\mathbf{x}'\mathbf{Ax} = 1x_1^2 + 4x_2^2 + 6x_1x_2.$$

Then

$$\frac{\partial \mathbf{x}'\mathbf{Ax}}{\partial \mathbf{x}} = \begin{bmatrix} 2x_1 + 6x_2 \\ 6x_1 + 8x_2 \end{bmatrix} = \begin{bmatrix} 2 & 6 \\ 6 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 2\mathbf{Ax}, \tag{A-131}$$

which is the general result when $\mathbf{A}$ is a symmetric matrix. If $\mathbf{A}$ is not symmetric, then

$$\frac{\partial(\mathbf{x}'\mathbf{Ax})}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}')\mathbf{x}. \tag{A-132}$$

Referring to the preceding double summation, we find that for each term, the coefficient on a_{ij} is $x_i x_j$. Therefore,

$$\frac{\partial(\mathbf{x}'\mathbf{Ax})}{\partial a_{ij}} = x_i x_j.$$

The square matrix whose $i\,j$th element is $x_i x_j$ is $\mathbf{xx}'$, so

$$\frac{\partial(\mathbf{x}'\mathbf{Ax})}{\partial \mathbf{A}} = \mathbf{xx}'. \tag{A-133}$$

Derivatives involving determinants appear in maximum likelihood estimation. From the cofactor expansion in (A-51),

$$\frac{\partial|\mathbf{A}|}{\partial a_{ij}} = (-1)^{i+j}|\mathbf{A}_{ij}| = c_{ij}$$

where $|\mathbf{C}_{ji}|$ is the jith cofactor in $\mathbf{A}$. The inverse of $\mathbf{A}$ can be computed using

$$\mathbf{A}_{ij}^{-1} = \frac{|\mathbf{C}_{ji}|}{|\mathbf{A}|}$$

(note the reversal of the subscripts), which implies that

$$\frac{\partial \ln |\mathbf{A}|}{\partial a_{ij}} = \frac{(-1)^{i+j}\,|\mathbf{A}_{ij}|}{|\mathbf{A}|},$$

or, collecting terms,

$$\frac{\partial \ln |\mathbf{A}|}{\partial \mathbf{A}} = \mathbf{A}^{-1\prime}.$$

Because the matrices for which we shall make use of this calculation will be symmetric in our applications, the transposition will be unnecessary.

A.8.2 OPTIMIZATION

Consider finding the x where $f(x)$ is maximized or minimized. Because $f'(x)$ is the slope of $f(x)$, either optimum must occur where $f'(x) = 0$. Otherwise, the function will be increasing or decreasing at x. This result implies the **first-order or necessary condition for an optimum** (maximum or minimum):

$$\frac{dy}{dx} = 0. \qquad \textbf{(A-134)}$$

For a maximum, the function must be concave; for a minimum, it must be convex. The **sufficient condition for an optimum** is.

$$\begin{aligned}\text{For a maximum, } \frac{d^2y}{dx^2} &< 0;\\ \text{for a minimum, } \frac{d^2y}{dx^2} &> 0.\end{aligned} \qquad \textbf{(A-135)}$$

Some functions, such as the sine and cosine functions, have many **local optima**, that is, many minima and maxima. A function such as $(\cos x)/(1 + x^2)$, which is a damped cosine wave, does as well but differs in that although it has many local maxima, it has one, at $x = 0$, at which $f(x)$ is greater than it is at any other point. Thus, $x = 0$ is the **global maximum**, whereas the other maxima are only **local maxima**. Certain functions, such as a quadratic, have only a single optimum. These functions are **globally concave** if the optimum is a maximum and **globally convex** if it is a minimum.

For maximizing or minimizing a function of several variables, the first-order conditions are

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}. \qquad \textbf{(A-136)}$$

This result is interpreted in the same manner as the necessary condition in the univariate case. At the optimum, it must be true that no small change in any variable leads to an improvement in the function value. In the single-variable case, d^2y/dx^2 must be positive for a minimum and negative for a maximum. The second-order condition for an optimum in the multivariate case is that, at the optimizing value,

$$\mathbf{H} = \frac{\partial^2 f(\mathbf{x})}{\partial \mathbf{x}\, \partial \mathbf{x}'} \qquad \textbf{(A-137)}$$

must be positive definite for a minimum and negative definite for a maximum.

In a single-variable problem, the second-order condition can usually be verified by inspection. This situation will not generally be true in the multivariate case. As discussed earlier, checking the definiteness of a matrix is, in general, a difficult problem. For most of the problems encountered in econometrics, however, the second-order condition will be implied by the structure of the problem. That is, the matrix **H** will usually be of such a form that it is always definite.

For an example of the preceding, consider the problem

$$\text{maximize}_{\mathbf{x}} R = \mathbf{a}'\mathbf{x} - \mathbf{x}'\mathbf{A}\mathbf{x},$$

where

$$\mathbf{a}' = (5 \quad 4 \quad 2),$$

and

$$\mathbf{A} = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 3 & 2 \\ 3 & 2 & 5 \end{bmatrix}.$$

Using some now familiar results, we obtain

$$\frac{\partial R}{\partial \mathbf{x}} = \mathbf{a} - 2\mathbf{A}\mathbf{x} = \begin{bmatrix} 5 \\ 4 \\ 2 \end{bmatrix} - \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 4 \\ 6 & 4 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \mathbf{0}. \qquad \textbf{(A-138)}$$

The solutions are

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 & 2 & 6 \\ 2 & 6 & 4 \\ 6 & 4 & 10 \end{bmatrix}^{-1} \begin{bmatrix} 5 \\ 4 \\ 2 \end{bmatrix} = \begin{bmatrix} 11.25 \\ 1.75 \\ -7.25 \end{bmatrix}.$$

The sufficient condition is that

$$\frac{\partial^2 R(\mathbf{x})}{\partial \mathbf{x}\, \partial \mathbf{x}'} = -2\mathbf{A} = \begin{bmatrix} -4 & -2 & -6 \\ -2 & -6 & -4 \\ -6 & -4 & -10 \end{bmatrix} \qquad \textbf{(A-139)}$$

must be negative definite. The three characteristic roots of this matrix are -15.746, -4, and -0.25403. Because all three roots are negative, the matrix is negative definite, as required.

In the preceding, it was necessary to compute the characteristic roots of the Hessian to verify the sufficient condition. For a general matrix of order larger than 2, this will normally require a computer. Suppose, however, that **A** is of the form

$$\mathbf{A} = \mathbf{B}'\mathbf{B},$$

where **B** is some known matrix. Then, as shown earlier, we know that **A** will always be positive definite (assuming that **B** has full rank). In this case, it is not necessary to calculate the characteristic roots of **A** to verify the sufficient conditions.

A.8.3 CONSTRAINED OPTIMIZATION

It is often necessary to solve an optimization problem subject to some constraints on the solution. One method is merely to "solve out" the constraints. For example, in the maximization problem considered earlier, suppose that the constraint $x_1 = x_2 - x_3$ is imposed on the solution. For a single constraint such as this one, it is possible merely to substitute the right-hand side of this equation for x_1 in the objective function and solve the resulting problem as a function of the remaining two variables. For more general constraints, however, or when there is more than one constraint, the method of Lagrange multipliers provides a more straightforward method of solving the problem. We seek to

$$\begin{aligned} \text{maximize}_{\mathbf{x}}\ f(\mathbf{x}) \text{ subject to } c_1(\mathbf{x}) &= 0 \\ c_2(\mathbf{x}) &= 0, \\ &\cdots \\ c_J(\mathbf{x}) &= 0. \end{aligned} \tag{A-140}$$

The Lagrangean approach to this problem is to find the stationary points—that is, the points at which the derivatives are zero—of

$$L^*(\mathbf{x}, \boldsymbol{\lambda}) = f(\mathbf{x}) + \sum_{j=1}^{J} \lambda_j c_j(\mathbf{x}) = f(\mathbf{x}) + \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x}). \tag{A-141}$$

The solutions satisfy the equations

$$\begin{aligned} \frac{\partial L^*}{\partial \mathbf{x}} &= \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + \frac{\partial \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x})}{\partial \mathbf{x}} = \mathbf{0}(n \times 1), \\ \frac{\partial L^*}{\partial \boldsymbol{\lambda}} &= \mathbf{c}(\mathbf{x}) = \mathbf{0}\ (J \times 1). \end{aligned} \tag{A-142}$$

The second term in $\partial L^*/\partial \mathbf{x}$ is

$$\frac{\partial \boldsymbol{\lambda}'\mathbf{c}(\mathbf{x})}{\partial \mathbf{x}} = \frac{\partial \mathbf{c}(\mathbf{x})'\boldsymbol{\lambda}}{\partial \mathbf{x}} = \left[\frac{\partial \mathbf{c}(\mathbf{x})'}{\partial \mathbf{x}}\right]\boldsymbol{\lambda} = \mathbf{C}'\boldsymbol{\lambda}, \tag{A-143}$$

where $\mathbf{C}$ is the matrix of derivatives of the constraints with respect to $\mathbf{x}$. The jth row of the $J \times n$ matrix $\mathbf{C}$ is the vector of derivatives of the jth constraint, $c_j(\mathbf{x})$, with respect to $\mathbf{x}'$. Upon collecting terms, the first-order conditions are

$$\begin{aligned} \frac{\partial L^*}{\partial \mathbf{x}} &= \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} + \mathbf{C}'\boldsymbol{\lambda} = \mathbf{0}, \\ \frac{\partial L^*}{\partial \lambda} &= \mathbf{c}(\mathbf{x}) = \mathbf{0}. \end{aligned} \tag{A-144}$$

There is one very important aspect of the constrained solution to consider. In the unconstrained solution, we have $\partial f(\mathbf{x})/\partial \mathbf{x} = \mathbf{0}$. From (A-144), we obtain, for a constrained solution,

$$\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = -\mathbf{C}'\boldsymbol{\lambda}, \tag{A-145}$$

which will not equal $\mathbf{0}$ unless $\boldsymbol{\lambda} = \mathbf{0}$. This result has two important implications:

- The constrained solution cannot be superior to the unconstrained solution. This is implied by the nonzero gradient at the constrained solution. (That is, unless $\mathbf{C} = \mathbf{0}$ which could happen if the constraints were nonlinear. But, even if so, the solution is still not better than the unconstrained optimum.)
- If the Lagrange multipliers are zero, then the constrained solution will equal the unconstrained solution.

To continue the example begun earlier, suppose that we add the following conditions:

$$x_1 - x_2 + x_3 = 0,$$
$$x_1 + x_2 + x_3 = 0.$$

To put this in the format of the general problem, write the constraints as $\mathbf{c}(\mathbf{x}) = \mathbf{Cx} = \mathbf{0}$, where

$$\mathbf{C} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}.$$

The Lagrangean function is

$$R^*(\mathbf{x}, \boldsymbol{\lambda}) = \mathbf{a}'\mathbf{x} - \mathbf{x}'\mathbf{Ax} + \boldsymbol{\lambda}'\mathbf{Cx}.$$

Note the dimensions and arrangement of the various parts. In particular, $\mathbf{C}$ is a 2×3 matrix, with one row for each constraint and one column for each variable in the objective function. The vector of Lagrange multipliers thus has two elements, one for each constraint. The necessary conditions are

$$\mathbf{a} - 2\mathbf{Ax} + \mathbf{C}'\boldsymbol{\lambda} = \mathbf{0} \quad \text{(three equations)}, \tag{A-146}$$

and

$$\mathbf{Cx} = \mathbf{0} \quad \text{(two equations)}.$$

These may be combined in the single equation

$$\begin{bmatrix} -2\mathbf{A} & \mathbf{C}' \\ \mathbf{C} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \boldsymbol{\lambda} \end{bmatrix} = \begin{bmatrix} -\mathbf{a} \\ \mathbf{0} \end{bmatrix}.$$

Using the partitioned inverse of (A-74) produces the solutions

$$\boldsymbol{\lambda} = -[\mathbf{CA}^{-1}\mathbf{C}']^{-1}\,\mathbf{CA}^{-1}\mathbf{a} \tag{A-147}$$

and

$$\mathbf{x} = \frac{1}{2}\mathbf{A}^{-1}[\mathbf{I} - \mathbf{C}'(\mathbf{CA}^{-1}\mathbf{C}')^{-1}\mathbf{CA}^{-1}]\mathbf{a}. \tag{A-148}$$

The two results, (A-147) and (A-148), yield analytic solutions for $\boldsymbol{\lambda}$ and $\mathbf{x}$. For the specific matrices and vectors of the example, these are $\lambda = [-0.5\ -7.5]'$, and the constrained solution vector, $\mathbf{x}^* = [1.50\ -1.5]'$. Note that in computing the solution to this sort of problem, it is not necessary to use the rather cumbersome form of (A-148). Once λ is obtained from (A-147), the solution can be inserted in (A-146) for a much simpler computation. The solution

$$\mathbf{x} = \frac{1}{2}\mathbf{A}^{-1}\mathbf{a} + \frac{1}{2}\mathbf{A}^{-1}\mathbf{C}'\boldsymbol{\lambda}$$

suggests a useful result for the constrained optimum:

$$\text{constrained solution} = \text{unconstrained solution} + [2\mathbf{A}]^{-1}\,\mathbf{C}'\boldsymbol{\lambda}. \qquad \textbf{(A-149)}$$

Finally, by inserting the two solutions in the original function, we find that $R = 24.375$ and $R^* = 2.25$, which illustrates again that the constrained solution (in this *maximization* problem) is inferior to the unconstrained solution.

A.8.4 TRANSFORMATIONS

If a function is strictly monotonic, then it is a **one-to-one function**. Each y is associated with exactly one value of x, and vice versa. In this case, an **inverse function** exists, which expresses x as a function of y, written

$$y = f(x)$$

and

$$x = f^{-1}(y).$$

An example is the inverse relationship between the log and the exponential functions.

The slope of the inverse function,

$$J = \frac{dx}{dy} = \frac{df^{-1}(y)}{dy} = f^{-1'}(y),$$

is the **Jacobian** of the transformation from y to x. For example, if

$$y = a + bx,$$

then

$$x = -\frac{a}{b} + \left[\frac{1}{b}\right]y$$

is the inverse transformation and

$$J = \frac{dx}{dy} = \frac{1}{b}.$$

Looking ahead to the statistical application of this concept, we observe that if $y = f(x)$ were *vertical,* then this would no longer be a functional relationship. The same x would be associated with more than one value of y. In this case, at this value of x, we would find that $J = 0$, indicating a singularity in the function.

If $\mathbf{y}$ is a column vector of functions, $\mathbf{y} = \mathbf{f}(\mathbf{x})$, then

$$\mathbf{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}'} = \begin{bmatrix} \partial x_1/\partial y_1 & \partial x_1/\partial y_2 & \cdots & \partial x_1/\partial y_n \\ \partial x_2/\partial y_1 & \partial x_2/\partial y_2 & \cdots & \partial x_2/\partial y_n \\ & & \vdots & \\ \partial x_n/\partial y_1 & \partial x_n/\partial y_2 & \cdots & \partial x_n/\partial y_n \end{bmatrix}.$$

Consider the set of linear functions $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{f}(\mathbf{x})$. The inverse transformation is $\mathbf{x} = \mathbf{f}^{-1}(\mathbf{y})$, which will be

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

if **A** is nonsingular. If **A** is singular, then there is no inverse transformation. Let **J** be the matrix of partial derivatives of the inverse functions:

$$\mathbf{J} = \left[\frac{\partial x_i}{\partial y_j}\right].$$

The absolute value of the determinant of **J**,

$$\text{abs}(|\mathbf{J}|) = \text{abs}\left(\det\left(\left[\frac{\partial \mathbf{x}}{\partial \mathbf{y}'}\right]\right)\right),$$

is the Jacobian determinant of the transformation from **y** to **x**. In the nonsingular case,

$$\text{abs}(|\mathbf{J}|) = \text{abs}(|\mathbf{A}^{-1}|) = \frac{1}{\text{abs}(|\mathbf{A}|)}.$$

In the singular case, the matrix of partial derivatives will be singular and the determinant of the Jacobian will be zero. In this instance, the singular Jacobian implies that **A** is singular or, equivalently, that the transformations from **x** to **y** are functionally dependent. The singular case is analogous to the single-variable case.

Clearly, if the vector **x** is given, then $\mathbf{y} = \mathbf{Ax}$ can be computed from **x**. Whether **x** can be deduced from **y** is another question. Evidently, it depends on the Jacobian. If the Jacobian is not zero, then the inverse transformations exist, and we can obtain **x**. If not, then we cannot obtain **x**.

APPENDIX B

PROBABILITY AND DISTRIBUTION THEORY

B.1 INTRODUCTION

This appendix reviews the distribution theory used later in the book. A previous course in statistics is assumed, so most of the results will be stated without proof. The more advanced results in the later sections will be developed in greater detail.

B.2 RANDOM VARIABLES

We view our observation on some aspect of the economy as the **outcome** or realization of a random process that is almost never under our (the analyst's) control. In the current literature, the descriptive (and perspective laden) term **data generating process (DPG)** is often used for this underlying mechanism. The observed (measured) outcomes of the process are assigned unique numeric values. The assignment is one to one; each outcome