APPENDIX A

MATHEMATICAL SUMMARY

THIS APPENDIX SUMMARIZES essential background material concerning matrices and vectors. It is by no means a complete exposition of the subject [see, for example, Stewart (1998), Golub and Van Loan (1996), and Meyer (2000)] but concentrates mainly on those features useful in optimization.

A.1 DEFINITIONS

A matrix is an array of numbers, symbols, functions, and so on

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

An element of the matrix $\mathbf{A}$ is denoted by a_{ij}, where the subscript i corresponds to the row number and subscript j corresponds to the column number. Thus $\mathbf{A}$ in (A.1) has a total of n rows and m columns, and the dimensions of $\mathbf{A}$ are n by $m (n \times m)$. If $m = n$, $\mathbf{A}$ is called a "square" matrix. If all elements of $\mathbf{A}$ are zero except the main diagonal (a_{ii}, $i = 1, \ldots, n$), $\mathbf{A}$ is called a diagonal matrix. A diagonal matrix with each $a_{ii} = 1$ is called the identity matrix, abbreviated $\mathbf{I}$.

Vectors are a special type of matrix, defined as having one column and n rows. For example in (A.2) $\mathbf{x}$ has n components

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad n \times 1 \text{ matrix, a vector} \tag{A.2}$$

A vector can be thought of as a point in n-dimensional space, although the graphical representation of such a point, when the dimension of the vector is greater than 3, is not feasible. The general rules for matrix addition, subtraction, and multiplication described in Section A.2 apply also to vectors.

The transpose of a matrix or a vector is formed by assembling the elements of the first row of the matrix as the elements of the first column of the transposed matrix, the second row into the second column, and so on. In other words, a_{ij} in the original matrix $\mathbf{A}$ becomes the component a_{ji} in the transpose $\mathbf{A}^T$. Note that the position of the diagonal components (a_{ij}) are unchanged by transposition. If the dimension of $\mathbf{A}$ is $n \times m$, the dimension of $\mathbf{A}^T$ is $m \times n$ (m rows and n columns). If square matrices $\mathbf{A}$ and $\mathbf{A}^T$ are identical, $\mathbf{A}$ is called a symmetric matrix. The transpose of a vector $\mathbf{x}$ is a row

$$\mathbf{x}^T = [x_1 \quad x_2 \quad \cdots \quad x_n] \tag{A.3}$$

A.2 BASIC MATRIX OPERATIONS

First we present the rules for equality, addition, and multiplication of matrices.

Equality

$$\mathbf{A} = \mathbf{B} \qquad \text{if and only if} \qquad a_{ij} = b_{ij} \quad \text{for all } i \text{ and } j$$

Furthermore, both **A** and **B** must have the same dimensions (**A** and **B** are "conformable").

Addition

$$\mathbf{A} + \mathbf{B} = \mathbf{C} \quad \text{requires that the element} \quad c_{ij} = a_{ij} + b_{ij}, \quad \text{for all } i \text{ and } j$$

A, **B**, and **C** must all have the same dimensions.

Multiplication

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

If the matrix **A** has dimensions $n \times m$ and **B** has dimensions $q \times r$, then to obtain the product **AB** requires that $m = q$ (the number of columns of **A** equals the number of rows of **B**). The resulting matrix **C** is of dimension $n \times r$ and thus depends on the dimensions of both **A** and **B**. An element c_{ij} of **C** is obtained by summing the products of the elements of the ith row of **A** times the corresponding elements of the jth column of **B**:

$$c_{ij} = \sum_{k=1}^{m} a_{ik} b_{kj} \tag{A.4}$$

Note that the number of terms in the summation is m, corresponding to the number of columns of **A** and the rows of **B**. Matrix multiplication in general is not commutative as is the case with scalars, that is,

$$\mathbf{AB} \neq \mathbf{BA}$$

Often the validity of this rule is obvious because the matrix dimensions are not conformable, but even for square matrices commutation is not allowed.

Multiplication of a matrix by a scalar

Each component of the matrix is multiplied by the scalars,

$$s\mathbf{A} = \mathbf{B} \text{ is obtained by } s(a_{ij}) = b_{ij} \tag{A.5}$$

Transpose of a product of matrices

The transpose of a matrix product $(\mathbf{AB})^T$ is $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$. Likewise, $(\mathbf{ABC})^T = \mathbf{C}^T (\mathbf{AB})^T = \mathbf{C}^T \mathbf{B}^T \mathbf{A}^T$.

EXAMPLE A.1 MATRIX OPERATIONS

Consider a number of simple examples of these operations.

Multiplication:

For

$$\mathbf{A} = \underset{(3 \times 3)}{\begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}} \qquad \mathbf{B} = \underset{(3 \times 2)}{\begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}}$$

find **AB.**

Solution.

$$\underset{(3\times2)}{\mathbf{AB}} = \begin{bmatrix} 1(1) + 0(3) + 0(5) & 1(2) + 0(4) + 0(6) \\ 1(1) + 1(3) + 0(5) & 1(2) + 1(4) + 0(6) \\ 1(1) + 1(3) + 1(5) & 1(2) + 1(4) + 1(6) \end{bmatrix} = \begin{bmatrix} 1 & 2 \\ 4 & 6 \\ 9 & 12 \end{bmatrix}$$

Addition:

For

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 2 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 1 & 5 & 1 \\ 4 & 4 & 4 \\ 1 & 0 & 1 \end{bmatrix}$$

Find **A + B.**

Solution.

$$\mathbf{A} + \mathbf{B} = \begin{bmatrix} 1+1 & 0+5 & 2+1 \\ 1+4 & -1+4 & 0+4 \\ 0+1 & 0+0 & 0+1 \end{bmatrix} = \begin{bmatrix} 2 & 5 & 3 \\ 5 & 3 & 4 \\ 1 & 0 & 1 \end{bmatrix}$$

Subtraction:

For

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 2 & 6 \\ 1 & 3 \end{bmatrix}$$

find **A – B.**

Solution.

$$\mathbf{A} - \mathbf{B} = \begin{bmatrix} 1-2 & 1-6 \\ 1-1 & 1-3 \end{bmatrix} = \begin{bmatrix} -1 & -5 \\ 0 & -2 \end{bmatrix}$$

Transpose:

For

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} 2 & 1 \\ 3 & 4 \end{bmatrix}$$

find $(\mathbf{AB})^T$.

Solution.

$$(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T = \begin{bmatrix} 2 & 3 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 5 & 2 \\ 5 & 1 \end{bmatrix}$$

Multiplication of matrices by vectors:

A coordinate transformation can be performed by multiplying a matrix times a vector. If

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 2 & 0 & 3 \\ 4 & 8 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

find $\mathbf{y} = \mathbf{Ax}$.

$$\mathbf{y} = \begin{bmatrix} 1(1) + 1(1) + 2(1) \\ 2(1) + 0(1) + 3(1) \\ 4(1) + 8(1) + 4(1) \end{bmatrix} = \begin{bmatrix} 4 \\ 5 \\ 16 \end{bmatrix}$$

Note $\mathbf{y}$ has the same dimension as $\mathbf{x}$. We have transformed a point in three-dimensional space to another point in that same space.

Other commonly encountered vector–matrix products ($\mathbf{x}$ and $\mathbf{y}$ are n-component vectors) include

1. $\mathbf{x}^T\mathbf{x} = \sum_{i=1}^{n} x_i^2$ (a scalar) (A.6)

2. $\mathbf{x}^T\mathbf{y} = \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i$ (A.7)

   Equation (A.7) is referred to as the inner product, or dot product, of two vectors. If the two vectors are *orthogonal*, then $\mathbf{x}^T\mathbf{y} = 0$. In two or three dimensions, this means that the vectors $\mathbf{x}$ and $\mathbf{y}$ are perpendicular to each other.

3. $\mathbf{x}^T\mathbf{Ax}$ Here $\mathbf{A}$ is a square matrix of dimension $n \times n$ and the product is a scalar. If $\mathbf{A}$ is a diagonal matrix, then

$$\mathbf{x}^T\mathbf{Ax} = \sum_{i=1}^{n} a_{ii}x_i^2 \tag{A.8}$$

4. $$\mathbf{x}\mathbf{x}^T = \begin{bmatrix} x_1x_1 & x_1x_2 & \cdots & x_1x_n \\ \vdots & \vdots & & \vdots \\ x_nx_1 & & \cdots & x_nx_n \end{bmatrix} \tag{A.9}$$

Each vector has the dimensions ($n \times 1$) and the matrix is square ($n \times n$). Note that $\mathbf{x}\mathbf{x}^T$ is a matrix rather than a scalar (as with $\mathbf{x}^T\mathbf{x}$).

There is no matrix version of simple division, as with scalar quantities. Rather, the inverse of a matrix ($\mathbf{A}^{-1}$), which exists only for square matrices, is the closest analog to a divisor. An inverse matrix is defined such that $\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\,\mathbf{A} = \mathbf{I}$ (all three matrices are $n \times n$). In scalar algebra, the equation $a{\cdot}b = c$ can be solved for b by simply multiplying both sides of the equation by $1/a$. For a matrix equation, the analog of solving

$$\mathbf{AB} = \mathbf{C} \tag{A.10}$$

is to premultiply both sides by $\mathbf{A}^{-1}$:

$$\begin{aligned} \mathbf{A}^{-1}\mathbf{AB} &= \mathbf{A}^{-1}\mathbf{C} \\ \mathbf{IB} &= \mathbf{A}^{-1}\mathbf{C} \end{aligned} \tag{A.11}$$

Because $\mathbf{IB} = \mathbf{B}$, an explicit solution for $\mathbf{B}$ results. Note that the order of multiplication is critical because of the lack of commutation. Postmultiplication of both sides of Equation (A.10) by $\mathbf{A}^{-1}$ is allowable but does not lead to a solution for $\mathbf{B}$.

To get the inverse of a diagonal matrix, assemble the inverse of each element on the main diagonal. If

$$\mathbf{A} = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}$$

then

$$\mathbf{A}^{-1} = \begin{bmatrix} 1/a_{11} & 0 & 0 \\ 0 & 1/a_{22} & 0 \\ 0 & 0 & 1/a_{33} \end{bmatrix}$$

The proof is evident by multiplication: $\mathbf{A}\mathbf{A}^{-1} = \mathbf{I}$.

For a general square matrix of size 2×2 or 3×3, the procedure is more involved and is discussed later in Examples A.3 and A.7.

The *determinant* (denoted by det [**A**] or $|\mathbf{A}|$) is reasonably easy to calculate by hand for matrices up to size 3×3:

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

Another way to calculate the value of a determinant is to evaluate its cofactors. The cofactor of an element a_{ij} of the matrix is found by first deleting from the original matrix the ith row and jth column corresponding to that element; the resulting array is the minor (M_{ij}) for that element and has dimension $(n - 1) \times (n - 1)$. The cofactor is defined as

$$c_{ij} = (-1)^{i+j} \det M_{ij} \tag{A.13}$$

The determinant of the original matrix is calculated by either

1. $\sum_{j=1}^{n} a_{ij}c_{ij}$ (i fixed arbitrarily; row expansion) (A.14)

or

2. $\sum_{i=1}^{n} a_{ij}c_{ij}$ (j fixed arbitrarily; column expansion) (A.15)

For example, if

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

an expansion of the first row gives

$$\det[\mathbf{A}] = a_{11}c_{11} + a_{12}c_{12}$$
$$c_{11} = (-1)^{1+1}a_{22} = a_{22}$$
$$c_{12} = (-1)^{1+2}a_{21} = -a_{21}$$

so that

$$\det[\mathbf{A}] = a_{11}a_{22} - a_{12}a_{21}$$

EXAMPLE A.2 CALCULATE THE VALUE OF A DETERMINANT USING COFACTORS

Calculate the determinant

$$\det\begin{bmatrix} 1 & 2 & 1 \\ 2 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

using the first row as the expansion.

Solution.

$$\det[\mathbf{A}] = c_{11} + 2c_{12} + c_{13} = \det\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} - 2\det\begin{bmatrix} 2 & 1 \\ 0 & 1 \end{bmatrix} + \det\begin{bmatrix} 2 & 1 \\ 0 & 0 \end{bmatrix}$$

$$1 = 1 - 4 + 0 = -3$$

It is actually easier to use the third row because of its two zeros.

$$\det \mathbf{A} = c_{33} = (-1)^{3+3}\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} = -3$$

The *adjoint* of a matrix is constructed using the cofactors defined earlier. The elements $\bar{a}_{ij}$ of the adjoint matrix $\bar{\mathbf{A}}$ are defined as

$$\bar{a}_{ij} = c_{ji} \tag{A.16}$$

In other words, the adjoint matrix is the array composed of the transpose of the cofactors.

The adjoint of $\mathbf{A}$ can be used to directly calculate the inverse, $\mathbf{A}^{-1}$.

$$\mathbf{A}^{-1} = \frac{\text{adj}[\mathbf{A}]}{|\mathbf{A}|} \tag{A.17}$$

Note that the denominator of (A.17), the determinant of $\mathbf{A} \equiv |\mathbf{A}|$, is a scalar. If $|\mathbf{A}| = 0$, the inverse does not exist. A square matrix with determinant equal to zero is called a *singular* matrix. Conversely, for a nonsingular matrix $\mathbf{A}$, $\det \mathbf{A} \neq 0$.

EXAMPLE A.3 CALCULATION OF THE INVERSE OF A MATRIX

Consider the following matrix and find its inverse.

$$\mathbf{A} = \begin{bmatrix} 1 & 4 \\ 2 & 1 \end{bmatrix} \qquad |\mathbf{A}| = 1 - 8 = -7$$

Solution. The cofactors are

$$c_{11} = 1 \qquad c_{12} = -2 \qquad c_{21} = -4 \qquad c_{22} = 1$$

$$\text{adj}\ \mathbf{A} = \begin{bmatrix} 1 & -4 \\ -2 & 1 \end{bmatrix}$$

$$\mathbf{A}^{-1} = \frac{1}{-7}\begin{bmatrix} 1 & -4 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} \frac{-1}{7} & \frac{4}{7} \\ \frac{2}{7} & \frac{-1}{7} \end{bmatrix}$$

The use of Equation (A.17) for inversion is conceptually simple, but it is not a very efficient method for calculating the inverse matrix. A method based on use of row operations is discussed in Section A.3. For matrices of size larger than 3 × 3, we recommend that you use software such as MATLAB to find $\mathbf{A}^{-1}$.

Another use for the matrix inverse is to express one set of variables in terms of another, an important operation in constrained optimization (see Chapter 8). For example, suppose $\mathbf{x}$ and $\mathbf{z}$ are two n-vectors that are related by

$$\mathbf{z} = \mathbf{A}\mathbf{x} \tag{A.18}$$

Then, to express $\mathbf{x}$ in terms of $\mathbf{z}$, merely multiply both sides of (A.18) by $\mathbf{A}^{-1}$ (note that $\mathbf{A}$ must be $n \times n$):

$$\mathbf{A}^{-1}\mathbf{z} = \mathbf{x} \tag{A.19}$$

EXAMPLE A.4 RELATION OF VARIABLES

Suppose that

$$z_1 = x_1 + x_2$$

and

$$z_2 = 2x_1 + x_2$$

What are x_1 and x_2 in terms of z_1 and z_2?

Solution. Let

$$\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \quad \text{and} \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Therefore $\mathbf{z} = \mathbf{A}\mathbf{x}$, where

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

The inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{bmatrix} -1 & 1 \\ 2 & -1 \end{bmatrix}$$

hence $\mathbf{x} = \mathbf{A}^{-1}\mathbf{z}$ or

$$x_1 = -z_1 + z_2$$

$$x_2 = 2z_1 - z_2$$

The inverse matrix also can be employed in the solution of linear algebraic equations,

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{A.20}$$

which arise in many applications of engineering as well as in optimization theory. To have a unique solution to Equation (A.20), there must be the same number of independent equations as unknown variables. Note that the number of equations is equal to the number of rows of **A**, and the number of unknowns is equal to the number of columns of **A**.

With the inverse matrix, you can solve directly for **x**:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} \tag{A.21}$$

Although this is a conceptually convenient way to solve for **x**, it is not necessarily the most efficient method for doing so. We shall return to the matter of solving linear equations in Section A.4.

The final matrix characteristic covered here involves differentiation of function of a vector with respect to a vector. Suppose $f(\mathbf{x})$ is a scalar function of n variables $(x_1, x_2, \ldots, x_n)$. The first partial derivative of $f(\mathbf{x})$ with respect to **x** is

$$\frac{\partial f}{\partial \mathbf{x}} = \nabla_{\mathbf{x}} f = \left[\frac{\partial f}{\partial x_1} \; \frac{\partial f}{\partial x_2} \; \cdots \; \frac{\partial f}{\partial x_n}\right]^T$$

For a vector function $\mathbf{h}(\mathbf{x})$, such as occurs in a series of nonlinear multivariable constraints

$$\begin{aligned} h_1(x_1, x_2, \ldots, x_n) &= 0 \\ h_2(x_1, x_2, \ldots, x_n) &= 0 \\ &\vdots \\ h_m(x_1, x_2, \ldots, x_n) &= 0 \end{aligned}$$

the matrix of first partial derivatives, called the Jacobian matrix, is

$$\mathbf{J} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}} = \begin{bmatrix} \dfrac{\partial h_1}{\partial x_1} & \dfrac{\partial h_1}{\partial x_2} & \cdots & \dfrac{\partial h_1}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial h_m}{\partial x_1} & \dfrac{\partial h_m}{\partial x_2} & \cdots & \dfrac{\partial h_m}{\partial x_n} \end{bmatrix}$$

For a scalar function, the matrix of second derivatives, called the Hessian matrix, is

$$\mathbf{H}(\mathbf{x}) \equiv \nabla^2 f = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x_1^2} & \dfrac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \dfrac{\partial^2 f}{\partial x_1 \partial x_n} \\ \dfrac{\partial^2 f}{\partial x_2 \partial x_1} & \dfrac{\partial^2 f}{\partial x_2^2} & \cdots & \dfrac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & & & \vdots \\ \dfrac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \cdots & \dfrac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

The use of this matrix and its eigenvalue properties is discussed in several chapters. For continuously differentiable functions, **H** is symmetric.

A.3 LINEAR INDEPENDENCE AND ROW OPERATIONS

As mentioned earlier, singular matrices have a determinant of zero value. This outcome occurs when a row or column contains all zeros or when a row (or column) in the matrix is linearly dependent on one or more of the other rows (or columns). It can be shown that for a square matrix, row dependence implies column dependence. By definition the columns of **A**, $\mathbf{a}_j$, are linearly independent if

$$\sum_{j=1}^{n} d_j \mathbf{a}_j = \mathbf{0} \quad \text{only if } d_j = 0 \text{ for all } j \tag{A.22}$$

Conversely, linear dependence occurs when some nonzero set of values for d_j satisfies Equation (A.22). The *rank* of a matrix is defined as the number of linearly independent columns ($\leq n$).

EXAMPLE A.5 LINEAR INDEPENDENCE AND THE RANK OF A MATRIX

Calculate the rank of

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

Solution. Note that columns 1 and 3 are identical. Likewise the third row can be formed by multiplying the first row by 2. Equation (A.22) is

$$d_1 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + d_2 \begin{bmatrix} -1 \\ 0 \\ -2 \end{bmatrix} + d_3 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} = 0$$

One solution of (A.22) is $d_1 = 1$, $d_2 = 0$, $d_3 = -1$. Because a nontrivial (nonzero) solution exists, then the matrix has one dependent and two independent columns, and the rank ≤ 2 (here 2). The determinant is zero, as can be readily verified using Equation (A.12).

In general for a matrix, the determination of linear independence cannot be performed by inspection. For large matrices, rather than solving the set of linear equations (A.22), elementary row or column operations can be used to demonstrate linear independence. These operations involve adding some multiple of one row to another row, analogous to the types of algebraic operations (discussed later) that are used to solve simultaneous equations. The value of the determinant of **A** is invariant under these row (or column) operations. Implications with respect to linear independence and the use of determinants for equation-solving are discussed in Section A.4.

EXAMPLE A.6 USE OF ROW OPERATIONS

Use row operations to determine if the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 2 \\ 3 & 4 & 5 \end{bmatrix}$$

is nonsingular, that is, composed of linearly independent columns.

Solution. First create zeros in the a_{21} and a_{31} position by multiplication or addition. The necessary transformations are

1. Multiply row 1 by (–1); add to row 2

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 3 & 4 & 5 \end{bmatrix}$$

2. Multiply row 1 by (–3); add to row 3

$$\mathbf{C}_2 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 4 & 2 \end{bmatrix}$$

Next use row 2 to create a zero in a_{32}.

3. Multiply row 2 by (–2); add to row 3

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that neither rows 1 or 2 are changed in this step. The appearance of a row with all zero elements indicates that the matrix is singular (det [**A**] = 0).

Row operations can also be used to obtain an inverse matrix. Suppose we augment **A** with an identity matrix **I** of the same dimension; then multiply the augmented matrix by $\mathbf{A}^{-1}$:

$$\mathbf{A}^{-1}[\mathbf{A} \,\vdots\, \mathbf{I}] = [\mathbf{I} \,\vdots\, \mathbf{A}^{-1}] \tag{A.23}$$

If **A** is transformed by row operations to obtain **I**, $\mathbf{A}^{-1}$ occurs in the augmented part of the matrix.

EXAMPLE A.7 CALCULATION OF INVERSE MATRIX

Verify the results of Example A.3 using row operations.

Solution. Form the augmented matrix

$$\mathbf{C}_0 = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 2 & 1 & 0 & 1 \end{bmatrix}$$

Successive transformations would be

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 0 & -7 & -2 & 1 \end{bmatrix} \qquad \mathbf{C}_2 = \begin{bmatrix} 1 & 4 & 1 & 0 \\ 0 & 1 & \frac{2}{7} & -\frac{1}{7} \end{bmatrix}$$

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 0 & -\frac{1}{7} & \frac{4}{7} \\ 0 & 1 & \frac{2}{7} & -\frac{1}{7} \end{bmatrix}$$

Therefore the inverse of $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \begin{bmatrix} -\frac{1}{7} & \frac{4}{7} \\ \frac{2}{7} & -\frac{1}{7} \end{bmatrix}$$

A.4 SOLUTION OF LINEAR EQUATIONS

The need to solve sets of linear equations arises in many optimization applications. Consider Equation (A.20), where $\mathbf{A}$ is an $n \times n$ matrix corresponding to the coefficients in n equations in n unknowns. Because $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$, then from (A.17) $|\mathbf{A}|$ must be nonzero; $\mathbf{A}$ must have rank n, that is, no linearly dependent rows or columns exist, for a unique solution. Let us illustrate two cases where $|\mathbf{A}| = 0$:

$$2x_1 + 2x_2 = 6$$

$$x_1 + x_2 = 5$$

or

$$\begin{bmatrix} 2 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 6 \\ 5 \end{bmatrix}$$

It is obvious that only one linearly independent column or row exists, and $|\mathbf{A}|$ is zero. Note that there is no solution to this set of equations. As a second case, suppose $\mathbf{b}$ were changed to $\begin{bmatrix} 6 \\ 3 \end{bmatrix}$. Here an infinite number of solutions can be obtained, but no unique solution exists.

Degenerate cases such as those above are not frequently encountered. More often, $|\mathbf{A}| \neq 0$. Let

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

or

$$2x_1 + x_2 = 1 \tag{A.24a}$$

$$x_1 + 2x_2 = 0 \tag{A.24b}$$

By algebraic substitution, x_1 and x_2 can be found. Multiply Equation (A.24a) by (–0.5) and add this equation to (A.24b),

$$2x_1 + x_2 = 1 \tag{A.24c}$$

$$0 + 1.5x_2 = -0.5 \tag{A.24d}$$

Solve (A.24d) for $x_2 = -0.333$. This result can be substituted into (A.24c) to obtain $x_1 = 0.667$.

The steps employed in Equations (A.24) are equivalent to row operations. The use of row operations to simplify linear algebraic equations is the basis for Gaussian elimination (Golub and Van Loan, 1996). Gaussian elimination transforms the original matrix into upper triangular form, that is, all components of the matrix below the main diagonal are zero. Let us illustrate the process by solving a set of three equations in three unknowns for $\mathbf{x}$.

EXAMPLE A.8 SOLUTION OF SIMULTANEOUS LINEAR EQUATIONS

Solve for $\mathbf{x}$ given $\mathbf{A}$ and $\mathbf{b}$.

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 2 & 2 \\ 2 & 1 & 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

Solution. First a composite matrix from $\mathbf{A}$ and $\mathbf{b}$ is constructed:

$$\mathbf{C}_0 = [\mathbf{A} \,|\, \mathbf{b}] = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 2 & 2 & 0 \\ 2 & 1 & 1 & 2 \end{bmatrix}$$

Carry out row operations, keeping the first row intact; successive matrices are

$$\mathbf{C}_1 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & 1 & -1 \\ 2 & 1 & 1 & 2 \end{bmatrix} \qquad \mathbf{C}_2 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & 1 & -1 \\ 0 & 1 & -1 & 0 \end{bmatrix}$$

Next, with the second row in $\mathbf{C}_2$ kept intact, the upper triangular form is achieved by operating on the third row:

$$\mathbf{C}_3 = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 2 & 1 & -1 \\ 0 & 0 & -1.5 & 0.5 \end{bmatrix}$$

$\mathbf{C}_3$ can now be converted to the form of algebraic equations:

$$\begin{aligned} x_1 + x_3 &= 1 \\ 2x_2 + x_3 &= -1 \\ -1.5x_3 &= 0.5 \end{aligned}$$

which can be solved stage by stage starting with the last row to get $x_3 = -0.333$, $x_2 = -0.333$, $x_1 = 1.333$.

Gaussian elimination is a very efficient method for solving n equations in n unknowns, and this algorithm is readily available in many software packages. For solution of linear equations, this method is preferred computationally over the use of the matrix inverse. For hand calculations, Cramer's rule is also popular.

The determinant of $\mathbf{A}$ is unchanged by the row operations used in Gaussian elimination. Take the first three columns of $\mathbf{C}_3$ above. The determinant is simply the product of the diagonal terms. If none of the diagonal terms are zero when the matrix is reformulated as upper triangular, then $|\mathbf{A}| \neq 0$ and a solution exists. If $|\mathbf{A}| = 0$, there is no solution to the original set of equations.

A set of nonlinear equations can be solved by combining a Taylor series linearization with the linear equation-solving approach discussed above. For solving a single nonlinear equation, $h(x) = 0$, Newton's method applied to a function of a single variable is the well-known iterative procedure

$$x^{k+1} - x^k \equiv \Delta x^k = -\frac{h(x^k)}{dh(x^k)/dx} \tag{A.25}$$

or

$$\left[\frac{dh}{dx}\right]\bigg|_{x^k} (\Delta x^k) = -h(x^k)$$

where k is the iteration number and Δx^k is the correction to the previous value, x^k. Similarly, a set of nonlinear equations, $\mathbf{h(x)} = \mathbf{0}$, can be solved iteratively using Newton's method, by solving a set of linearized equations of the form $\mathbf{Ax} = \mathbf{b}$:

$$\left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]\bigg|_{\mathbf{x}^k} \cdot \Delta \mathbf{x}^k = -\mathbf{h}(\mathbf{x}^k) \tag{A.26}$$

Note that the Jacobian matrix $\partial\mathbf{h}/\partial\mathbf{x}$ on the left-hand side of Equation (A.26) is analogous to $\mathbf{A}$ in Equation (A.20), and $\Delta\mathbf{x}^k$ is analogous to $\mathbf{x}$. To compute the correction vector $\Delta\mathbf{x}$, $\partial\mathbf{h}/\partial\mathbf{x}$ must be nonsingular. However, there is no guarantee even then that Newton's method will converge to an $\mathbf{x}$ that satisfies $\mathbf{h(x)} = 0$.

In solving sets of simultaneous linear equations, the "condition" of the matrix is quite important. If some elements are quite large and some are quite small (but nonzero), numerical roundoff or truncation in a computer can have a significant effect on accuracy of the solution. A type of matrix is referred to as "ill conditioned" if it is nearly singular (equivalent to the scalar division by 0). A common measure of the degree of ill conditioning is the condition number, namely the ratio of the eigenvalues with largest (α_h) and smallest (α_l) modulus:

$$\text{Condition number} = \frac{|\alpha_h|}{|\alpha_l|} \tag{A.27}$$

The bigger the ratio, the worse the conditioning; a value of 1.0 is best. The calculation of eigenvalues are discussed in the next section. In general, as the dimension of the matrix increases, numerical accuracy of the elements is diminished. One technique to solve ill-conditioned sets of equations that has some advantages in speed and accuracy over Gaussian elimination is called "L-U decomposition" (Dongarra et al., 1979; Stewart, 1998), in which the original matrix is decomposed into upper and lower triangular forms.

A.5 EIGENVALUES, EIGENVECTORS

An $n \times n$ matrix has n eigenvalues. We define an n-vector $\mathbf{v}$, the eigenvector, which is associated with an eigenvalue e such that

$$\mathbf{Av} = e\mathbf{v} \tag{A.28}$$

Hence the product of the matrix $\mathbf{A}$ multiplying the eigenvector $\mathbf{v}$ is the same as the product obtained by multiplying the vector $\mathbf{v}$ by the scalar eigenvalue e. One eigenvector exists for each of the n eigenvalues. Eigenvalues and eigenvectors provide unambiguous information about the nature of functions used in optimization. If all eigenvalues of $\mathbf{A}$ are positive, then $\mathbf{A}$ is positive-definite. If all $e_i < 0$, then $\mathbf{A}$ is negative-definite. See Chapter 4 for a more complete discussion of definiteness and how it relates to convexity and concavity.

If we rearrange Equation (A.28) (note that the identity matrix must be introduced to maintain conformable matrices),

$$(\mathbf{A} - e\mathbf{I})\mathbf{v} = \mathbf{0} \tag{A.29}$$

$(\mathbf{A} - e\mathbf{I})$ in Equation (A.29) has the unknown variable e substracted from each diagonal element of $\mathbf{A}$. Equation (A.29) is a set of linear algebraic equations where $\mathbf{v}$ is the unknown vector. However, because the right-hand side of (A.29) is zero, either $\mathbf{v} = \mathbf{0}$ (the trivial solution), or a nonunique solution exists. For example in

$$2v_1 + 2v_2 = 0$$

$$v_1 + v_2 = 0$$

then det [**A**] $= 0$, and the solution is nonunique, that is, $v_1 = -v_2$. The equations are redundant. However, if one of the coefficients of v_1 or v_2 in Equation (A.29) changes, then the only solution is $v_1 = v_2 = 0$ (the trivial solution).

The determinant of $(\mathbf{A} - e\mathbf{I})$ must be zero for a nontrivial solution ($\mathbf{v} \neq \mathbf{0}$) to exist. Let us illustrate this idea with a (2×2) matrix:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \qquad (\mathbf{A} - e\mathbf{I}) = \begin{bmatrix} 1-e & 2 \\ 2 & 1-e \end{bmatrix}$$

$$\det\begin{bmatrix} 1-e & 2 \\ 2 & 1-e \end{bmatrix} = (1-e)^2 - 4 = e^2 - 2e - 3 = 0 \qquad \text{(A.30)}$$

Equation (A.30) determines values of e which yield a nontrivial solution. Factoring (A.30)

$$(e-3)(e+1) = 0 \qquad e = 3, -1$$

Therefore, the eigenvalues are 3 and -1. Note that for $e = 3$,

$$\mathbf{A} - e\mathbf{I} = \begin{bmatrix} -2 & 2 \\ 2 & -2 \end{bmatrix}$$

and for $e = -1$,

$$\mathbf{A} - e\mathbf{I} = \begin{bmatrix} 2 & 2 \\ 2 & 2 \end{bmatrix}$$

both of which are singular matrices.

For each eigenvalue there exists a corresponding eigenvector. For $e_1 = 3$, Equation (A.29) becomes

$$\begin{bmatrix} (1-3) & 2 \\ 2 & (1-3) \end{bmatrix}\begin{bmatrix} v_{11} \\ v_{12} \end{bmatrix} = 0$$

$$-2v_{11} + 2v_{12} = 0$$

$$2v_{11} - 2v_{12} = 0$$

Note that these equations are equivalent and cannot be solved uniquely; the solution to both equations is $v_{11} = v_{12}$. Thus, the eigenvector has direction but not length. The direction of the eigenvector can be specified by choosing v_{11} and calculating v_{12}. For example, let $v_{11} = 1$. Then

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

The magnitude of $\mathbf{v}_1$ cannot be determined uniquely. Similarly, for $e_2 = -1$,

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

is a solution of (A.29).

For a general $n \times n$ matrix, an nth-order polynomial results from solving det $(\mathbf{A} - e\mathbf{I}) = 0$. This polynomial will have n roots, and some of the roots may be imaginary numbers. A computer program can be used to generate the polynomial and factor it using a root-finding technique, such as Newton's method. However, more efficient iterative techniques can be found in computer software to calculate both e_i and $\mathbf{v}_i$ (Dongarra et al., 1979).

Principal minors

In Chapter 4 we discuss the definitions of convexity and concavity in terms of eigenvalues; an equivalent definition using determinants of principal minors is also provided. A principal minor of $\mathbf{A}$ of order k is a submatrix found by deleting any $n - k$ columns (and their corresponding rows) from the matrix. The leading principal minor of order k is found by deleting the last $n - k$ columns and rows. In Example A.2, the leading principal minor (order 1) is 1; the leading principal minor (order 2) is $\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$, and for order 3 the minor is the 3×3 matrix itself.

REFERENCES

Dongarra, J. J.; J. R. Bunch; C. B. Moler; and G. W. Stewart. *LINPACK User's Guide*, SIAM, Philadelphia, PA (1979).

Golub, G. H.; and C. F. Van Loan. *Matrix Computations*. Johns Hopkins, Baltimore, MD (1996).

Meyer, D. C. *Matrix Algebra and Applied Linear Algebra*. SIAM, Philadelphia, PA (2000).

Stewart, G. W. *Matrix Algorithms: Basic Decomposition*. SIAM, Philadelphia, PA (1998).

SUPPLEMENTARY REFERENCES

Amundson, N. R., *Mathematic Methods in Chemical Engineering: Matrices and Their Application*. Prentice-Hall, Englewood Cliffs, NJ (1966).

Campbell, H. G. *An Introduction to Matrices, Vectors, and Linear Programming*. Prentice-Hall, Englewood Cliffs, NJ (1977).

Daniel, J. W. *Applied Linear Algebra.* Prentice-Hall, Englewood Cliffs, NJ (1988).
Demmel, J. W. *Applied Numerical Linear Algebra.* SIAM, Philadelphia, PA (1997).

PROBLEMS

A.1 For

$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 & 4 \\ 1 & 3 \end{bmatrix}$$

Find
(a) $\mathbf{AB}$ and $\mathbf{BA}$ (compare)
(b) $\mathbf{A}^T\mathbf{B}$
(c) $\mathbf{A} + \mathbf{B}$
(d) $\mathbf{A} - \mathbf{B}$
(e) det $\mathbf{A}$, det $\mathbf{B}$
(f) Adj $\mathbf{A}$, Adj $\mathbf{B}$
(g) $\mathbf{A}^{-1}$, $\mathbf{B}^{-1}$ (verify the answer)

A.2 Solve $\mathbf{Ax} = \mathbf{b}$ for $\mathbf{x}$, where

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \\ 1 & 0 & 1 \end{bmatrix} \qquad \mathbf{b} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

Use
(a) Gaussian elimination and demonstrate that $\mathbf{A}$ is nonsingular. Check to see that the determinant does not change after each row operation.
(b) Use $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.
(c) Use Cramer's rule.

A.3 Suppose

$$\begin{aligned} z_1 &= 3x_1 + x_3 \\ z_2 &= x_1 + x_2 + x_3 \qquad \mathbf{z} = \mathbf{Ax} \\ z_3 &= 2x_2 + x_3 \end{aligned}$$

Find equations for x_1, x_2, and x_3 in terms of z_1, z_2, z_3. Use an algebraic method first; check the result using $\mathbf{A}^{-1}$.

A.4 For

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} \qquad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} \qquad \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

find the magnitude (norm) of each vector.
What is $\mathbf{x}_1^T\mathbf{x}_2$? $\mathbf{x}_1\mathbf{x}_2^T$? $\mathbf{x}_1^T\mathbf{A}\mathbf{x}_1$?
Find a vector $\mathbf{x}_3$ that is orthogonal to $\mathbf{x}_1(\mathbf{x}_1^T\mathbf{x}_3 = 0)$. Are $\mathbf{x}_1$, $\mathbf{x}_2$, and $\mathbf{x}_3$ linearly independent?

A.5 For

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -3 & 1 \end{bmatrix}$$

calculate det **A** using expansion by minors of the second row. Repeat with the third column.

A.6 Calculate the eigenvalues and eigenvectors of $\begin{bmatrix} 0 & 1 \\ 1 & 4 \end{bmatrix}$. Repeat for

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

A.7 Show that for a 2×2 symmetrical matrix, the eigenvalues must be real (do not contain imaginary components). Develop a 2×2 nonsymmetrical matrix which has complex eigenvalues.

A.8 A technique called LU decomposition can be used to solve sets of linear algebraic equations. **L** and **U** are lower and upper triangular matrices, respectively. A lower triangular matrix has zeros above the main diagonal; an upper triangular matrix has zeros below the main diagonal. Any matrix **A** can be formed by the product of **LU**.
(a) For

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 \\ 2 & 3 & 1 \\ 1 & 0 & 1 \end{bmatrix}$$

find some **L** and **U** that satisfy $\mathbf{LU} = \mathbf{A}$.
(b) If $\mathbf{Ax} = \mathbf{b}$, $\mathbf{LUx} = \mathbf{b}$ or $\mathbf{Ux} = \mathbf{L}^{-1}\mathbf{b} = \hat{\mathbf{b}}$.
Let

$$\mathbf{b} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$$

Calculate $\mathbf{L}^{-1}$ and $\hat{\mathbf{b}}$. Then solve for **x** using substitution from the upper triangular matrix **U**.

A.9 You are to solve the two nonlinear equations,

$$x_1^2 + x_2^2 = 8$$

$$x_1 x_2 = 4$$

using the Newton–Raphson method. Suggested starting points are (0, 1) and (4, 4).