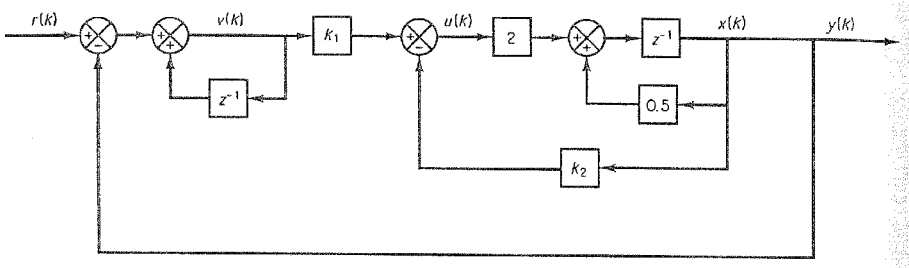


Figure 8–11 A servo system.

where

$$\mathbf{x}(k) = \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} = \begin{bmatrix} x(k) - x(\infty) \\ v(k) - v(\infty) \end{bmatrix}, \qquad w(k) = u(k) - u(\infty)$$

$$\mathbf{G} = \begin{bmatrix} 0.5 & 0 \\ -0.5 & 1 \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} 2 \\ -2 \end{bmatrix}, \qquad \mathbf{K} = [k_2 \quad -k_1]$$

Problem B–8–10

Referring to Problem B–8–9, it is desired to design the state feedback gain matrix $\mathbf{K} = [k_2 \quad -k_1]$ such that the system has reasonable step-response characteristics. Let us assume that we use the quadratic optimal control scheme.

Let us assume the following performance index:

$$J = \frac{1}{2}\sum_{k=0}^{\infty} [\mathbf{x}(k)^*\mathbf{Q}\mathbf{x}(k) + w(k)^*Rw(k)]$$

If $\mathbf{Q}$ and R are chosen to be positive definite, the resulting system is stable. For this problem, we choose

$$\mathbf{Q} = \begin{bmatrix} 100 & 0 \\ 0 & 1 \end{bmatrix}, \qquad R = 1$$

Note that the present $\mathbf{Q}$ and R are only one possible set. (Other positive definite $\mathbf{Q}$ and R may be chosen. The resulting system is stable but different for each different set of $\mathbf{Q}$ and R.)

Using the state-space representation shown in Problem B–8–9, determine matrix $\mathbf{K}$ with MATLAB. Write a MATLAB program. Using matrix $\mathbf{K}$ thus determined, obtain the unit-step response of the designed system with MATLAB. Plot $y(k)$ versus k and $v(k)$ versus k.

Appendix A

Vector-Matrix Analysis

A–1 DEFINITIONS

Matrices that we frequently encounter in the study of modern control theory are the symmetric matrix, skew-symmetric matrix, orthogonal matrix, Hermitian matrix, skew-Hermitian matrix, unitary matrix, and normal matrix. The following equations define these matrices:

$\mathbf{A}^T = \mathbf{A}$	A is symmetric
$\mathbf{A}^T = -\mathbf{A}$	A is skew-symmetric
$\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A} = \mathbf{I}$	A is orthogonal
$\mathbf{A}^* = \mathbf{A}$	A is Hermitian
$\mathbf{A}^* = -\mathbf{A}$	A is skew-Hermitian
$\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A} = \mathbf{I}$	A is unitary
$\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$ or $\mathbf{A}\mathbf{A}^T = \mathbf{A}^T\mathbf{A}$	A is normal

where the superscript * denotes the conjugate transpose and superscript T signifies the transpose.

A–2 DETERMINANTS

Determinants of a 2 × 2 Matrix, a 3 × 3 Matrix, and a 4 × 4 Matrix. For a 2×2 matrix $\mathbf{A}$, we have

$$|\mathbf{A}| = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} = a_1 b_2 - b_1 a_2$$

For a 3×3 matrix $\mathbf{A}$,

$$|\mathbf{A}| = \begin{vmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{vmatrix} = a_1 b_2 c_3 + b_1 c_2 a_3 + c_1 a_2 b_3 - c_1 b_2 a_3 - b_1 a_2 c_3 - a_1 b_3 c_2$$

For a 4×4 matrix $\mathbf{A}$,

$$\begin{aligned} |\mathbf{A}| &= \begin{vmatrix} a_1 & a_2 & a_3 & a_4 \\ b_1 & b_2 & b_3 & b_4 \\ c_1 & c_2 & c_3 & c_4 \\ d_1 & d_2 & d_3 & d_4 \end{vmatrix} \\ &= \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix} \begin{vmatrix} c_3 & c_4 \\ d_3 & d_4 \end{vmatrix} - \begin{vmatrix} a_1 & a_2 \\ c_1 & c_2 \end{vmatrix} \begin{vmatrix} b_3 & b_4 \\ d_3 & d_4 \end{vmatrix} \\ &\quad + \begin{vmatrix} a_1 & a_2 \\ d_1 & d_2 \end{vmatrix} \begin{vmatrix} b_3 & b_4 \\ c_3 & c_4 \end{vmatrix} + \begin{vmatrix} b_1 & b_2 \\ c_1 & c_2 \end{vmatrix} \begin{vmatrix} a_3 & a_4 \\ d_3 & d_4 \end{vmatrix} \\ &\quad - \begin{vmatrix} b_1 & b_2 \\ d_1 & d_2 \end{vmatrix} \begin{vmatrix} a_3 & a_4 \\ c_3 & c_4 \end{vmatrix} + \begin{vmatrix} c_1 & c_2 \\ d_1 & d_2 \end{vmatrix} \begin{vmatrix} a_3 & a_4 \\ b_3 & b_4 \end{vmatrix} \end{aligned} \tag{A–1}$$

(This expansion is called Laplace's expansion by the minors.)

Properties of the Determinant. The determinant of an $n \times n$ matrix has the following properties:

1. If two rows (or two columns) of the determinant are interchanged, only the sign of the determinant is changed.
2. The determinant is invariant under the addition of a scalar multiple of a row (or a column) to another row (or column).
3. If an $n \times n$ matrix has two identical rows (or columns), then the determinant is zero.
4. For an $n \times n$ matrix $\mathbf{A}$,
$$|\mathbf{A}^T| = |\mathbf{A}|, \qquad |\mathbf{A}^*| = |\overline{\mathbf{A}}|$$
5. The determinant of a product of two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ is the product of their determinants:
$$|\mathbf{AB}| = |\mathbf{A}|\,|\mathbf{B}| = |\mathbf{BA}|$$
6. If a row (or a column) is multiplied by a scalar k, then the determinant is multiplied by k.
7. If all elements of an $n \times n$ matrix are multiplied by k, then the determinant is multiplied by k^n; that is,
$$|k\mathbf{A}| = k^n |\mathbf{A}|$$
8. If the eigenvalues of $\mathbf{A}$ are λ_i $(i = 1, 2, \ldots, n)$, then
$$|\mathbf{A}| = \lambda_1 \lambda_2 \ldots \lambda_n$$
Hence, $|\mathbf{A}| \neq 0$ implies $\lambda_i \neq 0$ for $i = 1, 2, \ldots, n$. (For details of the eigenvalue, see Section A–6.)

9. If matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, respectively, then
$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{D}|, \qquad \text{if } |\mathbf{A}| \neq 0 \text{ and } |\mathbf{D}| \neq 0 \tag{A–2}$$
$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{vmatrix} = \begin{vmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = 0, \qquad \text{if } |\mathbf{A}| = 0 \text{ or } |\mathbf{D}| = 0 \text{ or } |\mathbf{A}| = |\mathbf{D}| = 0$$
Also,
$$\begin{vmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{vmatrix} = \begin{cases} |\mathbf{A}|\,|\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}|, & \text{if } |\mathbf{A}| \neq 0 \quad \text{(A–3)} \\ |\mathbf{D}|\,|\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}|, & \text{if } |\mathbf{D}| \neq 0 \quad \text{(A–4)} \end{cases}$$
[For the derivation of Equation (A–2), see Problem A–1. For derivations of Equations (A–3) and (A–4), refer to Problem A–2.]
10. For an $n \times m$ matrix $\mathbf{A}$ and an $m \times n$ matrix $\mathbf{B}$,
$$|\mathbf{I}_n + \mathbf{AB}| = |\mathbf{I}_m + \mathbf{BA}| \tag{A–5}$$
(For the proof, see Problem A–3.) In particular, for $m = 1$, that is, for an $n \times 1$ matrix $\mathbf{A}$ and a $1 \times n$ matrix $\mathbf{B}$, we have
$$|\mathbf{I}_n + \mathbf{AB}| = 1 + \mathbf{BA} \tag{A–6}$$
Equations (A–2) through (A–6) are useful in computing the determinants of matrices of large order.

A–3 INVERSION OF MATRICES

Nonsingular Matrix and Singular Matrix. A square matrix $\mathbf{A}$ is called a nonsingular matrix if a matrix $\mathbf{B}$ exists such that $\mathbf{BA} = \mathbf{AB} = \mathbf{I}$. If such a matrix $\mathbf{B}$ exists, then it is denoted by $\mathbf{A}^{-1}$. $\mathbf{A}^{-1}$ is called the *inverse* of $\mathbf{A}$. The inverse matrix $\mathbf{A}^{-1}$ exists if $|\mathbf{A}|$ is nonzero. If $\mathbf{A}^{-1}$ does not exist, $\mathbf{A}$ is said to be *singular*.

If $\mathbf{A}$ and $\mathbf{B}$ are nonsingular matrices, then the product $\mathbf{AB}$ is a nonsingular matrix and
$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$$
Also,
$$(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$$
and
$$(\mathbf{A}^*)^{-1} = (\mathbf{A}^{-1})^*$$

Properties of the Inverse Matrix. The inverse of a matrix has the following properties.

1. If k is a nonzero scalar and $\mathbf{A}$ is an $n \times n$ nonsingular matrix, then
$$(k\mathbf{A})^{-1} = \frac{1}{k}\mathbf{A}^{-1}$$

2. The determinant of $\mathbf{A}^{-1}$ is the inverse of the determinant of $\mathbf{A}$, or

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

This can be verified easily as follows:

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

Useful Formulas for Finding the Inverse of a Matrix

1. For a 2×2 matrix $\mathbf{A}$, where

$$\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad ad - bc \neq 0$$

the inverse matrix is given by

$$\mathbf{A}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

2. For a 3×3 matrix $\mathbf{A}$, where

$$\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}, \qquad |\mathbf{A}| \neq 0$$

the inverse matrix is given by

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\begin{bmatrix} \begin{vmatrix} e & f \\ h & i \end{vmatrix} & -\begin{vmatrix} b & c \\ h & i \end{vmatrix} & \begin{vmatrix} b & c \\ e & f \end{vmatrix} \\ -\begin{vmatrix} d & f \\ g & i \end{vmatrix} & \begin{vmatrix} a & c \\ g & i \end{vmatrix} & -\begin{vmatrix} a & c \\ d & f \end{vmatrix} \\ \begin{vmatrix} d & e \\ g & h \end{vmatrix} & -\begin{vmatrix} a & b \\ g & h \end{vmatrix} & \begin{vmatrix} a & b \\ d & e \end{vmatrix} \end{bmatrix}$$

3. If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are, respectively, an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, then

$$(\mathbf{A} + \mathbf{BDC})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{D}^{-1} + \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} \tag{A–7}$$

provided the indicated inverses exist. Equation (A–7) is commonly referred to as the *matrix inversion lemma*. (For the proof, see Problem A–4.)

If $\mathbf{D} = \mathbf{I}_m$, then Equation (A–7) simplifies to

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

In this last equation, if $\mathbf{B}$ and $\mathbf{C}$ are an $n \times 1$ matrix and a $1 \times n$ matrix, respectively, then

$$(\mathbf{A} + \mathbf{BC})^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{B}\mathbf{C}\mathbf{A}^{-1}}{1 + \mathbf{C}\mathbf{A}^{-1}\mathbf{B}} \tag{A–8}$$

Equation (A–8) is useful in that if an $n \times n$ matrix $\mathbf{X}$ can be written as $\mathbf{A} + \mathbf{BC}$, where $\mathbf{A}$ is an $n \times n$ matrix whose inverse is known and $\mathbf{BC}$ is a product of a column vector and a row vector, then $\mathbf{X}^{-1}$ can be obtained easily in terms of the known $\mathbf{A}^{-1}$, $\mathbf{B}$, and $\mathbf{C}$.

4. If $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are, respectively, an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, then

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix} \tag{A–9}$$

provided $|\mathbf{A}| \neq 0$ and $|\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B}| \neq 0$, or

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} + \mathbf{D}^{-1} \end{bmatrix} \tag{A–10}$$

provided $|\mathbf{D}| \neq 0$ and $|\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C}| \neq 0$. In particular, if $\mathbf{C} = \mathbf{0}$ or $\mathbf{B} = \mathbf{0}$, then Equations (A–9) and (A–10) can be simplified as follows:

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{D}^{-1} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \tag{A–11}$$

or

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C}\mathbf{A}^{-1} & \mathbf{D}^{-1} \end{bmatrix} \tag{A–12}$$

[For the derivation of Equations (A–9) through (A–12), refer to Problems A–5 and A–6.]

A–4 RULES OF MATRIX OPERATIONS

In this section we shall review some of the rules of algebraic operations with matrices and then give definitions of the derivative and the integral of matrices. Then the rules of differentiation of matrices will be presented.

Note that matrix algebra differs from ordinary number algebra in that matrix multiplication is not commutative and cancellation of matrices is not valid.

Multiplication of a Matrix by a Scalar. The product of a matrix and a scalar is a matrix in which each element is multiplied by the scalar. That is,

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

Multiplication of a Matrix by a Matrix. Multiplication of a matrix by a matrix is possible between matrices in which the number of columns in the first matrix is equal to the number of rows in the second. Otherwise, multiplication is not defined.

Consider the product of an $n \times m$ matrix $\mathbf{A}$ and an $m \times r$ matrix $\mathbf{B}$:

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1r} \\ b_{21} & b_{22} & \cdots & b_{2r} \\ \vdots & \vdots & & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mr} \end{bmatrix}$$

$$= \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1r} \\ c_{21} & c_{22} & \cdots & c_{2r} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nr} \end{bmatrix}$$

where

$$c_{ik} = \sum_{j=1}^{m} a_{ij} b_{jk}$$

Thus, multiplication of an $n \times m$ matrix by an $m \times r$ matrix yields an $n \times r$ matrix. It should be noted that, in general, matrix multiplication is not commutative; that is

$$\mathbf{AB} \neq \mathbf{BA} \qquad \text{in general}$$

For example,

$$\mathbf{AB} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} & a_{11}b_{12} + a_{12}b_{22} \\ a_{21}b_{11} + a_{22}b_{21} & a_{21}b_{12} + a_{22}b_{22} \end{bmatrix}$$

and

$$\mathbf{BA} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} b_{11}a_{11} + b_{12}a_{21} & b_{11}a_{12} + b_{12}a_{22} \\ b_{21}a_{11} + b_{22}a_{21} & b_{21}a_{12} + b_{22}a_{22} \end{bmatrix}$$

Thus, in general, $\mathbf{AB} \neq \mathbf{BA}$. Hence, the order of multiplication is significant and must be preserved. If $\mathbf{AB} = \mathbf{BA}$, matrices $\mathbf{A}$ and $\mathbf{B}$ are said to commute. In the preceding matrices $\mathbf{A}$ and $\mathbf{B}$, if, for example, $a_{12} = a_{21} = b_{12} = b_{21} = 0$, then $\mathbf{A}$ and $\mathbf{B}$ commute.

For $n \times n$ diagonal matrices $\mathbf{A}$ and $\mathbf{B}$,

$$\mathbf{AB} = [a_{ij}\delta_{ij}][b_{ij}\delta_{ij}] = \begin{bmatrix} a_{11}b_{11} & & & 0 \\ & a_{22}b_{22} & & \\ & & \ddots & \\ 0 & & & a_{nn}b_{nn} \end{bmatrix}$$

If $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are an $n \times m$ matrix, an $m \times r$ matrix, and an $r \times p$ matrix, respectively, then the following associativity law holds true:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC})$$

This may be proved as follows:

$$(i,k)\text{th element of } \mathbf{AB} = \sum_{j=1}^{m} a_{ij} b_{jk}$$

$$(j,h)\text{th element of } \mathbf{BC} = \sum_{k=1}^{r} b_{jk} c_{kh}$$

$$(i,h)\text{th element of } (\mathbf{AB})\mathbf{C} = \sum_{k=1}^{r} \left(\sum_{j=1}^{m} a_{ij} b_{jk} \right) c_{kh} = \sum_{j=1}^{m} \sum_{k=1}^{r} (a_{ij} b_{jk}) c_{kh}$$

$$= \sum_{j=1}^{m} \sum_{k=1}^{r} a_{ij} (b_{jk} c_{kh}) = \sum_{j=1}^{m} a_{ij} \left[\sum_{k=1}^{r} b_{jk} c_{kh} \right]$$

$$= (i,h)\text{th element of } \mathbf{A}(\mathbf{BC})$$

Since the associativity of multiplication of matrices holds true, we have

$$\mathbf{ABCD} = (\mathbf{AB})(\mathbf{CD}) = \mathbf{A}(\mathbf{BCD}) = (\mathbf{ABC})\mathbf{D}$$

$$\mathbf{A}^{m+n} = \mathbf{A}^m \mathbf{A}^n, \qquad m, n = 1, 2, 3, \ldots$$

If $\mathbf{A}$ and $\mathbf{B}$ are $n \times m$ matrices and $\mathbf{C}$ and $\mathbf{D}$ are $m \times r$ matrices, then the following distributivity law holds true:

$$(\mathbf{A} + \mathbf{B})(\mathbf{C} + \mathbf{D}) = \mathbf{AC} + \mathbf{AD} + \mathbf{BC} + \mathbf{BD}$$

This can be proved by comparing the (i,j)th element of $(\mathbf{A} + \mathbf{B})(\mathbf{C} + \mathbf{D})$ and the (i,j)th element of $(\mathbf{AC} + \mathbf{AD} + \mathbf{BC} + \mathbf{BD})$.

Remarks on Cancellation of Matrices. Cancellation of matrices is not valid in matrix algebra. Consider the product of two singular matrices $\mathbf{A}$ and $\mathbf{B}$. Take, for example,

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix} \neq \mathbf{0}, \qquad \mathbf{B} = \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} \neq \mathbf{0}$$

Then

$$\mathbf{AB} = \begin{bmatrix} 2 & 1 \\ 6 & 3 \end{bmatrix} \begin{bmatrix} 1 & -2 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \mathbf{0}$$

Clearly, $\mathbf{AB} = \mathbf{0}$ implies neither $\mathbf{A} = \mathbf{0}$ nor $\mathbf{B} = \mathbf{0}$. In fact, $\mathbf{AB} = \mathbf{0}$ implies one of the following three:

1. $\mathbf{A} = \mathbf{0}$.
2. $\mathbf{B} = \mathbf{0}$.
3. Both $\mathbf{A}$ and $\mathbf{B}$ are singular.

It can easily be proved that, if both $\mathbf{A}$ and $\mathbf{B}$ are nonzero matrices and $\mathbf{AB} = \mathbf{0}$, then both $\mathbf{A}$ and $\mathbf{B}$ must be singular. Assume that $\mathbf{B}$ is nonzero and $\mathbf{A}$ is not singular. Then $|\mathbf{A}| \neq 0$ and $\mathbf{A}^{-1}$ exists. Then we obtain

$$\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B} = \mathbf{0}$$

which contradicts the assumption that $\mathbf{B}$ is nonzero. In this way we can prove that both $\mathbf{A}$ and $\mathbf{B}$ must be singular if $\mathbf{A} \neq \mathbf{0}$ and $\mathbf{B} \neq \mathbf{0}$.

Similarly, notice that if $\mathbf{A}$ is singular then neither $\mathbf{AB} = \mathbf{AC}$ nor $\mathbf{BA} = \mathbf{CA}$ implies $\mathbf{B} = \mathbf{C}$. If, however, $\mathbf{A}$ is a nonsingular matrix, then $\mathbf{AB} = \mathbf{AC}$ implies $\mathbf{B} = \mathbf{C}$ and $\mathbf{BA} = \mathbf{CA}$ also implies $\mathbf{B} = \mathbf{C}$.

Derivative and Integral of a Matrix. The derivative of an $n \times m$ matrix $\mathbf{A}(t)$ is defined by the matrix whose (i,j)th element is the derivative of the (i,j)th element of the original matrix, provided that all the elements $a_{ij}(t)$ have derivatives with respect to t:

$$\frac{d}{dt}\mathbf{A}(t) = \begin{bmatrix} \frac{d}{dt}a_{11}(t) & \cdots & \frac{d}{dt}a_{1m}(t) \\ \vdots & & \vdots \\ \frac{d}{dt}a_{n1}(t) & \cdots & \frac{d}{dt}a_{nm}(t) \end{bmatrix}$$

In the case of an n-dimensional vector $\mathbf{x}(t)$,

$$\frac{d}{dt}\mathbf{x}(t) = \begin{bmatrix} \frac{d}{dt}x_1(t) \\ \vdots \\ \frac{d}{dt}x_n(t) \end{bmatrix}$$

Similarly, the integral of an $n \times m$ matrix $\mathbf{A}(t)$ with respect to t is defined by the matrix whose (i,j)th element is the integral of the (i,j)th element of the original matrix, or

$$\int \mathbf{A}(t)\,dt = \begin{bmatrix} \int a_{11}(t)\,dt & \cdots & \int a_{1m}(t)\,dt \\ \vdots & & \vdots \\ \int a_{n1}(t)\,dt & \cdots & \int a_{nm}(t)\,dt \end{bmatrix}$$

provided that the $a_{ij}(t)$'s are integrable as functions of t.

Differentiation of a Matrix. If the elements of matrices $\mathbf{A}$ and $\mathbf{B}$ are functions of t, then

$$\frac{d}{dt}(\mathbf{A} + \mathbf{B}) = \frac{d}{dt}\mathbf{A} + \frac{d}{dt}\mathbf{B} \tag{A–13}$$

$$\frac{d}{dt}(\mathbf{AB}) = \frac{d\mathbf{A}}{dt}\mathbf{B} + \mathbf{A}\frac{d\mathbf{B}}{dt} \tag{A–14}$$

If $k(t)$ is a scalar and is a function of t, then

$$\frac{d}{dt}[\mathbf{A}k(t)] = \frac{d\mathbf{A}}{dt}k(t) + \mathbf{A}\frac{dk(t)}{dt} \tag{A–15}$$

Also,

$$\int_a^b \frac{d\mathbf{A}}{dt}\mathbf{B}\,dt = \mathbf{AB}\Big|_a^b - \int_a^b \mathbf{A}\frac{d\mathbf{B}}{dt}\,dt \tag{A–16}$$

It is important to note that the derivative of $\mathbf{A}^{-1}$ is given by

$$\frac{d}{dt}\mathbf{A}^{-1} = -\mathbf{A}^{-1}\frac{d\mathbf{A}}{dt}\mathbf{A}^{-1} \tag{A–17}$$

Equation (A–17) can be derived easily by differentiating $\mathbf{AA}^{-1}$ with respect to t. Since

$$\frac{d}{dt}\mathbf{AA}^{-1} = \frac{d\mathbf{A}}{dt}\mathbf{A}^{-1} + \mathbf{A}\frac{d\mathbf{A}^{-1}}{dt}$$

and also

$$\frac{d}{dt}\mathbf{AA}^{-1} = \frac{d}{dt}\mathbf{I} = \mathbf{0}$$

we obtain

$$\mathbf{A}\frac{d\mathbf{A}^{-1}}{dt} = -\frac{d\mathbf{A}}{dt}\mathbf{A}^{-1}$$

or

$$\mathbf{A}^{-1}\mathbf{A}\frac{d\mathbf{A}^{-1}}{dt} = \frac{d\mathbf{A}^{-1}}{dt} = -\mathbf{A}^{-1}\frac{d\mathbf{A}}{dt}\mathbf{A}^{-1}$$

which is the desired result.

Derivatives of a Scalar Function with Respect to a Vector. If $J(\mathbf{x})$ is a scalar function of a vector $\mathbf{x}$, then

$$\frac{\partial J}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial J}{\partial x_1} \\ \vdots \\ \frac{\partial J}{\partial x_n} \end{bmatrix}, \qquad \frac{\partial^2 J}{\partial \mathbf{x}^2} = \begin{bmatrix} \frac{\partial^2 J}{\partial^2 x_1} & \frac{\partial^2 J}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_1 \partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 J}{\partial x_n \partial x_1} & \frac{\partial^2 J}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 J}{\partial x_n^2} \end{bmatrix}$$

Also, for a scalar function $V(\mathbf{x}(t))$, we have

$$\frac{d}{dt}V(\mathbf{x}(t)) = \left(\frac{\partial V}{\partial \mathbf{x}}\right)^T \frac{d\mathbf{x}}{dt}$$

Jacobian. If an $m \times 1$ matrix $\mathbf{f}(\mathbf{x})$ is a vector function of an n-vector $\mathbf{x}$ (*note:* an n-vector is meant as an n-dimensional vector), then

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{A–18}$$

Such an $n \times m$ matrix is called a *Jacobian*.

Notice that, by using this definition of the Jacobian, we have

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{Ax} = \mathbf{A}^T \tag{A–19}$$

The fact that Equation (A–19) holds true can be easily seen from the following example. If A and x are given by

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

then

$$\mathbf{Ax} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}$$

and

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{Ax} = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_2}{\partial x_1} \\ \dfrac{\partial f_1}{\partial x_2} & \dfrac{\partial f_2}{\partial x_2} \\ \dfrac{\partial f_1}{\partial x_3} & \dfrac{\partial f_2}{\partial x_3} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{bmatrix} = \mathbf{A}^T$$

Also, we have the following useful formula. For an $n \times n$ real matrix A and a real n-vector x,

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{Ax} = \mathbf{Ax} + \mathbf{A}^T\mathbf{x} \tag{A–20}$$

In addition, if matrix A is a real symmetric matrix, then

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{Ax} = 2\mathbf{Ax}$$

Note that if A is an $n \times n$ Hermitian matrix and x is a complex n-vector then

$$\frac{\partial}{\partial \bar{\mathbf{x}}}\mathbf{x}^*\mathbf{Ax} = \mathbf{Ax} \tag{A–21}$$

[For derivations of Equations (A–20) and (A–21), see Problem A–7.]

For an $n \times m$ real matrix A, a real n-vector x, and a real m-vector y, we have

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{Ay} = \mathbf{Ay} \tag{A–22}$$

$$\frac{\partial}{\partial \mathbf{y}}\mathbf{x}^T\mathbf{Ay} = \mathbf{A}^T\mathbf{x} \tag{A–23}$$

Similarly, for an $n \times m$ complex matrix A, a complex n-vector x, and a complex m-vector y, we have

$$\frac{\partial}{\partial \bar{\mathbf{x}}}\mathbf{x}^*\mathbf{Ay} = \mathbf{Ay} \tag{A–24}$$

$$\frac{\partial}{\partial \mathbf{y}}\mathbf{x}^*\mathbf{Ay} = \mathbf{A}^T\bar{\mathbf{x}} \tag{A–25}$$

[For derivations of Equations (A–22) through (A–25), refer to Problem A–8.] Note that Equation (A–25) is equivalent to the following equation:

$$\overline{\frac{\partial}{\partial \mathbf{y}}\mathbf{x}^*\mathbf{Ay}} = \mathbf{A}^*\mathbf{x}$$

A–5 VECTORS AND VECTOR ANALYSIS

Linear Dependence and Independence of Vectors. Vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are said to be *linearly independent* if the equation

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \cdots + c_n\mathbf{x}_n = \mathbf{0}$$

where $c_1, c_2, \ldots, c_n$ are constants, implies that $c_1 = c_2 = \cdots = c_n = 0$. Conversely, vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are said to be *linearly dependent* if and only if $\mathbf{x}_i$ can be expressed as a linear combination of $\mathbf{x}_j$ $(j = 1, 2, \ldots, n; j \neq i)$.

It is important to note that if vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent and vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n, \mathbf{x}_{n+1}$ are linearly dependent, then $\mathbf{x}_{n+1}$ can be expressed as a unique linear combination of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$.

Necessary and Sufficient Conditions for Linear Independence of Vectors. It can be proved that the necessary and sufficient conditions for n-vectors $\mathbf{x}_i$ $(i = 1, 2, \ldots, m)$ to be linearly independent are that

1. $m \leq n$.
2. There exists at least one nonzero m-column determinant of the $n \times m$ matrix whose columns consist of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m$.

Hence, for n vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ the necessary and sufficient condition for linear independence is

$$|\mathbf{A}| \neq 0$$

where A is the $n \times n$ matrix whose ith column is made up of the components of $\mathbf{x}_i$ $(i = 1, 2, \ldots, n)$.

Inner Product. Any rule that assigns to each pair of vectors x and y in a vector space a scalar quantity is called an *inner product* or *scalar product* and is given the symbol $\langle \mathbf{x}, \mathbf{y} \rangle$, provided that the following four axioms are satisfied:

1. $$\langle \mathbf{y}, \mathbf{x} \rangle = \overline{\langle \mathbf{x}, \mathbf{y} \rangle}$$

   where the bar denotes the conjugate of a complex number

2. $$\langle c\mathbf{x}, \mathbf{y} \rangle = \bar{c}\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, \bar{c}\mathbf{y} \rangle$$

   where c is a complex number

3. $$\langle \mathbf{x} + \mathbf{y}, \mathbf{z} + \mathbf{w} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle + \langle \mathbf{x}, \mathbf{w} \rangle + \langle \mathbf{y}, \mathbf{w} \rangle$$

4. $$\langle \mathbf{x}, \mathbf{x} \rangle > 0, \qquad \text{for } \mathbf{x} \neq \mathbf{0}$$

In any finite-dimensional vector space, there are many different definitions of the inner product, all satisfying the four axioms.

In this book, unless the contrary is stated, we shall adopt the following definition of the inner product: The inner product of a pair of n-vectors $\mathbf{x}$ and $\mathbf{y}$ in a vector space V is given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \cdots + \bar{x}_n y_n = \sum_{i=1}^{n} \bar{x}_i y_i \qquad \text{(A–26)}$$

where the summation is a complex number and where the $\bar{x}_i$'s are the complex conjugates of the x_i's. This definition clearly satisfies the four axioms. The inner product can then be expressed as follows:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^* \mathbf{y}$$

where $\mathbf{x}^*$ denotes the conjugate transpose of $\mathbf{x}$. Also,

$$\langle \mathbf{x}, \mathbf{y} \rangle = \overline{\langle \mathbf{y}, \mathbf{x} \rangle} = \overline{\mathbf{y}^* \mathbf{x}} = \mathbf{y}^T \bar{\mathbf{x}} = \mathbf{x}^* \mathbf{y} \qquad \text{(A–27)}$$

The inner product of two n-vectors $\mathbf{x}$ and $\mathbf{y}$ with real components is therefore given by

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{i=1}^{n} x_i y_i \qquad \text{(A–28)}$$

In this case, clearly we have

$$\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}, \qquad \text{for real vectors } \mathbf{x} \text{ and } \mathbf{y}$$

It is noted that the real or complex vector $\mathbf{x}$ is said to be *normalized* if $\langle \mathbf{x}, \mathbf{x} \rangle = 1$. It is also noted that, for an n-vector $\mathbf{x}$, $\mathbf{x}^*\mathbf{x}$ is a nonnegative scalar, but $\mathbf{x}\mathbf{x}^*$ is an $n \times n$ matrix. That is,

$$\begin{aligned} \mathbf{x}^*\mathbf{x} = \langle \mathbf{x}, \mathbf{x} \rangle &= \bar{x}_1 x_1 + \bar{x}_2 x_2 + \cdots + \bar{x}_n x_n \\ &= |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2 \end{aligned}$$

and

$$\mathbf{x}\mathbf{x}^* = \begin{bmatrix} x_1\bar{x}_1 & x_1\bar{x}_2 & \cdots & x_1\bar{x}_n \\ x_2\bar{x}_1 & x_2\bar{x}_2 & \cdots & x_2\bar{x}_n \\ \vdots & \vdots & & \vdots \\ x_n\bar{x}_1 & x_n\bar{x}_2 & \cdots & x_n\bar{x}_n \end{bmatrix}$$

Notice that, for an $n \times n$ complex matrix $\mathbf{A}$ and complex n-vectors $\mathbf{x}$ and $\mathbf{y}$, the inner product of $\mathbf{x}$ and $\mathbf{A}\mathbf{y}$ and that of $\mathbf{A}^*\mathbf{x}$ and $\mathbf{y}$ are the same, or

$$\langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle = \mathbf{x}^*\mathbf{A}\mathbf{y}, \qquad \langle \mathbf{A}^*\mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^*\mathbf{A}\mathbf{y}$$

Similarly, for an $n \times n$ real matrix $\mathbf{A}$ and real n-vectors $\mathbf{x}$ and $\mathbf{y}$, the inner product of $\mathbf{x}$ and $\mathbf{A}\mathbf{y}$ and that of $\mathbf{A}^T\mathbf{x}$ and $\mathbf{y}$ are the same, or

$$\langle \mathbf{x}, \mathbf{A}\mathbf{y} \rangle = \mathbf{x}^T\mathbf{A}\mathbf{y}, \qquad \langle \mathbf{A}^T\mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T\mathbf{A}\mathbf{y}$$

Unitary Transformation. If $\mathbf{A}$ is a unitary matrix (that is, if $\mathbf{A}^{-1} = \mathbf{A}^*$), then the inner product $\langle \mathbf{x}, \mathbf{x} \rangle$ is invariant under the linear transformation $\mathbf{x} = \mathbf{A}\mathbf{y}$, because

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

Such a transformation $\mathbf{x} = \mathbf{A}\mathbf{y}$, where $\mathbf{A}$ is a unitary matrix, which transforms $\sum_{i=1}^{n} \bar{x}_i x_i$ into $\sum_{i=1}^{n} \bar{y}_i y_i$, is called a *unitary transformation*.

Orthogonal Transformation. If $\mathbf{A}$ is an orthogonal matrix (that is, if $\mathbf{A}^{-1} = \mathbf{A}^T$), then the inner product $\langle \mathbf{x}, \mathbf{x} \rangle$ is invariant under the linear transformation $\mathbf{x} = \mathbf{A}\mathbf{y}$, because

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

Such a transformation $\mathbf{x} = \mathbf{A}\mathbf{y}$, which transforms $\sum_{i=1}^{n} x_i^2$ into $\sum_{i=1}^{n} y_i^2$, is called an *orthogonal transformation*.

Norms of a Vector. Once we define the inner product, we can use this inner product to define norms of a vector $\mathbf{x}$. The concept of a norm is somewhat similar to that of the absolute value. A norm is a function that assigns to every vector $\mathbf{x}$ in a given vector space a real number denoted by $\|\mathbf{x}\|$ such that

1. $\|\mathbf{x}\| > 0$, for $\mathbf{x} \neq \mathbf{0}$
2. $\|\mathbf{x}\| = 0$, if and only if $\mathbf{x} = \mathbf{0}$
3. $\|k\mathbf{x}\| = |k| \, \|\mathbf{x}\|$, where k is a scalar and $|k|$ is the absolute value of k
4. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$, for all $\mathbf{x}$ and $\mathbf{y}$
5. $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\|$ (Schwarz inequality)

Several different definitions of norms are commonly used in the literature. However, the following definition is widely used. A norm of a vector is defined as the nonnegative square root of $\langle \mathbf{x}, \mathbf{x} \rangle$:

$$\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2} = (\mathbf{x}^*\mathbf{x})^{1/2} = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} \qquad \text{(A–29)}$$

If $\mathbf{x}$ is a real vector, the quantity $\|\mathbf{x}\|^2$ can be interpreted geometrically as the square of the distance from the origin to the point represented by the vector $\mathbf{x}$. Note that

$$\|\mathbf{x} - \mathbf{y}\| = \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle^{1/2} = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2}$$

The five properties of norms listed earlier may be obvious, except perhaps the last two inequalities. These two inequalities may be proved as follows. From the definitions of the inner product and the norm, we have

$$\begin{aligned} \|\lambda\mathbf{x} + \mathbf{y}\|^2 &= \langle \lambda\mathbf{x} + \mathbf{y}, \lambda\mathbf{x} + \mathbf{y} \rangle = \langle \lambda\mathbf{x}, \lambda\mathbf{x} \rangle + \langle \mathbf{y}, \lambda\mathbf{x} \rangle + \langle \lambda\mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\ &= \bar{\lambda}\lambda\|\mathbf{x}\|^2 + \lambda\langle \mathbf{y}, \mathbf{x} \rangle + \bar{\lambda}\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \\ &= \bar{\lambda}(\lambda\|\mathbf{x}\|^2 + \langle \mathbf{x}, \mathbf{y} \rangle) + \lambda\overline{\langle \mathbf{x}, \mathbf{y} \rangle} + \|\mathbf{y}\|^2 \geq 0 \end{aligned}$$

If we choose

$$\lambda = -\frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\|\mathbf{x}\|^2}, \qquad \text{for } \mathbf{x} \neq \mathbf{0}$$

then

$$\lambda\overline{\langle \mathbf{x}, \mathbf{y} \rangle} + \|\mathbf{y}\|^2 = -\frac{\langle \mathbf{x}, \mathbf{y} \rangle \overline{\langle \mathbf{x}, \mathbf{y} \rangle}}{\|\mathbf{x}\|^2} + \|\mathbf{y}\|^2 \geq 0$$

and

$$\|\mathbf{x}\|^2\|\mathbf{y}\|^2 \geq \langle \mathbf{x}, \mathbf{y} \rangle \overline{\langle \mathbf{x}, \mathbf{y} \rangle} = |\langle \mathbf{x}, \mathbf{y} \rangle|^2, \qquad \text{for } \mathbf{x} \neq \mathbf{0}$$

For $\mathbf{x} = \mathbf{0}$, clearly,

$$\|\mathbf{x}\|^2\|\mathbf{y}\|^2 = |\langle \mathbf{x}, \mathbf{y} \rangle|^2$$

Therefore, we obtain the Schwarz inequality,

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \, \|\mathbf{y}\| \tag{A–30}$$

By use of the Schwarz inequality, we obtain the following inequality:

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{A–31}$$

This can be proved easily, since

$$\begin{aligned} \|\mathbf{x} + \mathbf{y}\|^2 &= \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle \\ &= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 + \langle \mathbf{x}, \mathbf{y} \rangle + \overline{\langle \mathbf{x}, \mathbf{y} \rangle} + \|\mathbf{y}\|^2 \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2 \operatorname{Re} \langle \mathbf{x}, \mathbf{y} \rangle \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2|\langle \mathbf{x}, \mathbf{y} \rangle| \\ &\leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\|\mathbf{x}\| \, \|\mathbf{y}\| \\ &= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \end{aligned}$$

Equations (A–26) through (A–31) are useful in modern control theory.

As stated earlier, different definitions of norms are used in the literature. Three such definitions of norms follow.

1. A norm $\|\mathbf{x}\|$ may be defined as follows:

$$\|\mathbf{x}\| = [(\mathbf{Tx})^*(\mathbf{Tx})]^{1/2} = (\mathbf{x}^*\mathbf{T}^*\mathbf{Tx})^{1/2} = (\mathbf{x}^*\mathbf{Qx})^{1/2}$$
$$= \left[\sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}\bar{x}_i x_j\right]^{1/2} \geq 0$$

   The matrix $\mathbf{Q} = \mathbf{T}^*\mathbf{T}$ is Hermitian, since $\mathbf{Q}^* = \mathbf{T}^*\mathbf{T} = \mathbf{Q}$. The norm $\|\mathbf{x}\| = (\mathbf{x}^*\mathbf{Qx})^{1/2}$ is a generalized form of $(\mathbf{x}^*\mathbf{x})^{1/2}$, which can be written as $(\mathbf{x}^*\mathbf{Ix})^{1/2}$.

2. A norm may be defined as the sum of the magnitudes of all the components x_i:

$$\|\mathbf{x}\| = \sum_{i=1}^{n} |x_i|$$

3. A norm may be defined as the maximum of the magnitudes of all the components x_i:

$$\|\mathbf{x}\| = \max_i \{|x_i|\}$$

   It can be shown that the various norms just defined are equivalent. Among these definitions of norms, norm $(\mathbf{x}^*\mathbf{x})^{1/2}$ is most commonly used in explicit calculations.

Norms of a Matrix. The concept of norms of a vector can be extended to matrices. There are several different definitions of norms of a matrix. Some of them follow.

1. A norm $\|\mathbf{A}\|$ of an $n \times n$ matrix $\mathbf{A}$ may be defined by

$$\|\mathbf{A}\| = \min k$$

   such that

$$\|\mathbf{Ax}\| \leq k\|\mathbf{x}\|$$

   For the norm $(\mathbf{x}^*\mathbf{x})^{1/2}$, this definition is equivalent to

$$\|\mathbf{A}\|^2 = \max_{\mathbf{x}} \{\mathbf{x}^*\mathbf{A}^*\mathbf{Ax}; \mathbf{x}^*\mathbf{x} = 1\}$$

   which means that $\|\mathbf{A}\|^2$ is the maximum of the "absolute value" of the vector $\mathbf{Ax}$ when $\mathbf{x}^*\mathbf{x} = 1$.

2. A norm of an $n \times n$ matrix $\mathbf{A}$ may be defined by

$$\|\mathbf{A}\| = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|$$

   where $|a_{ij}|$ is the absolute value of a_{ij}.

3. A norm may be defined by

$$\|\mathbf{A}\| = \left(\sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2\right)^{1/2}$$

4. Another definition of a norm is given by

$$\|\mathbf{A}\| = \max_i \left(\sum_{j=1}^{n} |a_{ij}|\right)$$

Note that all definitions of norms of an $n \times n$ matrix $\mathbf{A}$ have the following properties:

1. $\|\mathbf{A}\| = \|\mathbf{A}^*\|$ or $\|\mathbf{A}\| = \|\mathbf{A}^T\|$
2. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$
3. $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \, \|\mathbf{B}\|$
4. $\|\mathbf{Ax}\| \leq \|\mathbf{A}\| \, \|\mathbf{x}\|$

Orthogonality of Vectors. If the inner product of two vectors $\mathbf{x}$ and $\mathbf{y}$ is zero, or $\langle \mathbf{x}, \mathbf{y} \rangle = 0$, then vectors $\mathbf{x}$ and $\mathbf{y}$ are said to be *orthogonal to each other*. For example, vectors

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \mathbf{x}_3 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$$

are orthogonal in pairs and thus form an orthogonal set.

In an n-dimensional vector space, vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ defined by

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \qquad \ldots, \qquad \mathbf{x}_n = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

satisfy the conditions $\langle \mathbf{x}_i, \mathbf{x}_j \rangle = \delta_{ij}$, or

$$\langle \mathbf{x}_i, \mathbf{x}_i \rangle = 1$$

$$\langle \mathbf{x}_i, \mathbf{x}_j \rangle = 0, \qquad i \neq j$$

where $i, j = 1, 2, \ldots, n$. Such a set of vectors is said to be *orthonormal*, since the vectors are orthogonal to each other and each vector is normalized.

A nonzero vector $\mathbf{x}$ can be normalized by dividing $\mathbf{x}$ by $\|\mathbf{x}\|$. The normalized vector $\mathbf{x}/\|\mathbf{x}\|$ is a unit vector. Unit vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ form an orthonormal set if they are orthogonal in pairs.

Consider a unitary matrix $\mathbf{A}$. By partitioning $\mathbf{A}$ into column vectors $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$, we have

$$\mathbf{A}^*\mathbf{A} = \begin{bmatrix} \mathbf{A}_1^* \\ \hline \mathbf{A}_2^* \\ \hline \vdots \\ \hline \mathbf{A}_n^* \end{bmatrix} [\mathbf{A}_1 \vdots \mathbf{A}_2 \vdots \cdots \vdots \mathbf{A}_n]$$

$$= \begin{bmatrix} \mathbf{A}_1^*\mathbf{A}_1 & \mathbf{A}_1^*\mathbf{A}_2 & \cdots & \mathbf{A}_1^*\mathbf{A}_n \\ \mathbf{A}_2^*\mathbf{A}_1 & \mathbf{A}_2^*\mathbf{A}_2 & \cdots & \mathbf{A}_2^*\mathbf{A}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{A}_n^*\mathbf{A}_1 & \mathbf{A}_n^*\mathbf{A}_2 & \cdots & \mathbf{A}_n^*\mathbf{A}_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

it follows that

$$\mathbf{A}_i^*\mathbf{A}_i = \langle \mathbf{A}_i, \mathbf{A}_i \rangle = 1$$

$$\mathbf{A}_i^*\mathbf{A}_j = \langle \mathbf{A}_i, \mathbf{A}_j \rangle = 0, \qquad i \neq j$$

Thus, we see that the column vectors (or row vectors) of a unitary matrix $\mathbf{A}$ are orthonormal. The same is true for orthogonal matrices, since they are unitary.

A–6 EIGENVALUES, EIGENVECTORS, AND SIMILARITY TRANSFORMATION

In this section we shall first review important properties of the rank of a matrix and then give definitions of eigenvalues and eigenvectors. Finally, we shall discuss Jordan canonical forms, similarity transformation, and the trace of an $n \times n$ matrix.

Rank of a Matrix. A matrix $\mathbf{A}$ is called of rank m if the maximum number of linearly independent rows (or columns) is m. Hence, if there exists an $m \times m$ submatrix $\mathbf{M}$ of $\mathbf{A}$ such that $|\mathbf{M}| \neq 0$ and the determinant of every $r \times r$ submatrix (where $r \geq m + 1$) of $\mathbf{A}$ is zero, then the rank of $\mathbf{A}$ is m. [Note that, if the determinant of every $(m + 1) \times (m + 1)$ submatrix of $\mathbf{A}$ is zero, then any determinant of order s (where $s > m + 1$) is zero, since any determinant of order $s > m + 1$ can be expressed as a linear sum of determinants of order $m + 1$.]

Properties of Rank of a Matrix. We shall list important properties of the rank of a matrix in the following.

1. The rank of a matrix is invariant under the interchange of two rows (or columns), or the addition of a scalar multiple of a row (or column) to another row (or column), or the multiplication of any row (or column) by a nonzero scalar.
2. For an $n \times m$ matrix $\mathbf{A}$,
$$\text{rank } \mathbf{A} \leq \min(n, m)$$
3. For an $n \times n$ matrix $\mathbf{A}$, a necessary and sufficient condition for rank $\mathbf{A} = n$ is that $|\mathbf{A}| \neq 0$.
4. For an $n \times m$ matrix $\mathbf{A}$,
$$\text{rank } \mathbf{A}^* = \text{rank } \mathbf{A} \qquad \text{or} \qquad \text{rank } \mathbf{A}^T = \text{rank } \mathbf{A}$$
5. The rank of a product of two matrices $\mathbf{AB}$ cannot exceed the rank of $\mathbf{A}$ or the rank of $\mathbf{B}$; that is,
$$\text{rank } \mathbf{AB} \leq \min(\text{rank } \mathbf{A}, \text{rank } \mathbf{B})$$
Hence, if $\mathbf{A}$ is an $n \times 1$ matrix and $\mathbf{B}$ is a $1 \times m$ matrix, then rank $\mathbf{AB} = 1$ unless $\mathbf{AB} = \mathbf{0}$. If a matrix has rank 1, then this matrix can be expressed as a product of a column vector and a row vector.
6. For an $n \times n$ matrix $\mathbf{A}$ (where $|\mathbf{A}| \neq 0$) and an $n \times m$ matrix $\mathbf{B}$,
$$\text{rank } \mathbf{AB} = \text{rank } \mathbf{B}$$
Similarly, for an $m \times m$ matrix $\mathbf{A}$ (where $|\mathbf{A}| \neq 0$) and an $n \times m$ matrix $\mathbf{B}$,
$$\text{rank } \mathbf{BA} = \text{rank } \mathbf{B}$$

Eigenvalues of a Square Matrix. For an $n \times n$ matrix $\mathbf{A}$, the determinant

$$|\lambda\mathbf{I} - \mathbf{A}|$$

is called the *characteristic polynomial* of $\mathbf{A}$. It is an nth-degree polynomial in λ. The characteristic equation is given by

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

If the determinant $|\lambda\mathbf{I} - \mathbf{A}|$ is expanded, the characteristic equation becomes

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{vmatrix}$$
$$= \lambda^n + a_1\lambda^{n-1} + \cdots + a_{n-1}\lambda + a_n = 0$$

The n roots of the characteristic equation are called the *eigenvalues* of $\mathbf{A}$. They are also called the *characteristic roots*.

It is noted that an $n \times n$ real matrix $\mathbf{A}$ does not necessarily possess real eigenvalues. However, for an $n \times n$ real matrix $\mathbf{A}$, the characteristic equation $|\lambda\mathbf{I} - \mathbf{A}| = 0$ is a polynomial with real coefficients, and therefore any complex eigenvalues must occur in conjugate pairs; that is, if $\alpha + j\beta$ is an eigenvalue of $\mathbf{A}$, then $\alpha - j\beta$ is also an eigenvalue of $\mathbf{A}$.

There is an important relationship between the eigenvalues of an $n \times n$ matrix $\mathbf{A}$ and those of $\mathbf{A}^{-1}$. If we assume the eigenvalues of $\mathbf{A}$ to be λ_i and those of $\mathbf{A}^{-1}$ to be μ_i, then

$$\mu_i = \lambda_i^{-1}, \qquad i = 1, 2, \ldots, n$$

That is, if λ_i is an eigenvalue of $\mathbf{A}$, then λ_i^{-1} is an eigenvalue of $\mathbf{A}^{-1}$. To prove this, notice that the characteristic equation for matrix $\mathbf{A}$ can be written as

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

or

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

By assumption, the characteristic equation for the inverse matrix $\mathbf{A}^{-1}$ is

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

By comparing the last two equations, we see that

$$\mu = \lambda^{-1}$$

Hence, if λ is an eigenvalue of $\mathbf{A}$, then $\mu = \lambda^{-1}$ is an eigenvalue of $\mathbf{A}^{-1}$.

Finally, note that it is possible to prove that, for two square matrices $\mathbf{A}$ and $\mathbf{B}$,

$$|\lambda\mathbf{I} - \mathbf{AB}| = |\lambda\mathbf{I} - \mathbf{BA}|$$

(For the proof, see Problem A–9.)

Eigenvectors of an $n \times n$ Matrix. Any nonzero vector $\mathbf{x}_i$ such that

$$\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$$

is said to be an *eigenvector* associated with an eigenvalue λ_i of $\mathbf{A}$, where $\mathbf{A}$ is an $n \times n$ matrix. Since the components of $\mathbf{x}_i$ are determined from n linear homogeneous algebraic equations within a constant factor, if $\mathbf{x}_i$ is an eigenvector, then for any scalar $\alpha \neq 0$, $\alpha\mathbf{x}_i$ is also an eigenvector. The eigenvector is said to be a *normalized* eigenvector if its length or absolute value is unity.

Similar Matrices. The $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be *similar* if a nonsingular matrix $\mathbf{P}$ exists such that

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

The matrix $\mathbf{B}$ is said to be obtained from $\mathbf{A}$ by a *similarity transformation*, in which $\mathbf{P}$ is the transformation matrix. Notice that $\mathbf{A}$ can be obtained from $\mathbf{B}$ by a similarity transformation with a transformation matrix $\mathbf{P}^{-1}$, since

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

Diagonalization of Matrices. If an $n \times n$ matrix $\mathbf{A}$ has n distinct eigenvalues, then there are n linearly independent eigenvectors. If matrix $\mathbf{A}$ has a multiple eigenvalue of multiplicity k, then there are at least one and not more than k linearly independent eigenvectors associated with this eigenvalue.

If an $n \times n$ matrix has n linearly independent eigenvectors, it can be diagonalized by a similarity transformation. However, a matrix that does not have a complete set of n linearly independent eigenvectors cannot be diagonalized. Such a matrix can be transformed into a Jordan canonical form.

Jordan Canonical Form. A $k \times k$ matrix $\mathbf{J}$ is said to be in the Jordan canonical form if

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{p_1} & & & \mathbf{0} \\ & \mathbf{J}_{p_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{J}_{p_s} \end{bmatrix}$$

where the $\mathbf{J}_{p_i}$'s are $p_i \times p_i$ matrices of the form

$$\mathbf{J}_{p_i} = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda & 1 \\ 0 & 0 & 0 & \cdots & 0 & \lambda \end{bmatrix}$$

The matrices $\mathbf{J}_{p_i}$ are called p_ith-order Jordan blocks. Note that the λ in $\mathbf{J}_{p_i}$ and that in $\mathbf{J}_{p_j}$ may or may not be the same, and that

$$p_1 + p_2 + \cdots + p_s = k$$

For example, in a 7×7 matrix $\mathbf{J}$, if $p_1 = 3, p_2 = 2, p_3 = 1, p_4 = 1$, and the eigenvalues of $\mathbf{J}$ are $\lambda_1, \lambda_1, \lambda_1, \lambda_1, \lambda_1, \lambda_6, \lambda_7$, then the Jordan canonical form may be given by

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_3(\lambda_1) & & & \mathbf{0} \\ & \mathbf{J}_2(\lambda_1) & & \\ & & \mathbf{J}_1(\lambda_6) & \\ \mathbf{0} & & & \mathbf{J}_1(\lambda_7) \end{bmatrix} = \left[\begin{array}{ccc|cc|c|c} \lambda_1 & 1 & 0 & & & & 0 \\ 0 & \lambda_1 & 1 & & & & \\ 0 & 0 & \lambda_1 & & & & \\ \hline & & & \lambda_1 & 1 & & \\ & & & 0 & \lambda_1 & & \\ \hline & & & & & \lambda_6 & \\ \hline 0 & & & & & & \lambda_7 \end{array}\right]$$

Notice that a diagonal matrix is a special case of the Jordan canonical form.

Jordan canonical forms have the properties that the elements on the main diagonal of the matrix are the eigenvalues of A and that the elements immediately above (or below) the main diagonal are either 1 or 0 and all other elements are zeros.

The determination of the exact form of the Jordan block may not be simple. To illustrate some possible structures, consider a 3×3 matrix having a triple eigenvalue of λ_1. Then any one of the following Jordan canonical forms is possible:

$$\begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix}, \quad \left[\begin{array}{cc:c} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 0 \\ \hdashline 0 & 0 & \lambda_1 \end{array}\right], \quad \left[\begin{array}{c:cc} \lambda_1 & 0 & 0 \\ \hdashline 0 & \lambda_1 & 0 \\ 0 & 0 & \lambda_1 \end{array}\right]$$

Each of the three preceding matrices has the same characteristic equation $(\lambda - \lambda_1)^3 = 0$. The first one corresponds to the case where there exists only one linearly independent eigenvector, since by denoting the first matrix by A and solving the following equation for $\mathbf{x}$,

$$(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x} = \mathbf{0}$$

we obtain only one eigenvector:

$$\mathbf{x} = \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix}, \qquad a = \text{nonzero constant}$$

The second and third of these matrices have, respectively, two and three linearly independent eigenvectors. (Notice that only the diagonal matrix has three linearly independent eigenvectors.)

As we have seen, if a $k \times k$ matrix A has a k-multiple eigenvalue, then the following can be shown:

1. If the rank of $\lambda\mathbf{I} - \mathbf{A}$ is $k - s$ (where $1 \leq s \leq k$), then there exist s linearly independent eigenvectors associated with λ.
2. There are s Jordan blocks corresponding to the s eigenvectors.
3. The sum of the orders p_i of the Jordan blocks equals the multiplicity k.

Therefore, as demonstrated in the preceding three 3×3 matrices, even if the multiplicity of the eigenvalue is the same, the number of Jordan blocks and their orders may be different depending on the structure of matrix A.

Similarity Transformation When an $n \times n$ Matrix Has Distinct Eigenvalues. If n eigenvalues of A are distinct, there exists one eigenvector associated with each eigenvalue λ_i. It can be proved that such n eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are linearly independent.

Let us define an $n \times n$ matrix P such that

$$\mathbf{P} = [\mathbf{P}_1 \vdots \mathbf{P}_2 \vdots \cdots \vdots \mathbf{P}_n] = [\mathbf{x}_1 \vdots \mathbf{x}_2 \vdots \cdots \vdots \mathbf{x}_n]$$

where column vector $\mathbf{P}_i$ is equal to column vector $\mathbf{x}_i$, or

$$\mathbf{P}_i = \mathbf{x}_i, \qquad i = 1, 2, \ldots, n$$

Matrix P defined in this way is nonsingular, and $\mathbf{P}^{-1}$ exists. Noting that eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ satisfy the equations

$$\begin{aligned} \mathbf{A}\mathbf{x}_1 &= \lambda_1 \mathbf{x}_1 \\ \mathbf{A}\mathbf{x}_2 &= \lambda_2 \mathbf{x}_2 \\ &\vdots \\ \mathbf{A}\mathbf{x}_n &= \lambda_n \mathbf{x}_n \end{aligned}$$

we may combine these n equations into one, as follows:

$$\mathbf{A}[\mathbf{x}_1 \vdots \mathbf{x}_2 \vdots \cdots \vdots \mathbf{x}_n] = [\mathbf{x}_1 \vdots \mathbf{x}_2 \vdots \cdots \vdots \mathbf{x}_n] \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

or, in terms of matrix P,

$$\mathbf{A}\mathbf{P} = \mathbf{P} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

By premultiplying this last equation by $\mathbf{P}^{-1}$, we obtain

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

Thus, matrix A is transformed into a diagonal matrix by a similarity transformation.

The process that transforms matrix A into a diagonal matrix is called the *diagonalization* of matrix A.

As noted earlier, a scalar multiple of eigenvector $\mathbf{x}_i$ is also an eigenvector, since $\alpha\mathbf{x}_i$ satisfies the following equation:

$$\mathbf{A}(\alpha\mathbf{x}_i) = \lambda_i(\alpha\mathbf{x}_i)$$

Consequently, we may choose an α such that the transformation matrix P becomes as simple as possible.

To summarize, if the eigenvalues of an $n \times n$ matrix A are distinct, then there are exactly n eigenvectors and they are linearly independent. A transformation matrix P that transforms A into a diagonal matrix can be constructed from such n linearly independent eigenvectors.

Similarity Transformation When an $n \times n$ Matrix Has Multiple Eigenvalues. Let us assume that an $n \times n$ matrix A involves a k-multiple eigenvalue λ_1 and other eigenvalues $\lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_n$ that are all distinct and different from λ_1. That is, the eigenvalues of A are

$$\lambda_1, \lambda_1, \ldots, \lambda_1, \lambda_{k+1}, \lambda_{k-2}, \ldots, \lambda_n$$

We shall first consider the case where the rank of $\lambda_1 \mathbf{I} - \mathbf{A}$ is $n - 1$. For such a case there exists only one Jordan block for the multiple eigenvalue λ_1, and there is only one eigenvector associated with this multiple eigenvalue. The order of the Jordan block is k, which is the same as the order of multiplicity of the eigenvalue λ_1.

Note that, when an $n \times n$ matrix $\mathbf{A}$ does not possess n linearly independent eigenvectors, it cannot be diagonalized, but can be reduced to a Jordan canonical form.

In the present case, only one linearly independent eigenvector exists for λ_1. We shall now investigate whether it is possible to find $k - 1$ vectors that are somehow associated with this eigenvalue and that are linearly independent of the eigenvectors. Without proof, we shall show that this is possible. First, note that the eigenvector $\mathbf{x}_1$ is a vector that satisfies the equation

$$(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x}_1 = \mathbf{0}$$

so that $\mathbf{x}_1$ is annihilated by $\mathbf{A} - \lambda_1 \mathbf{I}$. Since we do not have enough vectors that are annihilated by $\mathbf{A} - \lambda_1 \mathbf{I}$, we seek vectors that are annihilated by $(\mathbf{A} - \lambda_1 \mathbf{I})^2$, $(\mathbf{A} - \lambda_1 \mathbf{I})^3$, and so on, until we obtain $k - 1$ vectors. The $k - 1$ vectors determined in this way are called *generalized eigenvectors*.

Let us define the desired $k - 1$ generalized eigenvectors as $\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_k$. Then these $k - 1$ generalized eigenvectors can be determined from the equations

$$\begin{aligned}
(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x}_1 &= \mathbf{0} \\
(\mathbf{A} - \lambda_1 \mathbf{I})^2 \mathbf{x}_2 &= \mathbf{0} \\
&\vdots \\
(\mathbf{A} - \lambda_1 \mathbf{I})^k \mathbf{x}_k &= \mathbf{0}
\end{aligned} \tag{A–32}$$

which can be rewritten as

$$\begin{aligned}
(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x}_1 &= \mathbf{0} \\
(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x}_2 &= \mathbf{x}_1 \\
&\vdots \\
(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x}_k &= \mathbf{x}_{k-1}
\end{aligned}$$

Notice that

$$(\mathbf{A} - \lambda_1 \mathbf{I})^{k-1}\mathbf{x}_k = (\mathbf{A} - \lambda_1 \mathbf{I})^{k-2}\mathbf{x}_{k-1} = \cdots = (\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{x}_2 = \mathbf{x}_1$$

or

$$(\mathbf{A} - \lambda_1 \mathbf{I})^{k-1}\mathbf{x}_k = \mathbf{x}_1 \tag{A–33}$$

The eigenvector $\mathbf{x}_1$ and the $k - 1$ generalized eigenvectors $\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_k$ determined in this way form a set of k linearly independent vectors.

A proper way to determine the generalized eigenvectors is to start with $\mathbf{x}_k$. That is, we first determine the $\mathbf{x}_k$ that will satisfy Equation (A–32) and at the same time will yield a nonzero vector $(\mathbf{A} - \lambda_1 \mathbf{I})^{k-1}\mathbf{x}_k$. Any such nonzero vector can be considered as a possible eigenvector $\mathbf{x}_1$. Therefore, to find eigenvector $\mathbf{x}_1$, we apply a row reduction process to $(\mathbf{A} - \lambda_1 \mathbf{I})^k$ and find k linearly independent vectors satisfying Equation (A–32). Then these vectors are tested to find one that yields a nonzero vector on the right-hand side of Equation (A–33). (Note that if we start with $\mathbf{x}_1$ then we must make arbitrary choices at each step along the way to determine $\mathbf{x}_2$, $\mathbf{x}_3, \ldots, \mathbf{x}_k$. This is time consuming and inconvenient. For this reason, this approach is not recommended.)

To summarize what we have discussed so far, the eigenvector $\mathbf{x}_1$ and the generalized eigenvectors $\mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_k$ satisfy the following equations:

$$\begin{aligned}
\mathbf{A}\mathbf{x}_1 &= \lambda_1 \mathbf{x}_1 \\
\mathbf{A}\mathbf{x}_2 &= \mathbf{x}_1 + \lambda_1 \mathbf{x}_2 \\
&\vdots \\
\mathbf{A}\mathbf{x}_k &= \mathbf{x}_{k-1} + \lambda_1 \mathbf{x}_k
\end{aligned}$$

The eigenvectors $\mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \ldots, \mathbf{x}_n$ associated with distinct eigenvalues λ_{k+1}, $\lambda_{k+2}, \ldots, \lambda_n$, respectively, can be determined from

$$\begin{aligned}
\mathbf{A}\mathbf{x}_{k+1} &= \lambda_{k+1}\mathbf{x}_{k+1} \\
\mathbf{A}\mathbf{x}_{k+2} &= \lambda_{k+2}\mathbf{x}_{k+2} \\
&\vdots \\
\mathbf{A}\mathbf{x}_n &= \lambda_n \mathbf{x}_n
\end{aligned}$$

Now define

$$\mathbf{S} = [\mathbf{S}_1 \vdots \mathbf{S}_2 \vdots \cdots \vdots \mathbf{S}_n] = [\mathbf{x}_1 \vdots \mathbf{x}_2 \vdots \cdots \vdots \mathbf{x}_n]$$

where the n column vectors of $\mathbf{S}$ are linearly independent. Thus, matrix $\mathbf{S}$ is nonsingular. Then, combining the preceding eigenvector equations and generalized eigenvector equations into one, we obtain

$$\mathbf{A}[\mathbf{x}_1 \vdots \mathbf{x}_2 \vdots \cdots \vdots \mathbf{x}_k \vdots \mathbf{x}_{k+1} \vdots \cdots \vdots \mathbf{x}_n]$$

$$= [\mathbf{x}_1 \vdots \mathbf{x}_2 \vdots \cdots \vdots \mathbf{x}_k \vdots \mathbf{x}_{k+1} \vdots \cdots \vdots \mathbf{x}_n] \left[\begin{array}{cccc:ccc}
\lambda_1 & 1 & & 0 & & & 0 \\
 & \lambda_1 & 1 & & & & \\
 & & \ddots & \ddots & & & \\
 & & & 1 & & & \\
0 & & & \lambda_1 & 0 & & \\
\hdashline
 & & & 0 & \lambda_{k+1} & & 0 \\
 & & & & & \ddots & \\
0 & & & & 0 & & \lambda_n
\end{array}\right]$$

Hence,

$$\mathbf{AS} = \mathbf{S}\left[\begin{array}{c:ccc}
\mathbf{J}_k(\lambda_1) & & & 0 \\
\hdashline
 & \lambda_{k+1} & & \\
 & & \ddots & \\
0 & & & \lambda_n
\end{array}\right]$$

By premultiplying this last equation by $\mathbf{S}^{-1}$, we obtain

$$\mathbf{S}^{-1}\mathbf{AS} = \begin{bmatrix} \mathbf{J}_k(\lambda_1) & & 0 \\ & \lambda_{k+1} & \\ & & \ddots \\ 0 & & \lambda_n \end{bmatrix}$$

In the preceding discussion we considered the case where the rank of $\lambda_1\mathbf{I} - \mathbf{A}$ was $n - 1$. Next we shall consider the case where the rank of $\lambda_1\mathbf{I} - \mathbf{A}$ is $n - s$ (where $2 \leq s \leq n$). Since we assumed that matrix $\mathbf{A}$ involves the k-multiple eigenvalue λ_1 and other eigenvalues $\lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_n$ that are all distinct and different from λ_1, we have s linearly independent eigenvectors associated with eigenvalue λ_1. Hence, there are s Jordan blocks corresponding to eigenvalue λ_1.

For notational convenience, let us define the s linearly independent eigenvectors associated with eigenvalue λ_1 as $\mathbf{v}_{11}, \mathbf{v}_{21}, \ldots, \mathbf{v}_{s1}$. We shall define the generalized eigenvectors associated with $\mathbf{v}_{i1}$ as $\mathbf{v}_{i2}, \mathbf{v}_{i3}, \ldots, \mathbf{v}_{ip_i}$, where $i = 1, 2, \ldots, s$. Then there are altogether k such vectors (eigenvectors and generalized eigenvectors), which are

$$\mathbf{v}_{11}, \mathbf{v}_{12}, \ldots, \mathbf{v}_{1p_1}, \mathbf{v}_{21}, \mathbf{v}_{22}, \ldots, \mathbf{v}_{2p_2}, \ldots, \mathbf{v}_{s1}, \mathbf{v}_{s2}, \ldots, \mathbf{v}_{sp_s}$$

The generalized eigenvectors are determined from

$$\begin{aligned} (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{11} &= \mathbf{0}, & \cdots \quad (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{s1} &= \mathbf{0} \\ (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{12} &= \mathbf{v}_{11}, & \cdots \quad (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{s2} &= \mathbf{v}_{s1} \\ &\vdots & & \vdots \\ (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{1p_1} &= \mathbf{v}_{1p_1-1}, & \cdots \quad (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{sp_s} &= \mathbf{v}_{sp_s-1} \end{aligned}$$

where the s eigenvectors $\mathbf{v}_{11}, \mathbf{v}_{21}, \ldots, \mathbf{v}_{s1}$ are linearly independent and

$$p_1 + p_2 + \cdots + p_s = k$$

Note that $p_1, p_2, \ldots, p_s$ represent the order of each of the s Jordan blocks. (For the determination of the generalized eigenvectors, we follow the method discussed earlier. For an example showing the details of such a determination, see Problem A–11.)

Let us define an $n \times k$ matrix consisting of $\mathbf{v}_{11}, \mathbf{v}_{12}, \ldots, \mathbf{v}_{sp_s}$ as

$$\begin{aligned} \mathbf{S}(\lambda_1) &= [\mathbf{v}_{11} \vdots \mathbf{v}_{12} \vdots \cdots \vdots \mathbf{v}_{1p_1} \vdots \cdots \vdots \mathbf{v}_{s1} \vdots \mathbf{v}_{s2} \vdots \cdots \vdots \mathbf{v}_{sp_s}] \\ &= [\mathbf{x}_1 \vdots \mathbf{x}_2 \vdots \cdots \vdots \mathbf{x}_{p_1} \vdots \cdots \vdots \mathbf{x}_k] \\ &= [\mathbf{S}_1 \vdots \mathbf{S}_2 \vdots \cdots \vdots \mathbf{S}_k] \end{aligned}$$

and define

$$\begin{aligned} \mathbf{S} &= [\mathbf{S}(\lambda_1) \vdots \mathbf{S}_{k+1} \vdots \mathbf{S}_{k+2} \vdots \cdots \vdots \mathbf{S}_n] \\ &= [\mathbf{S}_1 \vdots \mathbf{S}_2 \vdots \cdots \vdots \mathbf{S}_n] \end{aligned}$$

where

$$\mathbf{S}_{k+1} = \mathbf{x}_{k+1}, \qquad \mathbf{S}_{k+2} = \mathbf{x}_{k+2}, \qquad \ldots, \qquad \mathbf{S}_n = \mathbf{x}_n$$

Note that $\mathbf{x}_{k+1}, \mathbf{x}_{k+2}, \ldots, \mathbf{x}_n$ are eigenvectors associated with eigenvalues $\lambda_{k+1}, \lambda_{k+2}, \ldots, \lambda_n$, respectively. Matrix $\mathbf{S}$ defined in this way is nonsingular. Now we obtain

$$\mathbf{AS} = \mathbf{S}\begin{bmatrix} \mathbf{J}_{p_1}(\lambda_1) & & 0 & & \\ & \mathbf{J}_{p_2}(\lambda_1) & & & \\ & & \ddots & & \\ 0 & & \mathbf{J}_{p_s}(\lambda_1) & 0 & \\ & & & \lambda_{k+1} & \\ & & 0 & & \ddots \\ & & & & \lambda_n \end{bmatrix}$$

where $\mathbf{J}_{p_i}(\lambda_1)$ is in the form

$$\mathbf{J}_{p_i}(\lambda_1) = \begin{bmatrix} \lambda_1 & 1 & & 0 \\ & \lambda_1 & 1 & \\ & & \ddots & \\ & & & 1 \\ 0 & & & \lambda_1 \end{bmatrix}$$

which is a $p_i \times p_i$ matrix. Hence,

$$\mathbf{S}^{-1}\mathbf{AS} = \begin{bmatrix} \mathbf{J}_{p_1}(\lambda_1) & & 0 & & \\ & \mathbf{J}_{p_2}(\lambda_1) & & & \\ & & \ddots & & \\ 0 & & \mathbf{J}_{p_s}(\lambda_1) & 0 & \\ & & & \lambda_{k+1} & \\ & & 0 & & \ddots \\ & & & & \lambda_n \end{bmatrix}$$

Thus, as we have shown, by using a set of n linearly independent vectors (eigenvectors and generalized eigenvectors), any $n \times n$ matrix can be reduced to a Jordan canonical form by a similarity transformation.

Similarity Transformation When an $n \times n$ Matrix Is Normal. First, recall that a matrix is normal if it is a real symmetric, a Hermitian, a real skew-symmetric, a skew-Hermitian, an orthogonal, or a unitary matrix.

Assume that an $n \times n$ normal matrix has a k-multiple eigenvalue λ_1 and that its other $n - k$ eigenvalues are distinct and different from λ_1. Then the rank of $\mathbf{A} - \lambda_1\mathbf{I}$ becomes $n - k$. (Refer to Problem A–12 for the proof.) If the rank of $\mathbf{A} - \lambda_1\mathbf{I}$ is $n - k$, there are k linearly independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ that satisfy the equation

$$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{x}_i = \mathbf{0}, \qquad i = 1, 2, \ldots, k$$

Therefore, there exist k Jordan blocks for eigenvalue λ_1. Since the number of Jordan blocks is the same as the multiplicity number of eigenvalue λ_1, all k Jordan blocks become first order. Since the remaining $n - k$ eigenvalues are distinct, the eigenvectors associated with these eigenvalues are linearly independent. Hence, the $n \times n$ normal matrix possesses altogether n linearly independent eigenvectors, and the Jordan canonical form of the normal matrix becomes a diagonal matrix.

It can be proved that if A is an $n \times n$ normal matrix, then, regardless of whether or not the eigenvalues include multiple eigenvalues, there exists an $n \times n$ unitary matrix U such that

$$\mathbf{U}^{-1}\mathbf{A}\mathbf{U} = \mathbf{U}^*\mathbf{A}\mathbf{U} = \mathbf{D} = \text{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_n)$$

where D is a diagonal matrix with n eigenvalues as diagonal elements.

Trace of an $n \times n$ Matrix. The trace of an $n \times n$ matrix A is defined as follows:

$$\text{trace of } \mathbf{A} = \text{tr}\,\mathbf{A} = \sum_{i=1}^{n} a_{ii}$$

The trace of an $n \times n$ matrix A has the following properties:

1. $$\text{tr}\,\mathbf{A}^T = \text{tr}\,\mathbf{A}$$
2. For $n \times n$ matrices A and B,
$$\text{tr}\,(\mathbf{A} + \mathbf{B}) = \text{tr}\,\mathbf{A} + \text{tr}\,\mathbf{B}$$
3. If the eigenvalues of A are denoted by $\lambda_1, \lambda_2, \ldots, \lambda_n$, then
$$\text{tr}\,\mathbf{A} = \lambda_1 + \lambda_2 + \cdots + \lambda_n \qquad \text{(A–34)}$$
4. For an $n \times m$ matrix A and an $m \times n$ matrix B, regardless of whether $\mathbf{AB} = \mathbf{BA}$ or $\mathbf{AB} \neq \mathbf{BA}$, we have
$$\text{tr}\,\mathbf{AB} = \text{tr}\,\mathbf{BA} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij} b_{ji}$$
If $m = 1$, then by writing A and B as a and b, respectively, we have
$$\text{tr}\,\mathbf{ab} = \mathbf{ba}$$
Hence, for an $n \times m$ matrix C, we have
$$\mathbf{a}^T\mathbf{C}\mathbf{a} = \text{tr}\,\mathbf{a}\mathbf{a}^T\mathbf{C}$$

Note that Equation (A–34) may be proved as follows. By use of a similarity transformation, we have

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D} = \text{diagonal matrix}$$

or

$$\mathbf{S}^{-1}\mathbf{A}\mathbf{S} = \mathbf{J} = \text{Jordan canonical form}$$

That is,

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1} \qquad \text{or} \qquad \mathbf{A} = \mathbf{S}\mathbf{J}\mathbf{S}^{-1}$$

Hence, by using property 4 listed here, we have

$$\text{tr}\,\mathbf{A} = \text{tr}\,\mathbf{P}\mathbf{D}\mathbf{P}^{-1} = \text{tr}\,\mathbf{P}^{-1}\mathbf{P}\mathbf{D} = \text{tr}\,\mathbf{D} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Similarly,

$$\text{tr}\,\mathbf{A} = \text{tr}\,\mathbf{S}\mathbf{J}\mathbf{S}^{-1} = \text{tr}\,\mathbf{S}^{-1}\mathbf{S}\mathbf{J} = \text{tr}\,\mathbf{J} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$$

Invariant Properties Under Similarity Transformation. If an $n \times n$ matrix A can be reduced to a similar matrix that has a simple form, then important properties of A can be readily observed. A property of a matrix is said to be invariant if it is possessed by all similar matrices. For example, the determinant and the characteristic polynomial are invariant under a similarity transformation, as shown in the following. Suppose that $\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{B}$. Then

$$\begin{aligned}|\mathbf{B}| &= |\mathbf{P}^{-1}\mathbf{A}\mathbf{P}| = |\mathbf{P}^{-1}|\,|\mathbf{A}|\,|\mathbf{P}| = |\mathbf{A}|\,|\mathbf{P}^{-1}|\,|\mathbf{P}| = |\mathbf{A}|\,|\mathbf{P}^{-1}\mathbf{P}| \\ &= |\mathbf{A}|\,|\mathbf{I}| = |\mathbf{A}|\end{aligned}$$

and

$$\begin{aligned}|\lambda\mathbf{I} - \mathbf{B}| &= |\lambda\mathbf{I} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P}| = |\mathbf{P}^{-1}(\lambda\mathbf{I})\mathbf{P} - \mathbf{P}^{-1}\mathbf{A}\mathbf{P}| \\ &= |\mathbf{P}^{-1}(\lambda\mathbf{I} - \mathbf{A})\mathbf{P}| = |\mathbf{P}^{-1}|\,|\lambda\mathbf{I} - \mathbf{A}|\,|\mathbf{P}| \\ &= |\lambda\mathbf{I} - \mathbf{A}|\,|\mathbf{P}^{-1}|\,|\mathbf{P}| = |\lambda\mathbf{I} - \mathbf{A}|\end{aligned}$$

Notice that the trace of a matrix is also invariant under similarity transformation, as was shown earlier:

$$\text{tr}\,\mathbf{A} = \text{tr}\,\mathbf{P}^{-1}\mathbf{A}\mathbf{P}$$

The property of symmetry of a matrix, however, is not invariant.

Notice that only invariant properties of matrices present intrinsic characteristics of the class of similar matrices. To determine the invariant properties of a matrix A, we examine the Jordan canonical form of A, since the similarity of two matrices can be defined in terms of the Jordan canonical form: The necessary and sufficient condition for $n \times n$ matrices A and B to be similar is that the Jordan canonical form of A and that of B be identical.

A–7 QUADRATIC FORMS

Quadratic Forms. For an $n \times n$ real symmetric matrix A and a real n-vector x, the form

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} x_i x_j, \qquad a_{ji} = a_{ij}$$

is called a *real quadratic form* in x_i. Frequently, a real quadratic form is called simply a *quadratic form*. Note that $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is a real scalar quantity.

Any real quadratic form can always be written as $\mathbf{x}^T\mathbf{A}\mathbf{x}$. For example,

$$x_1^2 - 2x_1x_2 + 4x_1x_3 + x_2^2 + 8x_3^2 = \begin{bmatrix} x_1 & x_2 & x_3 \end{bmatrix} \begin{bmatrix} 1 & -1 & 2 \\ -1 & 1 & 0 \\ 2 & 0 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

It is worthwhile to mention that, for an $n \times n$ real matrix A, if we define

$$\mathbf{B} = \tfrac{1}{2}(\mathbf{A} + \mathbf{A}^T) \qquad \text{and} \qquad \mathbf{C} = \tfrac{1}{2}(\mathbf{A} - \mathbf{A}^T)$$

then

$$\mathbf{A} = \mathbf{B} + \mathbf{C}$$

Notice that

$$\mathbf{B}^T = \mathbf{B} \qquad \text{and} \qquad \mathbf{C}^T = -\mathbf{C}$$

Hence, an $n \times n$ real matrix $\mathbf{A}$ can be expressed as a sum of a real symmetric and a real skew-symmetric matrix. Since $\mathbf{x}^T\mathbf{C}\mathbf{x}$ is a real scalar quantity, we have

$$\mathbf{x}^T\mathbf{C}\mathbf{x} = (\mathbf{x}^T\mathbf{C}\mathbf{x})^T = \mathbf{x}^T\mathbf{C}^T\mathbf{x} = -\mathbf{x}^T\mathbf{C}\mathbf{x}$$

Consequently, we have

$$\mathbf{x}^T\mathbf{C}\mathbf{x} = 0$$

This means that a quadratic form for a real skew-symmetric matrix is zero. Hence,

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

and we see that the real quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ involves only the symmetric component $\mathbf{x}^T\mathbf{B}\mathbf{x}$. This is the reason why the real quadratic form is defined only for a real symmetric matrix.

For a Hermitian matrix $\mathbf{A}$ and a complex n-vector $\mathbf{x}$, the form

$$\mathbf{x}^*\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\bar{x}_i x_j, \qquad a_{ji} = \bar{a}_{ij}$$

is called a *complex quadratic form*, or Hermitian form. Notice that the scalar quantity $\mathbf{x}^*\mathbf{A}\mathbf{x}$ is real, because

$$\overline{\mathbf{x}^*\mathbf{A}\mathbf{x}} = \mathbf{x}^T\overline{\mathbf{A}}\overline{\mathbf{x}} = (\mathbf{x}^T\overline{\mathbf{A}}\overline{\mathbf{x}})^T = \overline{\mathbf{x}}^T\overline{\mathbf{A}}^T\mathbf{x} = \mathbf{x}^*\mathbf{A}\mathbf{x}$$

Bilinear Forms. For an $n \times m$ real matrix $\mathbf{A}$, a real n-vector $\mathbf{x}$, and a real m-vector $\mathbf{y}$, the form

$$\mathbf{x}^T\mathbf{A}\mathbf{y} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}x_i y_j$$

is called a *real bilinear form* in x_i and y_j. $\mathbf{x}^T\mathbf{A}\mathbf{y}$ is a real scalar quantity.

For an $n \times m$ complex matrix $\mathbf{A}$, a complex n-vector $\mathbf{x}$, and a complex m-vector $\mathbf{y}$, the form

$$\mathbf{x}^*\mathbf{A}\mathbf{y} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}\bar{x}_i y_j$$

is called a *complex bilinear form*. $\mathbf{x}^*\mathbf{A}\mathbf{y}$ is a complex scalar quantity.

Definiteness and Semidefiniteness. A quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a real symmetric matrix (or a Hermitian form $\mathbf{x}^*\mathbf{A}\mathbf{x}$, where $\mathbf{A}$ is a Hermitian matrix), is said to be positive definite if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} > 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} > 0), \quad \text{for } \mathbf{x} \neq \mathbf{0}$$
$$\mathbf{x}^T\mathbf{A}\mathbf{x} = 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} = 0), \quad \text{for } \mathbf{x} = \mathbf{0}$$

$\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or $\mathbf{x}^*\mathbf{A}\mathbf{x}$) is said to be positive semidefinite if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} \geq 0), \quad \text{for } \mathbf{x} \neq \mathbf{0}$$
$$\mathbf{x}^T\mathbf{A}\mathbf{x} = 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} = 0), \quad \text{for } \mathbf{x} = \mathbf{0}$$

$\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or $\mathbf{x}^*\mathbf{A}\mathbf{x}$) is said to be negative definite if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} < 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} < 0), \quad \text{for } \mathbf{x} \neq \mathbf{0}$$
$$\mathbf{x}^T\mathbf{A}\mathbf{x} = 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} = 0), \quad \text{for } \mathbf{x} = \mathbf{0}$$

$\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or $\mathbf{x}^*\mathbf{A}\mathbf{x}$) is said to be negative semidefinite if

$$\mathbf{x}^T\mathbf{A}\mathbf{x} \leq 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} \leq 0), \quad \text{for } \mathbf{x} \neq \mathbf{0}$$
$$\mathbf{x}^T\mathbf{A}\mathbf{x} = 0 \quad (\text{or } \mathbf{x}^*\mathbf{A}\mathbf{x} = 0), \quad \text{for } \mathbf{x} = \mathbf{0}$$

If $\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or $\mathbf{x}^*\mathbf{A}\mathbf{x}$) can be of either sign, then $\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or $\mathbf{x}^*\mathbf{A}\mathbf{x}$) is said to be indefinite.

Note that if $\mathbf{x}^T\mathbf{A}\mathbf{x}$ or $\mathbf{x}^*\mathbf{A}\mathbf{x}$ is positive (or negative) definite we say that $\mathbf{A}$ is a positive (or negative) definite matrix. Similarly, matrix $\mathbf{A}$ is called a positive (or negative) semidefinite matrix if $\mathbf{x}^T\mathbf{A}\mathbf{x}$ or $\mathbf{x}^*\mathbf{A}\mathbf{x}$ is positive (or negative) semidefinite; matrix $\mathbf{A}$ is called an indefinite matrix if $\mathbf{x}^T\mathbf{A}\mathbf{x}$ or $\mathbf{x}^*\mathbf{A}\mathbf{x}$ is indefinite.

Note also that the eigenvalues of an $n \times n$ real symmetric or Hermitian matrix are real. (For the proof, see Problem A–13.) It can be shown that an $n \times n$ real symmetric or Hermitian matrix $\mathbf{A}$ is a positive definite matrix if all eigenvalues λ_i $(i = 1, 2, \ldots, n)$ are positive. Matrix $\mathbf{A}$ is positive semidefinite if all eigenvalues are nonnegative, or $\lambda_i \geq 0$ $(i = 1, 2, \ldots, n)$, and at least one of them is zero.

Notice that if $\mathbf{A}$ is a positive definite matrix then $|\mathbf{A}| \neq 0$, because all eigenvalues are positive. Hence, the inverse matrix always exists for a positive definite matrix.

In the process of determining the stability of an equilibrium state, we frequently encounter a scalar function $V(\mathbf{x})$. A scalar function $V(\mathbf{x})$, which is a function of $x_1, x_2, \ldots, x_n$, is said to be positive definite if

$$V(\mathbf{x}) > 0, \quad \text{for } \mathbf{x} \neq \mathbf{0}$$
$$V(\mathbf{0}) = 0$$

$V(\mathbf{x})$ is said to be positive semidefinite if

$$V(\mathbf{x}) \geq 0, \quad \text{for } \mathbf{x} \neq \mathbf{0}$$
$$V(\mathbf{0}) = 0$$

If $-V(\mathbf{x})$ is positive definite (or positive semidefinite), then $V(\mathbf{x})$ is said to be negative definite (or negative semidefinite).

Necessary and sufficient conditions for the quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or the Hermitian form $\mathbf{x}^*\mathbf{A}\mathbf{x}$) to be positive definite, negative definite, positive semidefinite, or negative semidefinite have been given by J. J. Sylvester. Sylvester's criteria follow.

Sylvester's Criterion for Positive Definiteness of a Quadratic Form or Hermitian Form. A necessary and sufficient condition for a quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or a Hermitian form $\mathbf{x}^*\mathbf{A}\mathbf{x}$), where $\mathbf{A}$ is an $n \times n$ real symmetric matrix (or Hermitian

matrix), to be positive definite is that the determinant of $\mathbf{A}$ be positive and the successive principal minors of the determinant of $\mathbf{A}$ (the determinants of the $k \times k$ matrices in the top-left corner of matrix $\mathbf{A}$, where $k = 1, 2, \ldots, n-1$) be positive; that is, we must have

$$a_{11} > 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} > 0, \quad \ldots, \quad |\mathbf{A}| > 0$$

where

$$a_{ij} = a_{ji}, \quad \text{for real symmetric matrix } \mathbf{A}$$
$$a_{ij} = \bar{a}_{ji}, \quad \text{for Hermitian matrix } \mathbf{A}$$

Sylvester's Criterion for Negative Definiteness of a Quadratic Form or Hermitian Form. A necessary and sufficient condition for a quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or a Hermitian form $\mathbf{x}^*\mathbf{A}\mathbf{x}$), where $\mathbf{A}$ is an $n \times n$ real symmetric matrix (or Hermitian matrix), to be negative definite is that the determinant of $\mathbf{A}$ be positive if n is even and negative if n is odd, and that the successive principal minors of even order be positive and the successive principal minors of odd order be negative; that is, we must have

$$a_{11} < 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} < 0, \quad \ldots$$
$$|\mathbf{A}| > 0 \quad (n \text{ even})$$
$$|\mathbf{A}| < 0 \quad (n \text{ odd})$$

where

$$a_{ij} = a_{ji}, \quad \text{for real symmetric matrix } \mathbf{A}$$
$$a_{ij} = \bar{a}_{ji}, \quad \text{for Hermitian matrix } \mathbf{A}$$

[This condition can be derived by requiring that $\mathbf{x}^T(-\mathbf{A})\mathbf{x}$ be positive definite.]

Sylvester's Criterion for Positive Semidefiniteness of a Quadratic Form or Hermitian Form. A necessary and sufficient condition for a quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or a Hermitian form $\mathbf{x}^*\mathbf{A}\mathbf{x}$), where $\mathbf{A}$ is a real symmetric matrix (or a Hermitian matrix), to be positive semidefinite is that $\mathbf{A}$ be singular ($|\mathbf{A}| = 0$) and all the principal minors be nonnegative:

$$a_{ii} \geq 0, \quad \begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix} \geq 0, \quad \begin{vmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{vmatrix} \geq 0, \quad \ldots, \quad |\mathbf{A}| = 0$$

where $i < j < k$ and

$$a_{ij} = a_{ji}, \quad \text{for real symmetric matrix } \mathbf{A}$$
$$a_{ij} = \bar{a}_{ji}, \quad \text{for Hermitian matrix } \mathbf{A}$$

(It is important to point out that in the positive semidefiniteness test or negative semidefiniteness test we must check the signs of all the principal minors, not just successive principal minors. See Problem A–15.)

Sylvester's Criterion for Negative Semidefiniteness of a Quadratic Form or a Hermitian Form. A necessary and sufficient condition for a quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$ (or a Hermitian form $\mathbf{x}^*\mathbf{A}\mathbf{x}$), where $\mathbf{A}$ is an $n \times n$ real symmetric matrix (or Hermitian matrix), to be negative semidefinite is that $\mathbf{A}$ be singular ($|\mathbf{A}| = 0$) and that all the principal minors of even order be nonnegative and those of odd order be nonpositive:

$$a_{ii} \leq 0, \quad \begin{vmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{vmatrix} \geq 0, \quad \begin{vmatrix} a_{ii} & a_{ij} & a_{ik} \\ a_{ji} & a_{jj} & a_{jk} \\ a_{ki} & a_{kj} & a_{kk} \end{vmatrix} \leq 0, \quad \ldots, \quad |\mathbf{A}| = 0$$

where $i < j < k$ and

$$a_{ij} = a_{ji}, \quad \text{for real symmetric matrix } \mathbf{A}$$
$$a_{ij} = \bar{a}_{ji}, \quad \text{for Hermitian matrix } \mathbf{A}$$

A–8 PSEUDOINVERSES

The concept of pseudoinverses of a matrix is a generalization of the notion of an inverse. It is useful for finding a "solution" to a set of algebraic equations in which the number of unknown variables and the number of independent linear equations are not equal.

In what follows, we shall consider pseudoinverses that enable us to determine minimum norm solutions.

Minimum Norm Solution That Minimizes $\|\mathbf{x}\|$. Consider a linear algebraic equation

$$x_1 + 5x_2 = 1$$

Since we have two variables and only one equation, no unique solution exists. Instead, there exist an infinite number of solutions. Graphically, any point on line $x_1 + 5x_2 = 1$, as shown in Figure A–1, is a possible solution. However, if we decide to pick the point that is closest to the origin, the solution becomes unique.

Consider the vector-matrix equation

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{A–35}$$

where $\mathbf{A}$ is an $n \times m$ matrix, $\mathbf{x}$ is an m-vector, and $\mathbf{b}$ is an n-vector. We assume that $m > n$ (that is, the number of unknown variables is greater than the number of equations) and that the equation has an infinite number of solutions. Let us find the unique solution $\mathbf{x}$ that is located closest to the origin or that has the minimum norm $\|\mathbf{x}\|$.

Let us define the minimum norm solution as $\mathbf{x}^\circ$. That is, $\mathbf{x}^\circ$ satisfies the condition that $\mathbf{A}\mathbf{x}^\circ = \mathbf{b}$ and $\|\mathbf{x}^\circ\| \leq \|\mathbf{x}\|$ for all $\mathbf{x}$ that satisfy $\mathbf{A}\mathbf{x} = \mathbf{b}$. This means that

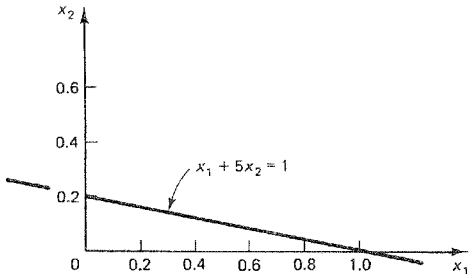


Figure A–1 Line $x_1 + 5x_2 = 1$ on the x_1x_2 plane.

the solution point $\mathbf{x}^\circ$ is nearest to the origin of the m-dimensional space among all possible solutions of Equation (A–35). We shall obtain such a minimum norm solution in the following.

Right Pseudoinverse Matrix. For a vector-matrix equation

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A}$ is an $n \times m$ matrix having rank n, $\mathbf{x}$ is an m-vector, and $\mathbf{b}$ is an n-vector, the solution that minimizes the norm $\|\mathbf{x}\|$ is given by

$$\mathbf{x}^\circ = \mathbf{A}^{RM}\mathbf{b}$$

where $\mathbf{A}^{RM} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$.

This can be proved as follows. First, note that norm $\|\mathbf{x}\|$ can be written as follows:

$$\|\mathbf{x}\| = \|\mathbf{x} - \mathbf{x}^\circ + \mathbf{x}^\circ\| = \|\mathbf{x}^\circ\| + \|\mathbf{x} - \mathbf{x}^\circ\| + 2(\mathbf{x}^\circ)^T(\mathbf{x} - \mathbf{x}^\circ)$$

The last term, $2(\mathbf{x}^\circ)^T(\mathbf{x} - \mathbf{x}^\circ)$, can be shown to be zero, since

$$\begin{aligned}(\mathbf{x}^\circ)^T(\mathbf{x} - \mathbf{x}^\circ) &= [\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}]^T[\mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}] \\ &= \mathbf{b}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{A}[\mathbf{x} - \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}] \\ &= \mathbf{b}^T(\mathbf{A}\mathbf{A}^T)^{-1}[\mathbf{A}\mathbf{x} - (\mathbf{A}\mathbf{A}^T)(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{b}] \\ &= \mathbf{b}^T(\mathbf{A}\mathbf{A}^T)^{-1}(\mathbf{b} - \mathbf{b}) \\ &= 0\end{aligned}$$

Hence,

$$\|\mathbf{x}\| = \|\mathbf{x}^\circ\| + \|\mathbf{x} - \mathbf{x}^\circ\|$$

which can be rewritten as

$$\|\mathbf{x}\| - \|\mathbf{x}^\circ\| = \|\mathbf{x} - \mathbf{x}^\circ\|$$

Since $\|\mathbf{x} - \mathbf{x}^\circ\| \geq 0$, we obtain

$$\|\mathbf{x}\| \geq \|\mathbf{x}^\circ\|$$

Thus, we have shown that $\mathbf{x}^\circ$ is the solution that gives the minimum norm $\|\mathbf{x}\|$.

The matrix $\mathbf{A}^{RM} = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}$ that yields the minimum norm solution ($\|\mathbf{x}^\circ\|$ = minimum) is called the *right pseudoinverse* or *minimal right inverse* of $\mathbf{A}$.

Summary On the Right Pseudoinverse Matrix. The right pseudoinverse $\mathbf{A}^{RM}$ gives the solution $\mathbf{x}^\circ = \mathbf{A}^{RM}\mathbf{b}$ that minimizes the norm, or gives $\|\mathbf{x}^\circ\|$ = minimum. Note that the right pseudoinverse $\mathbf{A}^{RM}$ is an $m \times n$ matrix, since $\mathbf{A}$ is an $n \times m$ matrix and

$$\begin{aligned}\mathbf{A}^{RM} &= \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} \\ &= (m \times n \text{ matrix})(n \times n \text{ matrix})^{-1} \\ &= m \times n \text{ matrix}, \qquad m > n\end{aligned}$$

Notice that the dimension of $\mathbf{A}\mathbf{A}^T$ is smaller than the dimension of vector $\mathbf{x}$, which is m. Notice also that the right pseudoinverse $\mathbf{A}^{RM}$ possesses the property that it is indeed an "inverse" matrix if premultiplied by $\mathbf{A}$:

$$\mathbf{A}\mathbf{A}^{RM} = \mathbf{A}[\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}] = \mathbf{A}\mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1} = \mathbf{I}_n$$

Solution That Minimizes $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$. Consider a vector-matrix equation

$$\mathbf{A}\mathbf{x} = \mathbf{b} \tag{A–36}$$

where $\mathbf{A}$ is an $n \times m$ matrix, $\mathbf{x}$ is an m-vector, and $\mathbf{b}$ is an n-vector. Here we assume that $n > m$. That is, the number of unknown variables is smaller than the number of equations. In the classical sense, there may or may not exist any solution.

If no solution exists, we may wish to find a unique "solution" that minimizes the norm $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$. Let us define a "solution" to Equation (A–36) that will minimize $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ as $\mathbf{x}^\circ$. In other words, $\mathbf{x}^\circ$ satisfies the condition

$$\|\mathbf{A}\mathbf{x} - \mathbf{b}\| \geq \|\mathbf{A}\mathbf{x}^\circ - \mathbf{b}\|, \qquad \text{for all } \mathbf{x}$$

Note that $\mathbf{x}^\circ$ is not a solution in the classical sense, since it does not satisfy the original vector-matrix equation $\mathbf{A}\mathbf{x} = \mathbf{b}$. Therefore, we may call $\mathbf{x}^\circ$ an "approximate solution," in that it minimizes norm $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$. We shall obtain such an approximate solution in the following.

Left Pseudoinverse Matrix. For a vector-matrix equation

$$\mathbf{A}\mathbf{x} = \mathbf{b}$$

where $\mathbf{A}$ is an $n \times m$ matrix having rank m, $\mathbf{x}$ is an m-vector, and $\mathbf{b}$ is an n-vector, the vector $\mathbf{x}^\circ$ that minimizes the norm $\|\mathbf{A}\mathbf{x} - \mathbf{b}\|$ is given by

$$\mathbf{x}^\circ = \mathbf{A}^{LM}\mathbf{b} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

where $\mathbf{A}^{LM} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$.

To verify this, first note that

$$\begin{aligned}\|\mathbf{A}\mathbf{x} - \mathbf{b}\| &= \|\mathbf{A}(\mathbf{x} - \mathbf{x}^\circ) + \mathbf{A}\mathbf{x}^\circ - \mathbf{b}\| \\ &= \|\mathbf{A}(\mathbf{x} - \mathbf{x}^\circ)\| + \|\mathbf{A}\mathbf{x}^\circ - \mathbf{b}\| + 2[\mathbf{A}(\mathbf{x} - \mathbf{x}^\circ)]^T(\mathbf{A}\mathbf{x}^\circ - \mathbf{b})\end{aligned}$$

The last term can be shown to be zero as follows:

$$
\begin{aligned}
[\mathbf{A}(\mathbf{x}-\mathbf{x}^\circ)]^T(\mathbf{A}\mathbf{x}^\circ-\mathbf{b}) &= (\mathbf{x}-\mathbf{x}^\circ)^T\mathbf{A}^T[\mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T-\mathbf{I}_n]\mathbf{b}\\
&= (\mathbf{x}-\mathbf{x}^\circ)^T[(\mathbf{A}^T\mathbf{A})(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T-\mathbf{A}^T]\mathbf{b}\\
&= (\mathbf{x}-\mathbf{x}^\circ)^T(\mathbf{A}^T-\mathbf{A}^T)\mathbf{b}\\
&= 0
\end{aligned}
$$

Hence,

$$\|\mathbf{A}\mathbf{x}-\mathbf{b}\| = \|\mathbf{A}(\mathbf{x}-\mathbf{x}^\circ)\| + \|\mathbf{A}\mathbf{x}^\circ-\mathbf{b}\|$$

Noting that $\|\mathbf{A}(\mathbf{x}-\mathbf{x}^\circ)\| \geq 0$, we obtain

$$\|\mathbf{A}\mathbf{x}-\mathbf{b}\| - \|\mathbf{A}\mathbf{x}^\circ-\mathbf{b}\| = \|\mathbf{A}(\mathbf{x}-\mathbf{x}^\circ)\| \geq 0$$

or

$$\|\mathbf{A}\mathbf{x}-\mathbf{b}\| \geq \|\mathbf{A}\mathbf{x}^\circ-\mathbf{b}\|$$

Thus,

$$\mathbf{x}^\circ = \mathbf{A}^{LM}\mathbf{b} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}$$

minimizes $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|$.

The matrix $\mathbf{A}^{LM} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is called the *left pseudoinverse* or *minimal left inverse* of matrix $\mathbf{A}$. Note that $\mathbf{A}^{LM}$ is indeed the inverse matrix of $\mathbf{A}$, in that if postmultiplied by $\mathbf{A}$ it will give an identity matrix $\mathbf{I}_m$:

$$\mathbf{A}^{LM}\mathbf{A} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{A} = (\mathbf{A}^T\mathbf{A})^{-1}(\mathbf{A}^T\mathbf{A}) = \mathbf{I}_m$$

EXAMPLE PROBLEMS AND SOLUTIONS

Problem A–1

Show that if matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, respectively, and if $|\mathbf{A}| \neq 0$ and $|\mathbf{D}| \neq 0$, then

$$\begin{vmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{0} & \mathbf{D}\end{vmatrix} = \begin{vmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{C} & \mathbf{D}\end{vmatrix} = |\mathbf{A}|\,|\mathbf{D}| \neq 0, \qquad \text{if } |\mathbf{A}| \neq 0 \text{ and } |\mathbf{D}| \neq 0$$

Solution Since matrix $\mathbf{A}$ is nonsingular, we have

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{0} & \mathbf{D}\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{A}^{-1}\mathbf{B}\\ \mathbf{0} & \mathbf{I}\end{bmatrix}$$

Hence,

$$\begin{vmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{0} & \mathbf{D}\end{vmatrix} = \begin{vmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{vmatrix}\begin{vmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{vmatrix}\begin{vmatrix}\mathbf{I} & \mathbf{A}^{-1}\mathbf{B}\\ \mathbf{0} & \mathbf{I}\end{vmatrix} = |\mathbf{A}|\,|\mathbf{D}|$$

Similarly, since $\mathbf{D}$ is nonsingular, we get

$$\begin{vmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{C} & \mathbf{D}\end{vmatrix} = \begin{vmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}\end{vmatrix}\begin{vmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{0} & \mathbf{D}\end{vmatrix}\begin{vmatrix}\mathbf{I} & \mathbf{0}\\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}\end{vmatrix} = |\mathbf{A}|\,|\mathbf{D}|$$

Problem A–2

Show that if matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$ are an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, respectively, then

$$\begin{vmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{vmatrix} = \begin{cases}|\mathbf{A}|\,|\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}|, & \text{if } |\mathbf{A}| \neq 0\\ |\mathbf{D}|\,|\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}|, & \text{if } |\mathbf{D}| \neq 0\end{cases}$$

Solution if $|\mathbf{A}| \neq 0$, the matrix

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix}$$

can be written as a product of two matrices:

$$\begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{C} & \mathbf{I}_m\end{bmatrix} \quad \text{and} \quad \begin{bmatrix}\mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B}\\ \mathbf{0} & \mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\end{bmatrix}$$

or

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \begin{bmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{C} & \mathbf{I}_m\end{bmatrix}\begin{bmatrix}\mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B}\\ \mathbf{0} & \mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\end{bmatrix}$$

Hence,

$$\begin{aligned}
\begin{vmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{vmatrix} &= \begin{vmatrix}\mathbf{A} & \mathbf{0}\\ \mathbf{C} & \mathbf{I}_m\end{vmatrix}\begin{vmatrix}\mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B}\\ \mathbf{0} & \mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}\end{vmatrix}\\
&= |\mathbf{A}|\,|\mathbf{I}_m|\,|\mathbf{I}_n|\,|\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}|\\
&= |\mathbf{A}|\,|\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}|
\end{aligned}$$

Similarly, if $|\mathbf{D}| \neq 0$, then

$$\begin{bmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{bmatrix} = \begin{bmatrix}\mathbf{I}_n & \mathbf{B}\\ \mathbf{0} & \mathbf{D}\end{bmatrix}\begin{bmatrix}\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0}\\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m\end{bmatrix}$$

and therefore

$$\begin{aligned}
\begin{vmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{vmatrix} &= \begin{vmatrix}\mathbf{I}_n & \mathbf{B}\\ \mathbf{0} & \mathbf{D}\end{vmatrix}\begin{vmatrix}\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C} & \mathbf{0}\\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m\end{vmatrix}\\
&= |\mathbf{I}_n|\,|\mathbf{D}|\,|\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}|\,|\mathbf{I}_m|\\
&= |\mathbf{D}|\,|\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}|
\end{aligned}$$

Problem A–3

For an $n \times m$ matrix $\mathbf{A}$ and an $m \times n$ matrix $\mathbf{B}$, show that

$$|\mathbf{I}_n + \mathbf{A}\mathbf{B}| = |\mathbf{I}_m + \mathbf{B}\mathbf{A}|$$

Solution Consider the following matrix:

$$\begin{bmatrix}\mathbf{I}_n & -\mathbf{A}\\ \mathbf{B} & \mathbf{I}_m\end{bmatrix}$$

Referring to Problem A–2,

$$\begin{vmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{C} & \mathbf{D}\end{vmatrix} = \begin{cases}|\mathbf{A}|\,|\mathbf{D}-\mathbf{C}\mathbf{A}^{-1}\mathbf{B}|, & \text{if } |\mathbf{A}| \neq 0\\ |\mathbf{D}|\,|\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}|, & \text{if } |\mathbf{D}| \neq 0\end{cases}$$

Hence,

$$\begin{vmatrix} \mathbf{I}_n & -\mathbf{A} \\ \mathbf{B} & \mathbf{I}_m \end{vmatrix} = \begin{cases} |\mathbf{I}_n|\,|\mathbf{I}_m + \mathbf{BA}| = |\mathbf{I}_m + \mathbf{BA}| \\ |\mathbf{I}_m|\,|\mathbf{I}_n + \mathbf{AB}| = |\mathbf{I}_n + \mathbf{AB}| \end{cases}$$

and we have

$$|\mathbf{I}_n + \mathbf{AB}| = |\mathbf{I}_m + \mathbf{BA}|$$

Problem A–4

If A, B, C, and D are, respectively, an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, then we have the following matrix inversion lemma:

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

where we assume the indicated inverses to exist. Prove this matrix inversion lemma.

Solution Let us premultiply both sides of the equation by (A + BDC):

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

or

$$\begin{aligned} \mathbf{I} &= \mathbf{I} + \mathbf{BDCA}^{-1} - \mathbf{B}(\mathbf{D}^{-1} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} - \mathbf{BDCA}^{-1}\mathbf{B}(\mathbf{D}^{-1} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \\ &= \mathbf{I} + \mathbf{BDCA}^{-1} - (\mathbf{B} + \mathbf{BDCA}^{-1}\mathbf{B})(\mathbf{D}^{-1} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \\ &= \mathbf{I} + \mathbf{BDCA}^{-1} - \mathbf{BD}(\mathbf{D}^{-1} + \mathbf{CA}^{-1}\mathbf{B})(\mathbf{D}^{-1} + \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} \\ &= \mathbf{I} + \mathbf{BDCA}^{-1} - \mathbf{BDCA}^{-1} \\ &= \mathbf{I} \end{aligned}$$

Hence, we have proved the matrix inversion lemma.

Problem A–5

Prove that if A, B, C, and D are, respectively, an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, then

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \tag{A–37}$$

provided $|\mathbf{A}| \neq 0$ and $|\mathbf{D}| \neq 0$.

Prove also that

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{CA}^{-1} & \mathbf{D}^{-1} \end{bmatrix} \tag{A–38}$$

provided $|\mathbf{A}| \neq 0$ and $|\mathbf{D}| \neq 0$.

Solution Note that

$$\begin{bmatrix} \mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B} - \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

Hence, Equation (A–37) is proved. Similarly,

$$\begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{CA}^{-1} & \mathbf{D}^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C} + \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m \end{bmatrix}$$

Hence, we have proved Equation (A–38).

Problem A–6

Prove that if A, B, C, and D are, respectively, an $n \times n$, an $n \times m$, an $m \times n$, and an $m \times m$ matrix, then

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{bmatrix}$$

provided $|\mathbf{A}| \neq 0$ and $|\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| \neq 0$.

Prove also that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} + \mathbf{D}^{-1} \end{bmatrix}$$

provided $|\mathbf{D}| \neq 0$ and $|\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}| \neq 0$.

Solution First, note that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_m \end{bmatrix}\begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{bmatrix} \tag{A–39}$$

By taking the inverse of both sides of Equation (A–39), we obtain

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_m \end{bmatrix}^{-1}$$

By referring to Problem A–5, we find

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{I}_n & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ \mathbf{0} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{bmatrix}$$

and

$$\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_m \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\mathbf{CA}^{-1} & \mathbf{I}_m \end{bmatrix}$$

Hence,

$$\begin{aligned} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{I}_n & \mathbf{A}^{-1}\mathbf{B} \\ \mathbf{0} & \mathbf{D} - \mathbf{CA}^{-1}\mathbf{B} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{A} & \mathbf{0} \\ \mathbf{C} & \mathbf{I}_m \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \mathbf{I}_n & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ \mathbf{0} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{bmatrix}\begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ -\mathbf{CA}^{-1} & \mathbf{I}_m \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{bmatrix} \end{aligned}$$

provided $|\mathbf{A}| \neq 0$ and $|\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B}| \neq 0$.

Similarly, notice that

$$\begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_n & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}\begin{bmatrix} \mathbf{A} - \mathbf{BD}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix} \tag{A–40}$$

By taking the inverse of both sides of Equation (A–40) and referring to Problem A–5, we obtain

$$\begin{aligned} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{bmatrix}^{-1} &= \begin{bmatrix} \mathbf{A} - \mathbf{BD}^{-1}\mathbf{C} & \mathbf{0} \\ \mathbf{D}^{-1}\mathbf{C} & \mathbf{I}_m \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{I}_n & \mathbf{B} \\ \mathbf{0} & \mathbf{D} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & \mathbf{0} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & \mathbf{I}_m \end{bmatrix}\begin{bmatrix} \mathbf{I}_n & -\mathbf{BD}^{-1} \\ \mathbf{0} & \mathbf{D}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} + \mathbf{D}^{-1} \end{bmatrix} \end{aligned}$$

provided $|\mathbf{D}| \neq 0$ and $|\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C}| \neq 0$.

Problem A–7

For an $n \times n$ real matrix $\mathbf{A}$ and real n-vectors $\mathbf{x}$ and $\mathbf{y}$, show that

(a) $$\frac{\partial}{\partial \mathbf{x}}\mathbf{y}^T\mathbf{x} = \mathbf{y}$$

(b) $$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x}$$

For an $n \times n$ Hermitian matrix $\mathbf{A}$ and a complex n-vector $\mathbf{x}$, show that

(c) $$\frac{\partial}{\partial \bar{\mathbf{x}}}\mathbf{x}^*\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x}$$

Solution

(a) Note that

$$\mathbf{y}^T\mathbf{x} = y_1x_1 + y_2x_2 + \cdots + y_nx_n$$

which is a scalar quantity. Hence,

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{y}^T\mathbf{x} = \begin{bmatrix} \dfrac{\partial}{\partial x_1}\mathbf{y}^T\mathbf{x} \\ \vdots \\ \dfrac{\partial}{\partial x_n}\mathbf{y}^T\mathbf{x} \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \mathbf{y}$$

(b) Notice that

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_ix_j$$

which is a scalar quantity. Hence,

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} = \begin{bmatrix} \dfrac{\partial}{\partial x_1}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}x_ix_j\right) \\ \vdots \\ \dfrac{\partial}{\partial x_n}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}x_ix_j\right) \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1}^{n} a_{1j}x_j + \sum\limits_{i=1}^{n} a_{i1}x_i \\ \vdots \\ \sum\limits_{j=1}^{n} a_{nj}x_j + \sum\limits_{i=1}^{n} a_{in}x_i \end{bmatrix}$$

$$= \mathbf{A}\mathbf{x} + \mathbf{A}^T\mathbf{x}$$

which is Equation (A–20)

If matrix $\mathbf{A}$ is a real symmetric matrix, then

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{x} = 2\mathbf{A}\mathbf{x}, \qquad \text{if } \mathbf{A} = \mathbf{A}^T$$

(c) For a Hermitian matrix $\mathbf{A}$, we have

$$\mathbf{x}^*\mathbf{A}\mathbf{x} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\bar{x}_ix_j$$

and

$$\frac{\partial}{\partial \bar{\mathbf{x}}}\mathbf{x}^*\mathbf{A}\mathbf{x} = \begin{bmatrix} \dfrac{\partial}{\partial \bar{x}_1}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}\bar{x}_ix_j\right) \\ \vdots \\ \dfrac{\partial}{\partial \bar{x}_n}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}\bar{x}_ix_j\right) \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1}^{n} a_{1j}x_j \\ \vdots \\ \sum\limits_{j=1}^{n} a_{nj}x_j \end{bmatrix} = \mathbf{A}\mathbf{x}$$

which is Equation (A–21).

Note that

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^*\mathbf{A}\mathbf{x} = \begin{bmatrix} \dfrac{\partial}{\partial x_1}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}\bar{x}_ix_j\right) \\ \vdots \\ \dfrac{\partial}{\partial x_n}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{n} a_{ij}\bar{x}_ix_j\right) \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{n} a_{i1}\bar{x}_i \\ \vdots \\ \sum\limits_{i=1}^{n} a_{in}\bar{x}_i \end{bmatrix} = \mathbf{A}^T\bar{\mathbf{x}}$$

Therefore,

$$\overline{\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^*\mathbf{A}\mathbf{x}} = \mathbf{A}^*\mathbf{x} = \mathbf{A}\mathbf{x}$$

Problem A–8

For an $n \times m$ complex matrix $\mathbf{A}$, a complex n-vector $\mathbf{x}$, and a complex m-vector $\mathbf{y}$, show that

(a) $$\frac{\partial}{\partial \bar{\mathbf{x}}}\mathbf{x}^*\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{y}$$

(b) $$\frac{\partial}{\partial \bar{\mathbf{y}}}\mathbf{x}^*\mathbf{A}\mathbf{y} = \mathbf{A}^T\bar{\mathbf{x}}$$

Solution

(a) Notice that

$$\mathbf{x}^*\mathbf{A}\mathbf{y} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ij}\bar{x}_iy_j$$

Hence,

$$\frac{\partial}{\partial \bar{\mathbf{x}}}\mathbf{x}^*\mathbf{A}\mathbf{y} = \begin{bmatrix} \dfrac{\partial}{\partial \bar{x}_1}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} a_{ij}\bar{x}_iy_j\right) \\ \vdots \\ \dfrac{\partial}{\partial \bar{x}_n}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} a_{ij}\bar{x}_iy_j\right) \end{bmatrix} = \begin{bmatrix} \sum\limits_{j=1}^{m} a_{1j}y_j \\ \vdots \\ \sum\limits_{j=1}^{m} a_{nj}y_j \end{bmatrix} = \mathbf{A}\mathbf{y}$$

which is Equation (A–24).

(b) Notice that

$$\frac{\partial}{\partial \mathbf{y}}\mathbf{x}^*\mathbf{A}\mathbf{y} = \begin{bmatrix} \dfrac{\partial}{\partial y_1}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} a_{ij}\bar{x}_iy_j\right) \\ \vdots \\ \dfrac{\partial}{\partial y_m}\left(\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m} a_{ij}\bar{x}_iy_j\right) \end{bmatrix} = \begin{bmatrix} \sum\limits_{i=1}^{n} a_{i1}\bar{x}_i \\ \vdots \\ \sum\limits_{i=1}^{n} a_{im}\bar{x}_i \end{bmatrix} = \mathbf{A}^T\bar{\mathbf{x}}$$

which is Equation (A–25).

Similarly, for an $n \times m$ real matrix $\mathbf{A}$, a real n-vector $\mathbf{x}$, and a real m-vector $\mathbf{y}$, we have

$$\frac{\partial}{\partial \mathbf{x}}\mathbf{x}^T\mathbf{A}\mathbf{y} = \mathbf{A}\mathbf{y}, \qquad \frac{\partial}{\partial \mathbf{y}}\mathbf{x}^T\mathbf{A}\mathbf{y} = \mathbf{A}^T\mathbf{x}$$

which are Equations (A–22) and (A–23), respectively.

Problem A–9

Given two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$, prove that the eigenvalues of $\mathbf{AB}$ and those of $\mathbf{BA}$ are the same, even if $\mathbf{AB} \neq \mathbf{BA}$.

Solution First, we shall consider the case where $\mathbf{A}$ (or $\mathbf{B}$) is nonsingular. In this case,

$$|\lambda\mathbf{I} - \mathbf{BA}| = |\lambda\mathbf{I} - \mathbf{A}^{-1}(\mathbf{AB})\mathbf{A}| = |\mathbf{A}^{-1}(\lambda\mathbf{I} - \mathbf{AB})\mathbf{A}| = |\mathbf{A}^{-1}||\lambda\mathbf{I} - \mathbf{AB}||\mathbf{A}| = |\lambda\mathbf{I} - \mathbf{AB}|$$

Next we shall consider the case where both $\mathbf{A}$ and $\mathbf{B}$ are singular. There exist $n \times n$ nonsingular matrices $\mathbf{P}$ and $\mathbf{Q}$ such that

$$\mathbf{PAQ} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

where $\mathbf{I}_r$ is the $r \times r$ identity matrix and r is the rank of $\mathbf{A}$, $r < n$. We have

$$|\lambda\mathbf{I} - \mathbf{BA}| = |\lambda\mathbf{I} - \mathbf{Q}^{-1}\mathbf{BAQ}| = |\lambda\mathbf{I} - \mathbf{Q}^{-1}\mathbf{BP}^{-1}\mathbf{PAQ}|$$
$$= \left|\lambda\mathbf{I} - \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix}\begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right|$$

where

$$\mathbf{Q}^{-1}\mathbf{BP}^{-1} = \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix}$$

Then

$$|\lambda\mathbf{I} - \mathbf{BA}| = \left|\lambda\mathbf{I} - \begin{bmatrix} \mathbf{G}_{11} & \mathbf{0} \\ \mathbf{G}_{21} & \mathbf{0} \end{bmatrix}\right| = \begin{vmatrix} \lambda\mathbf{I}_r - \mathbf{G}_{11} & \mathbf{0} \\ -\mathbf{G}_{21} & \lambda\mathbf{I}_{n-r} \end{vmatrix}$$
$$= |\lambda\mathbf{I}_r - \mathbf{G}_{11}||\lambda\mathbf{I}_{n-r}|$$

Also,

$$|\lambda\mathbf{I} - \mathbf{AB}| = |\lambda\mathbf{I} - \mathbf{PABP}^{-1}| = |\lambda\mathbf{I} - \mathbf{PAQQ}^{-1}\mathbf{BP}^{-1}|$$
$$= \left|\lambda\mathbf{I} - \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix}\right|$$
$$= \left|\lambda\mathbf{I} - \begin{bmatrix} \mathbf{G}_{11} & \mathbf{G}_{12} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}\right|$$
$$= \begin{vmatrix} \lambda\mathbf{I}_r - \mathbf{G}_{11} & -\mathbf{G}_{12} \\ \mathbf{0} & \lambda\mathbf{I}_{n-r} \end{vmatrix}$$
$$= |\lambda\mathbf{I}_r - \mathbf{G}_{11}||\lambda\mathbf{I}_{n-r}|$$

Hence, we have proved that

$$|\lambda\mathbf{I} - \mathbf{BA}| = |\lambda\mathbf{I} - \mathbf{AB}|$$

or that the eigenvalues of $\mathbf{AB}$ and $\mathbf{BA}$ are the same regardless of whether $\mathbf{AB} = \mathbf{BA}$ or $\mathbf{AB} \neq \mathbf{BA}$.

Problem A–10

Show that the following 2×2 matrix $\mathbf{A}$ has two distinct eigenvalues and that the eigenvectors are linearly independent of each other:

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

Then normalize the eigenvectors.

Solution The eigenvalues are obtained from

$$|\lambda\mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda - 1 & -1 \\ 0 & \lambda - 2 \end{vmatrix} = (\lambda - 1)(\lambda - 2) = 0$$

as

$$\lambda_1 = 1 \quad \text{and} \quad \lambda_2 = 2$$

Thus, matrix $\mathbf{A}$ has two distinct eigenvalues.

There are two eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$ associated with λ_1 and λ_2, respectively. If we define

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix}$$

then the eigenvector $\mathbf{x}_1$ can be found from

$$\mathbf{Ax}_1 = \lambda_1\mathbf{x}_1$$

or

$$(\lambda_1\mathbf{I} - \mathbf{A})\mathbf{x}_1 = \mathbf{0}$$

Noting that $\lambda_1 = 1$, we have

$$\begin{bmatrix} 1-1 & -1 \\ 0 & 1-2 \end{bmatrix}\begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which gives

$$x_{11} = \text{arbitrary constant} \quad \text{and} \quad x_{21} = 0$$

Hence, eigenvector $\mathbf{x}_1$ may be written as

$$\mathbf{x}_1 = \begin{bmatrix} x_{11} \\ x_{21} \end{bmatrix} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix}$$

where $c_1 \neq 0$ is an arbitrary constant.

Similarly, for the eigenvector $\mathbf{x}_2$, we have

$$\mathbf{Ax}_2 = \lambda_2\mathbf{x}_2$$

or

$$(\lambda_2\mathbf{I} - \mathbf{A})\mathbf{x}_2 = \mathbf{0}$$

Noting that $\lambda_2 = 2$, we obtain

$$\begin{bmatrix} 2-1 & -1 \\ 0 & 2-2 \end{bmatrix}\begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

from which we get

$$x_{12} - x_{22} = 0$$

Hence, the eigenvector associated with $\lambda_2 = 2$ may be selected as

$$\mathbf{x}_2 = \begin{bmatrix} x_{12} \\ x_{22} \end{bmatrix} = \begin{bmatrix} c_2 \\ c_2 \end{bmatrix}$$

where $c_2 \neq 0$ is an arbitrary constant.

The two eigenvectors are therefore given by

$$\mathbf{x}_1 = \begin{bmatrix} c_1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} c_2 \\ c_2 \end{bmatrix}$$

The fact that eigenvectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are linearly independent can be seen from the fact that the determinant of the matrix $[\mathbf{x}_1\ \mathbf{x}_2]$ is nonzero:

$$\begin{vmatrix} c_1 & c_2 \\ 0 & c_2 \end{vmatrix} \neq 0$$

To normalize the eigenvectors, we choose $c_1 = 1$ and $c_2 = 1/\sqrt{2}$, or

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \mathbf{x}_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

Clearly, the absolute value of each eigenvector becomes unity and therefore the eigenvectors are normalized.

Problem A–11

Obtain a transformation matrix $\mathbf{T}$ that transforms the matrix

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

into a Jordan canonical form.

Solution The characteristic equation is

$$|\lambda\mathbf{I} - \mathbf{A}| = \left|\begin{array}{cc|cc} \lambda & -1 & 0 & -3 \\ 0 & \lambda+1 & -1 & -1 \\ \hline 0 & 0 & \lambda & -1 \\ 0 & 0 & 1 & \lambda+2 \end{array}\right| = \begin{vmatrix} \lambda & -1 \\ 0 & \lambda+1 \end{vmatrix} \begin{vmatrix} \lambda & -1 \\ 1 & \lambda+2 \end{vmatrix}$$

$$= (\lambda+1)^3\lambda = 0$$

Hence, matrix $\mathbf{A}$ involves eigenvalues

$$\lambda_1 = -1, \qquad \lambda_2 = -1, \qquad \lambda_3 = -1, \qquad \lambda_4 = 0$$

For the multiple eigenvalue -1, we have

$$\lambda_1\mathbf{I} - \mathbf{A} = \begin{bmatrix} -1 & -1 & 0 & -3 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & -1 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

which is of rank 2, or rank $(4 - 2)$. From the rank condition we see that there must be two Jordan blocks for eigenvalue -1, that is, one $p_1 \times p_1$ Jordan block and one $p_2 \times p_2$ Jordan block, where $p_1 + p_2 = 3$. Notice that for $p_1 + p_2 = 3$ there is only one combination (2 and 1) for the orders of p_1 and p_2. Let us choose

$$p_1 = 2 \qquad \text{and} \qquad p_2 = 1$$

Then there are one eigenvector and one generalized eigenvector for Jordan block $\mathbf{J}_{p1}$ and one eigenvector for Jordan block $\mathbf{J}_{p2}$.

Let us define an eigenvector and a generalized eigenvector for Jordan block $\mathbf{J}_{p1}$ as $\mathbf{v}_{11}$ and $\mathbf{v}_{12}$, respectively, and an eigenvector for Jordan block $\mathbf{J}_{p2}$ as $\mathbf{v}_{21}$. Then there must be vectors $\mathbf{v}_{11}$, $\mathbf{v}_{12}$, and $\mathbf{v}_{21}$ that satisfy the following equations:

$$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{11} = \mathbf{0}, \qquad (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{21} = \mathbf{0}$$

$$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{12} = \mathbf{v}_{11}$$

For $\lambda_1 = -1$, $\mathbf{A} - \lambda_1\mathbf{I}$ can be given as follows:

$$\mathbf{A} - \lambda_1\mathbf{I} = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -1 & -1 \end{bmatrix}$$

Noting that

$$(\mathbf{A} - \lambda_1\mathbf{I})^2 = \begin{bmatrix} 1 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

we determine vector $\mathbf{v}_{12}$ to be such that it will satisfy the equation

$$(\mathbf{A} - \lambda_1\mathbf{I})^2\mathbf{v}_{12} = \mathbf{0}$$

and at the same time will make $(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{12}$ nonzero. An example of such a generalized eigenvector $\mathbf{v}_{12}$ can be found to be

$$\mathbf{v}_{12} = \begin{bmatrix} -a \\ 0 \\ 0 \\ a \end{bmatrix}, \qquad a = \text{arbitrary nonzero constant}$$

The eigenvector $\mathbf{v}_{11}$ is then found to be a nonzero vector $(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{12}$:

$$\mathbf{v}_{11} = (\mathbf{A} - \lambda_1\mathbf{I})\mathbf{v}_{12} = \begin{bmatrix} 2a \\ a \\ a \\ -a \end{bmatrix}$$

Since a is an arbitrary nonzero constant, let us choose $a = 1$. Then we have

$$\mathbf{v}_{11} = \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \end{bmatrix} \qquad \text{and} \qquad \mathbf{v}_{12} = \begin{bmatrix} -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Next, we determine $\mathbf{v}_{21}$ so that $\mathbf{v}_{21}$ and $\mathbf{v}_{11}$ are linearly independent. For $\mathbf{v}_{21}$ we may choose

$$\mathbf{v}_{21} = \begin{bmatrix} b + 3c \\ -b \\ c \\ -c \end{bmatrix}$$

where b and c are arbitrary constants. Let us choose, for example, $b = 1$ and $c = 0$. Then

$$\mathbf{v}_{21} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$$

Clearly, $\mathbf{v}_{11}$, $\mathbf{v}_{12}$, and $\mathbf{v}_{21}$ are linearly independent. Let us define

$$\mathbf{v}_{11} = \mathbf{x}_1, \qquad \mathbf{v}_{12} = \mathbf{x}_2, \qquad \mathbf{v}_{21} = \mathbf{x}_3$$

and

$$T(\lambda_1) = [v_{11} \vdots v_{12} \vdots v_{21}] = [x_1 \vdots x_2 \vdots x_3] = \begin{bmatrix} 2 & -1 & 1 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ -1 & 1 & 0 \end{bmatrix}$$

For the distinct eigenvalue $\lambda_4 = 0$, the eigenvector x_4 can be determined from

$$(A - \lambda_4 I)x_4 = 0$$

Noting that

$$A - \lambda_4 I = A = \begin{bmatrix} 0 & 1 & 0 & 3 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & -2 \end{bmatrix}$$

we find

$$x_4 = \begin{bmatrix} d \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

where $d \neq 0$ is an arbitrary constant. By choosing $d = 1$, we have

$$T(\lambda_4) = x_4 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Thus, the transformation matrix T can be written as

$$T = [T(\lambda_1) \vdots T(\lambda_4)] = \begin{bmatrix} 2 & -1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix}$$

Then

$$T^{-1}AT = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & -1 & 1 & 0 \\ 1 & 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 3 \\ 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & -2 \end{bmatrix} \begin{bmatrix} 2 & -1 & 1 & 1 \\ 1 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \end{bmatrix}$$

$$= \left[\begin{array}{cc:cc} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ \hdashline 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] = \text{diag}[J_2(-1), J_1(-1), J_1(0)]$$

Problem A–12

Assume that an $n \times n$ normal matrix A has a k-multiple eigenvalue λ_1. Prove that the rank of $A - \lambda_1 I$ is $n - k$.

Solution Suppose that the rank of $A - \lambda_1 I$ is $n - m$. Then the equation

$$(A - \lambda_1 I)x = 0 \qquad \text{(A–41)}$$

will have m linearly independent vector solutions. Let us choose m such vectors so that they are orthogonal to each other and normalized. That is, vectors $x_1, x_2, \ldots, x_m$ will satisfy Equation (A–41) and will be orthonormal.

Let us consider $n - m$ vectors $x_{m+1}, x_{m+2}, \ldots, x_n$ such that all n vectors

$$x_1, x_2, \ldots, x_n$$

will be orthonormal to each other. Then matrix U, defined by

$$U = [x_1 \vdots x_2 \vdots \cdots \vdots x_n]$$

is a unitary matrix.

Since for $1 \leq i \leq m$, we have

$$Ax_i = \lambda_1 x_i$$

and therefore we can write

$$AU = U\begin{bmatrix} \lambda_1 I_m & B \\ 0 & C \end{bmatrix}$$

or

$$U^*AU = \begin{bmatrix} \lambda_1 I_m & B \\ 0 & C \end{bmatrix}$$

Noting that

$$\begin{aligned} \|Ax_i - \lambda x_i\|^2 &= \langle (A - \lambda I)x_i, (A - \lambda I)x_i \rangle \\ &= \langle (A^* - \bar{\lambda} I)(A - \lambda I)x_i, x_i \rangle \\ &= \langle (A - \lambda I)(A^* - \bar{\lambda} I)x_i, x_i \rangle \\ &= \langle (A^* - \bar{\lambda} I)x_i, (A^* - \bar{\lambda} I)x_i \rangle \\ &= \|A^*x_i - \bar{\lambda} x_i\|^2 \\ &= 0 \end{aligned}$$

we have

$$A^*x_i = \bar{\lambda} x_i$$

Therefore, we can write

$$A^*U = U\begin{bmatrix} \bar{\lambda}_1 I_m & B_1 \\ 0 & C_1 \end{bmatrix}$$

or

$$U^*A^*U = \begin{bmatrix} \bar{\lambda}_1 I_m & B_1 \\ 0 & C_1 \end{bmatrix}$$

Hence,

$$\begin{bmatrix} \lambda_1 I_m & B \\ 0 & C \end{bmatrix} = U^*AU = (U^*A^*U)^* = \begin{bmatrix} \bar{\lambda}_1 I_m & B_1 \\ 0 & C_1 \end{bmatrix}^* = \begin{bmatrix} \lambda_1 I_m & 0 \\ B_1^* & C_1^* \end{bmatrix}$$

Comparing the left and right sides of this last equation, we obtain

$$B = 0$$

Hence, we get

$$\mathbf{A} = \mathbf{U}\begin{bmatrix} \lambda_1 \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}\mathbf{U}^*$$

Then

$$\mathbf{A} - \lambda\mathbf{I} = \mathbf{U}\begin{bmatrix} (\lambda_1 - \lambda)\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \lambda\mathbf{I}_{n-m} \end{bmatrix}\mathbf{U}^*$$

The determinant of this last equation is

$$|\mathbf{A} - \lambda\mathbf{I}| = (\lambda_1 - \lambda)^m |\mathbf{C} - \lambda\mathbf{I}_{n-m}| \tag{A-42}$$

On the other hand, we have

$$\operatorname{rank}(\mathbf{A} - \lambda_1\mathbf{I}) = n - m = \operatorname{rank}\left\{\mathbf{U}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \lambda_1\mathbf{I}_{n-m} \end{bmatrix}\mathbf{U}^*\right\}$$

$$= \operatorname{rank}\begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \lambda_1\mathbf{I}_{n-m} \end{bmatrix} = \operatorname{rank}(\mathbf{C} - \lambda_1\mathbf{I}_{n-m})$$

Hence, we conclude that the rank of $\mathbf{C} - \lambda_1\mathbf{I}_{n-m}$ is $n - m$. Consequently,

$$|\mathbf{C} - \lambda_1\mathbf{I}_{n-m}| \neq 0$$

and from Equation (A–42), λ_1 is shown to be the m-multiple eigenvalue of $|\mathbf{A} - \lambda\mathbf{I}| = 0$. Since λ_1 is the k-multiple eigenvalue of $\mathbf{A}$, we must have $m = k$. Therefore, the rank of $\mathbf{A} - \lambda_1\mathbf{I}$ is $n - k$.

Note that, since the rank of $\mathbf{A} - \lambda_1\mathbf{I}$ is $n - k$, the equation

$$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{x}_i = \mathbf{0}$$

will have k linearly independent eigenvectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$.

Problem A–13

Prove that the eigenvalues of an $n \times n$ Hermitian matrix and of an $n \times n$ real symmetric matrix are real. Prove also that the eigenvalues of a skew-Hermitian matrix and of a real skew-symmetric matrix are either zero or purely imaginary.

Solution Let us define any eigenvalue of an $n \times n$ Hermitian matrix $\mathbf{A}$ by $\lambda = \alpha + j\beta$. There exists a vector $\mathbf{x} \neq \mathbf{0}$ such that

$$\mathbf{A}\mathbf{x} = (\alpha + j\beta)\mathbf{x}$$

The conjugate transpose of this last equation is

$$\mathbf{x}^*\mathbf{A}^* = (\alpha - j\beta)\mathbf{x}^*$$

Since $\mathbf{A}$ is Hermitian $\mathbf{A}^* = \mathbf{A}$. Therefore, we obtain

$$\mathbf{x}^*\mathbf{A}\mathbf{x} = (\alpha - j\beta)\mathbf{x}^*\mathbf{x}$$

On the other hand, since $\mathbf{A}\mathbf{x} = (\alpha + j\beta)\mathbf{x}$, we have

$$\mathbf{x}^*\mathbf{A}\mathbf{x} = (\alpha + j\beta)\mathbf{x}^*\mathbf{x}$$

Hence, we obtain

$$[(\alpha - j\beta) - (\alpha + j\beta)]\mathbf{x}^*\mathbf{x} = 0$$

or

$$-2j\beta\mathbf{x}^*\mathbf{x} = 0$$

Since $\mathbf{x}^*\mathbf{x} \neq 0$ (for $\mathbf{x} \neq \mathbf{0}$), we conclude that

$$\beta = 0$$

This proves that any eigenvalue of an $n \times n$ Hermitian matrix $\mathbf{A}$ is real. It follows that the eigenvalues of a real symmetric matrix are also real, since it is Hermitian.

To prove the second half of the problem, notice that if $\mathbf{B}$ is skew-Hermitian, then $j\mathbf{B}$ is Hermitian. Hence, the eigenvalues of $j\mathbf{B}$ are real, which implies that the eigenvalues of $\mathbf{B}$ are either zero or purely imaginary.

The eigenvalues of a real skew-symmetric matrix are also either zero or purely imaginary, since a real skew-symmetric matrix is skew-Hermitian.

Note that, in the real skew-symmetric matrix, purely imaginary eigenvalues always occur in conjugate pairs, since the coefficients of the characteristic equation are real. Note also that an $n \times n$ real skew-symmetric matrix is singular if n is odd, since such a matrix must include at least one zero eigenvalue.

Problem A–14

Examine whether or not the following 3×3 matrix $\mathbf{A}$ is positive definite:

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

Solution We shall demonstrate three different ways to test the positive definiteness of matrix $\mathbf{A}$.

1. We may first apply Sylvester's criterion for positive definiteness of a quadratic form $\mathbf{x}^T\mathbf{A}\mathbf{x}$. For the given matrix $\mathbf{A}$, we have

$$2 > 0, \qquad \begin{vmatrix} 2 & 2 \\ 2 & 6 \end{vmatrix} > 0, \qquad \begin{vmatrix} 2 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 1 \end{vmatrix} > 0$$

   Thus, the successive principal minors are all positive. Hence, matrix $\mathbf{A}$ is positive definite.

2. We may examine the positive definiteness of $\mathbf{x}^T\mathbf{A}\mathbf{x}$. Since

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = [x_1 x_2 x_3]\begin{bmatrix} 2 & 2 & -1 \\ 2 & 6 & 0 \\ -1 & 0 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_2 \end{bmatrix}$$

$$= 2x_1^2 + 4x_1x_2 - 2x_1x_3 + 6x_2^2 + x_3^2$$

$$= (x_1 - x_3)^2 + (x_1 + 2x_2)^2 + 2x_2^2$$

   we find that $\mathbf{x}^T\mathbf{A}\mathbf{x}$ is positive except at the origin ($\mathbf{x} = \mathbf{0}$). Hence, we conclude that matrix $\mathbf{A}$ is positive definite.

3. We may examine the eigenvalues of matrix $\mathbf{A}$. Note that

$$|\lambda\mathbf{I} - \mathbf{A}| = \lambda^3 - 9\lambda^2 + 15\lambda - 2$$

$$= (\lambda - 2)(\lambda - 0.1459)(\lambda - 6.8541)$$

   Hence,

$$\lambda_1 = 2, \qquad \lambda_2 = 0.1459, \qquad \lambda_3 = 6.8541$$

   Since all eigenvalues are positive, we conclude that $\mathbf{A}$ is a positive definite matrix.

Problem A–15

Examine whether the following matrix $\mathbf{A}$ is positive semidefinite:

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

Solution In the positive semidefiniteness test, we need to examine the signs of all principal minors in addition to the sign of the determinant of the given matrix, which must be zero; that is, $|\mathbf{A}|$ must be equal to 0.

For the 3×3 matrix

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

there are six principal minors:

$$a_{11}, \quad a_{22}, \quad a_{33}, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \quad \begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix}$$

We need to examine the signs of all six principal minors and the sign of $|\mathbf{A}|$.

For the given matrix $\mathbf{A}$,

$$a_{11} = 1 > 0$$

$$a_{22} = 4 > 0$$

$$a_{33} = 0$$

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = \begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 0$$

$$\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} 4 & 2 \\ 2 & 0 \end{vmatrix} = -4 < 0$$

$$\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33} \end{vmatrix} = \begin{vmatrix} 1 & 1 \\ 1 & 0 \end{vmatrix} = -1 < 0$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 0 \end{vmatrix} = 0$$

Clearly, two of the principal minors are negative. Hence, we conclude that matrix $\mathbf{A}$ is not positive semidefinite.

It is important to note that, had we tested the signs of only the successive principal minors and the determinant of $\mathbf{A}$,

$$1 > 0, \quad \begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix} = 0, \quad |\mathbf{A}| = \begin{vmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 0 \end{vmatrix} = 0$$

we would have reached the wrong conclusion that matrix $\mathbf{A}$ is positive semidefinite.

In fact, for the given matrix $\mathbf{A}$,

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

$$= (\lambda - 5.8541)\lambda(\lambda + 0.8541)$$

and so the eigenvalues are

$$\lambda_1 = 5.8541, \quad \lambda_2 = 0, \quad \lambda_3 = -0.8541$$

For matrix $\mathbf{A}$ to be positive semidefinite, all eigenvalues must be nonnegative and at least one of them must be zero. Clearly, matrix $\mathbf{A}$ is an indefinite matrix.

Appendix B

z Transform Theory

B–1 INTRODUCTION

This appendix first presents useful theorems of the z transform theory that were not treated in Chapter 2. Then we discuss details of the inversion integral method for finding the inverse z transform. Finally, we present the modified z transform method. At the end of this appendix (in the Example Problems and Solutions section), we discuss some of the interesting problems dealing with the z transformation, not treated in Chapter 2.

B–2 USEFUL THEOREMS OF THE z TRANSFORM THEORY

In this section we present some of the useful theorems of the z transform theory that were not discussed in Chapter 2.

Complex Differentiation. In the region of convergence a power series in z may be differentiated with respect to z any number of times to get a convergent series. The derivatives of $X(z)$ converge in the same region as $X(z)$.

Consider

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-k}$$

which converges in a certain region in the z plane. Differentiating $X(z)$ with respect to z, we obtain

$$\frac{d}{dz}X(z) = \sum_{k=0}^{\infty} (-k)x(k)z^{-k-1}$$

Multiplying both sides of this last equation by $-z$ gives

$$-z\frac{d}{dz}X(z) = \sum_{k=0}^{\infty} kx(k)z^{-k} \tag{B–1}$$

Thus, we have

$$\mathcal{Z}[kx(k)] = -z\frac{d}{dz}X(z) \tag{B–2}$$

Similarly, by differentiating both sides of Equation (B–1) with respect to z, we have

$$\frac{d}{dz}\left[-z\frac{d}{dz}X(z)\right] = \sum_{k=0}^{\infty}(-k^2)x(k)z^{-k-1}$$

Multiplying both sides of this last equation by $-z$, we obtain

$$-z\frac{d}{dz}\left[-z\frac{d}{dz}X(z)\right] = \sum_{k=0}^{\infty}k^2x(k)z^{-k}$$

or

$$\mathcal{Z}[k^2x(k)] = \left(-z\frac{d}{dz}\right)^2 X(z)$$

The operation $\left(-z\frac{d}{dz}\right)^2$ implies that we apply the operator $-z\frac{d}{dz}$ twice. Similarly, by repeating this process we have

$$\mathcal{Z}[k^m x(k)] = \left(-z\frac{d}{dz}\right)^m X(z) \tag{B–3}$$

Such complex differentiation enables us to obtain new z transform pairs from the known z transform pairs.

Example B–1

The z transform of the unit-step sequence $1(k)$ is given by

$$\mathcal{Z}[1(k)] = \frac{1}{1-z^{-1}}$$

Obtain the z transform of the unit-ramp sequence $x(k)$, where

$$x(k) = k$$

by using the complex differentiation theorem.

$$\mathcal{Z}[x(k)] = \mathcal{Z}[k] = \mathcal{Z}[k\cdot 1(k)] = -z\frac{d}{dz}\left(\frac{1}{1-z^{-1}}\right) = \frac{z^{-1}}{(1-z^{-1})^2}$$

Complex Integration. Consider the sequence

$$g(k) = \frac{x(k)}{k}$$

where $x(k)/k$ is finite for $k = 0$. The z transform of $x(k)/k$ is given by

$$\mathcal{Z}\left[\frac{x(k)}{k}\right] = \int_z^{\infty}\frac{X(z_1)}{z_1}dz_1 + \lim_{k\to 0}\frac{x(k)}{k} \tag{B–4}$$

where $\mathcal{Z}[x(k)] = X(z)$.

To prove Equation (B–4), note that

$$\mathcal{Z}\left[\frac{x(k)}{k}\right] = G(z) = \sum_{k=0}^{\infty}\frac{x(k)}{k}z^{-k}$$

Differentiating this last equation with respect to z yields

$$\frac{d}{dz}G(z) = -\sum_{k=0}^{\infty}x(k)z^{-k-1} = -z^{-1}\sum_{k=0}^{\infty}x(k)z^{-k} = -\frac{X(z)}{z}$$

Integrating both sides of this last equation with respect to z from z to ∞ gives

$$\int_z^{\infty}\frac{d}{dz}G(z)\,dz = G(\infty) - G(z) = -\int_z^{\infty}\frac{X(z_1)}{z_1}dz_1$$

or

$$G(z) = \int_z^{\infty}\frac{X(z_1)}{z_1}dz_1 + G(\infty)$$

Noting that $G(\infty)$ is given by

$$G(\infty) = \lim_{z\to\infty}G(z) = g(0) = \lim_{k\to 0}\frac{x(k)}{k}$$

we have

$$\mathcal{Z}\left[\frac{x(k)}{k}\right] = \int_z^{\infty}\frac{X(z_1)}{z_1}dz_1 + \lim_{k\to 0}\frac{x(k)}{k}$$

Partial Differentiation Theorem. Consider a function $x(t,a)$ or $x(kT,a)$ that is z-transformable. Here a is a constant or an independent variable. Define the z transform of $x(t,a)$ or $x(kT,a)$ as $X(z,a)$. Thus,

$$\mathcal{Z}[x(t,a)] = \mathcal{Z}[x(kT,a)] = X(z,a)$$

The z transform of the partial derivative of $x(t,a)$ or $x(kT,a)$ with respect to a can be given by

$$\mathcal{Z}\left[\frac{\partial}{\partial a}x(t,a)\right] = \mathcal{Z}\left[\frac{\partial}{\partial a}x(kT,a)\right] = \frac{\partial}{\partial a}X(z,a) \tag{B–5}$$

This equation is called the partial differentiation theorem.

To prove this theorem, note that

$$\mathcal{Z}\left[\frac{\partial}{\partial a}x(t,a)\right] = \mathcal{Z}\left[\frac{\partial}{\partial a}x(kT,a)\right] = \sum_{k=0}^{\infty}\frac{\partial}{\partial a}x(kT,a)z^{-k}$$
$$= \frac{\partial}{\partial a}\sum_{k=0}^{\infty}x(kT,a)z^{-k} = \frac{\partial}{\partial a}X(z,a)$$

Example B–2

Consider

$$x(t,a) = t^2e^{-at}$$

Obtain the z transform of this function $x(t,a)$ by use of the partial differentiation theorem.

Notice that

$$\frac{\partial}{\partial a}(-te^{-at}) = t^2 e^{-at}$$

and

$$\mathcal{Z}[te^{-at}] = \frac{Te^{-aT}z^{-1}}{(1-e^{-aT}z^{-1})^2}$$

Then we have

$$\begin{aligned}\mathcal{Z}[x(t,a)] = \mathcal{Z}[t^2e^{-at}] &= \mathcal{Z}\left[\frac{\partial}{\partial a}(-te^{-at})\right] \\ &= \frac{\partial}{\partial a}\left[-\frac{Te^{-aT}z^{-1}}{(1-e^{-aT}z^{-1})^2}\right] \\ &= \frac{T^2e^{-aT}(1+e^{-aT}z^{-1})z^{-1}}{(1-e^{-aT}z^{-1})^3}\end{aligned}$$

Real Convolution Theorem. Consider the functions $x_1(t)$ and $x_2(t)$, where

$$x_1(t) = 0, \quad \text{for } t < 0$$
$$x_2(t) = 0, \quad \text{for } t < 0$$

Assume that $x_1(t)$ and $x_2(t)$ are z-transformable and their z transforms are $X_1(z)$ and $X_2(z)$, respectively. Then

$$X_1(z)X_2(z) = \mathcal{Z}\left[\sum_{h=0}^{k} x_1(hT)x_2(kT-hT)\right] \tag{B–6}$$

This equation is called the real convolution theorem.

To prove this theorem, notice that

$$\begin{aligned}\mathcal{Z}\left[\sum_{h=0}^{k} x_1(hT)x_2(kT-hT)\right] &= \sum_{k=0}^{\infty}\sum_{h=0}^{k} x_1(hT)x_2(kT-hT)z^{-k} \\ &= \sum_{k=0}^{\infty}\sum_{h=0}^{\infty} x_1(hT)x_2(kT-hT)z^{-k}\end{aligned}$$

where we used the condition that $x_2(kT-hT) = 0$ for $h > k$. Now define $m = k - h$. Then

$$\mathcal{Z}\left[\sum_{h=0}^{k} x_1(hT)x_2(kT-hT)\right] = \sum_{h=0}^{\infty} x_1(hT)z^{-h}\sum_{m=-h}^{\infty} x_2(mT)z^{-m}$$

Since $x_2(mT) = 0$ for $m < 0$, this last equation becomes

$$\mathcal{Z}\left[\sum_{h=0}^{k} x_1(hT)x_2(kT-hT)\right] = \sum_{h=0}^{\infty} x_1(hT)z^{-h}\sum_{m=0}^{\infty} x_2(mT)z^{-m} = X_1(z)X_2(z)$$

Complex Convolution Theorem. The following, known as the complex convolution theorem, is useful in obtaining the z transform of the product of two sequences $x_1(k)$ and $x_2(k)$.

Suppose both $x_1(k)$ and $x_2(k)$ are zero for $k < 0$. Assume that

$$X_1(z) = \mathcal{Z}[x_1(k)], \quad |z| > R_1$$
$$X_2(z) = \mathcal{Z}[x_2(k)], \quad |z| > R_2$$

where R_1 and R_2 are the radii of absolute convergence for $x_1(k)$ and $x_2(k)$, respectively. Then the z transform of the product of $x_1(k)$ and $x_2(k)$ can be given by

$$\mathcal{Z}[x_1(k)x_2(k)] = \frac{1}{2\pi j}\oint_C \zeta^{-1}X_2(\zeta)X_1(\zeta^{-1}z)\,d\zeta \tag{B–7}$$

where $R_2 < |\zeta| < |z|/R_1$.

To prove this theorem, let us take the z transform of $x_1(k)x_2(k)$:

$$\mathcal{Z}[x_1(k)x_2(k)] = \sum_{k=0}^{\infty} x_1(k)x_2(k)z^{-k} \tag{B–8}$$

The series on the right-hand side of Equation (B–8) converges for $|z| > R$, where R is the radius of absolute convergence for $x_1(k)x_2(k)$. From Equation (2–23), we have

$$\begin{aligned}x_2(k) &= \frac{1}{2\pi j}\oint_C X_2(z)z^{k-1}\,dz \\ &= \frac{1}{2\pi j}\oint_C X_2(\zeta)\zeta^{k-1}\,d\zeta\end{aligned} \tag{B–9}$$

Substituting Equation (B–9) into Equation (B–8), we obtain

$$\mathcal{Z}[x_1(k)x_2(k)] = \frac{1}{2\pi j}\sum_{k=0}^{\infty}\oint_C X_2(\zeta)\zeta^{k-1}x_1(k)z^{-k}\,d\zeta$$

Noting that Equation (B–8) converges uniformly for the region $|z| > R$, we may interchange the order of summation and integration. Then

$$\mathcal{Z}[x_1(k)z_2(k)] = \frac{1}{2\pi j}\oint_C \zeta^{-1}X_2(\zeta)\sum_{k=0}^{\infty} x_1(k)(\zeta^{-1}z)^{-k}\,d\zeta$$

Since

$$\sum_{k=0}^{\infty} x_1(k)(\zeta^{-1}z)^{-k} = X_1(\zeta^{-1}z)$$

we have

$$\mathcal{Z}[x_1(k)x_2(k)] = \frac{1}{2\pi j}\oint_C \zeta^{-1}X_2(\zeta)X_1(\zeta^{-1}z)\,d\zeta \tag{B–10}$$

where C is a contour (a circle with its center at the origin), which lies in the region given by $|\zeta| > R_2$ and $|\zeta^{-1}z| > R_1$, or

$$R_2 < |\zeta| < \frac{|z|}{R_1} \tag{B–11}$$

Parseval's Theorem. Suppose the z transforms of sequences $x_1(k)$ and $x_2(k)$ are such that

$$X_1(z) = \mathcal{Z}[x_1(k)], \qquad |z| > R_1 \text{ (where } R_1 < 1)$$
$$X_2(z) = \mathcal{Z}[x_2(k)], \qquad |z| > R_2$$

and inequality (B–11) is satisfied for $|z| = 1$, or

$$R_2 < |\zeta| < \frac{1}{R_1}$$

Then, by substituting $|z| = 1$ into Equation (8–10), we obtain the following equation:

$$\mathcal{Z}[x_1(k)x_2(k)]_{|z|=1} = \sum_{k=0}^{\infty} x_1(k)x_2(k) = \frac{1}{2\pi j}\oint_C \zeta^{-1} X_2(\zeta)X_1(\zeta^{-1})\, d\zeta$$

If we set $x_1(k) = x_2(k) = x(k)$ in this last equation, we get

$$\sum_{k=0}^{\infty} x^2(k) = \frac{1}{2\pi j}\oint_C \zeta^{-1} X(\zeta)X(\zeta^{-1})\, d\zeta$$
$$= \frac{1}{2\pi j}\oint_C z^{-1} X(z)X(z^{-1})\, dz \qquad \text{(B–12)}$$

Equation (B–12) is Parseval's theorem. This theorem is useful for obtaining the summation of $x^2(k)$.

B–3 INVERSE z TRANSFORMATION AND INVERSION INTEGRAL METHOD

If $X(z)$ is expanded into a power series in z^{-1},

$$X(z) = \sum_{k=0}^{\infty} x(kT)z^{-k} = x(0) + x(T)z^{-1} + x(2T)z^{-2} + \cdots + x(kT)z^{-k} + \cdots$$

or

$$X(z) = \sum_{k=0}^{\infty} x(k)z^{-k} = x(0) + x(1)z^{-1} + x(2)z^{-2} + \cdots + x(k)z^{-k} + \cdots$$

then the values of $x(kT)$ or $x(k)$ give the inverse z transform. If $X(z)$ is given in the form of a rational function, the expansion into an infinite power series in increasing powers of z^{-1} can be accomplished by simply dividing the numerator by the denominator. If the resulting series is convergent, the coefficients of the z^{-k} term in the series are the values $x(kT)$ of the time sequence. However, it is usually difficult to get the closed-form expressions.

The following formulas are sometimes useful in recognizing the closed-form expressions for finite or infinite series in z^{-1}:

$$(1 - az^{-1})^3 = 1 - 3az^{-1} + 3a^2z^{-2} - a^3z^{-3}$$
$$(1 - az^{-1})^4 = 1 - 4az^{-1} + 6a^2z^{-2} - 4a^3z^{-3} + a^4z^{-4}$$

$$(1 - az^{-1})^{-1} = 1 + az^{-1} + a^2z^{-2} + a^3z^{-3} + a^4z^{-4} + a^5z^{-5} + \cdots \qquad |z| > 1$$
$$(1 - az^{-1})^{-2} = 1 + 2az^{-1} + 3a^2z^{-2} + 4a^3z^{-3} + 5a^4z^{-4} + 6a^5z^{-5} + \cdots, \qquad |z| > 1$$
$$(1 - az^{-1})^{-3} = 1 + 3az^{-1} + 6a^2z^{-2} + 10a^3z^{-3} + 15a^4z^{-4} + 21a^5z^{-5} + 28a^6z^{-6} + \cdots \qquad |z| > 1$$
$$(1 - az^{-1})^{-4} = 1 + 4az^{-1} + 10a^2z^{-2} + 20a^3z^{-3} + 35a^4z^{-4} + 56a^5z^{-5} + 84a^6z^{-6} + 120a^7z^{-7} + \cdots \qquad |z| > 1$$

For a given z transform $X(z)$, if a closed-form expression for $x(k)$ is desired, we may use the partial-fraction-expansion method or the inversion integral method discussed in what follows.

Inversion Integral Method. The inversion integral method, based on the inversion integral, is the most general method for obtaining the inverse z transform. It is based on complex variable theory. (For a rigorous and complete derivation of the inversion integral, refer to a book on complex variable theory.) In presenting the inversion integral formula for the z transform, we need to review the residue theorem and its associated background material.

Review of Background Material in Deriving the Inversion Integral Formula. Suppose z_0 is an isolated singular point (pole) of $F(z)$. It can be seen that a positive number r_1 exists such that the function $F(z)$ is analytic at every point z for which $0 < |z - z_0| \le r_1$. Let us denote the circle with center at $z = z_0$ and radius r_1 as Γ_1. Define Γ_2 as any circle with center at $z = z_0$ and radius $|z - z_0| = r_2$ for which $r_2 \le r_1$. Circles Γ_1 and Γ_2 are shown in Figure B–1. Then the Laurent series expansion of $F(z)$ about pole $z = z_0$ may be given by

$$F(z) = \sum_{n=0}^{\infty} a_n(z - z_0)^n + \sum_{n=1}^{\infty} \frac{b_n}{(z - z_0)^n}$$

where coefficients a_n and b_n are given by

$$a_n = \frac{1}{2\pi j}\oint_{\Gamma_1} \frac{F(z)}{(z - z_0)^{n+1}} dz, \qquad n = 0, 1, 2, \ldots$$

$$b_n = \frac{1}{2\pi j}\oint_{\Gamma_2} \frac{F(z)}{(z - z_0)^{-n+1}} dz, \qquad n = 1, 2, 3, \ldots$$

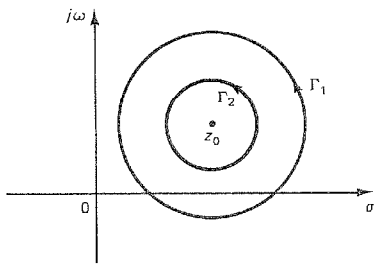


Figure B–1 Analytic region for function $F(z)$.

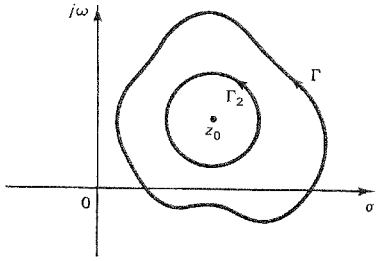


Figure B–2 Analytic region for function $F(z)$ as bonded by closed curve Γ.

Notice that the coefficient b_1 is given by

$$b_1 = \frac{1}{2\pi j}\oint_{\Gamma_2} F(z)\,dz \tag{B–13}$$

It can be proved that the value of the integral of Equation (B–13) is unchanged if Γ_1 is replaced by any closed curve Γ around z_0 such that $F(z)$ is analytic on and inside Γ except at pole $z = z_0$ (see Figure B–2). The closed curve Γ may extend outside the circle Γ_1. Then, by referring to the Cauchy–Goursat theorem, we have

$$\oint_{\Gamma} F(z)\,dz - \oint_{\Gamma_2} F(z)\,dz = 0$$

Thus, Equation (B–13) can be written as

$$b_1 = \frac{1}{2\pi j}\oint_{\Gamma} F(z)\,dz$$

The coefficient b_1 is called the *residue* of $F(z)$ at the pole z_0.

Next, let us assume that the closed curve Γ enclosed m isolated poles $z_1, z_2, \ldots, z_m$, as shown in Figure B–3. Notice that the function $F(z)$ is analytic in

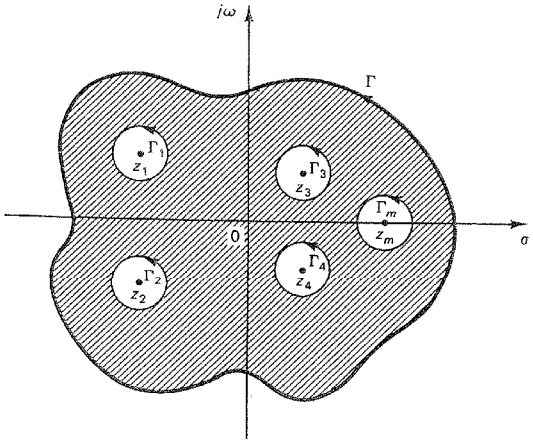


Figure B–3 Closed curve Γ enclosing m isolated poles $z_1, z_2, \ldots, z_m$.

the shaded region. According to the Cauchy–Goursat theorem, the integral of $F(z)$ over the shaded region is zero. The integral over the total shaded region is

$$\oint_{\Gamma} F(z)\,dz - \oint_{\Gamma_1} F(z)\,dz - \oint_{\Gamma_2} F(z)\,dz - \cdots - \oint_{\Gamma_m} F(z)\,dz = 0$$

where $\Gamma_1, \Gamma_2, \ldots, \Gamma_m$ are closed curves around the poles $z_1, z_2, \ldots, z_m$, respectively. Hence,

$$\begin{aligned}\oint_{\Gamma} F(z)\,dz &= \oint_{\Gamma_1} F(z)\,dz + \oint_{\Gamma_2} F(z)\,dz + \cdots + \oint_{\Gamma_m} F(z)\,dz \\ &= 2\pi j(b_{1_1} + b_{1_2} + \cdots + b_{1_m}) \\ &= 2\pi j(K_1 + K_2 + \cdots + K_m)\end{aligned} \tag{B–14}$$

where $K_1 = b_{1_1}, K_2 = b_{1_2}, \ldots, K_m = b_{1_m}$ are residues of $F(z)$ at poles $z_1, z_2, \ldots, z_m$, respectively.

Equation (B–14) is known as the *residue theorem*. It states that if a function $F(z)$ is analytic within and on a closed curve Γ, except at a finite number of poles $z_1, z_2, \ldots, z_m$ inside Γ, then the integral of $F(z)$ taken counterclockwise around Γ is equal to $2\pi j$ times the sum of the residues at poles $z_1, z_2, \ldots, z_m$.

Inversion Integral for the z Transform. We shall now use the Cauchy–Goursat theorem and the residue theorem to derive the inversion integral for the z transform.

From the definition of the z transform, we have

$$X(z) = \sum_{k=0}^{\infty} x(kT)z^{-k} = x(0) + x(T)z^{-1} + z(2T)z^{-2} + \cdots + x(kT)z^{-k} + \cdots$$

By multiplying both sides of this last equation by z^{k-1}, we obtain

$$X(z)z^{k-1} = x(0)z^{k-1} + x(T)z^{k-2} + x(2T)z^{k-3} + \cdots + x(kT)z^{-1} + \cdots \tag{B–15}$$

Notice that Equation (B–15) is the Laurent series expansion of $X(z)z^{k-1}$ around point $z = 0$.

Consider a circle C with its center at the origin of the z plane such that all poles of $X(z)z^{k-1}$ are inside it. Noting that the coefficient $x(kT)$ associated with the term z^{-1} in Equation (B–15) is the residue, we obtain

$$x(kT) = \frac{1}{2\pi j}\oint_C X(z)z^{k-1}\,dz \tag{B–16}$$

Equation (B–16) is the inversion integral for the z transform. The evaluation of the inversion integral can be done as presented next.

Let us define the poles of $X(z)z^{k-1}$ as $z_1, z_2, \ldots, z_m$. Since the closed curve C encloses all poles $z_1, z_2, \ldots, z_m$, then referring to Equation (B–14) we have

$$\begin{aligned}\oint_C X(z)z^{k-1}\,dz &= \oint_{C_1} X(z)z^{k-1}\,dz + \oint_{C_2} X(z)z^{k-1}\,dz + \cdots + \oint_{C_m} X(z)z^{k-1}\,dz \\ &= 2\pi j(K_1 + K_2 + \cdots + K_m)\end{aligned} \tag{B–17}$$

where $K_1, K_2, \ldots, K_m$ denote the residues of $X(z)z^{k-1}$ at poles $z_1, z_2, \ldots, z_m$, respectively, and $C_1, C_2, \ldots, C_m$ are small closed curves around the isolated poles $z_1, z_2, \ldots, z_m$, respectively.

Now we combine Equations (B–16) and (B–17) to obtain a very useful result. Since $X(z)z^{k-1}$ has m poles, that is, $z_1, z_2, \ldots, z_m$,

$$x(k) = x(kT) = K_1 + K_2 + \cdots + K_m$$
$$= \sum_{i=1}^{m} [\text{residue of } X(z)z^{k-1} \text{ at pole } z = z_i \text{ of } X(z)z^{k-1}] \qquad \text{(B–18)}$$

In evaluating residues, note that if the denominator of $X(z)z^{k-1}$ contains a simple pole $z = z_i$ then the corresponding residue K is

$$K = \lim_{z \to z_i} [(z - z_i)X(z)z^{k-1}]$$

If $X(z)z^{k-1}$ contains a multiple pole z_j of order q, then the residue K is given by

$$K = \frac{1}{(q-1)!} \lim_{z \to z_j} \frac{d^{q-1}}{dz^{q-1}} [(z - z_j)^q X(z)z^{k-1}]$$

Note that in this book we treat only one-sided z transforms. This implies that $x(k) = 0$ for $k < 0$. Hence, we restrict the values of k in Equation (B–17) to the nonnegative integer values.

If $X(z)$ has a zero of order r at the origin, then $X(z)z^{k-1}$ in Equations (B–17) will involve a zero of order $r + k - 1$ at the origin. If $r \geq 1$, then $r + k - 1 \geq 0$ for $k \geq 0$, and there is no pole at $z = 0$ in $X(z)z^{k-1}$. However, if $r \leq 0$, then there will be a pole at $z = 0$ for one or more nonnegative values of k. In such a case, separate inversion of Equation (B–17) is necessary for each of such values of k.

It should be noted that the inversion integral method, when evaluated by residues, is a very simple technique for obtaining the inverse z transform, provided that $X(z)z^{k-1}$ has no poles at the origin, $z = 0$. If, however, $X(z)z^{k-1}$ has a simple pole or a multiple pole at $z = 0$, then calculations may become cumbersome and the partial-fraction-expansion method may prove to be simpler to apply.

Comments on Calculating Residues. In obtaining the residues of a function $X(z)$, note that, regardless of the way we calculate the residues, the final result is the same. Therefore, we may use any method that is convenient for a given situation. As an example, consider the following function $X(z)$:

$$X(z) = \frac{2z^2 + 5z + 6}{(z+1)^3} + \frac{4z}{(z+1)^2} + \frac{5}{z+1}$$

We shall demonstrate three methods for calculating the residue of this function $X(z)$.

Method 1. The residue of this function may be obtained as the sum of the residues of the respective terms:

$$[\text{Residue } K \text{ of } X(z) \text{ at pole } z = -1]$$
$$= \frac{1}{(3-1)!} \lim_{z \to -1} \frac{d^2}{dz^2} \left[(z+1)^3 \frac{2z^2 + 5z + 6}{(z+1)^3} \right]$$
$$+ \frac{1}{(2-1)!} \lim_{z \to -1} \frac{d}{dz} \left[(z+1)^2 \frac{4z}{(z+1)^2} \right] + \lim_{z \to -1} \left[(z+1) \frac{5}{z+1} \right]$$
$$= \tfrac{1}{2} \lim_{z \to -1} (4) + \lim_{z \to -1} (4) + \lim_{z \to -1} (5) = 2 + 4 + 5$$
$$= 11$$

Method 2. If the three terms of $X(z)$ are combined into one as shown next,

$$X(z) = \frac{2z^2 + 5z + 6}{(z+1)^3} + \frac{4z}{(z+1)^2} + \frac{5}{z+1} = \frac{11z^2 + 19z + 11}{(z+1)^3}$$

then the residue can be calculated as follows:

$$[\text{Residue } K \text{ of } X(z) \text{ at pole } z = -1]$$
$$= \frac{1}{(3-1)!} \lim_{z \to -1} \frac{d^2}{dz^2} \left[(z+1)^3 \frac{11z^2 + 19z + 11}{(z+1)^3} \right]$$
$$= \tfrac{1}{2} \lim_{z \to -1} (22)$$
$$= 11$$

Method 3. If $X(z)$ is expanded into usual partial fractions as shown next,

$$X(z) = \frac{11z^2 + 19z + 11}{(z+1)^3} = \frac{3}{(z+1)^3} - \frac{3}{(z+1)^2} + \frac{11}{z+1}$$

then the residue of $X(z)$ is the coefficient of the term $1/(z+1)$. Thus,

$$[\text{Residue } K \text{ of } X(z) \text{ at pole } z = -1] = 11$$

B–4 MODIFIED z TRANSFORM METHOD

The modified z transform method is a modification of the z transform method. It is based on inserting a fictitious delay time at the output of the system, in addition to the insertion of the fictitious output sampler, and varying the amount of the fictitious delay time so that the output at any time between two consecutive sampling instants can be obtained.

The modified z transform method is useful not only in obtaining the response between two consecutive sampling instants, but also in obtaining the z transform of the process with pure delay or transportation lag. In addition, the modified z transform method is applicable to most sampling schemes.

Consider the system shown in Figure B–4(a). In this system a fictitious delay of $(1 - m)T$ seconds, where $0 \leq m \leq 1$ and T is the sampling period, is inserted at the output of the system. By varying m between 0 and 1, the output $y(t)$ at $t = kT - (1 - m)T$ (where $k = 1, 2, 3, \ldots$) may be obtained. Noting that $G^*(s)$ is given by

$$G^*(s) = \mathcal{L}[g(t)\delta_T(t)]$$

we define the modified pulse transfer function $G(z, m)$ by

$$\mathcal{Z}_m[G(s)] = G(z, m) = G^*(s, m)|_{s=(1/T)\ln z}$$
$$= \mathcal{L}[g(t - (1 - m)T)\delta_T(t)]|_{s=(1/T)\ln z} \qquad \text{(B–19)}$$

where the notation $\mathcal{Z}_m$ signifies the modified z transform.

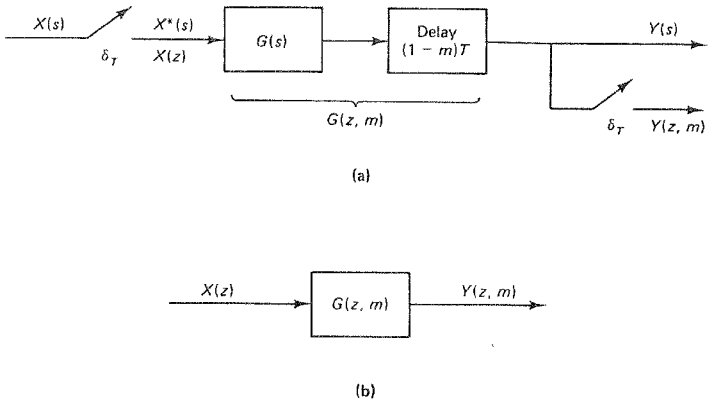


Figure B–4 (a) System with a fictitious delay time of $(1 - m)T$ sec; (b) modified pulse-transfer-function system with input $X(z)$ and output $Y(z, m)$.

Noting that

$$\mathcal{L}[g(t - (1 - m)T)\delta_T(t)] = \mathcal{L}[g(t - T + mT)\delta_T(t)]$$
$$= e^{-Ts}\mathcal{L}[g(t + mT)\delta_T(t)]$$

we have

$$G^*(s, m) = e^{-Ts}\mathcal{L}[g(t + mT)\delta_T(t)] \tag{B–20}$$

Since $\mathcal{L}[g(t + mT)\delta_T(t)]$ is the Laplace transform of the product of two time functions, by referring to Equation (3–19) it can be obtained as follows:

$$\mathcal{L}[g(t + mT)\delta_T(t)] = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} G(p)\frac{e^{mTp}}{1 - e^{-T(s-p)}}dp \tag{B–21}$$

The integration on the right-hand side of Equation (B–21) can be carried out in a way similar to that discussed in Section 3–3; that is, the convolution integral can be integrated in either the left half-plane or the right half-plane.

Let us consider the contour integration along the infinite semicircle in the left half-plane. Then

$$\mathcal{L}[g(t + mT)\delta_T(t)] = \sum\left[\text{residue of } \frac{G(s)e^{mTs}z}{z - e^{Ts}} \text{ at pole of } G(s)\right] \tag{B–22}$$

Hence, from Equations (B–19), (B–20), and (B–22), we obtain the modified z transform of $G(s)$ as follows:

$$G(z, m) = z^{-1}\sum\left[\text{residue of } \frac{G(s)e^{mTs}z}{z - e^{Ts}} \text{ at pole of } G(s)\right] \tag{B–23}$$

Note that the modified z transform $G(z, m)$ and the z transform $G(z)$ are related as follows:

$$G(z) = \lim_{m\to 0} zG(z, m) \tag{B–24}$$

Referring to Figure B–4(b), the output $Y(z, m)$ is obtained as follows:

$$Y(z, m) = G(z, m)X(z) \tag{B–25}$$

As in the case of the z transform, the modified z transform $Y(z, m)$ can be expanded into an infinite series in z^{-1}, as follows:

$$Y(z, m) = y_0(m)z^{-1} + y_1(m)z^{-2} + y_2(m)z^{-3} + \cdots \tag{B–26}$$

By multiplying both sides of Equation (B–26) by z, we have

$$zY(z, m) = y_0(m) + y_1(m)z^{-1} + y_2(m)z^{-2} + \cdots \tag{B–27}$$

where $y_k(m)$ represents the value of $y(t)$ between $t = kT$ and $t = (k + 1)T$ $(k = 0, 1, 2, \ldots)$, or

$$y_k(m) = y((k + m)T) \tag{B–28}$$

Note that if $y(k)$ is continuous then

$$\lim_{m\to 1} y_{k-1}(m) = \lim_{m\to 0} y_k(m) \tag{B–29}$$

The left-hand side of Equation (B–29) gives the values $y(0-), y(T-), y(2T-), \ldots$, and the right-hand side gives the values $y(0+), y(T+), y(2T+), \ldots$. If the output $y(kT)$ is continuous, then $y(kT-) = y(kT+)$.

Example B–3

Obtain the modified z transform of $G(s)$, where

$$G(s) = \frac{1}{s + a}$$

Referring to Equation (B–23), we obtain the modified z transform of $G(s)$ as follows:

$$G(z, m) = z^{-1}\left[\text{residue of } \frac{1}{s + a}\frac{e^{mTs}z}{z - e^{Ts}} \text{ at pole } s = -a\right]$$
$$= z^{-1}\left\{\lim_{s\to -a}\left[(s + a)\frac{1}{s + a}\frac{e^{mTs}z}{z - e^{Ts}}\right]\right\}$$
$$= z^{-1}\frac{e^{-maT}z}{z - e^{-aT}} = \frac{e^{-maT}z^{-1}}{1 - e^{-aT}z^{-1}}$$

Example B–4

Consider the systems shown in Figures B–5(a) and (b). Obtain the output $Y(z, m)$ of each system.

For the system shown in Figure B–5(a), we have

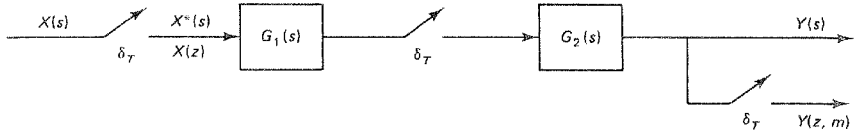
$$Y(z, m) = \mathcal{Z}_m[Y(s)] = G_2(z, m)G_1(z)X(z)$$

Note that

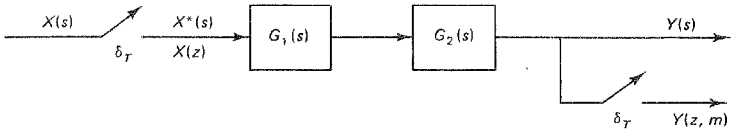
$$Y(z) = \mathcal{Z}[Y(s)] = G_2(z)G_1(z)X(z)$$

For the system shown in Figure B–5(b), we have

$$Y(z, m) = \mathcal{Z}_m[Y(s)] = G_1G_2(z, m)X(z)$$



(a)



(b)

Figure B–5 (a) System with a sampler between $G_1(s)$ and $G_2(s)$; (b) system with no sampler between $G_1(s)$ and $G_2(s)$.

where

$$G_1G_2(z,m) = \mathcal{Z}_m[G_1(s)G_2(s)]$$

Note that

$$Y(z) = G_1G_2(z)X(z)$$

Example B–5

Consider the system shown in Figure B–6. Obtain the modified z transform of $C(s)$. The output $C(z)$ is given by

$$C(z) = \frac{G(z)}{1 + GH(z)}R(z)$$

The modified z transform of $C(z)$ is given by

$$C(z,m) = \frac{G(z,m)}{1 + GH(z)}R(z) \qquad \text{(B–30)}$$

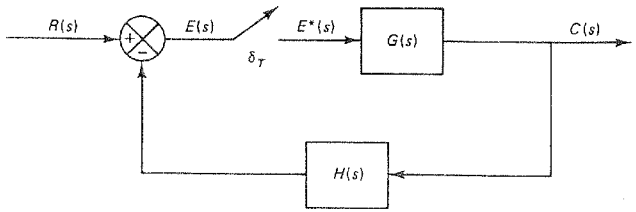


Figure B–6 Closed-loop control system.

Example B–6

Consider the system shown in Figure B–7. The sampling period T is 1 sec. or $T = 1$. Suppose that the system is subjected to a unit-step input. Obtain $c_k(m)$ for $m = 0.5$ and $k = 0, 1, 2, \ldots, 9$. Also, verify that Equation (B–24) holds true. The modified z transform of $G(s)$ is obtained from Equation (B–23) as follows:

$$\begin{aligned}
G(z,m) &= z^{-1}\sum\left[\text{residue of } \frac{G(s)e^{ms}z}{z - e^{s}} \text{ at pole of } G(s)\right] \\
&= z^{-1}(1 - z^{-1})\left\{\left[\text{residue of } \frac{1}{s^2(s+1)}\frac{e^{ms}z}{z - e^{s}} \text{ at double pole } s = 0\right]\right. \\
&\quad \left.+\left[\text{residue of } \frac{1}{s^2(s+1)}\frac{e^{ms}z}{z - e^{s}} \text{ at simple pole } s = -1\right]\right\} \\
&= z^{-1}(1 - z^{-1})\left\{\frac{1}{(2-1)!}\lim_{s\to 0}\frac{d}{ds}\left[s^2\frac{1}{s^2(s+1)}\frac{e^{ms}z}{z - e^{s}}\right]\right. \\
&\quad \left.+\lim_{s\to -1}\left[(s+1)\frac{1}{s^2(s+1)}\frac{e^{ms}z}{z - e^{s}}\right]\right\} \\
&= z^{-1}(1 - z^{-1})\left[\frac{mz^2 - mz - z^2 + 2z}{(z-1)^2} + \frac{e^{-m}z}{z - e^{-1}}\right] \\
&= \frac{(m-1)z^{-1} + (2-m)z^{-2}}{1 - z^{-1}} + \frac{e^{-m}z^{-1}(1 - z^{-1})}{1 - e^{-1}z^{-1}} \\
&= \frac{(m - 1 + e^{-m})z^{-1} + (2.3679 - 1.3679m - 2e^{-m})z^{-2} + [-0.3679(2 - m) + e^{-m}]z^{-3}}{(1 - z^{-1})(1 - 0.3679z^{-1})}
\end{aligned}$$

Referring to Equation (B–30) and noting that $R(z) = 1/(1 - z^{-1})$, we have

$$\begin{aligned}
C(z,m) &= \frac{G(z,m)}{1 + G(z)}\frac{1}{1 - z^{-1}} \\
&= \frac{(m - 1 + e^{-m})z^{-1} + (2.3679 - 1.3679m - 2e^{-m})z^{-2} + (-0.7358 + 0.3679m + e^{-m})z^{-3}}{1 - 2z^{-1} + 1.6321z^{-2} - 0.6321z^{-3}}
\end{aligned}$$

Hence, for $m = 0.5$ we have

$$C(z, 0.5) = \frac{0.1065z^{-1} + 0.4709z^{-2} + 0.05468z^{-3}}{1 - 2z^{-1} + 1.6321z^{-2} - 0.6321z^{-3}} \qquad \text{(B–31)}$$

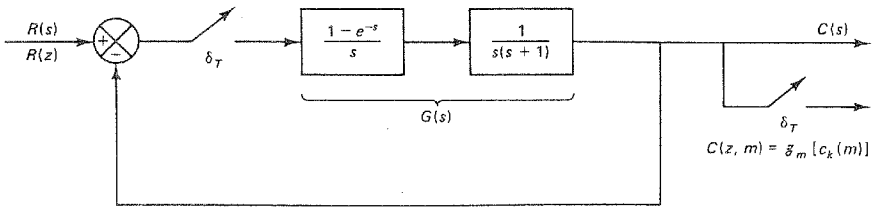


Figure B–7 Closed-loop control system.

By referring to Equation (B–27), Equation (B–31) can be expanded into an infinite series in z^{-1} as follows:

$$C(z, 0.5) = c_0(0.5)z^{-1} + c_1(0.5)z^{-2} + c_2(0.5)z^{-3} + \cdots$$

or

$$zC(z, 0.5) = c_0(0.5) + c_1(0.5)z^{-1} + c_2(0.5)z^{-2} + \cdots$$

where $c_k(0.5) = c((k + 0.5)T) = c(k + 0.5)$ and $k = 0, 1, 2, \ldots$. The values of $c_k(0.5)$ can easily be obtained with a digital computer. The computer solution for $k = 0, 1, 2, \ldots, 9$ is as follows:

$$c_0(0.5) = c(0.5) = 0.1065$$
$$c_1(0.5) = c(1.5) = 0.6839$$
$$c_2(0.5) = c(2.5) = 1.2487$$
$$c_3(0.5) = c(3.5) = 1.4485$$
$$c_4(0.5) = c(4.5) = 1.2913$$
$$c_5(0.5) = c(5.5) = 1.0078$$
$$c_6(0.5) = c(6.5) = 0.8236$$
$$c_7(0.5) = c(7.5) = 0.8187$$
$$c_8(0.5) = c(8.5) = 0.9302$$
$$c_9(0.5) = c(9.5) = 1.0447$$

These values give the response at the midpoints between pairs of consecutive sampling points. Note that by varying the value of m between 0 and 1 it is possible to find the response at any point between two consecutive sampling points, such as $c(1.2)$ and $c(2.8)$.

Finally, note that

$$G(z) = \mathcal{Z}[G(s)] = \mathcal{Z}\left[\frac{1 - e^{-s}}{s}\frac{1}{s(s + 1)}\right]$$
$$= \frac{(T - 1 + e^{-T})z^{-1} + (1 - e^{-T} - Te^{-T})z^{-2}}{(1 - z^{-1})(1 - e^{-T}z^{-1})}$$
$$= \frac{0.3679z^{-1} + 0.2642z^{-2}}{(1 - z^{-1})(1 - 0.3679z^{-1})}$$

and

$$\lim_{m \to 0} zG(z, m) = \frac{0.3679z^{-1} + 0.2642z^{-2}}{(1 - z^{-1})(1 - 0.3679z^{-1})}$$

Hence,

$$G(z) = \lim_{m \to 0} zG(z, m)$$

Clearly, Equation (B–24) holds true.

Summary. The main purpose of this section has been to present the modified z transform method for finding the response for any time between two consecutive sampling instants. It is noted that the modified z transform method can be used not only for such a purpose, but also for dealing with multirate sampling schemes.

EXAMPLE PROBLEMS AND SOLUTIONS

Problem B–1

Obtain the z transform of $1/k!$

Solution

$$\mathcal{Z}\left[\frac{1}{k!}\right] = \sum_{k=0}^{\infty}\frac{1}{k!}z^{-k}$$
$$= 1 + z^{-1} + \frac{1}{2!}z^{-2} + \frac{1}{3!}z^{-3} + \frac{1}{4!}z^{-4} + \cdots$$
$$= \exp(z^{-1})$$

Problem B–2

Obtain

$$\sum_{k=1}^{\infty}\left(\frac{1}{k}\right)z^{-k}$$

(This series looks like the z transform of $1/k$, but the k sequence begins here with 1 instead of 0.)

Solution Since

$$\sum_{k=0}^{\infty} z^{-k} = 1 + z^{-1} + z^{-2} + \cdots = \frac{1}{1 - z^{-1}}, \qquad |z| > 1$$

by multiplying both sides of this last equation by z^{-2}, we have

$$\sum_{k=0}^{\infty} z^{-k-2} = \frac{z^{-2}}{1 - z^{-1}}$$

Integrating this last equation with respect to z, we have

$$\int \sum_{k=0}^{\infty} z^{-k-2}\,dz = \int \frac{z^{-2}}{1 - z^{-1}}\,dz$$

or

$$\sum_{k=0}^{\infty}\frac{z^{-k-1}}{-k-1} = \ln(1 - z^{-1}) + \text{constant} \qquad \text{(B–32)}$$

where the constant in Equation (B–32) is zero. [To verify this, substitute ∞ for z in both sides of Equation (B–32).] Equation (B–32) can thus be rewritten as follows:

$$\sum_{k=1}^{\infty}\frac{z^{-k}}{-k} = \ln(1 - z^{-1}), \qquad |z| > 1$$

or

$$\sum_{k=1}^{\infty}\left(\frac{1}{k}\right)z^{-k} = -\ln(1 - z^{-1}), \qquad |z| > 1$$

Problem B–3

The first backward difference between $x(k)$ and $x(k - 1)$ is defined by

$$\nabla x(k) = x(k) - x(k - 1)$$

The second backward difference is defined by

$$\nabla^2 x(k) = \nabla[\nabla x(k)] = \nabla[x(k) - x(k-1)]$$
$$= \nabla x(k) - \nabla x(k-1)$$

and the third backward difference is defined by

$$\nabla^3 x(k) = \nabla^2 x(k) - \nabla^2 x(k-1)$$

Similarly, the mth backward difference is given by

$$\nabla^m x(k) = \nabla^{m-1} x(k) - \nabla^{m-1} x(k-1)$$

Obtain the z transforms of $\nabla x(k)$, $\nabla^2 x(k)$, $\nabla^3 x(k)$, and $\nabla^m x(k)$.

Solution The z transform of the first backward difference is obtained as follows:

$$\mathcal{Z}[\nabla x(k)] = \mathcal{Z}[x(k)] - \mathcal{Z}[x(k-1)]$$
$$= X(z) - z^{-1}X(z)$$
$$= (1 - z^{-1})X(z) \qquad \text{(B–33)}$$

Since

$$\nabla^2 x(k) = [x(k) - x(k-1)] - [x(k-1) - x(k-2)]$$
$$= x(k) - 2x(k-1) + x(k-2)$$

the z transform of $\nabla^2 x(k)$ is

$$\mathcal{Z}[\nabla^2 x(k)] = \mathcal{Z}[x(k)] - 2\,\mathcal{Z}[x(k-1)] + \mathcal{Z}[x(k-2)]$$
$$= X(z) - 2z^{-1}X(z) + z^{-2}X(z)$$
$$= (1 - z^{-1})^2 X(z) \qquad \text{(B–34)}$$

In this way we obtain

$$\mathcal{Z}[\nabla^3 x(k)] = (1 - z^{-1})^3 X(z)$$

Notice that the operation of taking the backward difference corresponds to multiplying $X(z)$ by $(1 - z^{-1})$. Thus, for the mth backward difference,

$$\nabla^m x(k) = \nabla^{m-1} x(k) - \nabla^{m-1} x(k-1)$$

we have

$$\mathcal{Z}[\nabla^m x(k)] = (1 - z^{-1})^m X(z) \qquad \text{(B–35)}$$

Problem B–4

The first forward difference between $x(k+1)$ and $x(k)$ is defined by

$$\Delta x(k) = x(k+1) - x(k)$$

The second forward difference is defined by

$$\Delta^2 x(k) = \Delta[\Delta x(k)] = \Delta[x(k+1) - x(k)]$$
$$= \Delta x(k+1) - \Delta x(k)$$

The third forward difference is defined by

$$\Delta^3 x(k) = \Delta^2 x(k+1) - \Delta^2 x(k)$$

and the mth forward difference is given by

$$\Delta^m x(k) = \Delta^{m-1} x(k+1) - \Delta^{m-1} x(k)$$

Obtain the z transforms of $\Delta x(k)$, $\Delta^2 x(k)$, $\Delta^3 x(k)$, and $\Delta^m x(k)$.

Solution The z transform of the first forward difference is given by

$$\mathcal{Z}[\Delta x(k)] = \mathcal{Z}[x(k+1)] - \mathcal{Z}[x(k)]$$
$$= zX(z) - zx(0) - X(z)$$
$$= (z-1)X(z) - zx(0) \qquad \text{(B–36)}$$

Since

$$\Delta^2 x(k) = [x(k+2) - x(k+1)] - [x(k+1) - x(k)]$$
$$= x(k+2) - 2x(k+1) + x(k)$$

the z transform of $\Delta^2 x(k)$ is

$$\mathcal{Z}[\Delta^2 x(k)] = \mathcal{Z}[x(k+2) - 2x(k+1) + x(k)]$$
$$= z^2 X(z) - z^2 x(0) - zx(1) - 2[zX(z) - zx(0)] + X(z)$$
$$= (z-1)^2 X(z) - z(z-1)x(0) - z\,\Delta x(0) \qquad \text{(B–37)}$$

where $\Delta x(0) = x(1) - x(0)$. The z transform of $\Delta^3 x(k)$ becomes

$$\mathcal{Z}[\Delta^3 x(k)] = \mathcal{Z}[x(k+3) - 3x(k+2) + 3x(k+1) - x(k)]$$
$$= (z-1)^3 X(z) - z(z-1)^2 x(0) - z(z-1)\,\Delta x(0) - z\,\Delta^2 x(0)$$

where $\Delta x(0) = x(1) - x(0)$ and $\Delta^2 x(0) = x(2) - 2x(1) + x(0)$. Similarly, for the mth forward difference

$$\Delta^m x(k) = \Delta^{m-1} x(k+1) - \Delta^{m-1} x(k)$$

we have

$$\mathcal{Z}[\Delta^m x(k)] = (z-1)^m X(z) - z \sum_{j=0}^{m-1} (z-1)^{m-j-1} \Delta^j x(0) \qquad \text{(B–38)}$$

Problem B–5

Solve the following difference equation:

$$(k+1)x(k+1) - x(k) = 0$$

where $x(k) = 0$ for $k < 0$ and $x(0) = 1$. Notice that this difference equation is of the time-varying kind. The solution of this type of difference equation may be obtained by use of the z transform. (It should be cautioned that, in general, the z transform approach to the solution of time-varying difference equations may not be successful.)

Solution First, note that

$$\mathcal{Z}[kx(k)] = -z\frac{d}{dz}X(z)$$

Since the original difference equation can be written as

$$kx(k) - x(k-1) = 0$$

the z transform of this last equation can be obtained as follows:

$$-z\frac{d}{dz}X(z) - z^{-1}X(z) = 0$$

or

$$z^2\frac{d}{dz}X(z) + X(z) = 0$$

from which we have

$$\frac{dX(z)}{X(z)} = -\frac{dz}{z^2}$$

or

$$\ln X(z) = \frac{1}{z} + \ln K$$

where K is a constant. Then $X(z)$ can be found from

$$X(z) = K \exp z^{-1}$$

Since $\exp z^{-1}$ may be expanded into the series

$$\exp z^{-1} = 1 + z^{-1} + \frac{1}{2!}z^{-2} + \frac{1}{3!}z^{-3} + \cdots, \qquad |z| > 0$$

we have

$$X(z) = K\left(1 + z^{-1} + \frac{1}{2!}z^{-2} + \frac{1}{3!}z^{-3} + \cdots\right)$$

from which we find the inverse z transform of $X(z)$ to be

$$x(k) = K\frac{1}{k!}, \qquad k = 0, 1, 2, \ldots$$

Since $x(0)$ is given as 1, we have

$$x(0) = K = 1$$

Thus, we have determined the unknown constant K. Hence, the solution to the given difference equation is

$$x(k) = \frac{1}{k!}, \qquad k = 0, 1, 2, \ldots$$

Problem B–6

Solve the following difference equation:

$$(k + 1)x(k + 1) - kx(k) = k + 1$$

where $x(k) = 0$ for $k \leq 0$.

Solution First note that by substituting $k = 0$ into the given difference equation, we have

$$x(1) = 1$$

Now define

$$y(k) = kx(k)$$

Then the given difference equation can be written as

$$y(k + 1) - y(k) = k + 1$$

Taking the z transform of this last equation, we have

$$zY(z) - zy(0) - Y(z) = \frac{z^{-1}}{(1 - z^{-1})^2} + \frac{1}{1 - z^{-1}}$$

Since $y(0) = 0$, we have

$$Y(z) = \frac{z^{-2}}{(1 - z^{-1})^3} + \frac{z^{-1}}{(1 - z^{-1})^2}$$

Referring to Problem A–2–8, we have

$$\mathcal{Z}^{-1}\left[\frac{z^{-2}}{(1 - z^{-1})^3}\right] = \frac{1}{2}(k^2 - k)$$

Hence, the inverse z transform of $Y(z)$ can be given by

$$y(k) = \tfrac{1}{2}(k^2 - k) + k = \tfrac{1}{2}(k^2 + k)$$

Then, $x(k)$ for $k = 1, 2, 3, \ldots$ is determined from

$$kx(k) = y(k) = \tfrac{1}{2}(k^2 + k)$$

as follows:

$$x(k) = \tfrac{1}{2}(k + 1), \qquad k = 1, 2, 3, \ldots$$

Problem B–7

Consider the system shown in Figure B–8. The sampling period is 2 sec, or $T = 2$. The input $x(t)$ is a Kronecker delta function $\delta_0(t)$; that is,

$$\delta_0(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases}$$

Obtain the response every 0.5 sec by using the modified z transform method.

Solution Since the input $x(t)$ is a Kronecker delta function, we have

$$X(z) = 1$$

The modified pulse transfer function $G(z, m)$ is obtained as follows. Referring to Equation (B–23),

$$G(z, m) = z^{-1}\left[\text{residue of } \frac{1}{s + 0.6931}\frac{e^{mTs}z}{z - e^{Ts}} \text{ at pole } s = -0.6931\right]$$

Noting that $T = 2$, we obtain

$$G(z, m) = z^{-1}\left\{\lim_{s \to -0.6931}\left[(s + 0.6931)\frac{1}{s + 0.6931}\frac{e^{2ms}z}{z - e^{2s}}\right]\right\}$$

$$= z^{-1}\frac{(e^{-1.3862})^m z}{z - e^{-1.3862}} = \frac{4^{-m}}{z - 0.25}$$

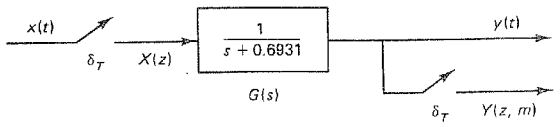


Figure B–8 Impulse-sampled system.

Hence, the output $Y(z,m)$ can be obtained as follows:

$$Y(z,m) = G(z,m)X(z) = \frac{4^{-m}}{z - 0.25}$$

Referring to Equation (B–27), we have

$$zY(z,m) = y_0(m) + y_1(m)z^{-1} + y_2(m)z^{-2} + \cdots$$

where $y_k(m) = y((k + m)T) = y(2k + 2m)$. In this problem $zY(z,m)$ can be expanded into an infinite series in z^{-1} as follows:

$$zY(z,m) = \frac{4^{-m}}{1 - 0.25z^{-1}}$$

$$= 4^{-m} + 4^{-m-1}z^{-1} + 4^{-m-2}z^{-2} + 4^{-m-3}z^{-3} + \cdots$$

Hence,

$$y_0(m) = 4^{-m}$$
$$y_1(m) = 4^{-m-1}$$
$$y_2(m) = 4^{-m-2}$$
$$y_3(m) = 4^{-m-3}$$
$$\vdots$$

To obtain the system output every 0.5 sec, we set $m = 0$, 0.25, 0.50, and 0.75. For $m = 0.25$, we obtain

$$y_0(0.25) = y(0.5) = 4^{-0.25} = 0.7071$$
$$y_1(0.25) = y(2.5) = 4^{-1.25} = 0.1768$$
$$y_2(0.25) = y(4.5) = 4^{-2.25} = 0.04419$$
$$y_3(0.25) = y(6.5) = 4^{-3.25} = 0.01105$$
$$\vdots$$

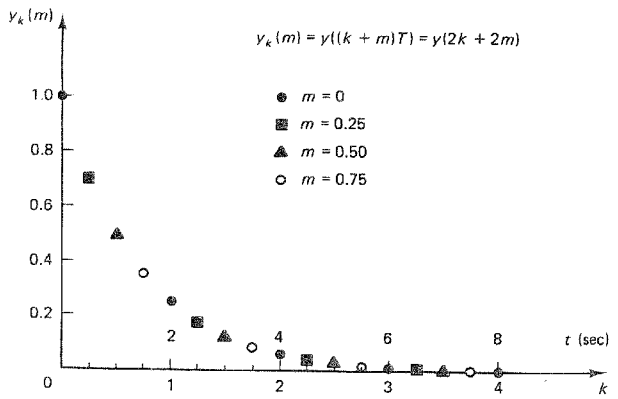


Figure B–9 Plot of $y_k(m)$ versus k for the system considered in Problem B–7.

Similarly, the values of $y_k(m)$ for $m = 0$, 0.5, and 0.75 can be calculated. The result is shown in Figure B–9 as a plot of $y_k(m)$ versus k.

Problem B–8

Obtain $C(z,m)$, the modified z transform of the output, of the system shown in Figure B–10.

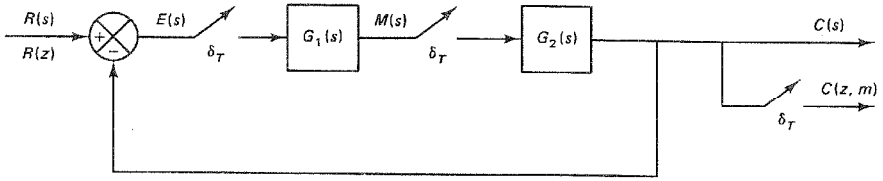


Figure B–10 Closed-loop discrete-time control system.

Solution From Figure B–10 we have

$$E(s) = R(s) - C(s)$$
$$M(s) = G_1(s)E^*(s)$$
$$C(s) = G_2(s)M^*(s)$$

Hence,

$$M^*(s) = G_1^*(s)E^*(s)$$

or

$$M(z) = G_1(z)E(z)$$

Also,

$$E^*(s) = R^*(s) - C^*(s) = R^*(s) - G_2^*(s)M^*(s)$$

or

$$E(z) = R(z) - G_2(z)M(z)$$

Therefore,

$$M(z) = G_1(z)[R(z) - G_2(z)M(z)]$$

from which we obtain

$$M(z) = \frac{G_1(z)R(z)}{1 + G_1(z)G_2(z)}$$

Since $C(z,m)$ can be given by $G_2(z,m)M(z)$, we have

$$C(z,m) = G_2(z,m)M(z) = \frac{G_1(z)G_2(z,m)}{1 + G_1(z)G_2(z)}R(z)$$

Appendix C

Pole Placement Design with Vector Control

C–1 INTRODUCTION

In Chapter 6 we presented the pole placement technique and state observer design when the control signal $u(k)$ was a scalar. If the control signal is a vector quantity (r-vector), however, we can expect to improve the system's response characteristics, because we have more freedom to choose control signals $u_1(k), u_2(k), \ldots, u_r(k)$. For example, in the case of the nth-order system with a scalar control, the deadbeat response can be achieved in at most n sampling periods. In the case of the vector control $\mathbf{u}(k)$, the deadbeat response can be achieved in less than n sampling periods.

It is noted that with the vector control it is possible to choose freely more than n parameters; that is, in addition to being able to place n closed-loop poles properly, we have the freedom to satisfy other requirements, if any, of the closed-loop system.

In the case of the vector control, however, the determination of the state feedback gain matrix $\mathbf{K}$ becomes more complex, as we shall see in this appendix.

C–2 PRELIMINARY DISCUSSIONS

Consider the system

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}\mathbf{u}(k) \tag{C–1}$$

where

$\mathbf{x}(k)$ = state vector (n-vector) at kth sampling instant

$\mathbf{u}(k)$ = control vector (r-vector) at kth sampling instant

$\mathbf{G} = n \times n$ matrix

$\mathbf{H} = n \times r$ matrix

We assume that the magnitudes of the r components of $\mathbf{u}(k)$ are unconstrained. As in the case of the system with a scalar control signal, it can be proved that a necessary and sufficient condition for arbitrary pole placement for the system defined by Equation (C–1) is that the system be completely state controllable.

Let us assume that the system defined by Equation (C–1) is completely state controllable. In the state feedback control scheme, the control vector $\mathbf{u}(k)$ is chosen as

$$\mathbf{u}(k) = -\mathbf{K}\mathbf{x}(k) \tag{C–2}$$

where $\mathbf{K}$ is the state feedback gain matrix. It is an $r \times n$ matrix. With state feedback the system becomes a closed-loop system and its state equation becomes

$$\mathbf{x}(k+1) = (\mathbf{G} - \mathbf{H}\mathbf{K})\mathbf{x}(k)$$

where we choose matrix $\mathbf{K}$ so that the eigenvalues of $\mathbf{G} - \mathbf{H}\mathbf{K}$ are the desired closed-loop poles $\mu_1, \mu_2, \ldots, \mu_n$.

Transforming State Equation Into Controllable Canonical Form. Consider the system defined by

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}_1 u(k) \tag{C–3}$$

where

$\mathbf{x}(k)$ = state vector (n-vector)

$u(k)$ = control signal (scalar)

$\mathbf{G} = n \times n$ matrix

$\mathbf{H}_1 = n \times 1$ matrix

Assume that the system is completely state controllable. Then the controllability matrix has its inverse. Define

$$[\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \cdots \vdots \mathbf{G}^{n-1}\mathbf{H}_1]^{-1} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_n \end{bmatrix}$$

where the $\mathbf{f}_i$'s are the row vectors. Then construct a transformation matrix $\mathbf{T}_1$ as follows:

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{f}_n \\ \mathbf{f}_n\mathbf{G} \\ \vdots \\ \mathbf{f}_n\mathbf{G}^{n-1} \end{bmatrix}^{-1} \tag{C–4}$$

where the $\mathbf{f}_n\mathbf{G}^k$ are row vectors ($k = 0, 1, 2, \ldots, n-1$). Then it can be shown that

$$\mathbf{T}_1^{-1}\mathbf{G}\mathbf{T}_1 = \begin{bmatrix} \mathbf{f}_n \\ \mathbf{f}_n\mathbf{G} \\ \vdots \\ \mathbf{f}_n\mathbf{G}^{n-1} \end{bmatrix} \mathbf{G} \begin{bmatrix} \mathbf{f}_n \\ \mathbf{f}_n\mathbf{G} \\ \vdots \\ \mathbf{f}_n\mathbf{G}^{n-1} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \tag{C–5}$$

and

$$\mathbf{T}_1^{-1}\mathbf{H}_1 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{C–6}$$

[See Problem C–1 for the derivation of Equations (C–5) and (C–6).]

Now if we define

$$\mathbf{x}(k) = \mathbf{T}_1\hat{\mathbf{x}}(k)$$

then Equation (C–3) becomes

$$\hat{\mathbf{x}}(k+1) = \mathbf{T}_1^{-1}\mathbf{G}\mathbf{T}_1\hat{\mathbf{x}}(k) + \mathbf{T}_1^{-1}\mathbf{H}_1 u(k)$$

or

$$\begin{bmatrix} \hat{x}_1(k+1) \\ \hat{x}_2(k+1) \\ \vdots \\ \hat{x}_{n-1}(k+1) \\ \hat{x}_n(k+1) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \begin{bmatrix} \hat{x}_1(k) \\ \hat{x}_2(k) \\ \vdots \\ \hat{x}_{n-1}(k) \\ \hat{x}_n(k) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u(k) \tag{C–7}$$

We have thus shown that the state equation, Equation (C–3), can be transformed into the controllable canonical form by use of the transformation matrix $\mathbf{T}_1$ defined by Equation (C–4).

Design Steps. In what follows we shall discuss the procedure for determining a state feedback gain matrix $\mathbf{K}$ such that the eigenvalues of $\mathbf{G} - \mathbf{H}\mathbf{K}$ are the desired values $\mu_1, \mu_2, \ldots, \mu_n$.

The state equation to be considered in the following was given by Equation (C–1):

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}\mathbf{u}(k)$$

We assume that the rank of the $n \times r$ matrix $\mathbf{H}$ is r. This last equation is equivalent to

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + [\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \cdots \vdots \mathbf{H}_r]\,\mathbf{u}(k)$$

where

$$[\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \cdots \vdots \mathbf{H}_r] = \mathbf{H}, \qquad \mathbf{H}_i = \begin{bmatrix} h_{1i} \\ h_{2i} \\ \vdots \\ h_{ni} \end{bmatrix}, \qquad i = 1, 2, \ldots, r$$

The procedure for designing the state feedback gain matrix $\mathbf{K}$ involves the following two steps:

Step 1. Extend the transformation process [the process that transforms the state equation given by Equation (C–3) into the state equation in the controllable canonical form given by Equation (C–7)] to the case where matrix $\mathbf{H}$ is an $n \times r$ matrix. That is, we transform the given state equation into a controllable canonical form by use of a transformation matrix $\mathbf{T}$, the exact form of which will be given later. By defining

$$\mathbf{x}(k) = \mathbf{T}\hat{\mathbf{x}}(k)$$

the original state equation, Equation (C–1), can be transformed into

$$\hat{\mathbf{x}}(k+1) = \mathbf{T}^{-1}\mathbf{G}\mathbf{T}\hat{\mathbf{x}}(k) + \mathbf{T}^{-1}\mathbf{H}\mathbf{u}(k) = \hat{\mathbf{G}}\hat{\mathbf{x}}(k) + \hat{\mathbf{H}}\mathbf{u}(k) \tag{C–8}$$

where $\hat{\mathbf{G}} = \mathbf{T}^{-1}\mathbf{G}\mathbf{T}$ is in a controllable canonical form and $\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H}$. (This controllable canonical form is slightly different from the usual form, as we shall see later.)

Step 2. By use of a state feedback gain matrix $\mathbf{K}$, the control vector can be given by

$$\mathbf{u}(k) = -\mathbf{K}\mathbf{x}(k) = -\mathbf{K}\mathbf{T}\hat{\mathbf{x}}(k)$$

and the system state equation becomes

$$\hat{\mathbf{x}}(k+1) = (\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{K}\mathbf{T})\hat{\mathbf{x}}(k)$$

We choose matrix $\mathbf{K}$ so that matrix $\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{K}\mathbf{T}$ will have the desired eigenvalues $\mu_1, \mu_2, \ldots, \mu_n$.

C–3 POLE PLACEMENT DESIGN

We shall first discuss the determination of a necessary transformation matrix $\mathbf{T}$ and then determine the state feedback gain matrix $\mathbf{K}$.

Consider the completely state controllable system defined by

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}\mathbf{u}(k) \tag{C–9}$$

where

$\mathbf{x}(k)$ = state vector (n-vector)

$\mathbf{u}(k)$ = control vector (r-vector)

$\mathbf{G} = n \times n$ matrix

$\mathbf{H} = [\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \cdots \vdots \mathbf{H}_r]$ – $n \times r$ matrix

We assume that the rank of matrix $\mathbf{H}$ is r. Thus, the component vectors $\mathbf{H}_1$, $\mathbf{H}_2, \ldots, \mathbf{H}_r$ of matrix $\mathbf{H}$ are linearly independent of each other. Since the system is assumed to be completely state controllable, the rank of the $n \times nr$ controllability matrix

$$[\mathbf{H} \vdots \mathbf{G}\mathbf{H} \vdots \cdots \vdots \mathbf{G}^{n-1}\mathbf{H}]$$

is n. The controllability matrix can be written in an expanded form as follows:

$$[\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \cdots \vdots \mathbf{H}_r \vdots \mathbf{GH}_1 \vdots \mathbf{GH}_2 \vdots \cdots \vdots \mathbf{GH}_r \vdots \cdots \vdots \mathbf{G}^{n-1}\mathbf{H}_1 \vdots \mathbf{G}^{n-1}\mathbf{H}_2 \vdots \cdots \vdots \mathbf{G}^{n-1}\mathbf{H}_r]$$

Let us choose n linearly independent vectors from this $n \times nr$ matrix. Let us begin from the left-hand side of this matrix. Since the first r vectors $\mathbf{H}_1, \mathbf{H}_2, \ldots, \mathbf{H}_r$ are linearly independent of each other, we choose these r vectors first. Then we examine $\mathbf{GH}_1$ if it is linearly independent of the r vectors already chosen. If it is, we have chosen $r + 1$ linearly independent vectors. Next, we examine $\mathbf{GH}_2, \mathbf{GH}_3, \ldots, \mathbf{GH}_r, \ldots$ in the order shown in the expanded controllability matrix until we find altogether n linearly independent vectors. (Since the rank of the controllability matrix is n, there always exist n linearly independent vectors.)

Once we have chosen n linearly independent vectors, we rearrange these vectors in the following way:

$$\mathbf{F} = [\mathbf{H}_1 \vdots \mathbf{GH}_1 \vdots \cdots \vdots \mathbf{G}^{n_1-1}\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \mathbf{GH}_2 \vdots \cdots \vdots \mathbf{G}^{n_2-1}\mathbf{H}_2 \vdots \cdots \vdots \mathbf{H}_r \vdots \mathbf{GH}_r \vdots \cdots \vdots \mathbf{G}^{n_r-1}\mathbf{H}_r] \qquad \text{(C–10)}$$

The numbers n_i are said to be *Kronecker invariant* and satisfy the equation

$$n_1 + n_2 + \cdots + n_r = n$$

We shall define the maximum of $n_1, n_2, \ldots, n_r$ as $n_{\min}$:

$$n_{\min} = \max(n_1, n_2, \ldots, n_r) \qquad \text{(C–11)}$$

We shall refer to this equation later in the discussion of deadbeat response. Next, we compute $\mathbf{F}^{-1}$ and define the η_i th row vector as $\mathbf{f}_i$, where

$$\eta_i = n_1 + n_2 + \cdots + n_i, \qquad i = 1, 2, \ldots, r$$

Then the required transformation matrix $\mathbf{T}$ can be given by

$$\mathbf{T} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_r \end{bmatrix}^{-1} \qquad \text{(C–12)}$$

where

$$\mathbf{S}_i = \begin{bmatrix} \mathbf{f}_i \\ \mathbf{f}_i\mathbf{G} \\ \vdots \\ \mathbf{f}_i\mathbf{G}^{n_i-1} \end{bmatrix}$$

Notice that the transformation matrix $\mathbf{T}$ given by Equation (C–12) is an extension of the transformation matrix given by Equation (C–4).

To simplify the presentation, in what follows we shall consider a simple case where $n = 4$ and $r = 2$. (In this case, only n_1 and n_2 are involved.) (Extension to more general cases is straightforward.) Then the transformation matrix $\mathbf{T}$ becomes a 4×4 matrix. The transformation matrix $\mathbf{T}$ given by Equation (C–12) becomes

$$\mathbf{T} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \end{bmatrix}^{-1}$$

where

$$\mathbf{S}_1 = \begin{bmatrix} \mathbf{f}_1 \\ \vdots \\ \mathbf{f}_1\mathbf{G}^{n_1-1} \end{bmatrix}, \qquad \mathbf{S}_2 = \begin{bmatrix} \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_2\mathbf{G}^{n_2-1} \end{bmatrix}$$

(Note that in the case of $n = 4$ there are three possibilities for the combinations of n_1 and n_2: $n_1 = 1, n_2 = 3$; $n_1 = 2, n_2 = 2$; and $n_1 = 3, n_2 = 1$.) For example, if $n_1 = 2$ and $n_2 = 2$, then matrices $\hat{\mathbf{G}}$ and $\hat{\mathbf{H}}$ become, respectively, as

$$\hat{\mathbf{G}} = \mathbf{T}^{-1}\mathbf{GT} = \left[\begin{array}{cc:cc} 0 & 1 & 0 & 0 \\ -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ \hdashline 0 & 0 & 0 & 1 \\ -a_{21} & -a_{22} & -a_{23} & -a_{24} \end{array}\right], \qquad \text{if } n_1 = 2, n_2 = 2 \qquad \text{(C–13)}$$

and

$$\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H} = \left[\begin{array}{cc} 0 & 0 \\ 1 & b_{12} \\ \hdashline 0 & 0 \\ 0 & 1 \end{array}\right], \qquad \begin{array}{l} \text{if } n_1 = 2, n_2 = 2 \\ \text{(Note: } b_{12} = \mathbf{f}_1\mathbf{GH}_2 = 0 \text{ in this case)} \end{array} \qquad \text{(C–14)}$$

(see Problem C–2). As another example, if $n_1 = 3$ and $n_2 = 1$, then

$$\hat{\mathbf{G}} = \mathbf{T}^{-1}\mathbf{GT} = \left[\begin{array}{ccc:c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ \hdashline -a_{21} & -a_{22} & -a_{23} & -a_{24} \end{array}\right], \qquad \text{if } n_1 = 3, n_2 = 1 \qquad \text{(C–15)}$$

and

$$\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H} = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ 1 & b_{12} \\ \hdashline 0 & 1 \end{array}\right], \qquad \begin{array}{l} \text{if } n_1 = 3, n_2 = 1 \\ \text{(Note: } b_{12} = \mathbf{f}_1\mathbf{G}^2\mathbf{H}_2 \text{ may or} \\ \qquad \text{may not be zero)} \end{array} \qquad \text{(C–16)}$$

(see Problem C–4). In what follows, we shall focus on the case where $n_1 = 2$ and $n_2 = 2$. (Other cases can be handled similarly. For example, for the case where $n_1 = 3$ and $n_2 = 1$, see Problems C–3, C–4, and C–5.) For the case where $n_1 = 2$ and $n_2 = 2$, matrix $\hat{\mathbf{G}} = \mathbf{T}^{-1}\mathbf{GT}$ can be given by Equation (C–13), and the characteristic equation is

$$\begin{aligned} |z\mathbf{I} - \hat{\mathbf{G}}| &= \begin{vmatrix} z & -1 & 0 & 0 \\ a_{11} & z + a_{12} & a_{13} & a_{14} \\ 0 & 0 & z & -1 \\ a_{21} & a_{22} & a_{23} & z + a_{24} \end{vmatrix} \\ &= \begin{vmatrix} z & -1 \\ a_{11} & z + a_{12} \end{vmatrix} \begin{vmatrix} z & -1 \\ a_{23} & z + a_{24} \end{vmatrix} + \begin{vmatrix} z & -1 \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} a_{13} & a_{14} \\ z & -1 \end{vmatrix} \\ &= (z^2 + a_{12}z + a_{11})(z^2 + a_{24}z + a_{23}) - (a_{22}z + a_{21})(a_{14}z + a_{13}) \\ &= 0 \end{aligned} \qquad \text{(C–17)}$$

where we have used Laplace's expansion by the minors. (See Appendix A for the details.) From Equation (C–17) the characteristic equation $|z\mathbf{I} - \hat{\mathbf{G}}| = 0$ becomes

$$|z\mathbf{I} - \hat{\mathbf{G}}| = \begin{vmatrix} z^2 + a_{12}z + a_{11} & a_{14}z + a_{13} \\ a_{22}z + a_{21} & z^2 + a_{24}z + a_{23} \end{vmatrix} = 0 \qquad \text{(C–18)}$$

The eigenvalues of $\hat{\mathbf{G}}$ can be determined by solving this characteristic equation.

Next, we shall determine the state feedback gain matrix $\mathbf{K}$ so that the eigenvalues of $\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{K}$ are $\mu_1, \mu_2, \ldots, \mu_n$, the desired values. Let us define a 2×2 matrix $\mathbf{B}$ such that

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}^{-1}$$

(Note that b_{12} is a constant appearing in $\hat{\mathbf{H}}$ matrix.) In the particular case where $n_1 = 2$ and $n_2 = 2$, the value of b_{12} is equal to 0. Thus, $\mathbf{B} = \mathbf{I}$. For more general cases, matrix $\mathbf{B}$ may not be the identity matrix.

Also, define a 2×4 matrix $\boldsymbol{\Delta}$ such that

$$\boldsymbol{\Delta} = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix} \qquad \text{(C–19)}$$

Then it will be seen that matrix $\mathbf{K}$ can be given by

$$\mathbf{K} = \mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}$$

and the control vector $\mathbf{u}(k)$ can be given by

$$\mathbf{u}(k) = -\mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}\mathbf{x}(k) = -\mathbf{B}\boldsymbol{\Delta}\hat{\mathbf{x}}(k)$$

Thus, the system state equation given by Equation (C–8) becomes

$$\hat{\mathbf{x}}(k+1) = \hat{\mathbf{G}}\hat{\mathbf{x}}(k) - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}\hat{\mathbf{x}}(k) = (\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta})\hat{\mathbf{x}}(k)$$

For the present case, matrix $\hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}$ becomes as follows:

$$\hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & 0 & 0 \\ \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ 0 & 0 & 0 & 0 \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix}$$

Hence,

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -a_{11} - \delta_{11} & -a_{12} - \delta_{12} & -a_{13} - \delta_{13} & -a_{14} - \delta_{14} \\ 0 & 0 & 0 & 1 \\ -a_{21} - \delta_{21} & -a_{22} - \delta_{22} & -a_{23} - \delta_{23} & -a_{24} - \delta_{24} \end{bmatrix}$$

Then, referring to Equation (C–18), the characteristic equation $|z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}|$ becomes

$$\begin{aligned} |z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}| &= \begin{vmatrix} z^2 + (a_{12} + \delta_{12})z + a_{11} + \delta_{11} & (a_{14} + \delta_{14})z + a_{13} + \delta_{13} \\ (a_{22} + \delta_{22})z + a_{21} + \delta_{21} & z^2 + (a_{24} + \delta_{24})z + a_{23} + \delta_{23} \end{vmatrix} \\ &= [z^2 + (a_{12} + \delta_{12})z + a_{11} + \delta_{11}][z^2 + (a_{24} + \delta_{24})z + a_{23} + \delta_{23}] \\ &\quad - [(a_{14} + \delta_{14})z + a_{13} + \delta_{13}][(a_{22} + \delta_{22})z + a_{21} + \delta_{21}] \\ &= 0 \end{aligned} \qquad \text{(C–20)}$$

We desire the eigenvalues of $\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}$ to be μ_1, μ_2, μ_3, and μ_4, or the desired characteristic equation to be

$$(z - \mu_1)(z - \mu_2)(z - \mu_3)(z - \mu_4) = z^4 + \alpha_1 z^3 + \alpha_2 z^2 + \alpha_3 z + \alpha_4 = 0 \qquad \text{(C–21)}$$

If we equate the coefficients of equal powers of z of the two characteristic equations, Equations (C–20) and (C–21), we obtain the following equations:

$$\begin{aligned} a_{12} + \delta_{12} + a_{24} + \delta_{24} &= \alpha_1 \\ a_{11} + \delta_{11} + (a_{12} + \delta_{12})(a_{24} + \delta_{24}) + a_{23} + \delta_{23} - (a_{14} + \delta_{14})(a_{22} + \delta_{22}) &= \alpha_2 \\ (a_{11} + \delta_{11})(a_{24} + \delta_{24}) + (a_{12} + \delta_{12})(a_{23} + \delta_{23}) & \\ \quad - (a_{13} + \delta_{13})(a_{22} + \delta_{22}) - (a_{21} + \delta_{21})(a_{14} + \delta_{14}) &= \alpha_3 \\ (a_{11} + \delta_{11})(a_{23} + \delta_{23}) - (a_{13} + \delta_{13})(a_{21} + \delta_{21}) &= \alpha_4 \end{aligned}$$

Notice that we have eight δ variables and four equations. Hence, the values of δ_{11}, δ_{12}, δ_{13}, δ_{14}, δ_{21}, δ_{22}, δ_{23}, and δ_{24} cannot be determined uniquely. There are many possible sets of values $\delta_{11}, \delta_{12}, \ldots, \delta_{24}$ and thus matrix $\boldsymbol{\Delta}$ is not unique. Any matrix $\boldsymbol{\Delta}$ whose elements satisfy the foregoing four equations is acceptable.

Once matrix $\boldsymbol{\Delta}$ is chosen, the required state feedback gain matrix $\mathbf{K}$ is given by

$$\mathbf{K} = \mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}$$

and the state feedback control vector is

$$\mathbf{u}(k) = -\mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}\mathbf{x}(k)$$

and the state equation given by Equation (C–9) becomes

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) - \mathbf{H}\mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}\mathbf{x}(k) = (\mathbf{G} - \mathbf{H}\mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1})\mathbf{x}(k)$$

As a matter of course, note that

$$|\mathbf{G} - \mathbf{H}\mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}| = |\mathbf{T}^{-1}||\mathbf{G} - \mathbf{H}\mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}||\mathbf{T}| = |\mathbf{T}^{-1}\mathbf{G}\mathbf{T} - \mathbf{T}^{-1}\mathbf{H}\mathbf{B}\boldsymbol{\Delta}| = |\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}|$$

For a given set of desired eigenvalues $\mu_1, \mu_2, \ldots, \mu_n$, we have the corresponding coefficients $\alpha_1, \alpha_2, \ldots, \alpha_n$ in the characteristic equation $|z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}| = 0$. For the given $\alpha_1, \alpha_2, \ldots, \alpha_n$, it is possible to choose a matrix $\boldsymbol{\Delta}$ that is not unique. (This means that we have some freedom to satisfy other requirements, *if any*.)

If the deadbeat response is desired, we require $\mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$. The desired characteristic equation given by Equation (C–21) becomes

$$z^4 = 0$$

Notice that if we choose, for example,

$$\Delta = \begin{bmatrix} -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ * & * & -a_{23} & -a_{24} \end{bmatrix} \tag{C–22}$$

where the elements indicated by asterisks are arbitrary constants, then $\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta$ becomes

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ ** & ** & 0 & 0 \end{bmatrix}$$

where the elements indicated by the double asterisks are arbitrary constants.

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ ** & 0 & 0 & 0 \\ 0 & ** & 0 & 0 \end{bmatrix}$$

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)^3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & ** & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Thus, the deadbeat response is obtained. Matrix Δ given by Equation (C–22) is not unique because different choices of elements can yield the deadbeat response. Hence, more than one state feedback gain matrix $\mathbf{K}$ exists that will yield the deadbeat response. This is expected, since we have two control signals $u_1(k)$ and $u_2(k)$ available, instead of just one control signal.

It is important to note that if we choose

$$\Delta = \begin{bmatrix} -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ -a_{21} & -a_{22} & -a_{23} & -a_{24} \end{bmatrix} \tag{C–23}$$

then

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)^2 = \mathbf{0}$$

Thus, $(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)^k$ becomes zero for $k = 2, 3, 4, \ldots$. The deadbeat response is achieved in two sampling periods. In fact, in general, by choosing the elements of Δ in the manner given by Equation (C–23), the deadbeat response can be achieved in $n_{\min}$ steps rather than n steps, where

$$n_{\min} = \max(n_1, n_2, \ldots, n_r)$$

Since $n_1 + n_2 + \cdots + n_r = n$, we note that $n_{\min}$ is always less than n.

Extension to the More General Case. Thus far, we have given detailed discussions for the case where $n = 4$ ($n_1 = n_2 = 2$) and $r = 2$. Extension of the preceding discussions to the more general case is straightforward. For example, consider the case where $n = 6$ and $r = 3$. For this case,

$$n_1 + n_2 + n_3 = 6$$

and we have several possible combinations of n_1, n_2, and n_3.

Now consider the case where $n_1 = 3$, $n_2 = 2$, and $n_3 = 1$. The modified 6×6 controllability matrix $\mathbf{F}$ for this case is

$$\mathbf{F} = [\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{G}^2\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \mathbf{G}\mathbf{H}_2 \vdots \mathbf{H}_3]$$

Define

$$\mathbf{F}^{-1} = \begin{bmatrix} *** \\ *** \\ \mathbf{f}_1 \\ *** \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{bmatrix} \begin{matrix} \} n_1 = 3 \\ \\ \\ \} n_2 = 2 \\ \\ \} n_3 = 1 \end{matrix}$$

where a row of asterisks denotes a row vector. Then the transformation matrix $\mathbf{T}$ can be formed as follows:

$$\mathbf{T} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \mathbf{S}_3 \end{bmatrix}^{-1}$$

where

$$\mathbf{S}_1 = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_1\mathbf{G} \\ \mathbf{f}_1\mathbf{G}^2 \end{bmatrix}, \qquad \mathbf{S}_2 = \begin{bmatrix} \mathbf{f}_2 \\ \mathbf{f}_2\mathbf{G} \end{bmatrix}, \qquad \mathbf{S}_3 = \mathbf{f}_3$$

Then the matrices $\hat{\mathbf{G}}$ and $\hat{\mathbf{H}}$ will have the following forms:

$$\hat{\mathbf{G}} = \left[\begin{array}{ccc:ccc} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ -a_{11} & -a_{12} & -a_{13} & -a_{14} & -a_{15} & -a_{16} \\ \hdashline 0 & 0 & 0 & 0 & 1 & 0 \\ -a_{21} & -a_{22} & -a_{23} & -a_{24} & -a_{25} & -a_{26} \\ \hdashline -a_{31} & -a_{32} & -a_{33} & -a_{34} & -a_{35} & -a_{36} \end{array}\right]$$

$$\hat{\mathbf{H}} = \left[\begin{array}{ccc} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & b_{12} & b_{13} \\ \hline 0 & 0 & 0 \\ 0 & 1 & b_{23} \\ \hline 0 & 0 & 1 \end{array}\right]$$

where $b_{12} = \mathbf{f}_1\mathbf{G}^2\mathbf{H}_2$, $b_{13} = \mathbf{f}_1\mathbf{G}^2\mathbf{H}_3$, and $b_{23} = \mathbf{f}_2\mathbf{G}\mathbf{H}_3$. These values may or may not be zero. (Notice that in matrix $\hat{\mathbf{G}}$ the principal minors are in the controllable canonical form.) The state feedback gain matrix $\mathbf{K}$ is given as follows:

$$\mathbf{K} = \mathbf{B}\Delta\mathbf{T}^{-1}$$

where

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} & b_{13} \\ 0 & 1 & b_{23} \\ 0 & 0 & 1 \end{bmatrix}^{-1}$$

and

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} & \delta_{15} & \delta_{16} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} & \delta_{25} & \delta_{26} \\ \delta_{31} & \delta_{32} & \delta_{33} & \delta_{34} & \delta_{35} & \delta_{36} \end{bmatrix}$$

Notice that

$$\hat{\mathbf{H}}\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & b_{12} & b_{13} \\ 0 & 0 & 0 \\ 0 & 1 & b_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & b_{12} & b_{13} \\ 0 & 1 & b_{23} \\ 0 & 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

The effect of postmultiplying matrix $\mathbf{B}$ to matrix $\hat{\mathbf{H}}$ is to eliminate the b_{ij} from the product matrix $\hat{\mathbf{H}}\mathbf{B}$.

Note that if $\mathbf{u}(k)$ is an r-vector the general form of $\mathbf{B}$ matrix is

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} & \cdots & b_{1r} \\ 0 & 1 & \cdots & b_{2r} \\ 0 & 0 & \cdots & b_{3r} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}^{-1} \tag{C–24}$$

where the constants b_{ij}'s are those that will appear in the $n \times r$ matrix $\hat{\mathbf{H}}$. (The elements of $\hat{\mathbf{H}}\mathbf{B}$ are either 0 or 1.)

Example C–1

Consider the system

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}\mathbf{u}(k)$$

where

$\mathbf{x}(k)$ = state vector (3-vector)

$\mathbf{u}(k)$ = control vector (2-vector)

and

$$\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.25 & 0 & 0.5 \end{bmatrix}, \qquad \mathbf{H} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$$

It is desired to determine the state feedback gain matrix $\mathbf{K}$ so that the response to the initial state $\mathbf{x}(0)$ is deadbeat. Note that with state feedback $\mathbf{u}(k) = -\mathbf{K}\mathbf{x}(k)$ the system equation becomes

$$\mathbf{x}(k+1) = (\mathbf{G} - \mathbf{H}\mathbf{K})\mathbf{x}(k) \tag{C–25}$$

We shall first examine the controllability matrix:

$$[\mathbf{H} \vdots \mathbf{G}\mathbf{H} \vdots \mathbf{G}^2\mathbf{H}] = [\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_2 \vdots \mathbf{G}^2\mathbf{H}_1 \vdots \mathbf{G}^2\mathbf{H}_2]$$

$$= \left[\begin{array}{cccccc} \boxed{0} & \boxed{1} & \boxed{0} & 0 & 1 & 0 \\ \boxed{0} & \boxed{0} & \boxed{1} & 0 & 0.5 & -0.25 \\ \boxed{1} & \boxed{0} & \boxed{0.5} & -0.25 & 0.25 & -0.125 \end{array}\right]$$

Clearly, the rank of this controllability matrix is 3. Therefore, arbitrary pole placement is possible. We now choose three linearly independent vectors starting from the left end. These vectors are shown enclosed by dashed lines. (The three linearly independent vectors chosen are $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{G}\mathbf{H}_1$.) Now we rearrange these three vectors according to Equation (C–10) and define matrix $\mathbf{F}$ as follows:

$$\mathbf{F} = [\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{H}_2]$$

We note that $n_1 = 2$ and $n_2 = 1$.

Rewriting matrix $\mathbf{F}$, we have

$$\mathbf{F} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0.5 & 0 \end{bmatrix}$$

The inverse of matrix $\mathbf{F}$ becomes

$$\mathbf{F}^{-1} = \begin{bmatrix} 0 & -0.5 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

We now define the η_ith row vector of $\mathbf{F}^{-1}$ as $\mathbf{f}_i$, where $\eta_1 = n_1$ and $\eta_2 = n_1 + n_2$. Since $n_1 = 2$ and $n_2 = 1$, the vectors $\mathbf{f}_1$ and $\mathbf{f}_2$ are the second and third row vectors, respectively. That is,

$$\mathbf{f}_1 = [0 \quad 1 \quad 0]$$

$$\mathbf{f}_2 = [1 \quad 0 \quad 0]$$

Next, define the transformation matrix $\mathbf{T}$ by

$$\mathbf{T} = \begin{bmatrix} \mathbf{S}_1 \\ \hline \mathbf{S}_2 \end{bmatrix}^{-1}$$

where

$$\mathbf{S}_1 = \begin{bmatrix} \mathbf{f}_1 \\ \hline \mathbf{f}_1\mathbf{G} \end{bmatrix}, \qquad \mathbf{S}_2 = \mathbf{f}_2$$

Hence,

$$\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

and

$$\mathbf{T}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

With this transformation matrix $\mathbf{T}$, we define

$$\mathbf{x}(k) = \mathbf{T}\hat{\mathbf{x}}(k)$$

Then

$$\begin{aligned} \mathbf{T}^{-1}\mathbf{G}\mathbf{T} = \hat{\mathbf{G}} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -0.25 & 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \\ &= \left[\begin{array}{cc:c} 0 & 1 & 0 \\ 0 & 0.5 & -0.25 \\ \hdashline 1 & 0 & 0 \end{array}\right] \end{aligned}$$

Also,

$$\begin{aligned} \mathbf{T}^{-1}\mathbf{H} = \hat{\mathbf{H}} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = \left[\begin{array}{cc} 0 & 0 \\ 1 & 0 \\ \hdashline 0 & 1 \end{array}\right] \end{aligned}$$

Next, we determine the state feedback gain matrix $\mathbf{K}$, where

$$\mathbf{K} = \mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1}$$

From Equation (C–24), matrix $\mathbf{B}$ for the present case is a 2×2 matrix. Noting that $b_{12} = 0$, we have

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

For the present case, $\boldsymbol{\Delta}$ is a 2×3 matrix:

$$\boldsymbol{\Delta} = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \end{bmatrix}$$

Now we determine matrix $\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}$:

$$\begin{aligned} \hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta} &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.5 & -0.25 \\ 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0.5 & -0.25 \\ 1 & 0 & 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 0 \\ \delta_{11} & \delta_{12} & \delta_{13} \\ \delta_{21} & \delta_{22} & \delta_{23} \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & 0 \\ -\delta_{11} & 0.5 - \delta_{12} & -0.25 - \delta_{13} \\ 1 - \delta_{21} & -\delta_{22} & -\delta_{23} \end{bmatrix} \end{aligned}$$

The characteristic equation $|z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}| = 0$ is given as follows:

$$\begin{aligned} |z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}| &= \begin{vmatrix} z & -1 & 0 \\ \delta_{11} & z - 0.5 + \delta_{12} & 0.25 + \delta_{13} \\ -1 + \delta_{21} & \delta_{22} & z + \delta_{23} \end{vmatrix} \\ &= 0 \end{aligned}$$

Since the deadbeat response is desired, the desired characteristic equation is

$$z^3 = 0$$

Note that the choice of the δ's is not unique and matrix $\boldsymbol{\Delta}$ is not unique. Suppose we choose the δ's so that

$$\begin{aligned} &\delta_{11} = 0, \quad \delta_{12} = 0.5, \quad \delta_{13} = -0.25 \\ &\delta_{21} = 1, \quad \delta_{22} = 0, \quad \delta_{23} = 0 \end{aligned}$$

Then

$$|z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta}| = \begin{vmatrix} z & -1 & 0 \\ 0 & z & 0 \\ 0 & 0 & z \end{vmatrix} = z^3 = 0$$

and thus

$$\boldsymbol{\Delta} = \begin{bmatrix} 0 & 0.5 & -0.25 \\ 1 & 0 & 0 \end{bmatrix}$$

is acceptable. Then matrix $\mathbf{K}$ is obtained as follows:

$$\begin{aligned} \mathbf{K} = \mathbf{B}\boldsymbol{\Delta}\mathbf{T}^{-1} &= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & -0.25 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} -0.25 & 0 & 0.5 \\ 0 & 1 & 0 \end{bmatrix} \end{aligned}$$

With this choice of matrix $\mathbf{K}$, $(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta})^k = 0$ for $k \geq n_{\min}$, where

$$n_{\min} = \max(n_1, n_2) = \max(2, 1) = 2$$

In fact,

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta})^2 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Thus,

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta})^k = \mathbf{0}, \qquad k = 2, 3, 4, \ldots$$

Note that

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta} = \mathbf{T}^{-1}\mathbf{G}\mathbf{T} - \mathbf{T}^{-1}\mathbf{H}\mathbf{B}\boldsymbol{\Delta} = \mathbf{T}^{-1}\mathbf{G}\mathbf{T} - \mathbf{T}^{-1}\mathbf{H}\mathbf{K}\mathbf{T} = \mathbf{T}^{-1}(\mathbf{G} - \mathbf{H}\mathbf{K})\mathbf{T}$$

Referring to Equation (C–25) and its solution $\mathbf{x}(k) = (\mathbf{G} - \mathbf{H}\mathbf{K})^k\mathbf{x}(0)$, we have $\mathbf{x}(k) = \mathbf{0}$ for $k = 2, 3, 4, \ldots$, since

$$\mathbf{G} - \mathbf{HK} = \mathbf{T}(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)\mathbf{T}^{-1}$$

$$(\mathbf{G} - \mathbf{HK})^2 = \mathbf{T}(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)\mathbf{T}^{-1}\mathbf{T}(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)\mathbf{T}^{-1} = \mathbf{T}(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta)^2\mathbf{T}^{-1} = \mathbf{0}$$

and

$$\mathbf{x}(k) = (\mathbf{G} - \mathbf{HK})^k\mathbf{x}(0) = \mathbf{0}, \qquad k = 2, 3, 4 \ldots$$

We have thus designed the state feedback gain matrix $\mathbf{K}$ so that the system's response to any initial state $\mathbf{x}(0)$ is deadbeat. The state $\mathbf{x}(k)$ can be transferred to the origin in at most two sampling periods. [Note that if the control signal $u(k)$ were a scalar then it would take at most three sampling periods, rather than at most two sampling periods, for deadbeat response.]

EXAMPLE PROBLEMS AND SOLUTIONS

Problem C–1

Consider the system given by

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}_1 u(k)$$

where

$\mathbf{x}(k)$ = state vector (n-vector)

$u(k)$ = control signal (scalar)

$\mathbf{G} = n \times n$ matrix

$\mathbf{H}_1 = n \times 1$ matrix

Assume that the system is completely state controllable.

Define

$$[\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \cdots \vdots \mathbf{G}^{n-1}\mathbf{H}_1]^{-1} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_n \end{bmatrix}$$

where the $\mathbf{f}_i$'s ($i = 1, 2, \ldots, n$) are row vectors. Define also

$$\mathbf{T}_1 = \begin{bmatrix} \mathbf{f}_n \\ \mathbf{f}_n\mathbf{G} \\ \vdots \\ \mathbf{f}_n\mathbf{G}^{n-1} \end{bmatrix}^{-1}$$

Show that

$$\mathbf{T}_1^{-1}\mathbf{G}\mathbf{T}_1 = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ -a_n & -a_{n-1} & -a_{n-2} & \cdots & -a_1 \end{bmatrix} \tag{C–26}$$

and

$$\mathbf{T}_1^{-1}\mathbf{H}_1 = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} \tag{C–27}$$

where the a_i's are the coefficients appearing in the characteristic polynomial of $\mathbf{G}$, or

$$|z\mathbf{I} - \mathbf{G}| = z^n + a_1 z^{n-1} + \cdots + a_{n-1}z + a_n$$

Solution We shall prove Equations (C–26) and (C–27) for the case where $n = 3$. (Extension of the derivation to an arbitrary positive integer n is straightforward.) Thus, we shall derive that

$$\mathbf{T}_1^{-1}\mathbf{G}\mathbf{T}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix} \tag{C–28}$$

Since

$$\mathbf{T}_1^{-1} = \begin{bmatrix} \mathbf{f}_3 \\ \mathbf{f}_3\mathbf{G} \\ \mathbf{f}_3\mathbf{G}^2 \end{bmatrix}$$

it is possible to rewrite Equation (C–28) as follows:

$$\mathbf{T}_1^{-1}\mathbf{G} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix}\begin{bmatrix} \mathbf{f}_3 \\ \mathbf{f}_3\mathbf{G} \\ \mathbf{f}_3\mathbf{G}^2 \end{bmatrix} \tag{C–29}$$

Now consider the conjugate transpose of the right-hand side of Equation (C–29). Noting that for physical systems the coefficients $a_1, a_2, \ldots, a_n$ of the characteristic polynomial are real, we have

$$[\mathbf{f}_3^* \vdots \mathbf{G}^*\mathbf{f}_3^* \vdots (\mathbf{G}^*)^2\mathbf{f}_3^*]\begin{bmatrix} 0 & 0 & -a_3 \\ 1 & 0 & -a_2 \\ 0 & 1 & -a_1 \end{bmatrix} = [\mathbf{G}^*\mathbf{f}_3^* \vdots (\mathbf{G}^*)^2\mathbf{f}_3^* \vdots -a_3\mathbf{f}_3^* - a_2\mathbf{G}^*\mathbf{f}_3^* - a_1(\mathbf{G}^*)^2\mathbf{f}_3^*]$$

Note that $\mathbf{G}^*$ satisfies its own characteristic equation:

$$\phi(\mathbf{G}^*) = (\mathbf{G}^*)^3 + a_1(\mathbf{G}^*)^2 + a_2\mathbf{G}^* + a_3\mathbf{I} = \mathbf{0}$$

Hence,

$$-[a_3\mathbf{I} + a_2\mathbf{G}^* + a_1(\mathbf{G}^*)^2]\mathbf{f}_3^* = (\mathbf{G}^*)^3\mathbf{f}_3^*$$

Consequently,

$$[\mathbf{f}_3^* \vdots \mathbf{G}^*\mathbf{f}_3^* \vdots (\mathbf{G}^*)^2\mathbf{f}_3^*]\begin{bmatrix} 0 & 0 & -a_3 \\ 1 & 0 & -a_2 \\ 0 & 1 & -a_1 \end{bmatrix} = [\mathbf{G}^*\mathbf{f}_3^* \vdots (\mathbf{G}^*)^2\mathbf{f}_3^* \vdots (\mathbf{G}^*)^3\mathbf{f}_3^*] = \mathbf{G}^*[\mathbf{f}_3^* \vdots \mathbf{G}^*\mathbf{f}_3^* \vdots (\mathbf{G}^*)^2\mathbf{f}_3^*]$$

Taking the conjugate transpose of both sides of this last equation, we obtain

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix}\begin{bmatrix} \mathbf{f}_3 \\ \mathbf{f}_3\mathbf{G} \\ \mathbf{f}_3\mathbf{G}^2 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_3 \\ \mathbf{f}_3\mathbf{G} \\ \mathbf{f}_3\mathbf{G}^2 \end{bmatrix}\mathbf{G} = \mathbf{T}_1^{-1}\mathbf{G}$$

which is Equation (C–29). Thus, we have shown that Equation (C–28) is true, or

$$\mathbf{T}_1^{-1}\mathbf{G}\mathbf{T}_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_3 & -a_2 & -a_1 \end{bmatrix}$$

Next, we shall show that

$$\mathbf{T}_1^{-1}\mathbf{H}_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Since

$$[\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{G}^2\mathbf{H}_1]^{-1} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{bmatrix}$$

we obtain

$$\mathbf{I} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{bmatrix}[\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{G}^2\mathbf{H}_1]$$

or

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{f}_1\mathbf{H}_1 & \mathbf{f}_1\mathbf{G}\mathbf{H}_1 & \mathbf{f}_1\mathbf{G}^2\mathbf{H}_1 \\ \mathbf{f}_2\mathbf{H}_1 & \mathbf{f}_2\mathbf{G}\mathbf{H}_1 & \mathbf{f}_2\mathbf{G}^2\mathbf{H}_1 \\ \mathbf{f}_3\mathbf{H}_1 & \mathbf{f}_3\mathbf{G}\mathbf{H}_1 & \mathbf{f}_3\mathbf{G}^2\mathbf{H}_1 \end{bmatrix}$$

Hence,

$$\mathbf{f}_3\mathbf{H}_1 = 0, \qquad \mathbf{f}_3\mathbf{G}\mathbf{H}_1 = 0, \qquad \mathbf{f}_3\mathbf{G}^2\mathbf{H}_1 = 1$$

By using these equations, we obtain

$$\mathbf{T}_1^{-1}\mathbf{H}_1 = \begin{bmatrix} \mathbf{f}_3 \\ \mathbf{f}_3\mathbf{G} \\ \mathbf{f}_3\mathbf{G}^2 \end{bmatrix}\mathbf{H}_1 = \begin{bmatrix} \mathbf{f}_3\mathbf{H}_1 \\ \mathbf{f}_3\mathbf{G}\mathbf{H}_1 \\ \mathbf{f}_3\mathbf{G}^2\mathbf{H}_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Note that the extension of the derivations presented here to the case of an arbitrary positive integer n can be made easily.

Problem C–2

Consider the system

$$\mathbf{x}(k+1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}\mathbf{u}(k)$$

where

$\mathbf{x}(k)$ = state vector (4-vector)

$\mathbf{u}(k)$ = control vector (2-vector)

and

$$\mathbf{G} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -1 & 2 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{H} = [\mathbf{H}_1 \vdots \mathbf{H}_2] = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Referring to Equation (C–10), obtain matrix $\mathbf{F}$. Then, by use of the transformation matrix $\mathbf{T}$ defined by Equation (C–12), determine matrices $\hat{\mathbf{G}} = \mathbf{T}^{-1}\mathbf{G}\mathbf{T}$ and $\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H}$. Finally, derive Equation (C–14).

Solution We shall first write the controllability matrix as follows:

$$[\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_2 \vdots \mathbf{G}^2\mathbf{H}_1 \vdots \mathbf{G}^2\mathbf{H}_2 \vdots \mathbf{G}^3\mathbf{H}_1 \vdots \mathbf{G}^3\mathbf{H}_2]$$

$$= \begin{bmatrix} \boxed{1} & \boxed{0} & \boxed{-1} & \boxed{0} & 2 & 0 & -5 & 2 \\ \boxed{0} & \boxed{0} & \boxed{1} & \boxed{0} & -3 & 2 & 11 & -4 \\ \boxed{0} & \boxed{0} & \boxed{0} & \boxed{2} & 3 & 0 & -6 & 4 \\ \boxed{0} & \boxed{1} & \boxed{1} & \boxed{1} & 0 & 1 & 2 & 1 \end{bmatrix}$$

We now choose four linearly independent vectors from this 4×8 matrix, starting from the left end. (These vectors are shown enclosed by dashed lines.) The four linearly independent vectors chosen are $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{G}\mathbf{H}_1$, and $\mathbf{G}\mathbf{H}_2$. Next, we rearrange these four vectors according to Equation (C–10) and define matrix $\mathbf{F}$ as follows:

$$\mathbf{F} = [\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \mathbf{G}\mathbf{H}_2]$$

(Note that in this case $n_1 = 2$ and $n_2 = 2$.) Thus,

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

The inverse of this matrix is given by

$$\mathbf{F}^{-1} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & -0.5 & 1 \\ 0 & 0 & 0.5 & 0 \end{bmatrix}$$

Since in this case $n_1 = 2$ and $n_2 = 2$, we define the second row vector of $\mathbf{F}^{-1}$ as $\mathbf{f}_1$ and the fourth row vector as $\mathbf{f}_2$. Then

$$\mathbf{f}_1 = [0 \quad 1 \quad 0 \quad 0]$$

$$\mathbf{f}_2 = [0 \quad 0 \quad 0.5 \quad 0]$$

The transformation matrix $\mathbf{T}$ is given by

$$\mathbf{T} = \begin{bmatrix} \mathbf{S}_1 \\ \hline \mathbf{S}_2 \end{bmatrix}^{-1}$$

where

$$\mathbf{S}_1 = \begin{bmatrix} \mathbf{f}_1 \\ \hline \mathbf{f}_1\mathbf{G} \end{bmatrix}, \qquad \mathbf{S}_2 = \begin{bmatrix} \mathbf{f}_2 \\ \hline \mathbf{f}_2\mathbf{G} \end{bmatrix}$$

Hence,

$$\mathbf{T} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_1\mathbf{G} \\ \mathbf{f}_2 \\ \mathbf{f}_2\mathbf{G} \end{bmatrix}^{-1} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & -0.5 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 2 & 1 & -2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ -0.5 & 0 & 1 & 1 \end{bmatrix}$$

With this transformation matrix T we obtain

$$\hat{\mathbf{G}} = \mathbf{T}^{-1}\mathbf{G}\mathbf{T} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & -0.5 & 1 \end{bmatrix}\begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -1 & 2 \\ 1 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 1 & -2 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ -0.5 & 0 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & -3 & 2 & 2 \\ 0 & 0 & 0 & 1 \\ 1.5 & 1.5 & -1 & 0 \end{bmatrix}$$

and

$$\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0 & 0.5 & -0.5 & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

Notice that, when $n_1 = n_2 = 2$, matrix $\hat{\mathbf{G}}$ has the form given by Equation (C–13) and matrix $\hat{\mathbf{H}}$ has the form given by Equation (C–14), or

$$\hat{\mathbf{G}} = \left[\begin{array}{cc:cc} 0 & 1 & 0 & 0 \\ -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ \hdashline 0 & 0 & 0 & 1 \\ -a_{21} & -a_{22} & -a_{23} & -a_{24} \end{array}\right], \qquad \hat{\mathbf{H}} = \left[\begin{array}{cc} 0 & 0 \\ 1 & b_{12} \\ \hdashline 0 & 0 \\ 0 & 1 \end{array}\right]$$

(Note that b_{12} is zero in this case.)

Finally, we shall derive Equation (C–14). Notice that

$$\mathbf{F}^{-1}\mathbf{F} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{f}_1 \\ \mathbf{m}_2 \\ \mathbf{f}_2 \end{bmatrix}[\mathbf{H}_1 \quad \mathbf{G}\mathbf{H}_1 \quad \mathbf{H}_2 \quad \mathbf{G}\mathbf{H}_2]$$

$$= \begin{bmatrix} \mathbf{m}_1\mathbf{H}_1 & \mathbf{m}_1\mathbf{G}\mathbf{H}_1 & \mathbf{m}_1\mathbf{H}_2 & \mathbf{m}_1\mathbf{G}\mathbf{H}_2 \\ \mathbf{f}_1\mathbf{H}_1 & \mathbf{f}_1\mathbf{G}\mathbf{H}_1 & \mathbf{f}_1\mathbf{H}_2 & \mathbf{f}_1\mathbf{G}\mathbf{H}_2 \\ \mathbf{m}_2\mathbf{H}_1 & \mathbf{m}_2\mathbf{G}\mathbf{H}_1 & \mathbf{m}_2\mathbf{H}_2 & \mathbf{m}_2\mathbf{G}\mathbf{H}_2 \\ \mathbf{f}_2\mathbf{H}_1 & \mathbf{f}_2\mathbf{G}\mathbf{H}_1 & \mathbf{f}_2\mathbf{H}_2 & \mathbf{f}_2\mathbf{G}\mathbf{H}_2 \end{bmatrix}$$

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

where $\mathbf{m}_1$ and $\mathbf{m}_2$ are the first row vector and the third row vector of $\mathbf{F}^{-1}$, respectively. Since $\mathbf{F}^{-1}\mathbf{F}$ is an identity matrix, we have $\mathbf{f}_1\mathbf{H}_1 = 0$, $\mathbf{f}_1\mathbf{H}_2 = 0$, $\mathbf{f}_1\mathbf{G}\mathbf{H}_1 = 1$, $\mathbf{f}_1\mathbf{G}\mathbf{H}_2 = 0$, $\mathbf{f}_2\mathbf{H}_1 = 0$, $\mathbf{f}_2\mathbf{H}_2 = 0$, $\mathbf{f}_2\mathbf{G}\mathbf{H}_1 = 0$, and $\mathbf{f}_2\mathbf{G}\mathbf{H}_2 = 1$. Thus, we have

$$\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_1\mathbf{G} \\ \mathbf{f}_2 \\ \mathbf{f}_2\mathbf{G} \end{bmatrix}[\mathbf{H}_1 \quad \mathbf{H}_2] = \begin{bmatrix} \mathbf{f}_1\mathbf{H}_1 & \mathbf{f}_1\mathbf{H}_2 \\ \mathbf{f}_1\mathbf{G}\mathbf{H}_1 & \mathbf{f}_1\mathbf{G}\mathbf{H}_2 \\ \mathbf{f}_2\mathbf{H}_1 & \mathbf{f}_2\mathbf{H}_2 \\ \mathbf{f}_2\mathbf{G}\mathbf{H}_1 & \mathbf{f}_2\mathbf{G}\mathbf{H}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}$$

which is Equation (C–14).

Problem C–3

Consider the system defined by Equation (C–8):

$$\hat{\mathbf{x}}(k+1) = \mathbf{T}^{-1}\mathbf{G}\mathbf{T}\hat{\mathbf{x}}(k) + \mathbf{T}^{-1}\mathbf{H}\mathbf{u}(k) = \hat{\mathbf{G}}\hat{\mathbf{x}}(k) + \hat{\mathbf{H}}\mathbf{u}(k)$$

where the transformation matrix T is defined by Equation (C–12). Assume that the matrix $\hat{\mathbf{G}}$ is given by Equation (C–15) and the matrix $\hat{\mathbf{H}}$ is given by Equation (C–16). That is,

$$\hat{\mathbf{G}} = \left[\begin{array}{ccc:c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ \hdashline -a_{21} & -a_{22} & -a_{23} & -a_{24} \end{array}\right], \qquad \hat{\mathbf{H}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & b_{12} \\ 0 & 1 \end{bmatrix}$$

Show that

$$|z\mathbf{I} - \hat{\mathbf{G}}| = \begin{vmatrix} z^3 + a_{13}z^2 + a_{12}z + a_{11} & a_{14} \\ a_{23}z^2 + a_{22}z + a_{21} & z + a_{24} \end{vmatrix}$$

and

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a_{11} - \delta_{11} & -a_{12} - \delta_{12} & -a_{13} - \delta_{13} & -a_{14} - \delta_{14} \\ -a_{21} - \delta_{21} & -a_{22} - \delta_{22} & -a_{23} - \delta_{23} & -a_{24} - \delta_{24} \end{bmatrix}$$

where

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}^{-1}, \qquad \boldsymbol{\Delta} = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix}$$

Show also that if we choose, for example,

$$\boldsymbol{\Delta} = \begin{bmatrix} -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ * & * & * & -a_{24} \end{bmatrix} \tag{C–30}$$

where the elements shown by asterisks are arbitrary constants, the system will exhibit the deadbeat response to any initial state $\mathbf{x}(0)$; that is,

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta})^k = \mathbf{0}, \qquad k = 4, 5, 6, \ldots$$

Show also that if we choose

$$\boldsymbol{\Delta} = \begin{bmatrix} -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ -a_{21} & -a_{22} & -a_{23} & -a_{24} \end{bmatrix} \tag{C–31}$$

then

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\boldsymbol{\Delta})^k = \mathbf{0}$$

for $k \geq n_{\min}$, where

$$n_{\min} = \max(n_1, n_2) = \max(3, 1) = 3$$

Solution For the case where $\hat{\mathbf{G}}$ is as given by Equation (C–15), we have

$$|z\mathbf{I} - \hat{\mathbf{G}}| = \begin{vmatrix} z & -1 & 0 & 0 \\ 0 & z & -1 & 0 \\ a_{11} & a_{12} & z + a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & z + a_{24} \end{vmatrix}$$

Expanding this determinant using the Laplace's expansion formula, we obtain

$$|z\mathbf{I} - \hat{\mathbf{G}}| = \begin{vmatrix} z & -1 \\ 0 & z \end{vmatrix} \begin{vmatrix} z + a_{13} & a_{14} \\ a_{23} & z + a_{24} \end{vmatrix} - \begin{vmatrix} z & -1 \\ a_{11} & a_{12} \end{vmatrix} \begin{vmatrix} -1 & 0 \\ a_{23} & z + a_{24} \end{vmatrix} + \begin{vmatrix} z & -1 \\ a_{21} & a_{22} \end{vmatrix} \begin{vmatrix} -1 & 0 \\ z + a_{13} & a_{14} \end{vmatrix}$$

$$= (z + a_{24})(z^3 + a_{13}z^2 + a_{12}z + a_{11}) - a_{14}(a_{23}z^2 + a_{22}z + a_{21})$$

Hence, the determinant $|z\mathbf{I} - \hat{\mathbf{G}}|$ may be written as follows:

$$|z\mathbf{I} - \hat{\mathbf{G}}| = \begin{vmatrix} z^3 + a_{13}z^2 + a_{12}z + a_{11} & a_{14} \\ a_{23}z^2 + a_{22}z + a_{21} & z + a_{24} \end{vmatrix} \tag{C–32}$$

Next, compute

$$\hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & b_{12} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix}$$

(Notice that the effect of postmultiplying matrix $\hat{\mathbf{H}}$ by matrix $\mathbf{B}$ is to eliminate b_{12} from the product matrix $\hat{\mathbf{H}}\mathbf{B}$.) Thus,

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -a_{11} - \delta_{11} & -a_{12} - \delta_{12} & -a_{13} - \delta_{13} & -a_{14} - \delta_{14} \\ -a_{21} - \delta_{21} & -a_{22} - \delta_{22} & -a_{23} - \delta_{23} & -a_{24} - \delta_{24} \end{bmatrix}$$

If we choose $\mathbf{\Delta}$ as given by Equation (C–30), then

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ * & * & * & 0 \end{bmatrix}$$

where the elements shown by asterisks are arbitrary constants. Notice that

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta})^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & * & * & 0 \end{bmatrix}$$

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta})^3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & * & 0 \end{bmatrix}$$

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta})^4 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Hence,

$$\mathbf{x}(k) = (\mathbf{G} - \mathbf{H}\mathbf{K})^k \mathbf{x}(0) = \mathbf{T}(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta})^k \mathbf{T}^{-1}\mathbf{x}(0) = \mathbf{0}, \qquad k \geq 4$$

We have thus seen that the deadbeat response is achieved by choosing $\mathbf{\Delta}$ as given by Equation (C–30).

However, if we choose $\mathbf{\Delta}$ as given by Equation (C–31), then the deadbeat response can be achieved in at most three sampling periods, because the asterisk appearing in $(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta})^3$ becomes zero and

$$\mathbf{x}(k) = \mathbf{T}(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta})^k \mathbf{T}^{-1}\mathbf{x}(0) = \mathbf{0}, \qquad k \geq n_{\min} = 3$$

Problem C–4

Consider the following system:

$$\mathbf{x}(k + 1) = \mathbf{G}\mathbf{x}(k) + \mathbf{H}\mathbf{u}(k)$$

where

$\mathbf{x}(k)$ = state vector (4-vector)

$\mathbf{u}(k)$ = control vector (2-vector)

and

$$\mathbf{G} = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -1 & 2 \\ 1 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{H} = [\mathbf{H}_1 \vdots \mathbf{H}_2] = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix}$$

By use of the state feedback control $\mathbf{u}(k) = -\mathbf{K}\mathbf{x}(k)$, we wish to place the closed-loop poles at the following locations:

$$z_1 = 0.5 + j0.5, \qquad z_2 = 0.5 - j0.5$$

$$z_3 = -0.2, \qquad z_4 = -0.8$$

Determine the required state feedback gain matrix $\mathbf{K}$. Then, using the given $\mathbf{G}$ and $\mathbf{H}$ matrices, derive Equation (C–16).

Solution We shall first examine the controllability matrix:

$$[\mathbf{H} \vdots \mathbf{G}\mathbf{H} \vdots \mathbf{G}^2\mathbf{H} \vdots \mathbf{G}^3\mathbf{H}] = [\mathbf{H}_1 \vdots \mathbf{H}_2 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_2 \vdots \mathbf{G}^2\mathbf{H}_1 \vdots \mathbf{G}^2\mathbf{H}_2 \vdots \mathbf{G}^3\mathbf{H}_1 \vdots \mathbf{G}^3\mathbf{H}_2]$$

$$= \left[\begin{array}{cc:cc:cc:cc} 0 & 1 & 1 & -1 & -3 & 2 & 11 & -5 \\ 1 & 0 & -2 & 1 & 8 & -3 & -22 & 11 \\ 0 & 0 & 3 & 0 & -3 & 3 & 15 & -6 \\ 1 & 0 & 1 & 1 & 2 & 0 & -1 & 2 \end{array}\right] \tag{C–33}$$

The rank of this controllability matrix is 4. Thus, arbitrary pole placement is possible. Four linearly independent vectors are chosen starting from the left end. (These vectors are shown enclosed by dashed lines.) The four linearly independent vectors chosen are $\mathbf{H}_1$, $\mathbf{H}_2$, $\mathbf{G}\mathbf{H}_1$, and $\mathbf{G}^2\mathbf{H}_1$. Now we rearrange these four vectors according to Equation (C–10) and define matrix $\mathbf{F}$ as follows:

$$\mathbf{F} = [\mathbf{H}_1 \vdots \mathbf{G}\mathbf{H}_1 \vdots \mathbf{G}^2\mathbf{H}_1 \vdots \mathbf{H}_2]$$

We note that $n_1 = 3$ and $n_2 = 1$ in this case. Rewriting matrix $\mathbf{F}$, we have

$$\mathbf{F} = \left[\begin{array}{ccc:c} 0 & 1 & -3 & 1 \\ 1 & -2 & 8 & 0 \\ 0 & 3 & -3 & 0 \\ 1 & 1 & 2 & 0 \end{array}\right] = \left[\begin{array}{c:c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]$$

Next, we compute $\mathbf{F}^{-1}$. Referring to Appendix A, we have

$$\mathbf{F}^{-1} = \begin{bmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}\mathbf{A}^{-1} & (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & -1 & -\frac{4}{3} & 2 \\ 0 & \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ 0 & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ 1 & \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix}$$

(The same result can be obtained easily by use of MATLAB.) Since $n_1 = 3$ and $n_2 = 1$, we choose the third row vector as $\mathbf{f}_1$ and the fourth row vector as $\mathbf{f}_2$. (Note that we define the η_ith row vector, where $\eta_i = n_1 + n_2 + \cdots + n_i$, as $\mathbf{f}_i$.) That is,

$$\mathbf{f}_1 = [0 \quad \tfrac{1}{3} \quad \tfrac{1}{3} \quad -\tfrac{1}{3}]$$

$$\mathbf{f}_2 = [1 \quad \tfrac{2}{3} \quad \tfrac{1}{3} \quad -\tfrac{2}{3}]$$

Next, we define the transformation matrix $\mathbf{T}$ by

$$\mathbf{T} = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \end{bmatrix}^{-1}$$

where

$$\mathbf{S}_1 = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_1\mathbf{G} \\ \mathbf{f}_1\mathbf{G}^2 \end{bmatrix}, \qquad \mathbf{S}_2 = [\mathbf{f}_2]$$

Hence,

$$\mathbf{T} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ 1 & \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix}^{-1} = \begin{bmatrix} -1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 3 & 3 & 0 & 0 \\ 1 & 3 & 1 & 0 \end{bmatrix}$$

With this transformation matrix $\mathbf{T}$, if we define

$$\mathbf{x}(k) = \mathbf{T}\hat{\mathbf{x}}(k)$$

then

$$\hat{\mathbf{G}} = \mathbf{T}^{-1}\mathbf{G}\mathbf{T} = \left[\begin{array}{ccc:c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 3 & -2 & 1 \\ \hdashline 2 & 0 & 0 & -1 \end{array}\right] \tag{C–34}$$

Also,

$$\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H} = \begin{bmatrix} 0 & \frac{1}{3} & \frac{1}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{2}{3} & -\frac{1}{3} & \frac{1}{3} \\ 1 & \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 1 & 0 \end{bmatrix} = \left[\begin{array}{cc} 0 & 0 \\ 0 & 0 \\ \hdashline 1 & 0 \\ 0 & 1 \end{array}\right] \tag{C–35}$$

Now we shall determine the state feedback gain matrix $\mathbf{K}$, where

$$\mathbf{K} = \mathbf{B}\Delta\mathbf{T}^{-1}$$

Referring to Equation (C–24) and noting that $b_{12} = 0$ in this case, matrix $\mathbf{B}$ is a 2×2 matrix given by

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{C–36}$$

For the present case, Δ is a 2×4 matrix:

$$\Delta = \begin{bmatrix} \delta_{11} & \delta_{12} & \delta_{13} & \delta_{14} \\ \delta_{21} & \delta_{22} & \delta_{23} & \delta_{24} \end{bmatrix}$$

Hence,

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\Delta = \left[\begin{array}{ccc:c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\delta_{11} & 3 - \delta_{12} & -2 - \delta_{13} & 1 - \delta_{14} \\ \hdashline 2 - \delta_{21} & -\delta_{22} & -\delta_{23} & -1 - \delta_{24} \end{array}\right]$$

Referring to Equation (C–32), we have

$$|z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\Delta| = \left|\begin{array}{c:c} z^3 + (2 + \delta_{13})z^2 + (-3 + \delta_{12})z + \delta_{11} & -1 + \delta_{14} \\ \hdashline \delta_{23}z^2 + \delta_{22}z + (-2 + \delta_{21}) & z + 1 + \delta_{24} \end{array}\right| = 0$$

This characteristic equation must be equal to the desired characteristic equation, which is

$$(z - 0.5 - j0.5)(z - 0.5 + j0.5)(z + 0.2)(z + 0.8) = z^4 - 0.34z^2 + 0.34z + 0.08 = 0$$

If we equate the coefficients of the equal powers of z of the two characteristic equations, we will have four equations for the determination of eight δ's. Hence, matrix Δ is not unique. Suppose we arbitrarily choose

$$\delta_{14} = 0, \qquad \delta_{22} = 0, \qquad \delta_{23} = 0, \qquad \delta_{24} = -1$$

Then

$$|z\mathbf{I} - \hat{\mathbf{G}} + \hat{\mathbf{H}}\mathbf{B}\Delta| = z^4 + (2 + \delta_{13})z^3 + (-3 + \delta_{12})z^2 + \delta_{11}z - 2 + \delta_{21} = 0$$

By equating this characteristic equation with the desired characteristic equation, we have

$$\delta_{11} = 0.34$$

$$\delta_{12} = 2.66$$

$$\delta_{13} = -2$$

$$\delta_{21} = 2.08$$

Thus,

$$\Delta = \begin{bmatrix} 0.34 & 2.66 & -2 & 0 \\ 2.08 & 0 & 0 & -1 \end{bmatrix}$$

Then matrix $\mathbf{K}$ is obtained as follows:

$$\mathbf{K} = \mathbf{B}\Delta\mathbf{T}^{-1} = \begin{bmatrix} 0 & -2.1067 & 0.7800 & 0.1067 \\ -1 & 0.02667 & 0.3600 & -0.02667 \end{bmatrix}$$

With the matrix $\mathbf{K}$ thus determined, state feedback control

$$u(k) = -\mathbf{K}\mathbf{x}(k)$$

will place the closed-loop poles at $z_1 = 0.5 + j0.5$, $z_2 = 0.5 - j0.5$, $z_3 = -0.2$, and $z_4 = -0.8$. It is noted that matrix $\mathbf{K}$ is not unique; there are many other possible matrices for $\mathbf{K}$.

Finally, we shall derive Equation (C–16). Notice that

$$\mathbf{F}^{-1}\mathbf{F} = \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{f}_1 \\ \mathbf{f}_2 \end{bmatrix} [\mathbf{H}_1 \quad \mathbf{GH}_1 \quad \mathbf{G}^2\mathbf{H}_1 \quad \mathbf{H}_2]$$

$$= \begin{bmatrix} \mathbf{m}_1\mathbf{H}_1 & \mathbf{m}_1\mathbf{GH}_1 & \mathbf{m}_1\mathbf{G}^2\mathbf{H}_1 & \mathbf{m}_1\mathbf{H}_2 \\ \mathbf{m}_2\mathbf{H}_1 & \mathbf{m}_2\mathbf{GH}_1 & \mathbf{m}_2\mathbf{G}^2\mathbf{H}_1 & \mathbf{m}_2\mathbf{H}_2 \\ \mathbf{f}_1\mathbf{H}_1 & \mathbf{f}_1\mathbf{GH}_1 & \mathbf{f}_1\mathbf{G}^2\mathbf{H}_1 & \mathbf{f}_1\mathbf{H}_2 \\ \mathbf{f}_2\mathbf{H}_1 & \mathbf{f}_2\mathbf{GH}_1 & \mathbf{f}_2\mathbf{G}^2\mathbf{H}_1 & \mathbf{f}_2\mathbf{H}_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

where $\mathbf{m}_1$ and $\mathbf{m}_2$ are the first row vector and second row vector of $\mathbf{F}^{-1}$, respectively. Since $\mathbf{F}^{-1}\mathbf{F}$ is an identity matrix, $\mathbf{f}_1\mathbf{H}_1 = 0$, $\mathbf{f}_1\mathbf{H}_2 = 0$, $\mathbf{f}_1\mathbf{GH}_1 = 0$, $\mathbf{f}_1\mathbf{G}^2\mathbf{H}_1 = 1$, $\mathbf{f}_2\mathbf{H}_1 = 0$, and $\mathbf{f}_2\mathbf{H}_2 = 1$. From Equation (C–33) we see that $\mathbf{GH}_2$ is linearly dependent on $\mathbf{H}_1$, $\mathbf{H}_2$, and $\mathbf{GH}_1$. Hence, $\mathbf{f}_1\mathbf{GH}_2 = \alpha\mathbf{f}_1\mathbf{H}_1 + \beta\mathbf{f}_1\mathbf{H}_2 + \gamma\mathbf{f}_1\mathbf{GH}_1 = 0$, where α, β, and γ are constants. Note that $\mathbf{f}_1\mathbf{G}^2\mathbf{H}_2$ may or may not be zero. Consequently,

$$\hat{\mathbf{H}} = \mathbf{T}^{-1}\mathbf{H} = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_1\mathbf{G} \\ \mathbf{f}_1\mathbf{G}^2 \\ \mathbf{f}_2 \end{bmatrix} [\mathbf{H}_1 \quad \mathbf{H}_2] = \begin{bmatrix} \mathbf{f}_1\mathbf{H}_1 & \mathbf{f}_1\mathbf{H}_2 \\ \mathbf{f}_1\mathbf{GH}_1 & \mathbf{f}_1\mathbf{GH}_2 \\ \mathbf{f}_1\mathbf{G}^2\mathbf{H}_1 & \mathbf{f}_1\mathbf{G}^2\mathbf{H}_2 \\ \mathbf{f}_2\mathbf{H}_1 & \mathbf{f}_2\mathbf{H}_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & b_{12} \\ 0 & 1 \end{bmatrix}$$

where $b_{12} = \mathbf{f}_1\mathbf{G}^2\mathbf{H}_2$. This last equation is Equation (C–16).

Problem C–5

Referring to Problem C–4, consider the same system. Suppose that we desire the deadbeat response to an arbitrary initial state $\mathbf{x}(0)$. Determine the state feedback gain matrix $\mathbf{K}$.

Solution Referring to Equations (C–34), (C–35), and (C–36), we have

$$\hat{\mathbf{G}} = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 3 & -2 & 1 \\ \hline 2 & 0 & 0 & -1 \end{array}\right], \qquad \hat{\mathbf{H}} = \left[\begin{array}{c} 0 \quad 0 \\ 0 \quad 0 \\ 1 \quad 0 \\ \hline 0 \quad 1 \end{array}\right]$$

$$\mathbf{B} = \begin{bmatrix} 1 & b_{12} \\ 0 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where b_{12} is zero. For the deadbeat response, we choose $\mathbf{\Delta}$ as follows:

$$\mathbf{\Delta} = \begin{bmatrix} -a_{11} & -a_{12} & -a_{13} & -a_{14} \\ -a_{21} & -a_{22} & -a_{23} & -a_{24} \end{bmatrix} = \begin{bmatrix} 0 & 3 & -2 & 1 \\ 2 & 0 & 0 & -1 \end{bmatrix}$$

where the a_{ij}'s are as defined in Equation (C–15). Then

$$\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

and we find

$$(\hat{\mathbf{G}} - \hat{\mathbf{H}}\mathbf{B}\mathbf{\Delta})^k = \mathbf{0}, \qquad k = 3, 4, 5, \ldots$$

The deadbeat response is reached in at most three sampling periods. [Note that in this problem $n_1 = 3$ and $n_2 = 1$. Hence, $n_{\min} = \max(n_1, n_2) = 3$.] The desired state feedback gain matrix $\mathbf{K}$ is obtained as follows:

$$\mathbf{K} = \mathbf{B}\mathbf{\Delta}\mathbf{T}^{-1}$$

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

$$= \begin{bmatrix} 1 & -\frac{5}{3} & 1 & -\frac{1}{3} \\ -1 & 0 & \frac{1}{3} & 0 \end{bmatrix}$$

With this matrix $\mathbf{K}$, the state feedback control

$$\mathbf{u}(k) = -\mathbf{K}\mathbf{x}(k)$$

will place the four closed-loop poles at the origin and thus will produce the deadbeat response to any initial state $\mathbf{x}(0)$.

References

A–1. Antoniou, A., *Digital Filters: Analysis and Design*. New York: McGraw-Hill Book Company, 1979.

A–2. Aseltine, J. A., *Transform Method in Linear System Analysis*. New York: McGraw-Hill Book Company, 1958.

A–3. Åström, K. J., and B. Wittenmark, *Computer Controlled Systems: Theory and Design*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1984.

B–1. Bellman, R., *Introduction to Matrix Analysis*. New York: McGraw-Hill Book Company, 1960.

B–2. Bristol, E. H., "Design and Programming Control Algorithms for DDC Systems," *Control Engineering*, **24**, Jan. 1977, pp. 24–26.

B–3. Butman, S., and R. Sivan (Sussman), "On Cancellations, Controllability and Observability," *IEEE Trans. Automatic Control*, **AC-9** (1964), pp. 317–18.

C–1. Cadzow, J. A., and H. R. Martens, *Discrete-Time and Computer Control Systems*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1970.

C–2. Chan, S. W., G. C. Goodwin, and K. S. Sin, "Convergence Properties of the Riccati Difference Equation in Optimal Filtering of Nonstabilizable Systems," *IEEE Trans. Automatic Control*, **AC-29** (1984), pp. 110–18.

C–3. Churchill, R. V., and J. W. Brown, *Complex Variables and Applications*, 4th ed., New York: McGraw-Hill Book Company, 1984.

D–1. Dorato, P., and A. H. Levis, "Optimal Linear Regulators: The Discrete-Time Case," *IEEE Trans. Automatic Control*, **AC-16** (1971), pp. 613–20.

E–1. Evans, W. R., "Control System Synthesis by Root Locus Method," *AIEE Trans. Part II*, **69** (1950), pp. 66–69.

F–1. Falb, P. L., and M. Athans, "A Direct Constructive Proof of the Criterion for Complete Controllability of Time-Invariant Linear Systems," *IEEE Trans. Automatic Control*, **AC-9** (1964), pp. 189–90.

F–2. Fortmann, T. E., "A Matrix Inversion Identity," *IEEE Trans. Automatic Control*, **AC-15** (1970), p. 599.

F–3. Franklin, G. F., J. D. Powell, and M. L. Workman, *Digital Control of Dynamic Systems*, 2nd ed., Reading, Mass.: Addison-Wesley Publishing Co., Inc., 1990.

F–4. Freeman, H., *Discrete-Time Systems*. New York: John Wiley & Sons, Inc., 1965.

G–1. Gantmacher, F. R., *Theory of Matrices*, Vols. I and II. New York: Chelsea Publishing Co., Inc., 1959.

G–2. Gopinath, B., "On the Control of Linear Multiple Input–Output Systems," *Bell Syst. Tech. J.*, **50** (1971), pp. 1063–81.

H–1. Hahn, W., *Theory and Application of Liapunov's Direct Method*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1963.

H–2. Halmos, P. R., *Finite Dimensional Vector Spaces*. Princeton, N.J.: D. Van Nostrand Company, 1958.

I–1. Ichikawa, K., *Theory for Design of Control Systems* (in Japanese). Tokyo: Gijutsu-Shoin, 1989.

J–1. Jerri, A. J., "The Shannon Sampling Theorem—Its Various Extensions and Applications: A Tutorial Review," *Proc. IEEE*, **65** (1977), pp. 1565–95.

J–2. Jury, E. I., "Hidden Oscillations in Sampled-Data Control Systems," *AIEE Trans. Part II*, **75** (1956), pp. 391–95.

J–3. Jury, E. I., *Sampled-Data Control Systems*. New York: John Wiley & Sons, Inc., 1958.

J–4. Jury, E. I., "Sampling Schemes in Sampled-Data Control Systems," *IRE Trans. Automatic Control*, **AC-6** (1961), pp. 88–90.

J–5. Jury, E. I., *Theory and Applications of the z Transform Method*. New York: John Wiley & Sons, Inc., 1964.

J–6. Jury, E. I., "A General z-Transform Formula for Sampled-Data Systems," *IEEE Trans. Automatic Control*, **AC-12** (1967), pp. 606–8.

J–7. Jury, E. I., "Sampled-Data Systems, Revisited: Reflections, Recollections, and Reassessments," *ASME J. Dynamic Systems, Measurement, and Control*, **102** (1980), pp. 208–16.

J–8. Jury, E. I., and J. Blanchard, "A Stability Test for Linear Discrete-Time Systems in Table Forms," *Proc. IRE*, **49** (1961), pp. 1947–48.

K–1. Kailath, T., *Linear Systems*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1980.

K–2. Kailath, T., and P. Frost, "An Innovations Approach to Least-Squares Estimation, Part II: Linear Smoothing in Additive White Noise," *IEEE Trans. Automatic Control*, **AC-13** (1968), pp. 655–60.

K–3. Kalman, R. E., "On the General Theory of Control Systems," *Proc. First Intern. Cong. IFAC*, Moscow, 1960. *Automatic and Remote Control*. London: Butterworth & Co., Ltd., 1961, pp. 481–92.

K–4. Kalman, R. E., and J. E. Bertram, "Control System Analysis and Design via the Second Method of Lyapunov: I. Continuous-Time Systems; II. Discrete-Time Systems," *ASME J. Basic Engineering, ser. D*, **82** (1960), pp. 371–93, 394–400.

K–5. Kalman, R. E., Y. C. Ho, and K. S. Narendra, "Controllability of Linear Dynamical Systems," *Contributions to Differential Equations*, **1** (1963), pp. 189–213.

K–6. Kanai, K., and N. Hori, *Introduction to Digital Control Systems* (in Japanese). Tokyo: Maki Shoten, 1992.

K–7. Katz, P., *Digital Control Using Microprocessors*. London: Prentice Hall International, Inc., 1981.

K–8. Kreindler, E., and P. E. Sarachik, "On the Concepts of Controllability and Observability of Linear Systems," *IEEE Trans. Automatic Control*, **AC-9** (1964), pp. 129–36.

K–9. Kuo, B. C., *Digital Control Systems*. New York: Holt, Rinehart and Winston, Inc., 1980.

L–1. LaSalle, J. P., and S. Lefschetz, *Stability by Liapunov's Direct Method with Applications*. New York: Academic Press, Inc., 1961.

L–2. Lee, E. B., and L. Markus, *Foundations of Optimal Control Theory*. New York: John Wiley & Sons, Inc., 1967.

L–3. Leondes, C. T., and M. Novak, "Reduced-Order Observers for Linear Discrete-Time Systems," *IEEE Trans. Automatic Control*, **AC-19** (1974), pp. 42–46.

L–4. Li, Y. T., J. L. Meiry, and R. E. Curry, "On the Ideal Sampler Approximation," *IEEE Trans. Automatic Control*, **AC-17** (1972), pp. 167–68.

L–5. Luenberger, D. G., "Observing the State of a Linear System," *IEEE Trans. Military Electronics*, **MIL-8** (1964), pp. 74–80.

L–6. Luenberger, D. G., "An Introduction to Observers," *IEEE Trans. Automatic Control*, **AC-16** (1971), pp. 596–602.

M–1. Melsa, J. L., and D. G. Schultz, *Linear Control Systems*. New York: McGraw-Hill Book Company, 1969.

M–2. Middleton, R. H., and G. C. Goodwin, *Digital Control and Estimation—A Unified Approach*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1990.

M–3. Mitra, S. K., and R. J. Sherwood, "Canonic Realizations of Digital Filters Using the Continued Fraction Expansion," *IEEE Trans. Audio and Electroacoustics*, **AU-20** (1972), pp. 185–94.

M–4. Mitra, S. K., and R. J. Sherwood, "Digital Ladder Networks," *IEEE Trans. Audio and Electroacoustics*, **AU-21** (1973), pp. 30–36.

N–1. Neuman, C. P., and C. S. Baradello, "Digital Transfer Functions for Microcomputer Control.," *IEEE Trans. Systems, Man, and Cybernetics*, **SMC-9** (1979), pp. 856–60.

N–2. Noble, B., and J. Daniel, *Applied Linear Algebra*, 2nd ed. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1977.

O–1. Ogata, K., *State Space Analysis of Control Systems*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1967.

O–2. Ogata, K., *Modern Control Engineering*, 2nd ed. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1990.

O–3. Ogata, K., *System Dynamics*, 2nd ed., Englewood Cliffs, N.J.: Prentice Hall, Inc., 1992.

O–4. Ogata, K., *Solving Control Engineering Problems with MATLAB*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1994.

O–5. Ogata, K., *Designing Linear Control Systems with MATLAB*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1994.

P–1. Pappas, T., A. J. Laub, and N. R. Sandell, Jr., "On the Numerical Solution of the Discrete-Time Algebraic Riccati Equation," *IEEE Trans. Automatic Control*, **AC-25** (1980), pp. 631–41.

P–2. Payne, H. J., and L. M. Silverman, "On the Discrete Time Algebraic Riccati Equation," *IEEE Trans. Automatic Control*, **AC-18** (1973), pp. 226–34.

P–3. Phillips, C. L., and H. T. Nagle, Jr., *Digital Control Systems Analysis and Design*. Englewood Cliffs, N.J.: Prentice Hall, Inc., 1984.

R–1. Ragazzini, J. R., and G. F. Franklin, *Sampled-Data Control Systems*. New York: McGraw-Hill Book Company, 1958.

R–2. Ragazzini, J. R., and L. A. Zadeh, "The Analysis of Sampled-Data Systems," *AIEE Trans. Part II*, **71** (1952), pp. 225–34.

S–1. Strang, G., *Linear Algebra and Its Applications*. New York: Academic Press, Inc., 1976.

T–1. Tou, J. T., *Digital and Sampled-Data Control Systems*. New York: McGraw-Hill Book Company, 1959.

T–2. Turnbull, H. W., and A. C. Aitken, *An Introduction to the Theory of Canonical Matrices*. London: Blackie and Son, Ltd., 1932.

V–1. Van Dooren, P., "A Generalized Eigenvalue Approach for Solving Riccati Equations," *SIAM J. Scientific and Statistical Computing*, **2** (1981), pp. 121–35.

W–1. Willems, J. C., and S. K. Mitter, "Controllability, Observability, Pole Allocation, and State Reconstruction," *IEEE Trans. Automatic Control*, **AC-16** (1971), pp. 582–95.

W–2. Wolovich, W. A., *Linear Multivariable Systems*. New York: Springer-Verlag, 1974.

W–3. Wonham, W. M., "On Pole Assignment in Multi-Input Controllable Linear Systems," *IEEE Trans. Automatic Control*, **AC-12** (1967), pp. 660–65.

Z–1. Zadeh, L. A., and C. A. Desoer, *Linear System Theory: The State Space Approach*. New York: McGraw-Hill Book Company, 1963.

Index

A

Absolute stability, 193
Ackermann's formula:
  for minimum-order observer design, 450, 454
  for observer design, 435–438, 440, 445, 496
  for pole placement, 408–412, 466, 493
Actuating error, 200
A/D converter, 15
  counter type, 15
  successive-approximation type, 15
Adjoint vector, 572
Alias, 98
Amplitude quantization, 8
Analog controller, 21
Analog multiplexer, 12
Analog signal, 1–2
Analog-to-digital conversion, 14
Analog-to-digital converter, 7
Analog transducer, 7
Analytical design method, 242–257
Angle:
  of arrival, 209
  of asymptote, 207–208
  of departure, 209
Aperture time, 14
Associativity law, 638
Asymptotic stability, 325
  in the large, 325

B

Backward difference:
  first, 697
  *m*th, 698
  second, 698
  third, 698
BIBO stability, 326
Bilinear form:
  complex, 660
  real, 660
Bilinear transformation, 191, 228, 231

Block diagram:
of continuous-time control system in state space, 296
of discrete-time control system in state space, 296
Bode diagram, 232–233
Bounded-input–bounded-output stability, 326
Breakaway point, 208–209
Break-in point, 208–209

C

Cancellation:
of poles with zeros, 210–211
Canonical forms:
controllable, 297–298, 300, 396, 398, 489
diagonal, 299–300, 399, 489
Jordan, 300, 302, 382, 390, 399–400, 651–652, 657, 659, 674
observable, 298–300, 398–399, 489
Cauchy-Goursat theorem, 688–689
Cayley-Hamilton theorem, 350, 380, 404, 408, 481, 485, 492
Characteristic polynomial, 649
Characteristic roots, 650
Clamper, 78
Coding, 6, 8
Coefficient quantization problem, 234
Compensation:
phase lag, 233
phase lag-lead, 233
phase lead, 233
Complementary strips, 175
Complete observability. (*See* Observability.)
Complete state controllability. (*See* Controllability.)
Complex convolution theorem, 684
Complex differentiation, 681–682
Complex integration, 682–683
Complex translation theorem, 34
Conformal mapping, 180
Constant-attenuation loci, 176–177
Constant-damping-ratio loci, 178–180
Constant-frequency loci, 176–178
Continuous-time analog signal, 1–2
Continuous-time quantized signal, 1–2
Contraction, 334–335, 367–368
Control energy, 622
Controllability, 377, 379
complete output, 385–386
complete state, 380–384, 387, 393, 406
matrix, 380, 401, 707–708
output, 387
in the z plane, 384
Controllable canonical form, 297–298, 300, 396, 398, 489
Convolution integral, 84–85
evaluation in the left half-plane of, 84–86
evaluation in the right half-plane of, 86–88
Convolution summation, 98, 100
Convolution theorem:
complex, 684
real, 684
Coprime polynomials, 518, 541
Covector, 572
c2d, 628
Current observer, 444

D

D/A converter:
using R-2R ladder circuit, 17–18
using weighted resistors, 16, 18
Data-acquisition process, 12
Data-acquisition system, 11–12
Data-distribution process, 12
Data-distribution system, 11
Data-hold, 6, 77
Data-hold circuits, 77

Deadbeat response, 242, 248, 411, 414–418, 435, 439–442, 444, 453–454, 470–471, 490, 494, 498, 502–505, 508, 550, 712–713, 715, 717–718, 723, 728–729
Decoder, 7
Decoding, 6
Definiteness:
negative, 661
positive, 660
Delay time, 194–195
Demultiplexer, 13
Derivative gain, 116
Derivative time, 115
Design:
based on analytical method, 242–257
based on frequency-response method, 225–242
based on pole placement, 402–421
based on pole placement with observed state feedback, 421–460
based on polynomial equations approach, 517–540
based on root-locus method, 204–225
Determinant, 633–635
properties of, 634–635
Diagonal canonical form, 299–300, 399, 489
Differentiation:
in the z plane, 165
Digital control system, 3, 5
Digital controller, 20
realization of, 122
Digital filter, 122
block diagram realization of, 122
direct programming of, 123–124, 133
ladder programming of, 128–135
parallel programming of, 127–128
parallel realization of, 163–165
series programming of, 126–127
series realization of, 163–165
Digital filter (*cont.*)
standard programming of, 124–125, 133–134
Digital integrator:
bilinear, 172
with delay, 171–172
without delay, 171
Digital PID control:
positional form, 116
velocity form, 117
Digital PID controller, 114–118
Digital signal, 2–3
Digital-to-analog converter, 7, 16
Digital transducer, 7
Diophantine equation, 518, 520–521, 523–525, 529, 533, 535, 547, 551, 555, 559
solution to, 520–521
Diophantus, 518
Direct division method, 40–42
Direct method of Liapunov, 322
Direct programming, 123
method, 336
Discrete-time control system, 3
Discrete-time signal, 2–3, 23
Discretization, 6, 394
of continuous-time state space equation, 314
Domain of attraction, 325
Double integrator system, 361–362, 439, 490, 513

E

Eigenvalue, 649–650, 678
Eigenvector, 650, 674
generalized, 654, 656
normalized, 650
Encoder, 7
Encoding, 6, 8
Equilibrium state, 324

Errors in A/D converters:
gain error, 16–17
linearity error, 16–17
offset error, 16–17
Euclidean norm, 324

F

Fibonacci series, 67–69
Filter, 603–604
Final value theorem, 36
Finite-impulse response filter, 135–137
First-order hold, 19, 80–82, 139–140
interpolative, 19–20
magnitude and phase characteristics of, 151–153
transfer function of, 80–82
Folding, 96
error, 97
frequency, 96
Format:
long, 318
short, 318
Forward difference, 322
first, 698–699
*m*th, 699
second, 698
third, 698
Frequency-response method, 225–242
Frequency spectrum:
of complementary components, 92
of ideal low-pass filter, 92–93
of primary component, 92
of sampled signal, 91–92
Full-order state observer, 426–444
Fundamental matrix, 303, 309

G

Gain crossover frequency, 274
Gain error of A/D converter, 16–17
Generalized eigenvector, 494, 496, 498, 654, 656, 674

H

Hermitian form, 660
Hermitian matrix, 633
Hidden instability, 334
Hidden oscillation, 98, 361
Higher-order hold circuits, 19, 82
Hold circuits, 17–18
Hold mode, 13
droop, 14

I

Ideal filter:
magnitude characteristics of, 92
unit-impulse response of, 93–94
Ideal low-pass filter, 92–93
Impulse sampler, 75–77, 83
Impulse sampling, 75, 77
Indefinite matrix, 661
Indefiniteness:
of scalar function, 661
Infinite-impulse response filter, 135
Initial value theorem, 35
Inner product, 643–645, 647
Instability, 325, 327–328
Integral gain, 116
Integral time, 115
Interpolative first-order hold, 19–20
Invariance:
of characteristic equation, 312
property, 401
Inverse *z* transform, 37, 687
computational method for obtaining, 42–46
difference equation approach to obtain, 46
direct division method for obtaining, 40–42, 62

Inverse *z* transform (*cont.*)
inversion integral method for obtaining, 50–52, 60–62, 64–66
MATLAB approach to obtain, 42–45
partial-fraction-expansion method for obtaining, 46–50, 64
Inverse *z* transformation, 37
Inverse of $z\mathbf{I} - \mathbf{G}$:
computation of, 304–309
Inversion integral:
for the *z* transform, 689
Inversion integral method, 50–52, 60–62, 64–66
Inverted pendulum control system, 596
Inverted pendulum system, 597, 625–628
Isolated equilibrium state, 324

J

Jacobian, 641
Jordan block, 383, 651–652
Jordan canonical form, 300, 302, 382, 390, 399–400, 651–652, 657, 659, 674
Jury stability table, 185, 187–188
Jury stability test, 185–190

K

Kalman, R. E., 377
Kronecker delta function, 42, 62
Kronecker delta input, 43, 103
Kronecker invariant, 708

L

Ladder programming, 128–135
Lag compensator, 224, 273–274
Lagrange multiplier, 570–572
Laplace's expansion by minors, 541, 634
Laurent series expansion, 141, 687, 689
Lead compensator, 262, 272–273
Left pseudoinverses matrix, 665–666
Liapunov:
direct method of, 322
first method of, 321
function, 322–323, 334, 591
second method of, 321–322
Liapunov stability analysis, 321–336
of continuous-time system, 329–332
of discrete-time system, 332–334
first method of, 321
second method of, 321–322
Liapunov theorems:
on asymptotic stability, 326–327
on instability, 327–328
on stability, 327
Liapunov's main stability theorem, 326, 363–365
Linear dependence of vectors, 643
Linear discrete-time state equation:
solution of time-invariant, 302–309
solution of time-varying, 309–310
Linear independence of vectors, 643
Linear system, 3
Linear time-varying discrete-time system, 309–310
Linearity error of A/D converter, 16–17

M

Mapping:
between *s* plane and *z* plane, 174–182
from *s* plane to *z* plane, 229
from *z* plane to *w* plane, 229
MATLAB programs:
for finding Fibonacci series, 68
for finding inverse *z* transform, 44, 63
for finding response to Kronecker delta input, 45
for finding unit-ramp response, 120, 260, 459, 531
for finding unit-step response, 119, 196, 240, 268, 421, 458, 530, 605–606

MATLAB programs (*cont.*)
for pole placement in the z plane, 500–501
for quadratic optimal control, 579, 590–591, 600–603, 615
Matrix:
cancellation of, 639
derivative of, 640
diagonalization of, 651, 653
differentiation of, 640
eigenvalue of, 649–650
eigenvector of, 650
exponential, 313
Hermitian, 633
indefinite, 661
integral of, 640
inverse, 635–637
inversion lemma, 573, 636, 668
multiplication by a matrix, 637
multiplication by a scalar, 637
negative definite, 661
negative semidefinite, 661
nonsingular, 635
norm of, 647
normal, 633
positive definite, 661
positive semidefinite, 661
rank of, 649
rules of operations of, 637–643
similar, 651
singular, 635
skew Hermitian, 633
skew symmetric, 633
stable, 365
symmetric, 633
trace of, 658
unitary, 645
Maximum overshoot, 195
Minimal left inverse, 666
Minimal polynomial, 350–354
Minimal right inverse, 665
Minimum control energy, 622
Minimum norm solution, 624
that minimizes $\|\mathbf{Ax} - \mathbf{b}\|$, 665
that minimizes $\|\mathbf{x}\|$, 663–665
Minimum-order observer, 446–450, 452–454, 469–470, 502–504
Model matching control system, 532–534, 536–537, 561
Modified z transform, 691–696
Monic polynomial, 518
Moving average filter, 136
Multiple-order sampling, 8
Multiple-rate sampling, 8
Multiplexer, 12

N

Negative definite matrix, 661
Negative definiteness:
of scalar function, 661
Negative semidefinite matrix, 661
Negative semidefiniteness:
of scalar function, 661
Nested programming method, 338, 343
Nilpotent matrix, 414–416
Nonminimum phase transfer function, 233
Nonrecursive filter, 136–138
Nonsingular matrix, 635
Norm, 645–647
Euclidean, 324
Normal matrix, 633
nth-order hold, 77
Nyquist frequency, 96

O

Observability, 377, 388
complete, 389–390
matrix, 389, 394, 401
in the z plane, 391–394
Observable canonical form, 298–300, 398–399, 489

Observation, 422
Observed-state feedback control system, 428, 434
with minimum-order observer, 447, 451–452
Observer error equation, 428, 443, 445, 450
Observer feedback gain matrix, 427, 434, 438, 442, 449–450, 496, 499
Observer poles, 428
Observer regulator, 502, 505, 543
Optimal control law:
minimum energy, 622–625
quadratic, 568–596
Optimal control system, 566, 568
Optimal control vector, 567
closed-loop form, 574
feedback form, 574
Optimal regulator system, 566
Orthogonal matrix, 633
Orthogonal set, 648
Orthogonal transformation, 645
Output controllability, 387

P

Parallel programming, 127–128
Parameter optimization problem, 591
Parseval's theorem, 686
Partial differentiation theorem, 683
Partial-fraction-expansion method, 46–50
Partial-fraction-expansion programming method, 339–341, 345
PD controller, 234
Peak time, 195
Performance index, 566
including cross term, 582
minimum value of, 575
Periodic sampling, 8
Phase lag compensation, 233
Phase lag compensator, 234
Phase lag-lead compensation, 233
Phase lag-lead compensator, 233
Phase lead compensation, 233
Phase lead compensator, 234, 237
Physical realizability:
condition for, 244–245
PI controller, 234
PID control action:
analog controller, 115
PID controller, 117–118, 121, 233–234
analog, 156–159
digital, 156–159
positional form, 116
velocity form, 117, 157, 159–160
Plant, 7
Pole assignment technique, 402
Pole placement, 408
design, 402–421, 707–718
design with vector control, 704–718
necessary and sufficient condition for, 402–408
Pole-zero cancellation, 211, 479–481
Pole, 39–40
Poly, 499
Polygonal hold, 19–20
Polynomial equations approach:
to design control systems, 525–532
to design regulator systems, 523–525
Polyvalm, 500
Positive definite matrix, 661
Positive definiteness:
of scalar function, 660–661
Positive semidefinite matrix, 661
Positive semidefiniteness:
of scalar function, 662–663
test, 680
Prediction observer, 428
design of, 430–444
full-order, 438
Primary strip, 175
Principle of duality, 392–394

Principle of superposition, 3
Process, 7
Proportional gain, 116
Pseudoinverse matrix:
  left, 665–666
  right, 664–665
Pulse transfer function, 98, 102, 104–118
  of cascaded elements, 108–110
  of closed-loop system, 110–111
  of digital controller, 111–118
  matrix, 310–312

Q

Quadratic form, 659–660
  complex, 660
  real, 659
Quadratic optimal control:
  of servo system, 596–609
  steady-state, 587–596
Quadratic optimal control problem:
  discretized, 580–582
  Liapunov approach to the solution of, 592–596
  steady-state, 592–594
Quadratic optimal regulator problem:
  Liapunov approach to the solution of, 591–592
  steady-state, 591–592
Quadratic performance index, 568
  with cross term, 582
Quantization, 1, 7
  error, 9
  level, 8–9
  noise, 9, 11, 126
  process, 4
Quantizer, 9–10

R

Radius of absolute convergence, 25
Random sampling, 8
Rank, 649
Rate time, 115
Reachability, 474–475
Real convolution theorem, 684
Real sampler, 78
Real translation theorem, 31
Recursive filter, 135, 137
Reduced-order observer, 446
Relative stability, 193, 195, 220
Reset time, 115
Residue, 50, 84–85, 145, 399, 688, 690–691
  theorem, 689
Response:
  to disturbance, 202
  between two consecutive sampling instants, 320–321
Riccati equation, 573–574
  steady-state, 588–589
Riccati transformation, 572–573
Right pseudoinverse, 623–624, 664–665
Rise time, 195
Root loci:
  asymptotes of, 207–208
  general rules for constructing, 207–210
Root locus, 206
Root-locus method, 205
  angle condition in, 206
  magnitude condition in, 206
Round-off error, 9
Routh stability criterion:
  bilinear transformation coupled with, 191–192, 258–259

S

Sample-and-hold, 6
Sample-and-hold circuit, 13–14
  hold mode operation of, 13–14
  tracking mode operation of, 13–14
Sampled-data control system, 3

Sampled-data signal, 2
Sampled-data transducer, 7
Sampling, 6
  frequency, 90
  process, 4
  theorem, 90–92
Scalar function:
  indefiniteness of, 322
  negative definiteness of, 322
  negative semidefiniteness of, 322
  positive definiteness of, 322
  positive semidefiniteness of, 322
Scalar product, 643
Schur-Cohn stability test, 185
Schwarz inequality, 645–646
Second method of Liapunov, 322
Second-order hold, 19
Series programming, 126–127
Servo system, 460
  with observed-state feedback, 465
  quadratic optimal control of, 596–609
  with state feedback, 464
  with state feedback and integral control, 460–461
Settling time, 195
Shannon's sampling theorem, 150–151
Shifting theorem, 31
Similar matrices, 651
Similarity transformation, 301, 311–312, 651–653, 657
  invariant properties under, 659
Singular matrix, 635
Sinusoidal pulse transfer function, 227–228
Skew-Hermitian matrix, 633
Skew-symmetric matrix, 633
  real, 679
Square matrix:
  eigenvalues of, 649–650
ss2tf, 603–604
Stability, 324
  Liapunov theorem on, 327
Stability analysis:
  of linear time-invariant system, 328
  by use of bilinear transformation and Routh stability criterion, 191–192
Stable matrix, 365
Staircase generator, 78
Standard programming, 124–126
Starred Laplace transform, 103–104
State, 294
State equation:
  solution of continuous-time, 312
  solution of linear time-invariant discrete-time, 302–309
  solution of linear time-varying discrete-time, 309–310
  z transform approach to the solution of discrete-time, 304–307
  zero-order hold equivalent of continuous-time, 315–317
State estimator, 422
State feedback gain matrix, 402–403, 410–414, 427, 492, 494
State observer, 422
  full-order, 422, 426–444
  minimum-order, 446–456
State observation:
  minimum-order, 422
  necessary and sufficient condition for, 422–425
  reduced-order, 422
State space, 294
State space representation:
  nonuniqueness of, 301
State transition matrix, 303, 305, 309–310
State variable, 294
State vector, 294
Static acceleration error constant, 200
Static error constants, 198–201
Static position error constant, 199
Static velocity error constant, 199
Steady-state actuating error, 198–200

Steady-state error, 196–197, 200–201
Steady-state quadratic optimal control law:
  Liapunov approach to, 591–594
Steady-state response, 193
Steady-state Riccati equation, 600, 621
Sylvester, J. J., 661
Sylvester matrix, 518–520, 524, 535, 540–541, 548, 551, 559
Sylvester criterion:
  for negative definiteness, 662
  for negative semidefiniteness, 663
  for positive definiteness, 661–662, 679
  for positive semidefiniteness, 662
Symmetric matrix, 633
  real, 679
System, 324

T

Time-invariant linear system, 3
Trace, 307, 658
Tracking mode, 13
Transducer, 7
Transfer function matrix:
  pulse, 310–312
Transient response, 193
  specifications, 193–195
Transportation lag, 280
Two-point boundary-value problem, 572
Type 1 servo system, 597
Type 1 system, 197–198
Type 2 system, 197–198
Type 0 system, 197–198

U

Underdetermined equation, 623
Uniform asymptotic stability, 365
  in the large, 364–365
Uniform stability, 324
Unit delay operator, 40
Unit impulses:
  train of, 75
Unit-ramp function, 26
Unit-step sequence, 26
Unitary matrix, 633, 648
Unitary transformation, 645

V

Vector:
  norm of, 645
  normalized, 644
  unit, 648
Vectors:
  linear dependence of, 643
  linear independence of, 643
  orthonormal, 648

W

w plane:
  design procedure in the, 228–242
w transformation, 228–229, 231
Weighting sequence, 100

Z

z transform, 24
  complex translation theorem of, 34
  convolution integral method for obtaining, 83
  of cosine function, 28
  definition of, 24
  of exponential function, 27
  final value theorem of, 36
  of first backward difference, 697–698
  of first forward difference, 698–699
  of function involving term $(1 - e^{-Ts})/s$, 88–90
  important properties of, 31
  initial value theorem of, 35

z transform (*cont.*)
  inverse, 37
  inversion integral for, 689
  linearity of, 31
  one-sided, 24–25
  of polynomial function, 27
  properties of, 38
  real translation theorem for, 31
  of second backward difference, 698
  of second forward difference, 698–699
  shifting theorem for, 31
  of sinusoidal function, 27
  table of, 29–30
z transform (*cont.*)
  two-sided, 25
  of unit-ramp function, 26
  of unit-step function, 25, 33
Zero, 39–40
Zero-order hold, 18–19, 78, 166
  Bode diagram of, 95
  frequency-response characteristics of, 94–96
  magnitude and phase characteristics of, 151–153
  transfer function of, 139