8

Functions of Square Matrices and the Cayley-Hamilton Theorem

8.1 INTRODUCTION

Functions of square matrices arise in connection with the solution of vector-matrix differential and difference equations. Some scalar-valued functions of matrices have already been considered, namely, $|\mathbf{A}|$, $\text{Tr}(\mathbf{A})$, $\|\mathbf{A}\|$, etc. In this chapter matrix-valued functions $f(\mathbf{A})$ of square matrices $\mathbf{A}$ are considered. These functions are themselves matrices of the same size as $\mathbf{A}$, and their element values depend upon the particular function as well as on the values of $\mathbf{A}$. This chapter is devoted to explaining when these functions can be defined, what these functions are, and how to compute them. Before beginning, it may be helpful to state clearly what they are *not*. If $\mathbf{A} = [a_{ij}]$, a matrix function $f(\mathbf{A})$ *is not* just the matrix made up of the elements $f(a_{ij})$ except in special cases.

Specific attention is given to the $n \times n$ matrix exponential function $f(\mathbf{A}) = e^{\mathbf{A}t}$ and to the matrix power function $f(\mathbf{A}) = \mathbf{A}^k$. Solutions of continuous-time state variable equations depend upon the matrix exponential. Solutions of discrete-time state equations depend in a similar way on powers of the $\mathbf{A}$ matrix. Methods of evaluating these two functions are stressed because of their importance in the analysis of control systems expressed in state variable format.

8.2 POWERS OF A MATRIX AND MATRIX POLYNOMIALS

A matrix is conformable with itself only if it is square. Only square, $n \times n$ matrices are considered in this chapter. The product $\mathbf{AA}$ will be referred to as $\mathbf{A}^2$ for obvious reasons. The product of k such factors, $\mathbf{AA}\cdots\mathbf{A}$, is defined as $\mathbf{A}^k$. By definition, $\mathbf{A}^0 = \mathbf{I}$. With these definitions all the usual rules of exponents apply. That is,

$$\mathbf{A}^m\mathbf{A}^n = (\underbrace{\mathbf{AA}\cdots\mathbf{A}}_{m \text{ factors}})(\underbrace{\mathbf{AA}\cdots\mathbf{A}}_{n \text{ factors}}) = A^{m+n}$$

$$(\mathbf{A}^m)^n = (\underbrace{\mathbf{AA}\cdots\mathbf{A}}_{m\text{ factors}})^n = (\underbrace{\mathbf{A}^n\mathbf{A}^n\cdots\mathbf{A}^n}_{m\text{ factors}})$$

$$= (\underbrace{\mathbf{AA}\cdots\mathbf{A}}_{n\text{ factors}})(\underbrace{\mathbf{AA}\cdots\mathbf{A}}_{n\text{ factors}})\cdots(\underbrace{\mathbf{AA}\cdots\mathbf{A}}_{n\text{ factors}}) = \mathbf{A}^{mn}$$

If $\mathbf{A}$ is nonsingular, then $(\mathbf{A}^{-1})^n = \mathbf{A}^{-n}$ and $\mathbf{A}^n\mathbf{A}^{-n} = \mathbf{A}^n(\mathbf{A}^{-1})^n = (\mathbf{AA}^{-1})^n = \mathbf{I}^n = \mathbf{I}$ or $\mathbf{A}^{n-n} = \mathbf{A}^0 = \mathbf{I}$ as agreed upon earlier.

The notion of matrix powers can be used to define matrix polynomials in a natural way. For example, if $P(x) = c_m x^m + c_{m-1}x^{m-1} + \cdots + c_1 x + c_0$ is an mth degree polynomial in the scalar variable x, then a corresponding matrix polynomial can be defined as

$$P(\mathbf{A}) = c_m\mathbf{A}^m + c_{m-1}\mathbf{A}^{m-1} + \cdots + c_1\mathbf{A} + c_0\mathbf{I}$$

If the scalar polynomial can be written in factored form as

$$P(x) = c(x - a_1)(x - a_2)\cdots(x - a_m)$$

then this is also true for the matrix polynomial

$$P(\mathbf{A}) = c(\mathbf{A} - \mathbf{I}a_1)(\mathbf{A} - \mathbf{I}a_2)\cdots(\mathbf{A} - \mathbf{I}a_m)$$

Thus a matrix polynomial of an $n \times n$ matrix $\mathbf{A}$ is just another $n \times n$ matrix whose elements depend on $\mathbf{A}$ as well as on the coefficients of the polynomial. A clear distinction should be made between the matrix polynomial function defined here as a combination of powers of a matrix $\mathbf{A}$, and the polynomial matrices (discussed in Chapter 4 and in Section 6.3.1), which are matrices whose elements are polynomials in some variable such as the Laplace transform variable s.

8.3 INFINITE SERIES AND ANALYTIC FUNCTIONS OF MATRICES

Let $\mathbf{A}$ be an $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$. Consider the infinite series in a scalar variable x,

$$\sigma(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k + \cdots$$

It is well known that a given infinite series may converge or diverge depending on the value of x. For example, the geometric series

$$1 + x + x^2 + x^3 + \cdots + x^k + \cdots$$

converges for $|x| < 1$ and diverges otherwise. Some infinite series are convergent for all values of x, such as

$$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^k}{k!} + \cdots$$

Because this infinite series is so widely useful, it is given a special name, e^x. The various test for convergence of series will not be considered here. The following theorem is of major importance.

Theorem 8.1. Let $\mathbf{A}$ be an $n \times n$ matrix with eigenvalues λ_i. If the infinite series $\sigma(x) = a_0 + a_1 x + a_2 x^2 + \cdots$ is convergent for each of the n values $x = \lambda_i$, then the corresponding matrix infinite series

$$\sigma(\mathbf{A}) = a_0 \mathbf{I} + a_1 \mathbf{A} + a_2 \mathbf{A}^2 + \cdots + a_k \mathbf{A}^k + \cdots = \sum_{k=0}^{\infty} a_k \mathbf{A}^k$$

converges.

Definition 8.1 [1]. A single-valued function $f(z)$, with z a complex scalar, is said to be *analytic* at a point z_0 if and only if its derivative exists at every point in some neighborhood of z_0. Points at which the function is not analytic are called *singular points*. For example, $f(z) = 1/z$ has $z = 0$ as its only singular point and is analytic at every other point.

The result, which makes Theorem 8.1 useful for the purposes of this book, is Theorem 8.2.

Theorem 8.2. If a function $f(z)$ is analytic (contains no singularities) at every point in some circle Ω in the complex plane, then $f(z)$ can be represented as a convergent power series (the Taylor series) at every point z inside Ω [1].

Taken together, Theorems 8.1 and 8.2 give Theorem 8.3.

Theorem 8.3. If $f(z)$ is any function which is analytic within a circle in the complex plane which contains all eigenvalues λ_i of $\mathbf{A}$, then a corresponding matrix function $f(\mathbf{A})$ can be defined by a convergent power series.

EXAMPLE 8.1 The function $e^{\alpha x}$ is analytic for all values of x. Thus it has a convergent series representation

$$e^{\alpha x} = 1 + \alpha x + \frac{\alpha^2 x^2}{2!} + \frac{\alpha^3 x^3}{3!} + \cdots + \frac{\alpha^k x^k}{k!} + \cdots$$

The corresponding matrix function is defined as

$$e^{\alpha \mathbf{A}} = \mathbf{I} + \alpha \mathbf{A} + \frac{\alpha^2 \mathbf{A}^2}{2!} + \frac{\alpha^3 \mathbf{A}^3}{3!} + \cdots + \frac{\alpha^k \mathbf{A}^k}{k!} + \cdots$$ ■

At any point within the circles of convergence, the power series representations for two functions $f(z)$ and $g(z)$ can be added, differentiated, or integrated term by term. Also, the product of the two series gives another series which converges to $f(z)g(z)$ [1]. Because of Theorems 8.1 and 8.2 the same properties are valid for analytic functions of the matrix $\mathbf{A}$. The usual cautions with matrix algebra must be observed, however. In particular, matrix multiplication is not commutative.

EXAMPLE 8.2 $e^{\mathbf{A}t} e^{\mathbf{B}t} = e^{(\mathbf{A}+\mathbf{B})t}$ if and only if $\mathbf{AB} = \mathbf{BA}$. This is easily verified by multiplying out the first few terms for $e^{\mathbf{A}t}$ and $e^{\mathbf{B}t}$. ■

EXAMPLE 8.3 Find $\frac{d}{dt}[e^{\mathbf{A}t}]$. The series representation is

$$e^{\mathbf{A}t} = \mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \frac{\mathbf{A}^3 t^3}{3!} + \cdots$$

Term-by-term differentiation gives

$$\frac{de^{\mathbf{A}t}}{dt} = \mathbf{A} + \frac{2\mathbf{A}^2 t}{2!} + \frac{3\mathbf{A}^3 t^2}{3!} + \cdots$$

Since $\mathbf{A}$ can be factored out on either the left or the right,

$$\frac{de^{\mathbf{A}t}}{dt} = \mathbf{A}\left[\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \cdots\right] = \left[\mathbf{I} + \mathbf{A}t + \frac{\mathbf{A}^2 t^2}{2!} + \cdots\right]\mathbf{A} = \mathbf{A}e^{\mathbf{A}t} = e^{\mathbf{A}t}\mathbf{A} \quad \blacksquare$$

EXAMPLE 8.4 Compute $\int_0^t e^{\mathbf{A}\tau}\, d\tau$. Using the series representation, term-by-term integration gives

$$\int_0^t e^{\mathbf{A}\tau}\, d\tau = \int_0^t \mathbf{I}\, d\tau + \mathbf{A}\int_0^t \tau\, d\tau + \frac{\mathbf{A}^2}{2!}\int_0^t \tau^2\, d\tau + \cdots = \mathbf{I}t + \frac{\mathbf{A}t^2}{2} + \frac{\mathbf{A}^2 t^3}{3!} + \cdots$$

Therefore, $\mathbf{A}\int_0^t e^{\mathbf{A}\tau}\, d\tau + \mathbf{I} = e^{\mathbf{A}t}$ or, if $\mathbf{A}^{-1}$ exists,

$$\int_0^t e^{\mathbf{A}\tau}\, d\tau = \mathbf{A}^{-1}[e^{\mathbf{A}t} - \mathbf{I}] = [e^{\mathbf{A}t} - \mathbf{I}]\mathbf{A}^{-1} \quad \blacksquare$$

Since the exponential function is analytic for all finite arguments, it is possible to define

$$e^{-\alpha\mathbf{A}} = \mathbf{I} - \alpha\mathbf{A} + \frac{\alpha^2\mathbf{A}^2}{2} - \frac{\alpha^3\mathbf{A}^3}{3!} + \cdots - \cdots$$

Since $\mathbf{A}$ commutes with itself,

$$e^{-\alpha\mathbf{A}} e^{\alpha\mathbf{A}} = e^{\alpha(\mathbf{A}-\mathbf{A})} = e^{\alpha[0]} = \mathbf{I}$$

which is analogous to the scalar result.

Although the exponential function of a matrix is the one which will be most useful in this text, many other functions can be defined, for example:

$$\sin\mathbf{A} = \mathbf{A} - \frac{\mathbf{A}^3}{3!} + \frac{\mathbf{A}^5}{5!} - \cdots \qquad \sinh\mathbf{A} = \mathbf{A} + \frac{\mathbf{A}^3}{3!} + \frac{\mathbf{A}^5}{5!} + \cdots$$

$$\cos\mathbf{A} = \mathbf{I} - \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^4}{4!} - \cdots \qquad \cosh\mathbf{A} = \mathbf{I} + \frac{\mathbf{A}^2}{2!} + \frac{\mathbf{A}^4}{4!} + \cdots$$

Using these definitions, it can be verifed that relations analogous to the scalar results hold. For example,

$$\sin^2\mathbf{A} + \cos^2\mathbf{A} = \mathbf{I}$$

$$\sin\mathbf{A} = \frac{e^{j\mathbf{A}} - e^{-j\mathbf{A}}}{2j}, \qquad \cos\mathbf{A} = \frac{e^{j\mathbf{A}} + e^{-j\mathbf{A}}}{2}$$

$$\cosh^2 \mathbf{A} - \sinh^2 \mathbf{A} = \mathbf{I}$$

$$\sinh \mathbf{A} = \frac{e^{\mathbf{A}} - e^{-\mathbf{A}}}{2}, \qquad \cosh \mathbf{A} = \frac{e^{\mathbf{A}} + e^{-\mathbf{A}}}{2}$$

Although all of the preceding matrix functions are defined in terms of their power series representations, the series converge to $n \times n$ matrices. It is known from Chapter 7 that every square matrix has a unique Jordan form $\mathbf{J} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$, from which $\mathbf{A} = \mathbf{M}\mathbf{J}\mathbf{M}^{-1}$. Thus (see Problem 8.5) for any integer k, $\mathbf{A}^k = \mathbf{M}\mathbf{J}^k\mathbf{M}^{-1}$. Therefore, $f(\mathbf{A}) = \mathbf{M}f(\mathbf{J})\mathbf{M}^{-1}$ for f any finite-degree polynomial or infinite series. Any analytic function satisfying the conditions of Theorem 8.2 can be expressed in terms of the modal matrix and the Jordan form. In those cases for which $\mathbf{A}$ is diagonalizable,

$$\mathbf{J} = \mathbf{\Lambda} = \text{Diag}[\lambda_1, \ldots, \lambda_n] \quad \text{and} \quad f(\mathbf{\Lambda}) = \text{Diag}[f(\lambda_1), \ldots, f(\lambda_n)]$$

The Jordan form decomposition of $\mathbf{A}$ may not be the most efficient method of computing $f(\mathbf{A})$ because it requires determination of the modal matrix. Section 8.5 develops methods of determining the closed form expressions for analytic functions of square matrices. However, the Jordan form–modal matrix approach does show several useful results in a simple fashion. For example, if $\{\lambda_i\}$ are the eigenvalues of $\mathbf{A}$, then $f(\lambda_i)$ are the eigenvalues of $f(\mathbf{A})$. This result is called Frobenius' theorem. As applied to the state equations, this indicates that if all eigenvalues of $\mathbf{A}$ are in the left-half complex plane (stable poles) then all eigenvalues of $e^{\mathbf{A}t}$ will have magnitudes less than one, that is, they are inside the unit circle. This is a result that could have been anticipated from the Chapter 2 discussion of Z-transforms and the stability regions for continuous and discrete systems.

8.4 THE CHARACTERISTIC POLYNOMIAL AND CAYLEY-HAMILTON THEOREM

Although arbitrary matrix polynomials have been discussed, one very special polynomial is the characteristic polynomial. If the characteristic polynomial for the matrix $\mathbf{A}$ is written as

$$|\mathbf{A} - \mathbf{I}\lambda| = (-\lambda)^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \cdots + c_1\lambda + c_0 = \Delta(\lambda)$$

then the corresponding matrix polynomial is

$$\Delta(\mathbf{A}) = (-1)^n\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + c_{n-2}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{A} + c_0\mathbf{I}$$

Cayley-Hamilton Theorem. Every matrix satisfies its own characteristic equation; that is, $\Delta(\mathbf{A}) = [\mathbf{0}]$.

Proof. (Valid when $\mathbf{A}$ is similar to a diagonal matrix. For the general case, see Problems 8.5 and 8.6.) A similarity transformation reduces $\mathbf{A}$ to the diagonal matrix $\mathbf{\Lambda}$, so

$$\mathbf{A} = \mathbf{M}\mathbf{\Lambda}\mathbf{M}^{-1}, \qquad \mathbf{A}^2 = \mathbf{M}\mathbf{\Lambda}^2\mathbf{M}^{-1}, \ldots, \mathbf{A}^k = \mathbf{M}\mathbf{\Lambda}^k\mathbf{M}^{-1}$$

Therefore,

$$\Delta(\mathbf{A}) = \mathbf{M}[(-1)^n \mathbf{\Lambda}^n + c_{n-1}\mathbf{\Lambda}^{n-1} + c_{n-2}\mathbf{\Lambda}^{n-2} + \cdots + c_1\mathbf{\Lambda} + c_0\mathbf{I}]\mathbf{M}^{-1}$$

Each term inside the brackets is a diagonal matrix. The sum of a typical i, i element is

$$(-\lambda_i)^n + c_{n-1}\lambda_i^{n-1} + \cdots + c_1\lambda_i + c_0$$

which is zero because λ_i is a root of the characteristic equation. Therefore,

$$\Delta(\mathbf{A}) = \mathbf{M}[\mathbf{0}]\mathbf{M}^{-1} = [\mathbf{0}]$$

EXAMPLE 8.5 Let $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$, $|\mathbf{A} - \mathbf{I}\lambda| = (3-\lambda)(2-\lambda) - 1 = \lambda^2 - 5\lambda + 5$. Then

$$\Delta(\mathbf{A}) = \mathbf{A}^2 - 5\mathbf{A} + 5\mathbf{I} = \begin{bmatrix} 10 & 5 \\ 5 & 5 \end{bmatrix} - 5\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix} + 5\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$ ■

Definition 8.2 [2]. The minimum polynomial of a square matrix **A** is the lowest-degree monic polynomial (that is, the coefficient of the highest power is normalized to one) which satisfies

$$m(\mathbf{A}) = [\mathbf{0}]$$

It would be slightly more efficient to use the minimum polynomial rather than the characteristic polynomial in some cases (in Section 8.5, for example). The minimum polynomial $m(\mathbf{A})$ and the characteristic polynomial $\Delta'(\mathbf{A})$ are often the same (if all λ_i are distinct, or in the simple degeneracy case). In factored form the only possible differences between $m(\mathbf{A})$ and $\Delta'(\mathbf{A})$ are the powers of the terms involving repeated roots. Since for the definitions given in Chapter 7 $\Delta(\lambda) = (-1)^n \Delta'(\lambda)$, then

$$\Delta(\lambda) = (-1)^n(\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_p)^{m_p}$$
$$m(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2} \cdots (\lambda - \lambda_p)^{k_p}$$

In all cases $k_i \le m_i$. In this text the characteristic polynomial will be used in most cases rather than the minimum polynomial because it is more familiar, and only occasionally requires a small amount of extra calculations (see Problem 8.18).

8.5 SOME USE OF THE CAYLEY-HAMILTON THEOREM

Matrix Inversion

Let **A** be an $n \times n$ matrix with the characteristic equation $\Delta(\lambda) = (-\lambda)^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0 = 0$. Recall that the constant $c_0 = \lambda_1\lambda_2\cdots\lambda_n = |\mathbf{A}|$ and is zero if and only if **A** is singular. Using $\Delta(\mathbf{A}) = (-1)^n\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \cdots + c_1\mathbf{A} + c_0\mathbf{I} = \mathbf{0}$, and assuming $\mathbf{A}^{-1}$ exists, multiplication by $\mathbf{A}^{-1}$ gives

$$(-1)^n\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{I} + c_0\mathbf{A}^{-1} = \mathbf{0}$$

or

$$\mathbf{A}^{-1} = \frac{-1}{c_0}[(-1)^n\mathbf{A}^{n-1} + c_{n-1}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{I}]$$

EXAMPLE 8.6 Let $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$, $\Delta(\lambda) = \lambda^2 - 5\lambda + 5$. Then

$$\Delta(\mathbf{A}) = \mathbf{A}^2 - 5\mathbf{A} + 5\mathbf{I} = \mathbf{0}, \qquad \mathbf{A}^{-1} = -\tfrac{1}{5}[\mathbf{A} - 5\mathbf{I}] = -\frac{1}{5}\begin{bmatrix} -2 & 1 \\ 1 & -3 \end{bmatrix}$$ ■

Reduction of a Polynomial in A *to One of Degree n − 1 or Less*

Let $P(x)$ be a scalar polynomial of degree m. Let $P_1(x)$ be another polynomial of degree n, where $n < m$. Then $P(x)$ can always be written $P(x) = Q(x)P_1(x) + R(x)$, where $Q(x)$ is a polynomial of degree $m - n$ and $R(x)$ is a remainder polynomial of degree $n - 1$ or less. For this scalar case, $Q(x)$ and $R(x)$ could be found by formally dividing $P(x)$ by $P_1(x)$, since this gives $P(x)/P_1(x) = Q(x) + R(x)/P_1(x)$.

EXAMPLE 8.7 Let $P(x) = 3x^4 + 2x^2 + x + 1$ and $P_1(x) = x^2 - 3$. Then it is easily verified that

$$P(x) = (3x^2 + 11)(x^2 - 3) + (x + 34)$$

so that $Q(x) = 3x^2 + 11$ and $R(x) = x + 34$. ■

Similarly, the matrix polynomial $P(\mathbf{A})$ can be written

$$P(\mathbf{A}) = Q(\mathbf{A})P_1(\mathbf{A}) + R(\mathbf{A})$$

since it is always defined in the same manner as its scalar counterpart. If the arbitrary polynomial P_1 used above is selected as the characteristic polynomial of $\mathbf{A}$, then the scalar version of P is

$$P(x) = Q(x)\Delta(x) + R(x)$$

Note that $\Delta(x) \neq 0$ except for those specific values $x = \lambda_i$, the eigenvalues. The matrix version of P is

$$P(\mathbf{A}) = Q(\mathbf{A})\Delta(\mathbf{A}) + R(\mathbf{A})$$

By the Cayley-Hamilton theorem, $\Delta(\mathbf{A}) = \mathbf{0}$, so $P(\mathbf{A}) = R(\mathbf{A})$. The coefficients of the matrix remainder polynomial R can be found by long division of the corresponding scalar polynomials. Alternatively, the Cayley-Hamilton theorem can be used to reduce each individual term in $P(\mathbf{A})$ to one of degree $n - 1$ or less.

EXAMPLE 8.8 Let $\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$, $\Delta(\lambda) = \lambda^2 - 5\lambda + 5$. Compute $P(\mathbf{A}) = \mathbf{A}^4 + 3\mathbf{A}^3 + 2\mathbf{A}^2 + \mathbf{A} + \mathbf{I}$.

Method 1: By long division,

$$\frac{P(x)}{\Delta(x)} = x^2 + 8x + 37 + \frac{146x - 184}{x^2 - 5x + 5} \quad \text{or} \quad P(x) = (x^2 + 8x + 37)\Delta(x) + (146x - 184)$$

Therefore, $R(x) = 146x - 184$ and the Cayley-Hamilton theorem guarantees that $P(\mathbf{A}) = R(\mathbf{A}) = 146\mathbf{A} - 184\mathbf{I}$.

Method 2: From the Cayley-Hamilton theorem,

$$\mathbf{A}^2 - 5\mathbf{A} + 5\mathbf{I} = [\mathbf{0}] \quad \text{or} \quad \mathbf{A}^2 = 5(\mathbf{A} - \mathbf{I})$$

Hence

$$\mathbf{A}^4 = \mathbf{A}^2\mathbf{A}^2 = 25(\mathbf{A}-\mathbf{I})(\mathbf{A}-\mathbf{I}) = 25(\mathbf{A}^2 - 2\mathbf{A} + \mathbf{I}) = 25[5(\mathbf{A}-\mathbf{I}) - 2\mathbf{A} + \mathbf{I}] = 25[3\mathbf{A} - 4\mathbf{I}]$$

$$\mathbf{A}^3 = \mathbf{A}(\mathbf{A}^2) = 5(\mathbf{A}^2 - \mathbf{A}) = 5[5(\mathbf{A}-\mathbf{I}) - \mathbf{A}] = 5[4\mathbf{A} - 5\mathbf{I}]$$

Thus

$$P(\mathbf{A}) = 25[3\mathbf{A} - 4\mathbf{I}] + 15[4\mathbf{A} - 5\mathbf{I}] + 10(\mathbf{A}-\mathbf{I}) + \mathbf{A} + \mathbf{I} = 146\mathbf{A} - 184\mathbf{I}$$ ■

Closed Form Solution for Analytic Functions of Matrices

Let $f(x)$ be a function which is analytic in a region Ω of the complex plane and let $\mathbf{A}$ be an $n \times n$ matrix whose eigenvalues $\lambda_i \in \Omega$. Then $f(x)$ has a power series representation

$$f(x) = \sum_{k=0}^{\infty} \alpha_k x^k$$

It is possible to regroup the infinite series for $f(x)$ so that

$$f(x) = \Delta(x) \sum_{k=0}^{\infty} \beta_k x^k + R(x)$$

The remainder R will have degree less than or equal to $n-1$. The analytic function of the square matrix $\mathbf{A}$ is defined by the same series as its scalar counterpart, but with $\mathbf{A}$ replacing x. Therefore, $f(\mathbf{A}) = R(\mathbf{A})$, since $\Delta(\mathbf{A})$ is always the null matrix.

Although the form of $R(x)$ is known to be

$$R(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_{n-1} x^{n-1}$$

it is clearly impossible to find the coefficients α_i by long division as in Example 8.8. However, if the n eigenvalues λ_i are distinct, n equations for determining the n α_i terms are available. Since $\Delta(\lambda_i) = 0$, setting $x = \lambda_i$ gives $f(\lambda_i) = R(\lambda_i)$, $i = 1, 2, \ldots, n$.

EXAMPLE 8.9 Find the closed form expression for sin **A** if $\mathbf{A} = \begin{bmatrix} -3 & 1 \\ 0 & -2 \end{bmatrix}$.

$\Delta(\lambda) = (-3-\lambda)(-2-\lambda)$, so $\lambda_1 = -3$, $\lambda_2 = -2$. Since $\mathbf{A}$ is a 2×2 matrix, it is known that R is of degree one (or less):

$$R(x) = \alpha_0 + \alpha_1 x$$

Also, using $x = \lambda_1$ and $x = \lambda_2$ gives

$$\sin\lambda_1 = R(\lambda_1) = \alpha_0 + \alpha_1\lambda_1$$

$$\sin\lambda_2 = R(\lambda_2) = \alpha_0 + \alpha_1\lambda_2$$

Solving gives

$$\alpha_0 = \frac{\lambda_1 \sin\lambda_2 - \lambda_2 \sin\lambda_1}{\lambda_1 - \lambda_2}, \qquad \alpha_1 = \frac{\sin\lambda_1 - \sin\lambda_2}{\lambda_1 - \lambda_2}$$

or $\alpha_0 = 3\sin(-2) - 2\sin(-3)$, $\alpha_1 = \sin(-2) - \sin(-3)$. Using these in $f(\mathbf{A}) = R(\mathbf{A})$ gives

$$\sin\mathbf{A} = \alpha_0\mathbf{I} + \alpha_1\mathbf{A} = \begin{bmatrix} \alpha_0 - 3\alpha_1 & \alpha_1 \\ 0 & \alpha_0 - 2\alpha_1 \end{bmatrix} = \begin{bmatrix} \sin(-3) & \sin(-2) - \sin(-3) \\ 0 & \sin(-2) \end{bmatrix}$$ ■

When λ_i is a repeated root, this procedure must be modified. Some of the equations $f(\lambda_i) = R(\lambda_i)$ will be repeated, so they do not form a set of n linearly independent equations. However, for λ_i a repeated root, $\left.\dfrac{d\Delta(\lambda)}{d\lambda}\right|_{\lambda=\lambda_i} = 0$ also, and so

$$\left.\frac{df(\lambda)}{d\lambda}\right|_{\lambda=\lambda_i} = \left.\frac{d\Delta}{d\lambda}\sum_{k=0}^{\infty} \beta_k \lambda_i^k + \Delta(\lambda_i)\frac{d}{d\lambda}[\Sigma\beta_k \lambda^k]\right|_{\lambda=\lambda_i} + \left.\frac{dR}{d\lambda}\right|_{\lambda=\lambda_i} = \left.\frac{dR}{d\lambda}\right|_{\lambda=\lambda_i}$$

For an eigenvalue with algebraic multiplicity m_i, the first $m_i - 1$ derivatives of Δ all vanish and thus

$$f(\lambda_i) = R(\lambda_i), \qquad \left.\frac{df}{d\lambda}\right|_{\lambda_i} = \left.\frac{dR}{d\lambda}\right|_{\lambda_i}, \qquad \left.\frac{d^2 f}{d\lambda^2}\right|_{\lambda_i} = \left.\frac{d^2 R}{d\lambda^2}\right|_{\lambda_i}, \ldots,$$

$$\left.\frac{d^{m_i-1} f}{d\lambda^{m_i-1}}\right|_{\lambda_i} = \left.\frac{d^{m_i-1} R}{d\lambda^{m_i-1}}\right|_{\lambda_i}$$

form a set of m_i linearly independent equations. Thus a full set of n equations is always available for finding the α_i coefficients of the remainder term R.

EXAMPLE 8.10 Find the closed form expression for $e^{\mathbf{A}t}$ if $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 27 & -27 & 9 \end{bmatrix}$.

We have

$$|\mathbf{A} - \mathbf{I}\lambda| = \Delta(\lambda) = -\lambda^3 + 9\lambda^2 - 27\lambda + 27 = (3 - \lambda)^3$$

Therefore, $\lambda_1 = \lambda_2 = \lambda_3 = 3$, and

$$e^{\mathbf{A}t} = R(\mathbf{A}) = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} + \alpha_2 \mathbf{A}^2$$

where

$$e^{3t} = \alpha_0 + 3\alpha_1 + 9\alpha_2$$

$$\left.\frac{de^{\lambda t}}{d\lambda}\right|_{\lambda=3} = \left.\frac{d}{d\lambda}[\alpha_0 + \lambda\alpha_1 + \lambda^2\alpha_2]\right|_{\lambda=3} \quad \text{or} \quad te^{3t} = \alpha_1 + 6\alpha_2$$

$$\left.\frac{d^2 e^{\lambda t}}{d\lambda^2}\right|_{\lambda=3} = \left.\frac{d^2}{d\lambda^2}[\alpha_0 + \lambda\alpha_1 + \lambda^2\alpha_2]\right|_{\lambda=3} \quad \text{or} \quad t^2 e^{3t} = 2\alpha_2$$

Solving for α_0, α_1, and α_2 gives

$$\alpha_2 = \tfrac{1}{2}t^2 e^{3t}, \qquad \alpha_1 = te^{3t} - 6\alpha_2 = (t - 3t^2)e^{3t},$$

$$\alpha_0 = e^{3t} - 3\alpha_1 - 9\alpha_2 = (1 - 3t + \tfrac{9}{2}t^2)e^{3t}$$

Using these coefficients in $R(\mathbf{A})$ gives

$$e^{\mathbf{A}t} = \begin{bmatrix} 1 - 3t + \frac{9}{2}t^2 & t - 3t^2 & \frac{1}{2}t^2 \\ \frac{27}{2}t^2 & 1 - 3t - 9t^2 & t + \frac{3}{2}t^2 \\ 27t + \frac{81}{2}t^2 & -27t - 27t^2 & 1 + 6t + \frac{9}{2}t^2 \end{bmatrix} e^{3t}$$

■

When some eigenvalues are repeated and others are simple roots, a full set of n independent equations are still available for computing the α_i coefficients.

The method presented above for computing functions of a matrix requires a knowledge of the eigenvalues. If the eigenvectors are also known, an alternative method can be used (see Problem 8.21). For the particular function $f(\mathbf{A}) = e^{At}$, a third alternative is available (see Problems 8.19 and 8.20). Finally, an aproximation can be obtained by truncating the infinite series after a finite number of terms.

8.6 SOLUTION OF THE UNFORCED STATE EQUATIONS

Complete solutions of the state equations are considered in the next chapter. In order to emphasize the importance of the matrix exponential and power functions, the initial condition response of the state equations are considered briefly here. Only the constant coefficient case is considered, that is the $\mathbf{A}$ matrix is not a function of time.

The Continuous-Time Case

When the control input $\mathbf{u}(t)$ is zero, the state vector $\mathbf{x}(t)$ evolves in time according to solutions of

$$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) \tag{8.1}$$

starting with initial conditions $\mathbf{x}(0)$. One classic method of solving differential equations is to guess a solution form, perhaps with adjustable parameters, and then see if it can be made to satisfy (a) the initial condition and (b) the differential equation. Here the "guess" is $\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0)$, and we merely verify that this is correct. In Example 8.4 $e^{[0]} = \mathbf{I}$ was introduced, so when $t = 0$ is substituted, the assumed solution reduces to $\mathbf{I}\mathbf{x}(0)$ and the initial conditions are satisfied. From Example 8.3 $d[e^{At}]/dt = \mathbf{A}e^{At}$, so that substitution of the assumed solution into Eq. (*8.1*) gives the self-consistent result $\dot{\mathbf{x}}(t) = \mathbf{A}e^{At}\mathbf{x}(0) = \mathbf{A}\mathbf{x}(t)$.

In Chapter 4 it was stated that the Laplace transform of a matrix of time functions can be calculated term by term on each element of the matrix. This provides an alternative approach to solving Eq. (*8.1*). The state vector $\mathbf{x}(t)$ is an example of a time-variable column matrix. Let $\mathscr{L}\{\mathbf{x}(t)\} = \mathbf{X}(s)$. Then $\mathscr{L}\{\dot{\mathbf{x}}(t)\} = s\mathbf{X}(s) - \mathbf{x}(0)$, so that Eq. (*8.1*) transforms to $s\mathbf{X}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{X}(s)$. Solving for $\mathbf{X}(s)$ gives $\mathbf{X}(s) = [s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{x}(0)$. The inverse Laplace transform then gives $\mathbf{x}(t) = \mathscr{L}^{-1}\{[s\mathbf{I} - \mathbf{A}]^{-1}\}\mathbf{x}(0)$. Not only is this the solution, but comparing it with the previous solution shows that

$$e^{\mathbf{A}t} = \mathscr{L}^{-1}\{[s\mathbf{I} - \mathbf{A}]^{-1}\}$$

(See also Problem 8.19.) One final method of solving Eq. (*8.1*) utilizes modal decoupling. The matrix $\mathbf{A}$ is assumed diagonalizable to keep the discussion simple. Equation (*8.1*) can be written as $\dot{\mathbf{x}} = \mathbf{M}\mathbf{\Lambda}\mathbf{M}^{-1}\mathbf{x}$. If a change of variables (change of basis) $\mathbf{M}^{-1}\mathbf{x} = \mathbf{w}$ is used, after premultiplying by $\mathbf{M}$ this becomes $\dot{\mathbf{w}} = \mathbf{\Lambda}\mathbf{w}$. Because $\mathbf{\Lambda}$ is diagonal this represents n scalar equations $\dot{w}_i = \lambda_i w_i$. Each of these equations has a solution $w_i(t) = e^{\lambda_i t} w_i(0)$. The entire set of solution components can be written as $\mathbf{w}(t) = \text{Diag}[e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t}]\mathbf{w}(0)$. The solution for $\mathbf{x}$ and not $\mathbf{w}$ is desired, but $\mathbf{x}(t) = \mathbf{M}\mathbf{w}(t)$. The initial conditions are presumed given for $\mathbf{x}(0)$ and not $\mathbf{w}(0)$, but $\mathbf{w}(0) = \mathbf{M}^{-1}\mathbf{x}(0)$, so the final solution is $\mathbf{x}(t) = \mathbf{M}\ \text{Diag}[e^{\lambda_1 t}, e^{\lambda_2 t}, \dots, e^{\lambda_n t}]\mathbf{M}^{-1}\mathbf{x}(0)$. Comparison with the previous solutions shows that $e^{\mathbf{A}t} = \mathbf{M}e^{\Lambda t}\mathbf{M}^{-1}$. This is an explicit verification of the result given at the end of Sec. 8.3 for a general function $f(\mathbf{A})$.

The Discrete-Time Case

The initial condition response of an unforced constant coefficient discrete-time system evolves according to

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) \tag{8.2}$$

The state at any general time point t_k is known to be (see Sec. 6.9) $\mathbf{x}(k) = \mathbf{A}^k\mathbf{x}(0)$. The Z-transform of any time-variable matrix, such as the column matrix $\mathbf{x}(k)$, is obtained by transforming each scalar element in the matrix. If $\mathbf{Z}\{\ \}$ represents the Z-transform operator and if the column of transformed elements is defined as $\mathbf{X}(z) = Z\{\mathbf{x}(k)\}$, then $Z\{\mathbf{x}(k+1)\} = z\mathbf{X}(z) - z\mathbf{x}(0)$, and the transformed Eq. (8.2) can be written as $(z\mathbf{I} - \mathbf{A})\mathbf{X}(z) = z\mathbf{x}(0)$. Solving for $\mathbf{X}(z)$ gives $\mathbf{X}(z) = [z\mathbf{I} - \mathbf{A}]^{-1}z\mathbf{x}(0)$. Using $Z^{-1}\{\ \}$ to represent the inverse Z-transform, the time domain solution is

$$\mathbf{x}(k) = Z^{-1}\{[z\mathbf{I} - \mathbf{A}]^{-1}z\}\mathbf{x}(0)$$

Comparison with the previous solution shows that the kth power of a matrix can be computed from $\mathbf{A}^k = Z^{-1}\{[z\mathbf{I} - \mathbf{A}]^{-1}z\}$, which is similar to the Laplace transform result for the continuous-time system. Writing $\mathbf{A} = \mathbf{M}\Lambda\mathbf{M}^{-1}$ and repeating the steps used in decoupling the continuous-time system shows that $\mathbf{A}^k = \mathbf{M}\Lambda^k\mathbf{M}^{-1}$ is an alternative way of computing the power of a matrix, as was already known from Sec. 8.3.

REFERENCES

1. Churchill, R. V.: *Introduction to Complex Variables and Applications,* 2nd ed., McGraw-Hill, New York, 1960.
2. Zadeh, L. A. and C. A. Desoer: *Linear System Theory, The State Space Approach,* McGraw-Hill, New York, 1963.
3. Moler, C. B. and C. F. Van Loan: "Nineteen Dubious Ways to Compute the Exponential of a Matrix," *SIAM Review,* Vol. 20, 1978, pp. 801–836.
4. Franklin, G. F. and J. D. Powell: *Digital Control of Dynamic Systems,* Addison-Wesley, Reading, Mass., 1980.
5. Brogan, W. L.: "Optimal Control Theory Applied to Systems Described by Partial Differential Equations," *Advances in Control Systems,* Vol. 6, C. T. Leondes, ed., Academic Press, New York, 1968.

ILLUSTRATIVE PROBLEMS

Inversion of Matrices and Reduction of Polynomials

8.1 If $\mathbf{A} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$, use the Cayley-Hamilton theorem to compute a. $\mathbf{A}^{-1}$ and b. $P(\mathbf{A}) = \mathbf{A}^5 + 16\mathbf{A}^4 + 32\mathbf{A}^3 + 16\mathbf{A}^2 + 4\mathbf{A} + \mathbf{I}$.

(a) $\Delta(\lambda) = |\mathbf{A} - \mathbf{I}\lambda| = \lambda^2 - 2\lambda + 2$. Therefore,

$$\mathbf{A}^2 - 2\mathbf{A} + 2\mathbf{I} = [\mathbf{0}] \quad \text{or} \quad \mathbf{A}^{-1} = -\tfrac{1}{2}[\mathbf{A} - 2\mathbf{I}] = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$$

(b) $P(\mathbf{A}) = R(\mathbf{A})$, where R is the remainder term in

$$\frac{P(x)}{\Delta(x)} = x^3 + 18x^2 + 66x + 112 + \frac{96x - 223}{\Delta(x)}$$

Therefore, $P(\mathbf{A}) = 96\mathbf{A} - 223\mathbf{I} = \begin{bmatrix} -127 & -96 \\ 96 & -127 \end{bmatrix}$.

8.2 $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$. Find a. $\mathbf{A}^{-1}$ and b. $P(\mathbf{A}) = \mathbf{A}^5 + \mathbf{A}^3 + \mathbf{A} + \mathbf{I}$.

(a) $\Delta(\lambda) = \lambda^2 - 5\lambda - 2$. Therefore, $\Delta(\mathbf{A}) = \mathbf{A}^2 - 5\mathbf{A} - 2\mathbf{I} = [\mathbf{0}]$ by the Cayley-Hamilton theorem and $\mathbf{A} - 5\mathbf{I} - 2\mathbf{A}^{-1} = [\mathbf{0}]$ or $\mathbf{A}^{-1} = \frac{1}{2}[\mathbf{A} - 5\mathbf{I}] = \begin{bmatrix} -2 & 1 \\ \frac{3}{2} & -\frac{1}{2} \end{bmatrix}$.

(b) $P(x)/\Delta(x) = x^3 + 5x^2 + 28x + 150 + [(807x + 301)/\Delta(x)]$. Therefore,

$$P(x) = \Delta(x)[x^3 + 5x^2 + 28x + 150] + \underbrace{807x + 301}_{R(x)}$$

and $P(\mathbf{A}) = R(\mathbf{A}) = 807\mathbf{A} + 301\mathbf{I} = \begin{bmatrix} 1108 & 1614 \\ 2421 & 3529 \end{bmatrix}$.

Functions with Singularities

8.3 Comment on the following functions in view of Theorems 8.1, 8.2, and 8.3:
(a) $f_a(x) = 1/(1-x)$, and **(b)** $f_b(x) = \tan x$.

(a) $f_a(x)$ has a singularity (pole) at $x = 1$, but is analytic elsewhere. In particular, it is analytic at all points x in the complex plane inside the circle $|x| < 1$, and for these points

$$f_a(x) = 1 + x + x^2 + x^3 + \cdots$$

If all the eigenvalues of $\mathbf{A}$ satisfy $|\lambda_i| < 1$, then $f_a(\mathbf{A}) = (\mathbf{I} - \mathbf{A})^{-1}$ exists and can be written as a convergent series

$$f_a(\mathbf{A}) = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \mathbf{A}^3 + \cdots$$

Note that $(\mathbf{I} - \mathbf{A})f_a(\mathbf{A}) = \mathbf{I} + (\mathbf{A} - \mathbf{A}) + (\mathbf{A}^2 - \mathbf{A}^2) + \cdots = \mathbf{I}$ as required. When $\mathbf{A}$ has $\lambda = 1$ as an eigenvalue, then $(\mathbf{I} - \mathbf{A})$ is singular and the inverse does not exist. If $\lambda = 1$ is not an eigenvalue, but at least one eigenvalue has a magnitude larger than unity, then $(\mathbf{I} - \mathbf{A})^{-1}$ exists but cannot be represented by the above infinite series.

(b) $f_b(x) = \sin x/\cos x$ has a singularity (pole) at each zero of $\cos x$, that is, at $x = \pm\pi/2$, $\pm 3\pi/2, \ldots$, but is analytic elsewhere. If all eigenvalues of $\mathbf{A}$ satisfy $|\lambda| < \pi/2$, then we could define $\tan \mathbf{A} = \mathbf{A} + \mathbf{A}^3/3 + 2\mathbf{A}^5/15 + \cdots$ and be assured that this series is convergent (see Problem 8.11).

Powers of a Jordan Block

8.4 Let $\mathbf{J}_1$ be an $m \times m$ Jordan block. Show that for any integer $k > 0$,

$$\mathbf{J}_1^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k(k-1)\lambda^{k-2} & \frac{1}{3!}k(k-1)(k-2)\lambda^{k-3} & \cdots & \frac{k!\lambda^{k-m+1}}{(k-m+1)!(m-1)!} \\ 0 & \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k(k-1)\lambda^{k-2} & & \\ 0 & 0 & \lambda^k & k\lambda^{k-1} & & \vdots \\ 0 & 0 & 0 & \lambda^k & & \\ 0 & 0 & 0 & & \ddots & k\lambda^{k-1} \\ \vdots & & & & & \\ 0 & 0 & 0 & & & \lambda^k \end{bmatrix}$$

With $k=1$, $\mathbf{J}_1$ satisfies the preceding equation. To see this, it must be recalled that $(k-m+1)! = \infty$ if $k-m+1<0$. Multiplying gives $\mathbf{J}_1$ squared:

$$\mathbf{J}_1 = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & & \\ 0 & 0 & \lambda & & \vdots \\ 0 & 0 & 0 & & \\ \vdots & & & & 1 \\ 0 & 0 & 0 & & \lambda \end{bmatrix}, \qquad \mathbf{J}_1^2 = \begin{bmatrix} \lambda^2 & 2\lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda^2 & 2\lambda & 1 & \ddots & \vdots \\ 0 & 0 & \lambda^2 & 2\lambda & \ddots & \\ \vdots & & & & \ddots & 1 \\ & & & & & 2\lambda \\ 0 & 0 & 0 & 0 & & \lambda \end{bmatrix}$$

Use induction, assuming the stated form holds for k, and show that it holds for $k+1$ by computing

$$\begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & & \\ 0 & 0 & \lambda & & \\ 0 & 0 & 0 & & \vdots \\ \vdots & & & & \\ 0 & 0 & 0 & & \lambda \end{bmatrix} \begin{bmatrix} \lambda^k & k\lambda^{k-1} & \frac{1}{2!}k(k-1)\lambda^{k-2} & \cdots & \\ 0 & \lambda^k & k\lambda^{k-1} & & \vdots \\ 0 & 0 & \lambda^k & \ddots & \\ 0 & 0 & 0 & \ddots & \\ \vdots & \vdots & & & k\lambda^{k-1} \\ 0 & 0 & & & \lambda^k \end{bmatrix} =$$

$$\begin{bmatrix} \lambda^{k+1} & (k+1)\lambda^k & \frac{1}{2!}k(k-1)\lambda^{k-1} + k\lambda^{k-1} & \frac{1}{3!}k(k-1)(k-2)\lambda^{k-2} + \frac{1}{2!}k(k-1)\lambda^{k-2} & \cdots \\ 0 & \lambda^{k+1} & (k+1)\lambda^k & & \\ 0 & 0 & \lambda^{k+1} & & \vdots \\ 0 & 0 & 0 & \vdots & \\ \vdots & & & & \\ 0 & 0 & 0 & & \lambda^{k+1} \end{bmatrix}$$

But

$$\left[\frac{1}{2!}k(k-1)+k\right]\lambda^{k-1} = \frac{k}{2!}(k-1+2)\lambda^{k-1} = \frac{(k+1)k}{2!}\lambda^{k-1}$$

$$\left[\frac{1}{3!}k(k-1)(k-2)+\frac{1}{2!}k(k-1)\right]\lambda^{k-2} = \frac{1}{3!}k(k-1)(k+1)\lambda^{k-2}$$

etc.

so the stated result holds.

Proof of Cayley-Hamilton Theorem

8.5 Prove the Cayley-Hamilton theorem when $\mathbf{A}$ is not diagonalizable.

Any square $\mathbf{A}$ can be reduced to Jordan canonical form $\mathbf{J} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$, so that $\mathbf{A} = \mathbf{M}\mathbf{J}\mathbf{M}^{-1}$, $\mathbf{A}^2 = \mathbf{M}\mathbf{J}\mathbf{M}^{-1}\mathbf{M}\mathbf{J}\mathbf{M}^{-1} = \mathbf{M}\mathbf{J}^2\mathbf{M}^{-1}$, $\mathbf{A}^k = \mathbf{M}\mathbf{J}^k\mathbf{M}^{-1}$. If the characteristic polynomial is $\Delta(\lambda) = c_n\lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$, then

$$\begin{aligned} \Delta(\mathbf{A}) &= c_n\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \cdots + c_1\mathbf{A} + c_0\mathbf{I} \\ &= \mathbf{M}[c_n\mathbf{J}^n + c_{n-1}\mathbf{J}^{n-1} + \cdots + c_1\mathbf{J} + c_0\mathbf{I}]\mathbf{M}^{-1} \end{aligned}$$

It is to be shown that the matrix polynomial inside the brackets sums to the zero matrix. The Jordan form is $\mathbf{J} = \text{diag}[\mathbf{J}_1, \mathbf{J}_2, \ldots, \mathbf{J}_p]$, where $\mathbf{J}_i$ are Jordan blocks. Also, $\mathbf{J}^k = \text{diag}[\mathbf{J}_1^k, \mathbf{J}_2^k, \ldots, \mathbf{J}_p^k]$. Therefore, to prove the theorem it is only necessary to show that $c_n\mathbf{J}_i^n + c_{n-1}\mathbf{J}_i^{n-1} + \cdots + c_1\mathbf{J}_i + c_0\mathbf{I} = [\mathbf{0}]$ for a typical block $\mathbf{J}_i$. Using the result of the previous problem, all terms below the main diagonal are zero. The sum of the terms in a typical main diagonal position is $c_n \lambda_i^n + c_{n-1}\lambda_i^{n-1} + \cdots + c_1\lambda_i + c_0$ and equals zero because this is just the characteristic polynomial evaluated with a root λ_i. A typical term of the sum in the diagonal just above the main diagonal is $nc_n\lambda_i^{n-1} + (n-1)c_{n-1}\lambda_i^{n-2} + \cdots + c_2\lambda_i + c_1$. This sum is zero since it equals $\left.\dfrac{d\Delta(\lambda)}{d\lambda}\right|_{\lambda=\lambda_i}$. The root λ_i is a multiple root, so

$$\Delta(\lambda) = c(\lambda - \lambda_i)^m(\lambda - \lambda_j)(\lambda - \lambda_k)\ldots, \quad \text{and} \quad \left.\frac{d\Delta}{d\lambda}\right|_{\lambda=\lambda_i} = 0$$

If $\mathbf{J}_i$ is an $r \times r$ block, λ_i is at least an rth-order root, so

$$\left.\frac{1}{2}\frac{d^2\Delta(\lambda)}{d\lambda^2}\right|_{\lambda=\lambda_i} = 0, \qquad \left.\frac{1}{3!}\frac{d^3\Delta(\lambda)}{d\lambda^3}\right|_{\lambda=\lambda_i} = 0, \ldots, \qquad \left.\frac{1}{(r-1)!}\frac{d^{r-1}\Delta(\lambda)}{d\lambda^{r-1}}\right|_{\lambda=\lambda_i} = 0$$

The successive diagonals above the main diagonal give terms which sum to these derivatives of the characteristic equation. Hence

$$c_n\mathbf{J}_i^n + \cdots + c_1\mathbf{J}_i + c_0\mathbf{I} = [\mathbf{0}]$$

and so

$$c_n\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + \cdots + c_1\mathbf{A} + c_0\mathbf{I} = [\mathbf{0}]$$

8.6 Give a general proof of the Cayley-Hamilton theorem without using the Jordan form.
A is an $n \times n$ matrix with n eigenvalues λ_i, some of which may be equal. The characteristic polynomial is

$$\Delta(\lambda) = |\mathbf{A} - \mathbf{I}\lambda| = (-\lambda)^n + c_{n-1}\lambda^{n-1} + c_{n-2}\lambda^{n-2} + \cdots + c_1\lambda + c_0 \tag{1}$$

We are to show that

$$\Delta(\mathbf{A}) = (-1)^n\mathbf{A}^n + c_{n-1}\mathbf{A}^{n-1} + c_{n-2}\mathbf{A}^{n-2} + \cdots + c_1\mathbf{A} + c_0\mathbf{I} = [\mathbf{0}] \tag{2}$$

Consider Adj[$\mathbf{A} - \mathbf{I}\lambda$]. Its elements are formed from $n-1 \times n-1$ determinants obtained by deleting a row and a column of $\mathbf{A} - \mathbf{I}\lambda$. Therefore, the highest power of λ that can be in any element of Adj[$\mathbf{A} - \mathbf{I}\lambda$] is λ^{n-1}. This means it is possible to write

$$\text{Adj}[\mathbf{A} - \mathbf{I}\lambda] = \mathbf{B}_{n-1}\lambda^{n-1} + \mathbf{B}_{n-2}\lambda^{n-2} + \cdots + \mathbf{B}_1\lambda + \mathbf{B}_0 \tag{3}$$

where the $\mathbf{B}_i$ terms are $n \times n$ matrices not containing λ, but are otherwise unknown. We use the known result

$$[\mathbf{A} - \mathbf{I}\lambda]\ \text{Adj}[\mathbf{A} - \mathbf{I}\lambda] = |\mathbf{A} - \mathbf{I}\lambda|\mathbf{I} \tag{4}$$

Substituting Eq. (*3*) into the left side of Eq. (*4*) gives

$$[\mathbf{A} - \mathbf{I}\lambda]\ \text{Adj}[\mathbf{A} - \mathbf{I}\lambda] = -\mathbf{B}_{n-1}\lambda^n + (\mathbf{A}\mathbf{B}_{n-1} - \mathbf{B}_{n-2})\lambda^{n-1} + (\mathbf{A}\mathbf{B}_{n-2} - \mathbf{B}_{n-3})\lambda^{n-2}$$
$$+ \cdots + (\mathbf{A}\mathbf{B}_2 - \mathbf{B}_1)\lambda^2 + (\mathbf{A}\mathbf{B}_1 - \mathbf{B}_0)\lambda + \mathbf{A}\mathbf{B}_0$$

Using Eq. (*1*) on the right side of Eq. (*4*) gives

$$\Delta(\lambda)\mathbf{I} = (-\lambda)^n\mathbf{I} + c_{n-1}\lambda^{n-1}\mathbf{I} + c_{n-2}\lambda^{n-2}\mathbf{I} + \cdots + c_1\lambda\mathbf{I} + c_0\mathbf{I}$$

The left side equals the right side, and the coefficients of like powers of λ on the two sides must be equal. This leads to the following set of equations:

$$-\mathbf{B}_{n-1} = (-1)^n\mathbf{I}$$

$$\mathbf{A}\mathbf{B}_{n-1} - \mathbf{B}_{n-2} = c_{n-1}\mathbf{I}$$

$$\mathbf{AB}_{n-2} - \mathbf{B}_{n-3} = c_{n-2}\mathbf{I}$$
$$\vdots$$
$$\mathbf{AB}_2 - \mathbf{B}_1 = c_2\mathbf{I}$$
$$\mathbf{AB}_1 - \mathbf{B}_0 = c_1\mathbf{I}$$
$$\mathbf{AB}_0 = c_0\mathbf{I}$$

If the first of these equations is premultiplied by $\mathbf{A}^n$, the second by $\mathbf{A}^{n-1}$, etc., the sum of the right-side terms is $\Delta(\mathbf{A})$. The left-side sum must also equal $\Delta(\mathbf{A})$, and takes the form

$$(-\mathbf{A}^n\mathbf{B}_{n-1} + \mathbf{A}^n\mathbf{B}_{n-1}) + (-\mathbf{A}^{n-1}\mathbf{B}_{n-2} + \mathbf{A}^{n-1}\mathbf{B}_{n-2}) + (-\mathbf{A}^{n-2}\mathbf{B}_{n-3} + \mathbf{A}^{n-2}\mathbf{B}_{n-3})$$
$$+ \cdots + (-\mathbf{A}^2\mathbf{B}_1 + \mathbf{A}^2\mathbf{B}_1) + (-\mathbf{AB}_0 + \mathbf{AB}_0) \equiv [\mathbf{0}]$$

Therefore, $\Delta(\mathbf{A}) = [\mathbf{0}]$.

Functions of a Jordan Block

8.7 If $\mathbf{J}_1$ is an $m \times m$ Jordan block with eigenvalue λ_1, find a general expression for the coefficients α_i for $e^{\mathbf{J}_1 t}$.

The characteristic equation is $\Delta(\lambda) = (\lambda - \lambda_1)^m$ and

$$e^{\mathbf{J}_1 t} = \alpha_0\mathbf{I} + \alpha_1\mathbf{J}_1 + \alpha_2\mathbf{J}_1^2 + \cdots + \alpha_{m-1}\mathbf{J}_1^{m-1}$$

where

$$e^{\lambda_1 t} = \alpha_0 + \alpha_1\lambda_1 + \alpha_2\lambda_1^2 + \cdots + \alpha_{m-1}\lambda_1^{m-1}$$

$$\left.\frac{de^{\lambda t}}{d\lambda}\right|_{\lambda_1} = te^{\lambda_1 t} = \alpha_1 + 2\alpha_2\lambda_1 + 3\alpha_3\lambda_1^2 + \cdots + (m-1)\alpha_{m-1}\lambda_1^{m-2}$$

More symmetry is achieved if the kth derivative term is divided by $1/k!$:

$$\tfrac{1}{2}t^2 e^{\lambda_1 t} = \alpha_2 + 3\alpha_3\lambda_1 + \cdots + \tfrac{1}{2}(m-1)(m-2)\alpha_{m-1}\lambda_1^{m-3}$$

$$\frac{1}{3!}t^3 e^{\lambda_1 t} = \alpha_3 + 4\lambda_1 + \cdots + \frac{1}{3!}(m-1)(m-2)(m-3)\alpha_{m-1}\lambda_1^{m-4}$$

$$\vdots$$

$$\frac{1}{(m-1)!}t^{m-1}e^{\lambda_1 t} = \alpha_{m-1}$$

or

$$\begin{bmatrix} 1 \\ t \\ \frac{1}{2}t^2 \\ \frac{1}{3!}t^3 \\ \vdots \\ \frac{1}{(m-1)!}t^{m-1} \end{bmatrix} e^{\lambda_1 t} = \begin{bmatrix} 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 & \cdots & \lambda_1^{m-1} \\ 0 & 1 & 2\lambda_1 & 3\lambda_1^2 & & (m-1)\lambda_1^{m-2} \\ 0 & 0 & 1 & 3\lambda_1 & & \frac{1}{2}(m-1)(m-2)\lambda_1^{m-3} \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & 0 & & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m-1} \end{bmatrix} \triangleq \mathbf{F}\boldsymbol{\alpha}$$

$$\boldsymbol{\alpha} = [\mathbf{F}]^{-1}\begin{bmatrix} 1 \\ t \\ \frac{1}{2}t^2 \\ \frac{1}{3!}t^3 \\ \vdots \\ \frac{1}{(m-1)!}t^{m-1} \end{bmatrix} e^{\lambda_1 t}$$

But using Gaussian elimination, for example, gives

$$\mathbf{F}^{-1} = \begin{bmatrix} 1 & -\lambda_1 & \lambda_1^2 & -\lambda_1^3 & \lambda_1^4 & \cdots & (-\lambda_1)^{m-1} \\ 0 & 1 & -2\lambda_1 & 3\lambda_1^2 & -4\lambda_1^3 & & (m-1)(-\lambda_1)^{m-2} \\ 0 & 0 & 1 & -3\lambda_1 & 6\lambda_1^2 & & \frac{1}{2}(m-1)(m-2)(-\lambda_1)^{m-3} \\ 0 & 0 & 0 & 1 & -4\lambda_1 & & \frac{1}{3!}(m-1)(m-2)(m-3)(-\lambda_1)^{m-4} \\ \vdots & & & & & \ddots & \vdots \\ & & & & & & -(m-1)\lambda_1 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

so

$$\alpha_0 = e^{\lambda_1 t}\left[1 - \lambda_1 t + \frac{\lambda_1^2 t^2}{2!} - \frac{\lambda_1^3 t^3}{3!} + \frac{\lambda_1^4 t^4}{4!} - \cdots + \frac{(-\lambda_1)^{m-1} t^{m-1}}{(m-1)!}\right]$$

$$\alpha_1 = e^{\lambda_1 t}\left[t - \lambda_1 t^2 + \frac{\lambda_1^2 t^3}{2} - \frac{\lambda_1^3 t^4}{3!} + \cdots + \frac{t^{m-1}(-\lambda_1)^{m-2}}{(m-2)!}\right]$$

$$\alpha_2 = e^{\lambda_1 t}\left[\tfrac{1}{2}t^2 - \frac{\lambda_1 t^3}{2} + \frac{\lambda_1^2 t^4}{4} + \cdots + \frac{t^{m-1}(-\lambda)^{m-3}}{2(m-3)!}\right]$$

$$\vdots$$

$$\alpha_{m-2} = e^{\lambda_1 t}\left[\frac{1}{(m-2)!}t^{m-2} - \frac{1}{(m-2)!}\lambda_1 t^{m-1}\right]$$

$$\alpha_{m-1} = e^{\lambda_1 t}\frac{t^{m-1}}{(m-1)!}$$

8.8 Use the results of the previous problem to find $e^{\mathbf{J}_1 t}$ if a. $\mathbf{J}_1$ is a 3×3 Jordan block, and b. $\mathbf{J}_1$ is a 4×4 Jordan block.

(a) $\mathbf{J}_1 = \begin{bmatrix} \lambda_1 & 1 & 0 \\ 0 & \lambda_1 & 1 \\ 0 & 0 & \lambda_1 \end{bmatrix}, \quad \mathbf{J}_1^2 = \begin{bmatrix} \lambda_1^2 & 2\lambda_1 & 1 \\ 0 & \lambda_1^2 & 2\lambda_1 \\ 0 & 0 & \lambda_1^2 \end{bmatrix}$

$$e^{\mathbf{J}_1 t} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{J}_1 + \alpha_2 \mathbf{J}_1^2$$

$$= \begin{bmatrix} \alpha_0 + \alpha_1\lambda_1 + \alpha_2\lambda_1^2 & \alpha_1 + 2\lambda_1\alpha_2 & \alpha_2 \\ 0 & \alpha_0 + \alpha_1\lambda_1 + \alpha_2\lambda_1^2 & \alpha_1 + 2\lambda_1\alpha_2 \\ 0 & 0 & \alpha_0 + \alpha_1\lambda_1 + \alpha_2\lambda_1^2 \end{bmatrix}$$

$$= e^{\lambda_1 t}\begin{bmatrix} 1 & t & \frac{1}{2}t^2 \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}$$

(b) $$\mathbf{J}_1 = \begin{bmatrix} \lambda_1 & 1 & 0 & 0 \\ 0 & \lambda_1 & 1 & 0 \\ 0 & 0 & \lambda_1 & 1 \\ 0 & 0 & 0 & \lambda_1 \end{bmatrix} \qquad \mathbf{J}_1^2 = \begin{bmatrix} \lambda_1^2 & 2\lambda_1 & 1 & 0 \\ 0 & \lambda_1^2 & 2\lambda_1 & 1 \\ 0 & 0 & \lambda_1^2 & 2\lambda_1 \\ 0 & 0 & 0 & \lambda_1^2 \end{bmatrix},$$

$$\mathbf{J}_1^3 = \begin{bmatrix} \lambda_1^3 & 3\lambda_1^2 & 3\lambda_1 & 1 \\ 0 & \lambda_1^3 & 3\lambda_1^2 & 3\lambda_1 \\ 0 & 0 & \lambda_1^3 & 3\lambda_1^2 \\ 0 & 0 & 0 & \lambda_1^3 \end{bmatrix}$$

so

$$e^{\mathbf{J}_1 t} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{J}_1 + \alpha_2 \mathbf{J}_1^2 + \alpha_3 \mathbf{J}_1^3$$
$$= e^{\lambda_1 t} \begin{bmatrix} 1 & t & \frac{1}{2}t^2 & \frac{1}{3!}t^3 \\ 0 & 1 & t & \frac{1}{2}t^2 \\ 0 & 0 & 1 & t \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The pattern illustrated by these two cases continues for any $m \times m$ Jordan block.

Some Matrix Identities

8.9 **(a)** If $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ -2 & -3 \end{bmatrix}$, compute the 2×2 matrices $\sin \mathbf{A}$ and $\cos \mathbf{A}$.

(b) Verify that $\sin^2 \mathbf{A} + \cos^2 \mathbf{A} = \mathbf{I}$.

(a) From Problem 7.2, $\Delta(\lambda) = \lambda^2 + 2\lambda + 1$ and $\lambda_1 = \lambda_2 = -1$. Now $\sin \mathbf{A} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A}$ and $\cos \mathbf{A} = \alpha_2 \mathbf{I} + \alpha_3 \mathbf{A}$, where

$$\left.\begin{aligned} \sin \lambda_1 &= \alpha_0 + \alpha_1 \lambda_1 \\ \frac{d}{d\lambda}(\sin \lambda)\Big|_{\lambda=\lambda_1} &= \cos \lambda_1 = \alpha_1 \end{aligned}\right\} \Rightarrow \begin{cases} \alpha_1 = \cos(-1) = \cos(1) \\ \alpha_0 = \sin \lambda_1 + \alpha_1 = \cos(1) - \sin(1) \end{cases}$$

and

$$\left.\begin{aligned} \cos \lambda_1 &= \alpha_2 + \lambda_1 \alpha_3 \\ \frac{d}{d\lambda}(\cos \lambda)\Big|_{\lambda=\lambda_1} &= -\sin \lambda_1 = \alpha_3 \end{aligned}\right\} \Rightarrow \begin{cases} \alpha_3 = \sin(1) \\ \alpha_2 = \cos(1) + \alpha_3 = \cos(1) + \sin(1) \end{cases}$$

Therefore,

$$\sin \mathbf{A} = \begin{bmatrix} 2\cos(1) - \sin(1) & 2\cos(1) \\ -2\cos(1) & -2\cos(1) - \sin(1) \end{bmatrix}$$

and

$$\cos \mathbf{A} = \begin{bmatrix} \cos(1) + 2\sin(1) & 2\sin(1) \\ -2\sin(1) & \cos(1) - 2\sin(1) \end{bmatrix}$$

(b) Multiplication gives

$$\sin^2 \mathbf{A} = \begin{bmatrix} \sin^2(1) - 4\cos(1)\sin(1) & -4\cos(1)\sin(1) \\ 4\cos(1)\sin(1) & \sin^2(1) + 4\cos(1)\sin(1) \end{bmatrix}$$

and

$$\cos^2 \mathbf{A} = \begin{bmatrix} \cos^2(1) + 4\cos(1)\sin(1) & 4\cos(1)\sin(1) \\ -4\cos(1)\sin(1) & \cos^2(1) - 4\cos(1)\sin(1) \end{bmatrix}$$

$$\text{so } \sin^2 \mathbf{A} + \cos^2 \mathbf{A} = \begin{bmatrix} \sin^2(1) + \cos^2(1) & 0 \\ 0 & \sin^2(1) + \cos^2(1) \end{bmatrix} = \mathbf{I}.$$

8.10 **(a)** Compute $e^{\mathbf{A}t}$, $e^{-\mathbf{A}t}$, $\sinh \mathbf{A}t$, and $\cosh \mathbf{A}t$ for $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$.

(b) Verify that $\sinh \mathbf{A}t = (e^{\mathbf{A}t} - e^{-\mathbf{A}t})/2$, $\cosh \mathbf{A}t = (e^{\mathbf{A}t} + e^{-\mathbf{A}t})/2$, and $\cosh^2 \mathbf{A}t - \sinh^2 \mathbf{A}t = \mathbf{I}$.

(a) The eigenvalues are $\lambda_1 = 0$ and $\lambda_2 = 2$, so $f(\mathbf{A}) = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} = \begin{bmatrix} \alpha_0 + \alpha_1 & \alpha_1 \\ \alpha_1 & \alpha_0 + \alpha_1 \end{bmatrix}$. The coefficients α_0 and α_1 are found from $f(\lambda_1) = \alpha_0$ and $f(\lambda_2) = \alpha_0 + 2\alpha_1$. Therefore, $\alpha_1 = [f(\lambda_2) - f(\lambda_1)]/2$ and $\alpha_0 + \alpha_1 = [f(\lambda_2) + f(\lambda_1)]/2$ for any analytic $f(\lambda)$. Letting $f(\lambda)$ be each of the four required functions gives, in turn,

$$e^{\mathbf{A}t} = \begin{bmatrix} \dfrac{e^{2t}+1}{2} & \dfrac{e^{2t}-1}{2} \\ \dfrac{e^{2t}-1}{2} & \dfrac{e^{2t}+1}{2} \end{bmatrix}, \qquad e^{-\mathbf{A}t} = \begin{bmatrix} \dfrac{e^{-2t}+1}{2} & \dfrac{e^{-2t}-1}{2} \\ \dfrac{e^{-2t}-1}{2} & \dfrac{e^{-2t+1}}{2} \end{bmatrix}$$

$$\sinh \mathbf{A}t = \begin{bmatrix} \dfrac{\sinh 2t}{2} & \dfrac{\sinh 2t}{2} \\ \dfrac{\sinh 2t}{2} & \dfrac{\sinh 2t}{2} \end{bmatrix}, \qquad \cosh \mathbf{A}t = \begin{bmatrix} \dfrac{\cosh 2t+1}{2} & \dfrac{\cosh 2t-1}{2} \\ \dfrac{\cosh 2t-1}{2} & \dfrac{\cosh 2t+1}{2} \end{bmatrix}$$

(b) Since $(e^{2t} - e^{-2t})/2 = \sinh 2t$ and $(e^{2t} + e^{-2t})/2 = \cosh 2t$, it is clear that $(e^{\mathbf{A}t} + e^{-\mathbf{A}t})/2 = \cosh \mathbf{A}t$ and $(e^{\mathbf{A}t} - e^{-\mathbf{A}t})/2 = \sinh \mathbf{A}t$. Computing

$$\cosh^2 \mathbf{A}t = \begin{bmatrix} \dfrac{\cosh^2 2t+1}{2} & \dfrac{\cosh^2 2t-1}{2} \\ \dfrac{\cosh^2 2t-1}{2} & \dfrac{\cosh^2 2t+1}{2} \end{bmatrix}$$

$$\sinh^2 \mathbf{A}t = \begin{bmatrix} \dfrac{\sinh^2 2t}{2} & \dfrac{\sinh^2 2t}{2} \\ \dfrac{\sinh^2 2t}{2} & \dfrac{\sinh^2 2t}{2} \end{bmatrix}$$

and using $\cosh^2 2t - \sinh^2 2t = 1$ gives $\cosh^2 \mathbf{A}t - \sinh^2 \mathbf{A}t = \mathbf{I}$.

8.11 Let $\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$.

(a) Find $\sin \mathbf{A}$, $\cos \mathbf{A}$, and $\tan \mathbf{A}$.
(b) Show that $\tan \mathbf{A} = (\sin \mathbf{A})(\cos \mathbf{A})^{-1}$.
(a) The eigenvalues are $\lambda_1 = \sqrt{2}$ and $\lambda_2 = -\sqrt{2}$. Since both eigenvalues satisfy $|\lambda_i| < \pi/2$, use of the results for $\tan \mathbf{A}$ of Problem 8.3 is justified. For all three functions, $f(\mathbf{A}) = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A}$, where $\alpha_0 = (1/2)[f(\sqrt{2}) + f(-\sqrt{2})]$ and $\alpha_1 = (1/\sqrt{2})[\alpha_0 - f(-\sqrt{2})]$, so

$$\sin \mathbf{A} = \frac{\sin \sqrt{2}}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \cos \mathbf{A} = (\cos \sqrt{2}) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

$$\tan \mathbf{A} = \frac{\tan \sqrt{2}}{\sqrt{2}} \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$$

(b) Using $(\cos \mathbf{A})^{-1} = (1/\cos \sqrt{2})\mathbf{I}$ shows that $\tan \mathbf{A} = (\sin \mathbf{A})(\cos \mathbf{A})^{-1}$.

Closed Form for Functions of a Matrix

8.12 If $\mathbf{A} = \begin{bmatrix} -2 & 2 \\ 1 & -3 \end{bmatrix}$, what is $\sin \mathbf{A}t$?

Since $\Delta(\lambda) = \lambda^2 + 5\lambda + 4 = (\lambda + 4)(\lambda + 1)$, the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = -4$. The general form of the solution is

$$\sin \mathbf{A}t = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A}$$

where

$$\left.\begin{aligned} \sin(-4t) &= \alpha_0 - 4\alpha_1 \\ \sin(-t) &= \alpha_0 - \alpha_1 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} \alpha_1 &= -\tfrac{1}{3}[\sin(-4t) - \sin(-t)] \\ \alpha_0 &= -\tfrac{1}{3}[\sin(-4t) - 4\sin(-t)] \end{aligned}\right.$$

Then $\sin \mathbf{A}t = \dfrac{1}{3}\begin{bmatrix} \sin(-4t) + 2\sin(-t) & -2\sin(-4t) + 2\sin(-t) \\ -\sin(-4t) + \sin(-t) & 2\sin(-4t) + \sin(-t) \end{bmatrix}$.

8.13 If $\mathbf{A} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -2 & 2 \\ 0 & 1 & -3 \end{bmatrix}$, find $e^{\mathbf{A}t}$.

Note that $\mathbf{A}$ is block diagonal and the lower block is the same matrix as in the previous problem. Since the algebra involved in finding the α's is the same when finding any analytic function, the answer can be written down by replacing $\sin \lambda_i t$ by $e^{\lambda_i t}$, so

$$e^{\mathbf{A}t} = \left[\begin{array}{c|cc} e^{2t} & 0 & 0 \\ \hline 0 & \dfrac{(e^{-4t} + 2e^{-t})}{3} & \dfrac{2(-e^{-4t} + e^{-t})}{3} \\ 0 & \dfrac{(-e^{-4t} + e^{-t})}{3} & \dfrac{(2e^{-4t} + e^{-t})}{3} \end{array}\right]$$

A useful partial check is that $e^{\mathbf{A}t}$ must always equal $\mathbf{I}$ when $t = 0$.

8.14 Compute $\begin{bmatrix} 1 & -1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}^k$ for any arbitrary integer k.

Computing $\Delta(\lambda) = (1 - \lambda)^3$ gives $\lambda_1 = \lambda_2 = \lambda_3 = 1$. Since $\mathbf{A}$ is 3×3, $\mathbf{A}^k$ can always be written as a polynomial of degree 2 (or less):

$$\mathbf{A}^k = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} + \alpha_2 \mathbf{A}^2 = \begin{bmatrix} \alpha_0 + \alpha_1 + \alpha_2 & -\alpha_1 - 2\alpha_2 & \alpha_1 + \alpha_2 \\ 0 & \alpha_0 + \alpha_1 + \alpha_2 & \alpha_1 + 2\alpha_2 \\ 0 & 0 & \alpha_0 + \alpha_1 + \alpha_2 \end{bmatrix}$$

where

$$(\lambda_1)^k = 1 = \alpha_0 + \alpha_1 + \alpha_2$$

$$\left.\frac{d(\lambda)^k}{d\lambda}\right|_{\lambda = \lambda_1} = k\lambda_1^{k-1} = k = \alpha_1 + 2\alpha_2$$

$$\left.\frac{d^2(\lambda)^k}{d\lambda^2}\right|_{\lambda = \lambda_1} = k(k-1)\lambda_1^{k-2} = k(k-1) = 2\alpha_2$$

Note that in this case all the individual α's need not be found. Combining the last two equations gives $\alpha_1 + \alpha_2 = k - k(k-1)/2 = k(3-k)/2$. All the combinations of α_i that are required in $\mathbf{A}^k$ are now available, and

$$\mathbf{A}^k = \begin{bmatrix} 1 & -k & k(3-k)/2 \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix}$$

8.15 Compute $e^{\mathbf{A}t}$ if $\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$.

We know that $e^{\mathbf{A}t} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} = \begin{bmatrix} \alpha_0 - \alpha_1 & \alpha_1 \\ \alpha_1 & \alpha_0 + \alpha_1 \end{bmatrix}$. Since this is the same **A** matrix as in Problem 8.11, the coefficients are

$$\alpha_0 = \tfrac{1}{2}[e^{\sqrt{2}t} + e^{-\sqrt{2}t}] = \cosh \sqrt{2}t, \quad \alpha_1 = \frac{e^{\sqrt{2}t} - e^{-\sqrt{2}t}}{2\sqrt{2}} = \frac{1}{\sqrt{2}} \sinh \sqrt{2}t$$

$$\text{so } e^{\mathbf{A}t} = \begin{bmatrix} \cosh \sqrt{2}t - \frac{1}{\sqrt{2}} \sinh \sqrt{2}t & \frac{1}{\sqrt{2}} \sinh \sqrt{2}t \\ \frac{1}{\sqrt{2}} \sinh \sqrt{2}t & \cosh \sqrt{2}t + \frac{1}{\sqrt{2}} \sinh \sqrt{2}t \end{bmatrix}.$$

8.16 Given $\mathbf{A} = \begin{bmatrix} 0 & -3 & 0 \\ 3 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. a. Find $\mathbf{A}^{-1}$ using the Cayley-Hamilton theorem. b. Compute $e^{\mathbf{A}t}$.

(a) The characteristic polynomial is $\Delta(\lambda) = (1+\lambda)(\lambda^2+9) = \lambda^3 + \lambda^2 + 9\lambda + 9$. Therefore $\Delta(\mathbf{A}) = [\mathbf{0}] = \mathbf{A}^3 + \mathbf{A}^2 + 9\mathbf{A} + 9\mathbf{I}$ so

$$\mathbf{A}^{-1} = -\frac{1}{9}[\mathbf{A}^2 + \mathbf{A} + 9\mathbf{I}] = \begin{bmatrix} 0 & \frac{1}{3} & 0 \\ -\frac{1}{3} & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

(b) The eigenvalues are roots of $\Delta(\lambda) = 0$, so $\lambda_1 = -1, \lambda_2 = 3j, \lambda_3 = -3j$. Since **A** is block diagonal,

$$e^{\mathbf{A}t} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{e^{\begin{bmatrix} 0 & -3 \\ 3 & 0 \end{bmatrix}t}} & \begin{matrix} 0 \\ 0 \end{matrix} \\ \hline 0 & 0 & e^{-t} \end{array}\right]$$

and

$$e^{\begin{bmatrix} 0 & -3 \\ 3 & 0 \end{bmatrix}t} = \alpha_0 \mathbf{I} + \alpha_1 \begin{bmatrix} 0 & -3 \\ 3 & 0 \end{bmatrix} = \begin{bmatrix} \alpha_0 & -3\alpha_1 \\ 3\alpha_1 & \alpha_0 \end{bmatrix}$$

where $e^{3jt} = \alpha_0 + \alpha_1 3j$ and $e^{-3jt} = \alpha_0 - \alpha_1 3j$. From this,

$$\alpha_0 = \tfrac{1}{2}[e^{3jt} + e^{-3jt}] = \cos 3t$$

$$\alpha_1 = \frac{1}{3}\left[\frac{e^{3jt} - e^{-3jt}}{2j}\right] = \tfrac{1}{3} \sin 3t$$

$$\text{and so } e^{\mathbf{A}t} = \begin{bmatrix} \cos 3t & -\sin 3t & 0 \\ \sin 3t & \cos 3t & 0 \\ 0 & 0 & e^{-t} \end{bmatrix}.$$

8.17 Find $e^{\mathbf{A}t}$ if $\mathbf{A} = \left[\begin{array}{cc|cc} 0 & -\Omega & a & 0 \\ \Omega & 0 & 0 & a \\ \hline 0 & 0 & 0 & -\Omega \\ 0 & 0 & \Omega & 0 \end{array}\right]$

The algebra can be simplified in this case if it is noted that

$$\mathbf{A} = \left[\begin{array}{c|c} \mathbf{B}_1 & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{B}_1 \end{array}\right] + \left[\begin{array}{c|c} \mathbf{0} & \mathbf{B}_2 \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right] \triangleq \mathbf{A}_1 + \mathbf{A}_2 \quad \text{and} \quad \mathbf{A}_1\mathbf{A}_2 = \mathbf{A}_2\mathbf{A}_1 = \left[\begin{array}{c|c} \mathbf{0} & a\mathbf{B}_1 \\ \hline \mathbf{0} & \mathbf{0} \end{array}\right]$$

Since $\mathbf{A}_1$ and $\mathbf{A}_2$ commute, $e^{(\mathbf{A}_1+\mathbf{A}_2)t} = e^{\mathbf{A}_1 t} e^{\mathbf{A}_2 t}$. But $\mathbf{A}_1$ is block diagonal, so

$$e^{\mathbf{A}_1 t} = \left[\begin{array}{c|c} e^{\mathbf{B}_1 t} & \mathbf{0} \\ \hline \mathbf{0} & e^{\mathbf{B}_1 t} \end{array}\right]$$

$e^{\mathbf{B}_1 t} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{B}_1$ and eigenvalues of $\mathbf{B}_1$ are $\pm j\Omega$:

$$\left.\begin{array}{l} e^{j\Omega t} = \alpha_0 + j\Omega\alpha_1 \\ e^{-j\Omega t} = \alpha_0 - j\Omega\alpha_1 \end{array}\right\} \Rightarrow e^{\mathbf{B}_1 t} = \begin{bmatrix} \cos\Omega t & -\sin\Omega t \\ \sin\Omega t & \cos\Omega t \end{bmatrix}$$

Consider $\mathbf{A}_2$: $|\mathbf{A}_2 - \mathbf{I}\lambda| = \lambda^4 = 0$. $\mathbf{A}_2$ has $\lambda_i = 0$ with algebraic multiplicity of four:

$$e^{\mathbf{A}_2 t} = \beta_0 \mathbf{I} + \beta_1 \mathbf{A}_2 + \beta_2 \mathbf{A}_2^2 + \beta_3 \mathbf{A}_2^3$$

Since $\mathbf{A}_2^2 = \mathbf{A}_2^3 = [\mathbf{0}]$, coefficients β_2 and β_3 are not needed. The remaining coefficients are found to be $e^{0t} = 1 = \beta_0$ and $\left.\dfrac{de^{\lambda t}}{d\lambda}\right|_{\lambda=0} = t = \beta_1$. Thus

$$e^{\mathbf{A}_2 t} = \left[\begin{array}{cc|cc} 1 & 0 & at & 0 \\ 0 & 1 & 0 & at \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array}\right]$$

Combining gives $e^{\mathbf{A}t} = e^{\mathbf{A}_1 t} e^{\mathbf{A}_2 t} = \left[\begin{array}{cc|cc} \cos\Omega t & -\sin\Omega t & at\cos\Omega t & -at\sin\Omega t \\ \sin\Omega t & \cos\Omega t & at\sin\Omega t & at\cos\Omega t \\ \hline 0 & 0 & \cos\Omega t & -\sin\Omega t \\ 0 & 0 & \sin\Omega t & \cos\Omega t \end{array}\right]$.

Minimum Polynomial

8.18 In computing $e^{\mathbf{A}_2 t}$ in the previous problem, it was found that only first-order terms in $\mathbf{A}_2$ were needed even though $\mathbf{A}_2$ was a 4×4 matrix. This is a case where the minimum polynomial is of lower order than the characteristic polynomial. Find the minimal polynomial of $\mathbf{A}_2$.

The minimal polynomial will have the same factors as the characteristic polynomial, but perhaps raised to smaller powers. (There could also be a sign difference, depending on how the characteristic polynomial is defined.)

Here $\Delta(\lambda) = \lambda^4$, so $m(\lambda) = \lambda^k$, where k is the smallest integer for which $m(\mathbf{A}) = [\mathbf{0}]$. From the results of the previous problem, $k = 2$ and therefore $m(\lambda) = \lambda^2$. If this had been known in advance, then it would have been known that $e^{\mathbf{A}_2 t} = \beta_0 \mathbf{I} + \beta_1 \mathbf{A}_2$ and a slight amount of matrix algebra would be avoided. This savings is usually offset by the effort required to find $m(\lambda)$, and in this text the minimum polynomial is seldom used.

Alternative Methods

8.19 It was shown in the text that $de^{\mathbf{A}t}/dt = \mathbf{A}e^{\mathbf{A}t}$ if $\mathbf{A}$ is a constant matrix. Use this result, plus the fact that $e^{\mathbf{A}\cdot 0} = \mathbf{I}$, to derive an alternative method of computing $e^{\mathbf{A}t}$.

Let $e^{\mathbf{A}t} = \mathbf{F}(t)$. Then $\mathbf{F}$ satisfies the matrix differential equation $\dot{\mathbf{F}} = \mathbf{A}\mathbf{F}$ with initial conditions $\mathbf{F}(0) = \mathbf{I}$. Laplace transforms can be used to solve this equation:

$$\mathcal{L}\{\dot{\mathbf{F}}(t)\} = s\mathbf{F}(s) - \mathbf{F}(0) = s\mathbf{F}(s) - \mathbf{I}$$

$$\mathcal{L}\{\mathbf{A}\mathbf{F}(t)\} = \mathbf{A}\mathcal{L}\{\mathbf{F}(t)\} = \mathbf{A}\mathbf{F}(s)$$

Therefore, $[s\mathbf{I}-\mathbf{A}]\mathbf{F}(s)=\mathbf{I}$ or $\mathbf{F}(s)=[s\mathbf{I}-\mathbf{A}]^{-1}$, and

$$e^{\mathbf{A}t}=\mathbf{F}(t)=\mathscr{L}^{-1}\{[s\mathbf{I}-\mathbf{A}]^{-1}\}$$

8.20 Use the method of the previous problem to compute $e^{\mathbf{A}t}$ for $\mathbf{A}=\begin{bmatrix}-2 & -2 & 0\\ 0 & 0 & 1\\ 0 & -3 & -4\end{bmatrix}$.

Form $s\mathbf{I}-\mathbf{A}=\begin{bmatrix}s+2 & 2 & 0\\ 0 & s & -1\\ 0 & 3 & s+4\end{bmatrix}$ and $|s\mathbf{I}-\mathbf{A}|=(s+2)(s^2+4s+3)$ $=(s+1)(s+2)(s+3)$. Then

$$\mathbf{F}(s)=[s\mathbf{I}-\mathbf{A}]^{-1}=\frac{1}{(s+1)(s+2)(s+3)}\begin{bmatrix}(s+1)(s+3) & -2(s+4) & -2\\ 0 & (s+2)(s+4) & s+2\\ 0 & -3(s+2) & s(s+2)\end{bmatrix}$$

The inverse Laplace transforms are computed term by term:

$$F_{11}(t)=\mathscr{L}^{-1}\left\{\frac{1}{s+2}\right\}=e^{-2t}; \qquad F_{21}=F_{31}=0$$

$$F_{12}(t)=\mathscr{L}^{-1}\left\{\frac{-2(s+4)}{(s+1)(s+2)(s+3)}\right\}=-3e^{-t}+4e^{-2t}-e^{-3t}$$

$$F_{13}(t)=\mathscr{L}^{-1}\left\{\frac{-2}{(s+1)(s+2)(s+3)}\right\}=-e^{-t}+2e^{-2t}-e^{-3t}$$

$$F_{22}(t)=\mathscr{L}^{-1}\left\{\frac{s+4}{(s+1)(s+3)}\right\}=\tfrac{3}{2}e^{-t}-\tfrac{1}{2}e^{-3t}$$

$$F_{32}(t)=\mathscr{L}^{-1}\left\{\frac{-3}{(s+1)(s+3)}\right\}=-\tfrac{3}{2}e^{-t}+\tfrac{3}{2}e^{-3t}$$

$$F_{23}(t)=\mathscr{L}^{-1}\left\{\frac{1}{(s+1)(s+3)}\right\}=\tfrac{1}{2}e^{-t}-\tfrac{1}{2}e^{-3t}$$

$$F_{33}(t)=\mathscr{L}^{-1}\left\{\frac{s}{(s+1)(s+3)}\right\}=-\tfrac{1}{2}e^{-t}+\tfrac{3}{2}e^{-3t}$$

8.21 Find the closed-form expression for $e^{\mathbf{A}t}$, where $\mathbf{A}=\begin{bmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ -27 & 54 & -36 & 10\end{bmatrix}$.

Writing $|\mathbf{A}-\mathbf{I}\lambda|=\Delta(\lambda)=\lambda^4-10\lambda^3+36\lambda^2-54\lambda+27=(\lambda-1)(\lambda-3)^3$ shows that $\lambda_1=1$, $\lambda_2=\lambda_3=\lambda_4=3$. For the repeated eigenvalue, $\text{rank}[\mathbf{A}-\mathbf{I}\lambda_2]=3$ and $q_2=1$. Thus the Jordan form is $\mathbf{J}=\left[\begin{array}{c|ccc}1 & 0 & 0 & 0\\ \hline 0 & 3 & 1 & 0\\ 0 & 0 & 3 & 1\\ 0 & 0 & 0 & 3\end{array}\right]$. Using the results of Problem 8.8,

$$e^{\mathbf{J}t} = \left[\begin{array}{c|ccc} e^t & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & e^{\mathbf{J}_2 t} & \\ 0 & & & \end{array}\right] = \begin{bmatrix} e^t & 0 & 0 & 0 \\ 0 & e^{3t} & te^{3t} & \frac{1}{2}t^2e^{3t} \\ 0 & 0 & e^{3t} & te^{3t} \\ 0 & 0 & 0 & e^{3t} \end{bmatrix}$$

and $e^{\mathbf{A}t} = \mathbf{M}e^{\mathbf{J}t}\mathbf{M}^{-1}$, where $\mathbf{M}$ is the modal matrix consisting of two eigenvectors and two generalized eigenvectors. The first column of $\text{Adj}[\mathbf{A} - \mathbf{I}\lambda]$ is

$$\begin{bmatrix} \lambda^2(10-\lambda) - 36\lambda + 54 \\ 27 \\ 27\lambda \\ 27\lambda^2 \end{bmatrix}, \quad \text{so with } \lambda = 1, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

and with $\lambda = 3$, $\mathbf{x}_2 = [1 \quad 3 \quad 9 \quad 27]^T$ is an eigenvector. Generalized eigenvectors are solutions of $\mathbf{A}\mathbf{x}_3 = 3\mathbf{x}_3 + \mathbf{x}_2$ and $\mathbf{A}\mathbf{x}_4 = 3\mathbf{x}_4 + \mathbf{x}_3$. Solutions are $\mathbf{x}_3 = [1 \quad 4 \quad 15 \quad 54]^T$ and $\mathbf{x}_4 = [0 \quad 1 \quad 7 \quad 36]^T$, so

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 1 & 3 & 4 & 1 \\ 1 & 9 & 15 & 7 \\ 1 & 27 & 54 & 36 \end{bmatrix} \quad \text{and} \quad \mathbf{M}^{-1} = \frac{1}{8}\begin{bmatrix} 27 & -27 & 9 & -1 \\ -85 & 133 & -55 & 7 \\ 66 & -106 & 46 & -6 \\ -36 & 60 & -28 & 4 \end{bmatrix}$$

Thus

$$e^{\mathbf{A}t} = \mathbf{M}e^{\mathbf{J}t}\mathbf{M}^{-1} =$$

$$\frac{1}{8}\begin{bmatrix} 27e^t + (-19 + 30t - 18t^2)e^{3t} & -27e^t + (27 - 46t + 30t^2)e^{3t} \\ 27e^t + (-27 + 54t - 54t^2)e^{3t} & -27e^t + (35 - 78t + 90t^2)e^{3t} \\ 27e^t + (-27 + 54t - 162t^2)e^{3t} & -27e^t + (27 - 54t + 270t^2)e^{3t} \\ 27e^t + (-27 - 162t - 486t^2)e^{3t} & -27e^t + (27 + 378t + 810t^2)e^{3t} \end{bmatrix}$$

$$\begin{bmatrix} 9e^t + (-9 + 18t - 14t^2)e^{3t} & -e^t + (1 - 2t + 2t^2)e^{3t} \\ 9e^t + (-9 + 26t - 42t^2)e^{3t} & -e^t + (1 - 2t + 6t^2)e^{3t} \\ 9e^t + (-1 - 6t - 126t^2)e^{3t} & -e^t + (1 + 6t + 18t^2)e^{3t} \\ 9e^t + (-9 - 270t - 378t^2)e^{3t} & -e^t + (9 + 54t + 54t^2)e^{3t} \end{bmatrix}$$

8.22 Another method of computing an analytic function $f(\mathbf{A})$ of an $n \times n$ matrix $\mathbf{A}$ with distinct eigenvalues is given by

$$f(\mathbf{A}) = \sum_{i=1}^{n} f(\lambda_i)\mathbf{Z}_i(\lambda)$$

where λ_i are the eigenvalues of $\mathbf{A}$ and the $n \times n$ $\mathbf{Z}_i$ matrices are given by

$$\mathbf{Z}_i(\lambda) = \frac{\prod_{\substack{j=1 \\ j \neq i}}^{n} (\mathbf{A} - \lambda_j\mathbf{I})}{\prod_{\substack{j=1 \\ j \neq i}}^{n} (\lambda_i - \lambda_j)}$$

This method is sometimes called Sylvester's expansion and sometimes it is referred to as Lagrange interpolation [2]. This is a restricted form, but it can be extended to the case of repeated eigenvalues. Use it to compute $e^{\mathbf{A}t}$ for

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

The eigenvalues are $\lambda_i = 1, -2, 3$ (see Problem 7.39).

$$\mathbf{Z}_1 = \frac{(\mathbf{A}+2\mathbf{I})(\mathbf{A}-3\mathbf{I})}{(1+2)(1-3)} = -\frac{1}{6}\begin{bmatrix} 4 & -2 & 3 \\ 1 & 3 & 1 \\ 1 & 3 & 1 \end{bmatrix}\begin{bmatrix} -1 & -2 & 3 \\ 1 & -2 & 1 \\ 1 & 3 & -4 \end{bmatrix} = -\frac{1}{6}\begin{bmatrix} -3 & 5 & -2 \\ 3 & -5 & 2 \\ 3 & -5 & 2 \end{bmatrix}$$

$$\mathbf{Z}_2 = \frac{(\mathbf{A}-\mathbf{I})(\mathbf{A}-3\mathbf{I})}{(-2-1)(-2-3)} = \frac{1}{15}\begin{bmatrix} 0 & 11 & -11 \\ 0 & 1 & -1 \\ 0 & -14 & 14 \end{bmatrix}$$

$$\mathbf{Z}_3 = \frac{(\mathbf{A}-\mathbf{I})(\mathbf{A}+2\mathbf{I})}{(3-1)(3+2)} = \frac{1}{10}\begin{bmatrix} 5 & 1 & 4 \\ 5 & 1 & 4 \\ 5 & 1 & 4 \end{bmatrix}$$

Therefore,

$$e^{\mathbf{A}t} = \frac{1}{6}e^{t}\begin{bmatrix} 3 & -5 & 2 \\ -3 & 5 & -2 \\ -3 & 5 & -2 \end{bmatrix} + \frac{1}{15}e^{-2t}\begin{bmatrix} 0 & 11 & -11 \\ 0 & 1 & -1 \\ 0 & -14 & 14 \end{bmatrix} + \frac{1}{10}e^{3t}\begin{bmatrix} 5 & 1 & 4 \\ 5 & 1 & 4 \\ 5 & 1 & 4 \end{bmatrix}$$

8.23 Find $\mathbf{A}^k$ for the $\mathbf{A}$ matrix of the previous problem.

The expansion matrices $\mathbf{Z}_i$ depend only on $\mathbf{A}$, not on the particular function of $\mathbf{A}$ that is being computed. They are the same as in the previous problem, so with $f(\mathbf{A}) = \mathbf{A}^k$, the result is

$$\mathbf{A}^k = \frac{1}{6}\begin{bmatrix} 3 & -5 & 2 \\ -3 & 5 & -2 \\ -3 & 5 & -2 \end{bmatrix} + \frac{(-2)^k}{15}\begin{bmatrix} 0 & 11 & -11 \\ 0 & 1 & -1 \\ 0 & -14 & 14 \end{bmatrix} + \frac{(3)^k}{10}\begin{bmatrix} 5 & 1 & 4 \\ 5 & 1 & 4 \\ 5 & 1 & 4 \end{bmatrix}$$

8.24 Find the closed form expression for $e^{\mathbf{A}t}$ if $\mathbf{A} = \begin{bmatrix} -13 & -15 \\ -15 & -13 \end{bmatrix}$. Investigate using the truncated infinite series to get a numerical approximation, with $t = 1$.

The eigenvalues of $\mathbf{A}$ are $\lambda = 2$ and -28. The closed form answer is

$$e^{\mathbf{A}t} = e^{2t}\begin{bmatrix} 0.5 & -0.5 \\ -0.5 & 0.5 \end{bmatrix} + e^{-28t}\begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$$

With $t = 1$ this becomes

$$e^{\mathbf{A}} \approx 0.5\begin{bmatrix} e^2 & -e^2 \\ -e^2 & e^2 \end{bmatrix} \approx \begin{bmatrix} 3.6945 & -3.6945 \\ -3.6945 & 3.6946 \end{bmatrix}$$

Using the first 30 terms in the infinite series gives the *bad* answer $e^{\mathbf{A}} = \begin{bmatrix} 2.435 & 2.435 \\ 2.435 & 2.435 \end{bmatrix} \times 10^{10}$. Keeping more terms does not help. The problem is that successive terms in the series depend upon $(\lambda_i t)^j/j!$. The alternating signs caused by $(-28)^j/j!$ cause loss of all significance, due to small differences of large numbers. A way to avoid this problem is to note that $e^{\mathbf{A}t} = [e^{\mathbf{A}t/m}]^m$. By picking some integer m such that $|\lambda|t/m$ is sufficiently small, the series for $e^{\mathbf{A}t/m}$ will converge rapidly. In fact the number of terms that need to be retained is K, where $[|\lambda|_{\max} t/m]^K/K! < \epsilon$. Note specifically that this depends on the largest-magnitude eigenvalue and on t/m. Once this truncated series is found, raising the answer to the mth power gives $e^{\mathbf{A}t}$. By selecting m as a power of 2, several successive doublings of $e^{\mathbf{A}t/m}$ efficiently give $e^{\mathbf{A}t}$. In this example it is found that $t/m = 0.25$ (i.e., $m = 4$) gives the correct answer, with about 20 terms retained in the series. Two successive doublings of $e^{\mathbf{A}t/m}$ are required at the end. If $t/m = 0.125$ ($m = 8$) is used, only 7 or 8 terms are needed in the series. This savings more than makes up for the extra matrix product (doubling) required to compute $e^{\mathbf{A}} = [e^{\mathbf{A}/8}]^8 = \{[(e^{\mathbf{A}/8})^2]^2\}^2$. (See Reference 3.)

PROBLEMS

8.25 Find the inverse of $\mathbf{A} = \begin{bmatrix} -1 & 2 & 0 \\ 1 & 1 & 0 \\ 2 & -1 & 2 \end{bmatrix}$ using the Cayley-Hamilton theorem.

8.26 Find $e^{\mathbf{A}t}$ if $\mathbf{A} = \begin{bmatrix} -1 & \frac{1}{2} \\ 0 & 1 \end{bmatrix}$.

8.27 Find $e^{\mathbf{A}t}$ with $\mathbf{A} = \begin{bmatrix} -3 & 2 \\ 0 & -3 \end{bmatrix}$.

8.28 Show that $e^{\begin{bmatrix} 0 & 1 \\ -3 & -4 \end{bmatrix}t} = \begin{bmatrix} \frac{1}{2}(3e^{-t} - e^{-3t}) & \frac{1}{2}(e^{-t} - e^{-3t}) \\ -\frac{3}{2}(e^{-t} - e^{-3t}) & \frac{1}{2}(3e^{-3t} - e^{-t}) \end{bmatrix}$.

8.29 Use $\mathbf{A} = \begin{bmatrix} 3 & 5 \\ \frac{16}{5} & 3 \end{bmatrix}$ to compute $e^{\mathbf{A}t}$.

8.30 Compute $e^{\mathbf{A}t}$ for $\mathbf{A} = \begin{bmatrix} -3 & 1 \\ 2 & -2 \end{bmatrix}$.

8.31 Find $\begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 1 \\ 0 & \frac{1}{2} & 2 \\ 0 & 0 & \frac{1}{2} \end{bmatrix}^k$.

8.32 Compute $e^{\mathbf{A}t}$ for $\mathbf{A} = \begin{bmatrix} -5 & -6 & 0 \\ 2 & 2 & 0 \\ 0 & 0 & -3 \end{bmatrix}$.

8.33 Find $e^{\mathbf{A}t}$ with $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{bmatrix}$.

8.34 Find $e^{\mathbf{A}t}$ for $\mathbf{A} = \begin{bmatrix} -10 & 0 & -10 & 0 \\ 0 & -0.7 & 9 & 0 \\ 0 & -1 & -0.7 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$.

8.35 Find a closed-form expression for

$$\mathbf{A}^k = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & -0.5 & 1.5 \end{bmatrix}^k$$

8.36 Show that $\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0.1653 & -0.9425 & 1.7085 \end{bmatrix}^k = (0.367955)^k \mathrm{E} + (0.665229)^k \mathrm{F} + (0.675317)^k \mathrm{G}$, where

$$\mathbf{E} = \begin{bmatrix} 4.916679{*}\mathrm{E}+00 & -1.467152{*}\mathrm{E}+01 & 1.094444{*}\mathrm{E}+01 \\ 1.809116{*}\mathrm{E}+00 & -5.398457{*}\mathrm{E}+00 & 4.027060{*}\mathrm{E}+00 \\ 6.656732{*}\mathrm{E}-01 & -1.986389{*}\mathrm{E}+00 & 1.481777{*}\mathrm{E}+00 \end{bmatrix}$$

$$\mathbf{F} = \begin{bmatrix} -8.285841^*\text{E}+01 & 3.478820^*\text{E}+02 & -3.334531^*\text{E}+02 \\ -5.511979^*\text{E}+01 & 2.314211^*\text{E}+02 & -2.218225^*\text{E}+02 \\ -3.666726^*\text{E}+01 & 1.539479^*\text{E}+02 & -1.475626^*\text{E}+02 \end{bmatrix}$$

$$\mathbf{G} = \begin{bmatrix} 7.894173^*\text{E}+01 & -3.332105^*\text{E}+02 & 3.225086^*\text{E}+02 \\ 5.331067^*\text{E}+01 & -2.250226^*\text{E}+02 & 2.177954^*\text{E}+02 \\ 3.600158^*\text{E}+01 & -1.519615^*\text{E}+02 & 1.470809^*\text{E}+02 \end{bmatrix}$$

8.37 Show that $\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ \frac{5}{6} & -\frac{13}{6} & -\frac{1}{3} \end{bmatrix}^k = \begin{bmatrix} (\frac{1}{2})^k & 1-(\frac{1}{2})^k & 0 \\ 0 & 1 & 0 \\ (\frac{1}{2})^k - (-\frac{1}{3})^k & 2(-\frac{1}{3})^k - (\frac{1}{2})^k - 1 & (-\frac{1}{3})^k \end{bmatrix}$.

8.38 The matrix $\mathbf{A} = \begin{bmatrix} 4 & -2 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 6 \end{bmatrix}$ is known to have eigenvalues $\lambda = \{6, 3+j, 3-j\}$ (see Problem 7.42). Find $\mathbf{A}_1 = e^{\mathbf{A}T}$ for $T = 0.2$, and then verify that the eigenvalues of $\mathbf{A}_1$ satisfy Frobenius' theorem.

8.39 Prove that $[e^{\mathbf{A}t}]^T = e^{\mathbf{F}t}$, where $\mathbf{F} = \mathbf{A}^T$.

8.40 The state of the unforced continuous-time system $\dot{\mathbf{x}} = \mathbf{Ax}$ is observed only at the periodic instants $t = 0, T, 2T, \ldots, kT, \ldots$. Use the exponent power law $[e^{\mathbf{A}T}]^k = e^{\mathbf{A}kT}$ to show that the initial-condition response can be described by either $\mathbf{x}(t_k) = e^{\mathbf{A}kT}\mathbf{x}(0)$ or $\mathbf{x}(k+1) = [\mathbf{A}_1]^k\mathbf{x}(0)$.

8.41 Use the modal decomposition of Sec. 8.6 to relate the approximate settling time T_s of a stable continuous-time system like the one in the previous problem to the eigenvalues of **A**. *Settling* is defined here as being within 2% of the final value.

8.42 If **A** has complex eigenvalues in the previous problem, the imaginary parts ω determine the modal frequencies of oscillation. This system is to be approximated by its states at discrete periodic times t_k (perhaps because of sampling instrumentation or digital computer control). What sampling time $T = t_{k+1} - t_k$ would you recommend? You may wish to reread Problem 2.21 at this point.

9

Analysis of Continuous- and Discrete-Time Linear State Equations

9.1 INTRODUCTION

The description of a physical system by a mathematical model was discussed in Chapter 1. The model often takes the form of a set of coupled differential equations of various orders. In other cases the original model takes the form of a set of discrete-time difference equations, as was the case when fitting empirical data with an ARMA model in Chapter 1. In Chapter 2, linear models were described by input-output transfer functions. The continuous-time case used Laplace transfer functions and the discrete-time case used Z-transforms. It was pointed out in Chapter 2 that the discrete-time model may represent an approximation of a continuous-time system or it may be necessitated because of sampling sensors or digital controllers. Chapter 3 developed state variable models for these same classes of physical systems, starting from either the coupled differential equation (difference equation) model or the Laplace transform (Z-transform) transfer function description. Figure 9.1 presents the modeling paradigm under discussion. It emphasizes that there are two distinct routes to the determination of a discrete-time approximate state model for a continuous-time system. This approximation problem is revisited in Sec. 9.8 because it is needed so frequently in control applications.

This chapter is devoted to the *solution* of the resulting vector-matrix state variable equations. It will be assumed here that the control input variable is known. In later chapters the typical control system design problem, the determination of the control inputs $\mathbf{u}(t)$ which will cause the states $\mathbf{x}(t)$ and/or outputs $\mathbf{y}(t)$ to behave as desired, will be discussed. The determination of the control for linear systems will be considered from two different points of view. In Chapter 13 the controller is designed to give specified closed-loop poles. This is the pole-placement problem. In Chapter 14 the controller is designed in order to minimize a quadratic cost function. This is a widely used subclass of optimal control theory. Before considering the controller design problem, the behavior of the system response for a known input should be

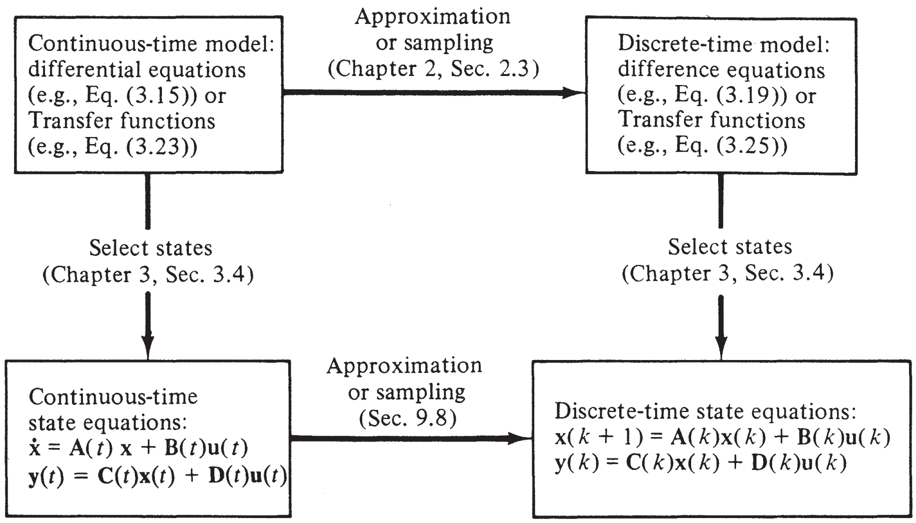


Figure 9.1 State variable modeling paradigm.

understood. The major effort is spent in solving the differential or difference equations for $\mathbf{x}(t)$ or $\mathbf{x}(k)$ in terms of $\mathbf{u}$. Then the output $\mathbf{y}$ is related to the state $\mathbf{x}$ and the input $\mathbf{u}$ in a simple algebraic fashion.

9.2 FIRST-ORDER SCALAR DIFFERENTIAL EQUATIONS

The familiar scalar differential equation

$$\dot{x} = a(t)x(t) + b(t)u(t) \tag{9.1}$$

is reviewed before considering the nth-order matrix case. When the input $u(t)$ is zero, the differential equation for x is said to be homogeneous. In this case,

$$\frac{dx}{dt} = a(t)x(t) \quad \text{or} \quad \frac{dx}{x} = a(t)\,dt$$

In the latter form the dependent variable x and the independent variable t are separated so that both sides of the equation represent an exact differential and can be integrated:

$$\int_{x(t_0)}^{x(t)} \frac{dx}{x} = \ln(x)\Big|_{x(t_0)}^{x(t)} = \int_{t_0}^{t} a(\tau)\,d\tau$$

or

$$\ln x(t) - \ln x(t_0) = \int_{t_0}^{t} a(\tau)\,d\tau$$

Using $\ln x(t) - \ln x(t_0) = \ln[x(t)/x(t_0)]$ and the fact that $e^{\ln x} = x$ gives

$$x(t) = x(t_0)e^{\int_{t_0}^{t} a(\tau)\,d\tau} \tag{9.2a}$$

In the particular case where $a(t) = a$ is constant, this reduces to

$$x(t) = x(t_0)e^{(t-t_0)a} \tag{9.2b}$$

As expected, the initial condition $x(t_0)$ must be specified before a unique solution $x(t)$ can be determined.

When the nonhomogeneous Eq. (*9.1*) is considered, a solution can still be obtained by reducing the equation to a form which can be easily integrated. One extra step is required first. Consider

$$\frac{d}{dt}[k(t)x(t)] = k(t)\dot{x}(t) + \dot{k}(t)x(t)$$

If Eq. (*9.1*) is multiplied by $k(t)$ and rearranged, the result is

$$k(t)\dot{x}(t) - k(t)a(t)x(t) = k(t)b(t)u(t)$$

The left-hand side can be made an exact differential provided a function $k(t)$ is selected that satisfies $\dot{k}(t) = -k(t)a(t)$. This requirement on $k(t)$ represents a first-order homogeneous equation of the type just considered. Its solution is $k(t) = k(t_0)e^{-\int_{t_0}^{t} a(\tau)\,d\tau}$. Using this, the nonhomogeneous equation for $x(t)$ can be written in terms of exact differentials,

$$d[k(t)x(t)] = k(t)b(t)u(t)\,dt$$

Carrying out the integration of both sides and solving for $x(t)$ gives

$$x(t) = \left[\frac{k(t_0)}{k(t)}\right]x(t_0) + \int_{t_0}^{t}\frac{k(\tau)}{k(t)}b(\tau)u(\tau)\,d\tau$$

Using the agreed-upon form for $k(t)$, the general solution becomes

$$x(t) = \{e^{\int_{t_0}^{t} a(\tau)\,d\tau}\}x(t_0) + \int_{t_0}^{t} e^{\int_{\tau}^{t} a(\zeta)d\zeta}\,b(\tau)u(\tau)\,d\tau \tag{9.3a}$$

If the coefficient a is constant, this reduces to

$$x(t) = e^{(t-t_0)a}x(t_0) + \int_{t_0}^{t} e^{(t-\tau)a}\,b(\tau)u(\tau)\,d\tau \tag{9.3b}$$

The last result can be derived directly using Laplace transforms and the convolution theorem (see Problem 9.1). A direct verification that this represents a solution of the differential equation is given in Problem 9.2.

When b is constant as well as a, Eq. (*9.3b*) indicates that $x(t)$ depends only on the time difference $t - t_0$ and so the starting time t_0 is often replaced by 0 for simplicity. When both a and b are constant, the system is said to be time-invariant because the response due to a given input is always the same regardless of the label attached to the starting time.

9.3 THE CONSTANT COEFFICIENT MATRIX CASE

The homogeneous set of n state equations

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}, \quad \mathbf{x}(t_0) \text{ given, } \mathbf{A} \text{ constant} \tag{9.4}$$

has a solution which is completely analogous to the scalar result of equation (*9.2b*):

$$\mathbf{x}(t) = e^{(t-t_0)\mathbf{A}}\mathbf{x}(t_0) \tag{9.5}$$

There are several methods of verifying that this is a solution to the state equations (see Problems 9.3 and 9.4). First, note that the initial conditions are satisfied. That is,

$$\mathbf{x}(t_0) = e^{(t_0-t_0)\mathbf{A}}\mathbf{x}(t_0) = e^{[\mathbf{0}]}\mathbf{x}(t_0) = \mathbf{x}(t_0)$$

Differentiating both sides of Eq. (*9.5*) and using the result of Example 8.3, it is easily verified that $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t)$. Since Eq. (*9.5*) satisfies the initial conditions and the differential equation, it represents a unique solution (see page 20 of Reference 1) of Eq. (*9.4*).

The nonhomogeneous set of state equations is now considered. The system matrix $\mathbf{A}$ is still constant, but $\mathbf{B}(t)$ may be time-varying. Components of $\mathbf{B}(t)\mathbf{u}(t)$ are assumed to be piecewise continuous to guarantee a unique solution (see page 74 of Reference 1):

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}(t)\mathbf{u}(t), \quad \mathbf{x}(t_0) \text{ given} \tag{9.6}$$

The technique used in solving the scalar equation is repeated with only minor dimensional modifications. Let $\mathbf{K}(t)$ be an $n \times n$ matrix. Premultiplying Eq. (*9.6*) by $\mathbf{K}(t)$ and rearranging gives

$$\mathbf{K}(t)\dot{\mathbf{x}}(t) - \mathbf{K}(t)\mathbf{A}\mathbf{x}(t) = \mathbf{K}(t)\mathbf{B}(t)\mathbf{u}(t)$$

Since $d[\mathbf{K}(t)\mathbf{x}(t)]/dt = \mathbf{K}\dot{\mathbf{x}} + \dot{\mathbf{K}}\mathbf{x}$, the left-hand side can be written as an exact (vector) differential provided $\dot{\mathbf{K}} = -\mathbf{K}(t)\mathbf{A}$. One such matrix is $\mathbf{K}(t) = e^{-(t-t_0)\mathbf{A}}$. Agreeing that this is the $\mathbf{K}$ matrix to be used, the differential equation can be written

$$d[\mathbf{K}(t)\mathbf{x}(t)] = \mathbf{K}(t)\mathbf{B}(t)\mathbf{u}(t)\,dt$$

Integration gives

$$\mathbf{K}(t)\mathbf{x}(t) - \mathbf{K}(t_0)\mathbf{x}(t_0) = \int_{t_0}^{t} \mathbf{K}(\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\,d\tau$$

The selected form for $\mathbf{K}$ always has an inverse, so

$$\mathbf{x}(t) = \mathbf{K}^{-1}(t)\mathbf{K}(t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{K}^{-1}(t)\mathbf{K}(\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\,d\tau$$

or

$$\mathbf{x}(t) = e^{(t-t_0)\mathbf{A}}\mathbf{x}(t_0) + \int_{t_0}^{t} e^{(t-\tau)\mathbf{A}}\mathbf{B}(\tau)\mathbf{u}(\tau)\,d\tau \tag{9.7}$$

This represents the solution for any system equation in the form of Eq. (*9.6*). Note that it is composed of a term depending only on the initial state and a convolution integral involving the input but not the initial state. These two terms are known by various names such as the homogeneous solution and the particular integral, the force-free response and the forced response, the zero input response and the zero state response, etc.

9.4 SYSTEM MODES AND MODAL DECOMPOSITION [2, 3]

Equation (*9.6*) is considered again. It is emphasized that the matrix **A** is constant, but $\mathbf{B}(t)$ may be time-varying. Assume that the n eigenvalues λ_i and n independent vectors, either eigenvectors or generalized eigenvectors, have been found for the matrix **A**. These vectors are denoted by ξ_i to avoid confusion with the state vector **x**. Since the set $\{\xi_i\}$ is linearly independent, it can be used as a basis for the state space Σ. Thus at any given time t, the state $\mathbf{x}(t)$ can be expressed as

$$\mathbf{x}(t) = q_1(t)\xi_1 + q_2(t)\xi_2 + \cdots + q_n(t)\xi_n \tag{9.8}$$

The time variation of **x** is contained in the expansion coefficients q_i since **A**, and hence the ξ_i, are constant. At any given time t, the vector $\mathbf{B}(t)\mathbf{u}(t) \in \Sigma$ and therefore it, too, can be expanded as

$$\mathbf{B}(t)\mathbf{u}(t) = \beta_1(t)\xi_1 + \beta_2(t)\xi_2 + \cdots + \beta_n(t)\xi_n$$

In fact, $\beta_i(t) = \langle \mathbf{r}_i, \mathbf{B}(t)\mathbf{u}(t)\rangle$, where $\{\mathbf{r}_i\}$ is the set of reciprocal basis vectors.

Using the above expansion, Eq. (*9.6*) becomes

$$\dot{q}_1\xi_1 + \dot{q}_2\xi_2 + \cdots + \dot{q}_n\xi_n = q_1\mathbf{A}\xi_1 + q_2\mathbf{A}\xi_2 + \cdots + q_n\mathbf{A}\xi_n + \beta_1\xi_1 + \beta_2\xi_2 + \cdots + \beta_n\xi_n$$

Assume for the moment that **A** is normal (see Sec. 5.12, page 184, and Problem 7.28, page 276) so that all ξ_i are eigenvectors rather than generalized eigenvectors. Then $\mathbf{A}\xi_i = \lambda_i\xi_i$, so that

$$(\dot{q}_1 - \lambda_1 q_1 - \beta_1)\xi_1 + (\dot{q}_2 - \lambda_2 q_2 - \beta_2)\xi_2 + \cdots + (\dot{q}_n - \lambda_n q_n - \beta_n)\xi_n = \mathbf{0}$$

Since the set $\{\xi_i\}$ is linearly independent, this requires that

$$\dot{q}_i = \lambda_i q_i + \beta_i \quad \text{for } i = 1, 2, \ldots, n$$

This demonstrates that when **A** is constant and has a full set of eigenvectors, the system is completely described by a set of n uncoupled scalar equations whose solutions are of the form

$$q_i(t) = e^{(t-t_0)\lambda_i} q_i(t_0) + \int_{t_0}^{t} e^{(t-\tau)\lambda_i}\beta_i(\tau)\,d\tau$$

Of course, $q_i(t_0) = \langle \mathbf{r}_i, \mathbf{x}(t_0)\rangle$. The state vector is given by

$$\mathbf{x}(t) = q_1(t)\xi_1 + q_2(t)\xi_2 + \cdots + q_n(t)\xi_n$$

The terms in this sum are called the system modes. The general response of a complicated system can be broken down into the sum of n simple modal responses.

It should be recognized that Eq. (*9.8*) can be written in terms of the modal matrix $\mathbf{M} = [\xi_1 \ldots \xi_n]$ as $\mathbf{x} = \mathbf{M}\mathbf{q}$, and as such, represents a change of basis. Using this notation, Eq. (*9.6*) is considered again:

$\dot{\mathbf{x}}$ becomes $\mathbf{M}\dot{\mathbf{q}}$, since **M** is constant

$\mathbf{A}\mathbf{x}$ becomes $\mathbf{A}\mathbf{M}\mathbf{q}$ and $\mathbf{B}\mathbf{u}$ remains unchanged

Therefore,

$$\mathbf{M}\dot{\mathbf{q}} = \mathbf{AMq} + \mathbf{Bu}$$

or

$$\dot{\mathbf{q}} = \mathbf{M}^{-1}\mathbf{AMq} + \mathbf{M}^{-1}\mathbf{Bu} = \mathbf{Jq} + \mathbf{B}_n\mathbf{u}$$

where $\mathbf{B}_n \triangleq \mathbf{M}^{-1}\mathbf{B}$ and the assumption regarding a full set of eigenvectors is dropped. $\mathbf{J}$ is the Jordan canonical form (or the diagonal matrix $\mathbf{\Lambda}$ in many cases). If the same change of basis is used in the output equation, then the system is described by the pair of *normal form* equations

$$\dot{\mathbf{q}} = \mathbf{Jq} + \mathbf{B}_n\mathbf{u} \tag{9.9}$$

$$\mathbf{y} = \mathbf{C}_n\mathbf{q} + \mathbf{Du} \tag{9.10}$$

where $\mathbf{C}_n \triangleq \mathbf{CM}$. One advantage of the normal form is that the state equations are as nearly uncoupled as possible. Each component of $\mathbf{q}$ is coupled to at most one other component because of the nature of the Jordan form matrix $\mathbf{J}$. The solution for Eq. (*9.9*) can be written as

$$\mathbf{q}(t) = e^{(t-t_0)\mathbf{J}}\mathbf{q}(t_0) + \int_{t_0}^{t} e^{(t-\tau)\mathbf{J}}\mathbf{B}_n(\tau)\mathbf{u}(\tau)\,d\tau$$

Relating this to the original state vector gives

$$\mathbf{x}(t) = \mathbf{Mq}(t) = \mathbf{M}e^{(t-t_0)\mathbf{J}}\mathbf{M}^{-1}\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{M}e^{(t-\tau)\mathbf{J}}\mathbf{M}^{-1}\mathbf{B}(\tau)\mathbf{u}(\tau)\,d\tau \tag{9.11}$$

The preceding equation used the fact that $\mathbf{M}^{-1}$ is the matrix of transposed reciprocal basis vectors, which means that

$$\mathbf{q}(t_0) = \mathbf{M}^{-1}\mathbf{x}(t_0)$$

Comparing Eqs. (*9.11*) and (*9.7*) shows that

$$e^{(t-t_0)\mathbf{A}} = \mathbf{M}e^{(t-t_0)\mathbf{J}}\mathbf{M}^{-1}$$

a result given earlier in Chapter 8.

Modal decomposition is useful because of the insight it gives regarding the intrinsic properties of the system. The properties of controllability, observability, stabilizability, and detectability (Chapter 11) are more easily understood and evaluated. The stability properties (Chapter 10) of the system are also more clearly revealed. Modal decomposition provides a simple geometrical picture for the motion of the state vector versus time. By retaining only the dominant modes, a high-order system can be approximated by a lower-order system.

It should be kept in mind that the modal decomposition technique is useful only if $\mathbf{A}$, and thus ξ_i, λ_i are constant. It is the invariance of the vector parts of $\mathbf{x}(t)$, that is, the ξ_i terms, that gives value to the method. If the modal matrix were time-varying and had to be continually reevaluated, most of the advantages of modal decomposition would be lost.

EXAMPLE 9.1 A system is described by

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix}=\begin{bmatrix}0 & 1\\ 8 & -2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix}+\begin{bmatrix}1\\ 1\end{bmatrix}u$$

$$y(t)=[4 \quad 1]\mathbf{x}(t)$$

The initial conditions are $\mathbf{x}(0)=[1 \quad -4]^T$. Assume that $u(t)=0$ and analyze this system.

With $\mathbf{A}=\begin{bmatrix}0 & 1\\ 8 & -2\end{bmatrix}$, the eigenvalues are $\lambda_1=-4, \lambda_2=2$.

The eigenvectors are $\xi_1=[1 \quad -4],^T \xi_2=[1 \quad 2]^T$.

The modal matrix and its inverse are $\mathbf{M}=\begin{bmatrix}1 & 1\\ -4 & 2\end{bmatrix}, \mathbf{M}^{-1}=\frac{1}{6}\begin{bmatrix}2 & -1\\ 4 & 1\end{bmatrix}$.

Any one of several methods gives

$$e^{\mathbf{A}t}=\frac{1}{6}\begin{bmatrix}2e^{-4t}+4e^{2t} & -e^{-4t}+e^{2t}\\ -8e^{-4t}+8e^{2t} & 4e^{-4t}+2e^{2t}\end{bmatrix}$$

so the homogeneous solution is $\mathbf{x}(t)=e^{\mathbf{A}t}\mathbf{x}(0)=[e^{-4t} \quad -4e^{-4t}]^T$, and the output is $y(t)=4e^{-4t}-4e^{-4t}=0$ for all t. ■

EXAMPLE 9.2 Modal decomposition is now applied in an attempt to gain insight into the unusual result of Example 9.1.

Since the eigenvalues are distinct, $\mathbf{M}^{-1}\mathbf{AM}=\mathbf{\Lambda}$ for this system, and Eq. (*9.9*) becomes

$$\begin{bmatrix}\dot{q}_1\\ \dot{q}_2\end{bmatrix}=\begin{bmatrix}-4 & 0\\ 0 & 2\end{bmatrix}\begin{bmatrix}q_1\\ q_2\end{bmatrix}+\begin{bmatrix}\frac{1}{6}\\ \frac{5}{6}\end{bmatrix}u$$

The initial conditions are $\mathbf{q}(0)=\mathbf{M}^{-1}\mathbf{x}(0)=[1 \quad 0]^T$. Equation (*9.10*) becomes $y=[0 \quad 6]\mathbf{q}$. The state vector $\mathbf{x}(t)$ can be written as the sum of two modes,

$$\mathbf{x}(t)=q_1(0)e^{-4t}\begin{bmatrix}1\\ -4\end{bmatrix}+q_2(0)e^{2t}\begin{bmatrix}1\\ 2\end{bmatrix}$$

The particular initial condition selected here has no component along the direction of mode 2, as evidenced by $q_2(0)=0$. Thus the second mode is not excited, since the input $u(t)$ has been assumed zero. The output of this system consists only of the second mode contribution, as evidenced by $\mathbf{C}_n=[0 \quad 6]$. Mode 1 contributes nothing to the output and mode 2 is not excited, so the output remains identically zero. ■

9.5 THE TIME-VARYING MATRIX CASE

The time-varying homogeneous state equations

$$\dot{\mathbf{x}}=\mathbf{A}(t)\mathbf{x} \tag{9.12}$$

are considered first. In order that this qualify as a valid state equation, it is required that there be a *unique* solution for every $\mathbf{x}(t_0)\in\Sigma$. This places some restriction on the kind of time variation allowed on the matrix $\mathbf{A}$. A *sufficient* condition for the existence of unique solutions is to require that all elements $a_{ij}(t)$ of $\mathbf{A}(t)$ be continuous. Weaker conditions may be found in textbooks on differential equations [1, 4].

Since $\dim(\Sigma) = n$, n linearly independent initial vectors $\mathbf{x}_i(t_0)$ can be found, and each one defines a unique solution of Eq. (*9.12*), called $\mathbf{x}_i(t)$, $t \geq t_0$. Define an $n \times n$ matrix $\mathbf{U}(t_0)$ with columns formed by the independent initial condition vectors $\mathbf{x}_i(t_0)$. (A particular set $\mathbf{U}(t_0) = I_n$ is sometimes used, but that restriction is unnecessary.) The n solutions corresponding to these initial conditions are used as the columns in forming an $n \times n$ matrix $\mathbf{U}(t) = [\mathbf{x}_1(t)] \quad \mathbf{x}_2(t) \quad \ldots \quad \mathbf{x}_n(t)]$. Any matrix $\mathbf{U}(t)$ satisfying

$$\dot{\mathbf{U}}(t) = \mathbf{A}(t)\mathbf{U}(t) \tag{9.13}$$

is called a *fundamental solution matrix*, provided that $|\mathbf{U}(t_0)| \neq 0$. Assuming that the fundamental solution matrix is available, the solution to Eq. (*9.12*) with an arbitrary initial condition vector $\mathbf{x}(t_0)$ is

$$\mathbf{x}(t) = \mathbf{U}(t)\mathbf{U}^{-1}(t_0)\mathbf{x}(t_0) \tag{9.14}$$

This is easily verified. Checking initial conditions,

$$\mathbf{x}(t_0) = \mathbf{U}(t_0)\mathbf{U}^{-1}(t_0)\mathbf{x}(t_0) = \mathbf{I}_n\,\mathbf{x}(t_0) = \mathbf{x}(t_0)$$

Checking to see that this solution satisfies the differential equation,

$$\dot{\mathbf{x}}(t) = \dot{\mathbf{U}}(t)\mathbf{U}^{-1}(t_0)\mathbf{x}(t_0) = \mathbf{A}(t)\mathbf{U}(t)\mathbf{U}^{-1}(t_0)\mathbf{x}(t_0) = \mathbf{A}(t)\mathbf{x}(t)$$

Both the initial conditions and the differential equation are satisfied, so this represents the unique solution to the homogeneous problem.

The nonhomogeneous time-varying state equation is solved in an analogous manner to the scalar and constant matrix cases. That is, the equation is reduced to exact differentials so that it can be integrated. Preliminary to this, it is noted that $\mathbf{U}^{-1}(t)$ can be shown to exist for all $t \geq t_0$ and that

$$\mathbf{U}(t)\mathbf{U}^{-1}(t) = \mathbf{I}_n \quad \text{so that} \quad \frac{d}{dt}(\mathbf{U}(t)\mathbf{U}^{-1}(t)) = [\mathbf{0}]$$

or

$$\frac{d\mathbf{U}}{dt}\mathbf{U}^{-1} + \mathbf{U}\frac{d\mathbf{U}^{-1}}{dt} = [\mathbf{0}] \quad \text{or} \quad \frac{d\mathbf{U}^{-1}}{dt} = -\mathbf{U}^{-1}\frac{d\mathbf{U}}{dt}\mathbf{U}^{-1}$$

Therefore,

$$\frac{d\mathbf{U}^{-1}}{dt} = -\mathbf{U}^{-1}(t)\mathbf{A}(t) \tag{9.15}$$

Note that the matrix $\mathbf{K}(t)$ of Sec. 9.3 is an example of $\mathbf{U}^{-1}(t)$. Premultiplying the time-varying version of Eq. (*9.6*) by $\mathbf{U}^{-1}(t)$, postmultiplying Eq. (*9.15*) by $\mathbf{x}(t)$, and adding the results gives

$$\mathbf{U}^{-1}(t)\dot{\mathbf{x}} + \frac{d\mathbf{U}^{-1}}{dt}\mathbf{x}(t) = \mathbf{U}^{-1}(t)\mathbf{B}(t)\mathbf{u}(t)$$

or

$$\frac{d}{dt}[\mathbf{U}^{-1}(t)\mathbf{x}(t)] = \mathbf{U}^{-1}(t)\mathbf{B}(t)\mathbf{u}(t)$$

The nonhomogeneous solution is obtained by integrating both sides from t_0 to t, that is,

$$\mathbf{U}^{-1}(t)\mathbf{x}(t) - \mathbf{U}^{-1}(t_0)\mathbf{x}(t_0) = \int_{t_0}^{t} \mathbf{U}^{-1}(\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\, d\tau$$

or

$$\mathbf{x}(t) = \mathbf{U}(t)\mathbf{U}^{-1}(t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{U}(t)\mathbf{U}^{-1}(\tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\, d\tau \qquad (9.16)$$

The result again takes the form of a term depending on the initial state and a convolution integral involving the input function. In fact, the first term is the same homogeneous solution given by Eq. (*9.14*). This result shows the *form* of the solution, but it may not be immediately useful. It assumes knowledge of the fundamental solution matrix $\mathbf{U}(t)$, and actually finding $\mathbf{U}$ has not yet been addressed.

9.6 THE TRANSITION MATRIX

The preceding results prompt the definition of an important matrix that can be associated with any linear system, namely, the *transition matrix:*

$$\mathbf{\Phi}(t, \tau) \triangleq \mathbf{U}(t)\mathbf{U}^{-1}(\tau) \qquad (9.17)$$

This $n \times n$ matrix is a linear transformation or mapping of Σ onto itself. That is, in the absence of any input $\mathbf{u}(t)$, given the state $\mathbf{x}(\tau)$ at any time τ, the state at any other time t is given by the mapping

$$\mathbf{x}(t) = \mathbf{\Phi}(t, \tau)\mathbf{x}(\tau)$$

The mapping of $\mathbf{x}(\tau)$ into itself requires that

$$\mathbf{\Phi}(\tau, \tau) = \mathbf{I}_n \quad \text{for any } \tau \qquad (9.18)$$

This is obviously true from Eq. (*9.17*). Differentiating $\mathbf{\Phi}(t, \tau)$ with respect to its first argument t gives

$$\frac{d\mathbf{\Phi}(t, \tau)}{dt} = \frac{d\mathbf{U}(t)}{dt}\mathbf{U}^{-1}(\tau) = \mathbf{A}(t)\mathbf{U}(t)\mathbf{U}^{-1}(\tau)$$

so

$$\frac{d\mathbf{\Phi}(t, \tau)}{dt} = \mathbf{A}(t)\mathbf{\Phi}(t, \tau) \qquad (9.19)$$

The set of differential equations (*9.19*), along with the initial condition, Eq. (*9.18*), is often considered as the definition for $\mathbf{\Phi}(t, \tau)$.

Two other important properties of the transition matrix are the semigroup property, mentioned in Chapter 3 while defining state,

$$\boldsymbol{\Phi}(t_2, t_0) = \boldsymbol{\Phi}(t_2, t_1)\boldsymbol{\Phi}(t_1, t_0) \quad \text{for any } t_0, t_1, t_2$$

and the relationship between $\boldsymbol{\Phi}^{-1}$ and $\boldsymbol{\Phi}$:

$$\boldsymbol{\Phi}^{-1}(t, t_0) = \boldsymbol{\Phi}(t_0, t) \quad \text{for any } t_0, t$$

Both of these propeties are immediately obvious if the definition of Eq. (*9.17*) is considered.

Methods of Computing the Transition Matrix

If the matrix **A** is constant, then

$$\boldsymbol{\Phi}(t, \tau) = e^{(t-\tau)\mathbf{A}} \qquad \text{(Compare Eqs. (9.7) and (9.16).)}$$

Therefore, all the methods of Chapter 8 are applicable for finding $\boldsymbol{\Phi}$, including

1. $\boldsymbol{\Phi}(t, 0) = \mathscr{L}^{-1}\{[\mathbf{I}s - \mathbf{A}]^{-1}\}$. $\boldsymbol{\Phi}(t, \tau)$ is then found by replacing t by $t - \tau$, since $\boldsymbol{\Phi}(t, \tau) = \boldsymbol{\Phi}(t - \tau, 0)$ when **A** is constant.
2. $\boldsymbol{\Phi}(t, \tau) = \alpha_0\mathbf{I} + \alpha_1\mathbf{A} + \cdots + \alpha_{n-1}\mathbf{A}^{n-1}$, where $e^{\lambda_i(t-\tau)} = \alpha_0 + \alpha_1\lambda_i + \cdots + \alpha_{n-1}\lambda_i^{n-1}$ and, if some eigenvalues are repeated, derivatives of the above expression with respect to λ must be used.
3. $\boldsymbol{\Phi}(t, \tau) = \mathbf{M}e^{\mathbf{J}(t-\tau)}\mathbf{M}^{-1}$, where **J** is the Jordan form (or the diagonal matrix $\boldsymbol{\Lambda}$), and **M** is the modal matrix.
4. $\boldsymbol{\Phi}(t, \tau) = \sum_{i=1}^{n} e^{\lambda_i(t-\tau)}\mathbf{Z}_i(\lambda)$, where the $n \times n$ matrices $\mathbf{Z}_i$ are defined in Problem 8.22.
5. $\boldsymbol{\Phi}(t, \tau) \cong \mathbf{I} + \mathbf{A}(t - \tau) + \frac{1}{2}\mathbf{A}^2(t - \tau)^2 + \frac{1}{3!}\mathbf{A}^3(t - \tau)^3 + \cdots$. This infinite series can be truncated after a finite number of terms to obtain an approximation for the transition matrix. See Problem 9.10 for a more efficient computational form of this series.

A modification of method 1, using signal flow graphs to avoid the matrix inversion, can also be used. Since $\phi_{ij}(s) \triangleq \mathscr{L}\{\phi_{ij}(t, 0)\}$ is the transfer function from the input to the jth integrator to the output of the ith integrator, that is, the ith state variable x_i, Mason's gain rule [5] can be used to write the components $\phi_{ij}(s)$ directly. Inverse Laplace transformations then give the elements of $\boldsymbol{\Phi}(t, 0)$.

When $\mathbf{A}(t)$ is time-varying, the choices for finding $\boldsymbol{\Phi}(t, \tau)$ are more restricted:

1. *Computer solution of* $\dot{\boldsymbol{\Phi}} = \mathbf{A}(t)\boldsymbol{\Phi}$ *with* $\boldsymbol{\Phi}(\tau, \tau) = \mathbf{I}$. This is expensive in terms of computer time if the transition matrix is required for all t and τ. It means solving the matrix differential equation many times, using a large set of different τ values as initial times.

2. *Let* $\mathbf{B}(t, \tau) = \int_\tau^t \mathbf{A}(\zeta)\, d\zeta$. Unlike the time-varying scalar case, $\boldsymbol{\Phi}(t, \tau) \neq e^{\mathbf{B}(t,\tau)}$ unless $\mathbf{B}(t, \tau)$ and $\mathbf{A}(t)$ commute. Unfortunately, they generally do not commute, but two cases for which they do are when **A** is constant and when **A** is diagonal. Whenever $\mathbf{BA} = \mathbf{AB}$, any method may be used for computing $\boldsymbol{\Phi}(t, \tau) = e^{\mathbf{B}(t,\tau)}$.

3. *Successive approximations may be used to obtain an approximate transition matrix, as derived in Problem 9.5:*

$$\mathbf{\Phi}(t, t_0) = \mathbf{I}_n + \int_{t_0}^{t} \mathbf{A}(\tau_0)\, d\tau_0 + \int_{t_0}^{t} \mathbf{A}(\tau_0) \int_{t_0}^{\tau_0} \mathbf{A}(\tau_1)\, d\tau_1\, d\tau_0$$

$$+ \int_{t_0}^{t} \mathbf{A}(\tau_0) \int_{t_0}^{\tau_0} \mathbf{A}(\tau_1) \int_{t_0}^{\tau_1} \mathbf{A}(\tau_2)\, d\tau_2\, d\tau_1\, d\tau_0 + \cdots$$

4. *In some special cases closed form solutions to the equations may be possible.*

In many cases it is necessary or desirable to select a set of discrete time points, t_k, such that $\mathbf{A}(t)$ can be approximated by a constant matrix over each interval $[t_k, t_{k+1}]$. Then a set of difference equations can be used to describe the state of the system at these discrete times. The approximating difference equation is derived in Sec. 9.8 and Problem 9.10. Solutions of this type of equation are discussed in Sec. 9.9.

9.7 SUMMARY OF CONTINUOUS-TIME LINEAR SYSTEM SOLUTIONS

The most general state space description of a linear system is given by

$$\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$$

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t)$$

The form of the solution for $\mathbf{x}(t)$ has been shown to be

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\, d\tau \tag{9.20}$$

Equation (*9.20*) is the explicit form of the (linear system) transformation

$$\mathbf{x}(t) = \mathbf{g}(\mathbf{x}(t_0), \mathbf{u}(t), t_0, t)$$

introduced in Chapter 3 when defining state. When the system matrix $\mathbf{A}$ is constant, the transition matrix can always be found in closed form, although it may be tedious to do so for high-order systems. In the time-varying case numerical solutions or approximations must be relied upon. When considering certain questions, it is valuable to know that a solution exists in the stated form, even if it cannot be easily computed.

When the solution for $\mathbf{x}(t)$ is used, the expression for the output becomes

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{C}(t)\mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\, d\tau + \mathbf{D}(t)\mathbf{u}(t)$$

or

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} [\mathbf{C}(t)\mathbf{\Phi}(t, \tau)\mathbf{B}(\tau) + \delta(t - \tau)\mathbf{D}(\tau)]\mathbf{u}(\tau)\, d\tau$$

The term inside the integral is an explicit expression for the weighting matrix $\mathbf{W}(t, \tau)$ used in the integral form of the input-output description for the system:

$$\mathbf{y}(t) = \int_{t_0}^{t} \mathbf{W}(t, \tau)\mathbf{u}(\tau)\, d\tau$$

It is seen that this input-output description is only valid when $\mathbf{x}(t_0) = \mathbf{0}$, the so-called zero state response case. This difficulty can be overcome by considering the initial state part of $\mathbf{y}(t)$ as having arisen because of some input between $t = -\infty$ and $t = t_0$. Then

$$\mathbf{y}(t) = \int_{-\infty}^{t} \mathbf{W}(t, \tau)\mathbf{u}(\tau)\, d\tau$$

9.8 DISCRETE-TIME MODELS OF CONTINUOUS-TIME SYSTEMS

A multivariable continuous-time system with r inputs $u_i(t)$ and m outputs $y_j(t)$ is considered. The mathematical model for such a system may originally be given in various forms, including transfer functions, coupled differential equations in y_j variables, or state variable format. These options correspond to the left side of Figure 9.1. There are at least three possible reasons for being interested in a discrete-time model of such a system, as suggested by Figure 9.2.

1. *Sampled outputs:* Sampling or time-shared sensors may provide output data only at discrete time points t_k. A scanning radar gives measurements to a target only once per scan cycle as the transmitted beam sweeps across the object being tracked. A digital voltmeter may be monitoring several signals via a multiplexed A/D input channel. No information is available between the sample times.

2. *Sampled inputs:* A digital controller calculates new values for the control inputs only once per control cycle. A zero-order hold converts the digital commands into a sequence of piecewise constant analog levels. These levels change only at the discrete time points t_k.

3. *Digital simulation:* Even though all the actual system input and output signals are continuous, a digital simulation may be desired to study the time response. This inherently involves discrete approximations of all the signals, and hence it is equivalent to a combination of sampled outputs and inputs. The goal is to pick a stepsize which is sufficiently small so that the continuous signals $\mathbf{u}$, $\mathbf{y}$, and $\mathbf{x}$ can be represented by piecewise constant approximations within an acceptable error.

Regardless of the reason for using the discrete model, the goal is to make the sampled values of system variables at times $t_0, t_1, \ldots, t_k, \ldots$ be an acceptably accurate representation of the corresponding continuous signals. Several methods of obtaining a discrete Z-transfer function from a continuous Laplace transfer function have been discussed in Chapter 2. See Problems 2.17 through 2.21 or the references on sampled-data control systems. The approximation of differential equations by difference

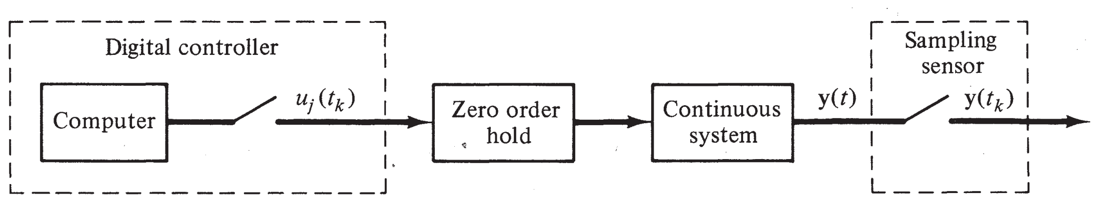


Figure 9.2

equations—or, equivalently, the various numerical integration techniques—is a standard topic in texts on numerical methods. Thus it is assumed that the top-level left-to-right transition in our modeling paradigm is understood. Chapter 3 gave details of finding state variable models from transfer functions and differential or difference equations. Thus both the continuous and sampled versions of the vertical transitions in Figure 9.1 have been explained. The last link in Figure 9.1, the lower-level horizontal transition from continuous-time to discrete-time state equations, is now discussed. Consider the system

$$\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \quad \text{with } \mathbf{x}(t_0) \text{ known}$$

Assume that $\{t_0, t_1, \ldots, t_k, \ldots\}$ is a set of discrete time points sufficiently close together so that during any interval $[t_k, t_{k+1}]$ the input vector $\mathbf{u}(t)$ can be approximated by $\mathbf{u}(t_k)$. Note that if the inputs are processed through a zero-order hold as part of a digital controller, then they are automatically constant over the sampling interval. Equation (*9.20*) can be used to write the solution at t_{k+1} by treating $\mathbf{x}(t_k)$ as the initial condition.

$$\mathbf{x}(t_{k+1}) = \mathbf{\Phi}(t_{k+1}, t_k)\mathbf{x}(t_k) + \int_{t_k}^{t_{k+1}} \mathbf{\Phi}(t_{k+1}, \tau)\mathbf{B}(\tau)\, d\tau\, \mathbf{u}(t_k) \tag{9.21}$$

Equation (*9.21*) is an approximating difference equation for the states. Note that the input $\mathbf{u}(\tau)$ has been replaced by $\mathbf{u}(t_k)$ and taken out of the integral sign because of its piecewise constant behavior. If $\mathbf{A}(t)$ and/or $\mathbf{B}(t)$ are also approximately constant over $[t_k, t_{k+1}]$, further simplifications can be made. For example, if $\mathbf{A}$ is (approximately) constant, then $\mathbf{\Phi}(t_{k+1}, t_k) = e^{\mathbf{A}T}$, where $T = t_{k+1} - t_k$ and where $\mathbf{A}$ is the value of $\mathbf{A}(t_k)$. If $\mathbf{B}$ is (approximately) constant over $[t_{k+1}, t_k]$, then it can be removed from the integral sign. This leads to a commonly used approximation for the discrete state equations,

$$\mathbf{x}(k+1) = \mathbf{A}_1\mathbf{x}(k) + \mathbf{B}_1\mathbf{u}(k) \tag{9.22}$$

where $\mathbf{A}_1$ and $\mathbf{B}_1$ are used in the discrete model to distinguish them from the continuous model matrices $\mathbf{A}$ and $\mathbf{B}$. The relationships are

$$\mathbf{A}_1 = e^{\mathbf{A}T} \quad \text{and} \quad \mathbf{B}_1 = \int e^{\mathbf{A}(t_{k+1} - \tau)}\, d\tau\, \mathbf{B}$$

Even though these results have been referred to as discrete *approximations,* they are exact for constant-coefficient systems whose inputs pass through a zero-order hold, as is common in digital controllers. A further analytical simplification can be made in the special case where $\mathbf{A}(t_k)^{-1}$ exists by using results from Chapter 8 for integrating the exponential matrix,

$$\mathbf{B}_1 = [\mathbf{A}_1(t_k) - \mathbf{I}]\mathbf{A}^{-1}(t_k)\mathbf{B}(t_k)$$

Although the sample times t_k are usually equally spaced, this is not required by Eq. (*9.21*). Reevaluation of $\mathbf{A}_1$ and $\mathbf{B}_1$ would be necessary for each cycle of Eq. (*9.22*) in the variable sample-rate case. Problem 9.10 gives an efficient algorithm for evaluating $\mathbf{A}_1$ and $\mathbf{B}_1$ using a truncated infinite series approximation.

EXAMPLE 9.3 Consider the second-order system $\ddot{y} + 3\dot{y} + 2y = u(t)$, which has the transfer function $y(s)/u(s) = 1/[s^2 + 3s + 2] = 1/[(s+1)(s+2)]$. Use this model as the starting point in the upper left corner of the paradigm of Figure 9.1. Obtain approximate discrete state models by going around both transition paths in Figure 9.1. Use $T = t_{k+1} - t_k = 0.2$ seconds. A continuous-state model is selected first. Recall from Chapter 3 that there are many different methods for picking states, and each method will give a different model. The controllable canonical form of the state equations is

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(t)$$

with $y(t) = x_1(t)$. Since $\mathbf{A}$ and $\mathbf{B}$ are constant and it is assumed that $u(t)$ will be piecewise constant over each sample period T, the discrete matrices $\mathbf{A}_1$ and $\mathbf{B}_1$ can be calculated as shown before. A truncated infinite series is used in the numerical evaluation (see Problem 9.10) and gives

$$\mathbf{x}(k+1) = \begin{bmatrix} 0.967141 & 0.148411 \\ -0.296821 & 0.521909 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 0.016429 \\ 0.148411 \end{bmatrix} u(k) \tag{9.23}$$

The sampled output equation is

$$y(k) = [1 \quad 0]\mathbf{x}(k)$$

To find the second form of the discrete model, a discrete Z-transfer function is found first. As discussed in Chapter 2, there are several ways of performing this step, such as approximating derivatives by forward or backward differences. In this example the exact conversion of the zero-order hold–continuous system combination gives

$$\begin{aligned} G(z) &= (1 - z^{-1})Z\{G(s)/s\} \\ &= (0.01643z + 0.013452)/(z^2 - 1.48905z + 0.548811) \end{aligned} \tag{9.24}$$

The denominator of this transfer function factors into $(z - e^{-T})(z - e^{-2T})$, where $T = 0.2$ has been used. There are many ways of picking a state model from this transfer function, as discussed in Chapter 3. Here the observable canonical form is used because then the discrete state x_1 will be the same physical variable as was the continuous state x_1. To see this, the simulation diagrams for the two models should be drawn. If this system model represents an armature-controlled dc motor, for example, x_1 is the motor shaft angle in both the preceding continuous- and discrete-state models and in the model from the current approach,

$$\begin{aligned} \begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} &= \begin{bmatrix} 1.48905 & 1 \\ -0.548811 & 0 \end{bmatrix} \begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 0.01643 \\ 0.013452 \end{bmatrix} u(k) \\ y(k) &= [1 \quad 0]\mathbf{x}(k) \end{aligned} \tag{9.25}$$

Equations (*9.23*) and (*9.25*) are both valid discrete models of the same system, with the same sampling rate and same assumptions, yet they are obviously different. Both models have the same input-output characteristics—namely, those described by the transfer function of Eq. (*9.24*). Their internal descriptions differ because different state variables were selected. Many other forms of the model could also be found. If a close correspondence with the continuous physical variables is desired, then the first procedure would be preferred. That is, if the continuous variable $x_2(t)$ represents the angular velocity of a motor shaft, then $x_2(k)$ of Eq. (*9.23*) is an approximation of that same variable. In Eq. (*9.25*) the meaning of x_2 is totally different and its time behavior, say for a step input, is completely different. ■

The sample rate used in developing a discrete model is often fixed by the sensor or controller cycle time. However, when a choice is still possible, such as early in the system design or in the case of digital simulation, a rough order of magnitude guide is useful. The piecewise constant approximations used before for **A** and **B** may be adequate if 6 to 10 samples occur per period of the highest-system modal frequency or per fastest time constant. The input frequency content also must be considered, and again the factor of 6 to 10 is a suggested starting range. Note that the theoretical lower limit on sample rate, the Nyquist rate of Sec. 2.5, is only two samples per period. Such slow sampling is never adequate in real systems. For a finer-tuned answer regarding sample rate, each case should be analyzed separately. The meaning of "acceptable" accuracy will be problem- and system-dependent. Furthermore, accuracy versus computer burden is a common design trade-off.

9.9 ANALYSIS OF CONSTANT COEFFICIENT DISCRETE-TIME STATE EQUATIONS

In this and subsequent sections the subscripts on the matrices $\mathbf{A}_1$ and $\mathbf{B}_1$ of the discrete state models will be dropped for convenience. It is assumed here that $\mathbf{A}(k)$ is constant, so the index k can be omitted. The homogeneous case is first considered:

$$\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k)$$

The initial conditions $\mathbf{x}(0)$ are assumed known, so that $\mathbf{x}(1) = \mathbf{A}\mathbf{x}(0)$. Using this in the difference equation gives $\mathbf{x}(2) = \mathbf{A}\mathbf{x}(1) = \mathbf{A}^2\mathbf{x}(0)$. Continuing this process, the solution at a general time t_k is expressed in terms of $\mathbf{x}(0)$ as

$$\mathbf{x}(k) = \mathbf{A}^k\mathbf{x}(0) \tag{9.26}$$

The methods of Chapter 8 can be used to determine $\mathbf{A}^k$ as a general function of k, so that repeated matrix multiplications are unnecessary.

The nonhomogeneous case is now considered. A sequence of input vectors $\mathbf{u}(0)$, $\mathbf{u}(1)$, $\mathbf{u}(2), \ldots$ is given, as well as the initial conditions $\mathbf{x}(0)$. Then

$$\mathbf{x}(1) = \mathbf{A}\mathbf{x}(0) + \mathbf{B}(0)\mathbf{u}(0)$$

$$\mathbf{x}(2) = \mathbf{A}\mathbf{x}(1) + \mathbf{B}(1)\mathbf{u}(1) = \mathbf{A}^2\mathbf{x}(0) + \mathbf{A}\mathbf{B}(0)\mathbf{u}(0) + \mathbf{B}(1)\mathbf{u}(1)$$

$$\mathbf{x}(3) = \mathbf{A}\mathbf{x}(2) + \mathbf{B}(2)\mathbf{u}(2) = \mathbf{A}^3\mathbf{x}(0) + \mathbf{A}^2\mathbf{B}(0)\mathbf{u}(0) + \mathbf{A}\mathbf{B}(1)\mathbf{u}(1) + \mathbf{B}(2)\mathbf{u}(2)$$

At a general time t_k, this leads to

$$\mathbf{x}(k) = \mathbf{A}^k\mathbf{x}(0) + \sum_{j=0}^{k-1} \mathbf{A}^{k-1-j}\mathbf{B}(j)\mathbf{u}(j) \tag{9.27}$$

A change in the dummy summation index allows this result to be written in the alternative form

$$\mathbf{x}(k) = \mathbf{A}^k\mathbf{x}(0) + \sum_{j=1}^{k} \mathbf{A}^{k-j}\mathbf{B}(j-1)\mathbf{u}(j-1) \tag{9.28}$$

Either of these forms may be used. The close analogy with the continuous-time system results is made more apparent by using the definition for the discrete system transition matrix. Whenever **A** is constant, the discrete transition matrix is given by

$$\mathbf{\Phi}(k, j) = \mathbf{A}^{k-j}$$

Then

$$\mathbf{x}(k) = \mathbf{\Phi}(k, 0)\mathbf{x}(0) + \sum_{j=1}^{k} \mathbf{\Phi}(k, j)\mathbf{B}(j-1)\mathbf{u}(j-1) \tag{9.29}$$

and the only difference from the continuous result is the replacement of the convolution integration by a discrete summation.

9.10 MODAL DECOMPOSITION

The system matrix $\mathbf{A}$ is still considered constant, and its eigenvalues and eigenvectors (or generalized eigenvectors) are λ_i and ξ_i, respectively. Then, if the change of basis $\mathbf{x}(k) = \mathbf{M}\mathbf{q}(k)$ is used, where $\mathbf{M} = [\xi_1 \quad \xi_2 \quad \cdots \quad \xi_n]$, the state equations reduce to

$$\mathbf{M}\mathbf{q}(k+1) = \mathbf{A}\mathbf{M}\mathbf{q}(k) + \mathbf{B}(k)\mathbf{u}(k)$$

or

$$\mathbf{q}(k+1) = \mathbf{J}\mathbf{q}(k) + \mathbf{B}_n(k)\mathbf{u}(k) \tag{9.30}$$

and

$$\mathbf{y}(k) = \mathbf{C}(k)\mathbf{M}\mathbf{q}(k) + \mathbf{D}(k)\mathbf{u}(k)$$

or

$$\mathbf{y}(k) = \mathbf{C}_n(k)\mathbf{q}(k) + \mathbf{D}(k)\mathbf{u}(k) \tag{9.31}$$

where $\mathbf{J} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M}$, $\mathbf{B}_n(k) = \mathbf{M}^{-1}\mathbf{B}(k)$, and $\mathbf{C}_n(k) = \mathbf{C}(k)\mathbf{M}$. Just as in the continuous case, the equations in $\mathbf{q}$ are as nearly uncoupled as possible and provide the same advantages. When $\mathbf{A}$ has a full set of eigenvectors, then $\mathbf{J}$ will be the diagonal matrix $\mathbf{\Lambda}$. A typical equation for the q_i components then takes the form

$$q_i(k+1) = \lambda_i q_i(k) + \langle \mathbf{r}_i, \mathbf{B}(k)\mathbf{u}(k)\rangle$$

The solution is

$$q_i(k) = \lambda_i^k q_i(0) + \sum_{j=1}^{k} \lambda_i^{k-j}\langle \mathbf{r}_i, \mathbf{B}(j-1)\mathbf{u}(j-1)\rangle$$

so that

$$\mathbf{q}(k) = \mathbf{\Lambda}^k\mathbf{q}(0) + \sum_{j=1}^{k} \mathbf{\Lambda}^{k-j}\mathbf{M}^{-1}\mathbf{B}(j-1)\mathbf{u}(j-1)$$

Using $\mathbf{x}(k) = \mathbf{M}\mathbf{q}(k)$ and $\mathbf{q}(0) = \mathbf{M}^{-1}\mathbf{x}(0)$ gives

$$\mathbf{x}(k) = \mathbf{M}\mathbf{\Lambda}^k\mathbf{M}^{-1}\mathbf{x}(0) + \sum_{j=1}^{k} \mathbf{M}\mathbf{\Lambda}^{k-j}\mathbf{M}^{-1}\mathbf{B}(j-1)\mathbf{u}(j-1)$$

This demonstrates again that $\mathbf{A}^k = \mathbf{M}\mathbf{\Lambda}^k\mathbf{M}^{-1}$ and provides a means of computing the transiton matrix, provided $\mathbf{A}$ has a full set of eigenvectors.

The modal decomposition technique provides geometrical insight into the system's structure. The behavior of $\mathbf{x}(k)$ versus the time index k can be represented as the vector sum of the eigenvectors ξ_i multiplied by the easily evaluated time-variable coefficients $q_i(k)$. That is,

$$\mathbf{x}(k) = \mathbf{Mq}(k) = \xi_1 q_1(k) + \xi_2 q_2(k) + \cdots + \xi_n q_n(k)$$

9.11 TIME-VARIABLE COEFFICIENTS

When $\mathbf{A}(k)$ is a time-variable matrix, then the solution technique of Sec. 9.9 must be modified slightly. Rather than the powers of $\mathbf{A}$, products of $\mathbf{A}$ evaluated at successive time points k are obtained. That is, the solution for $\mathbf{x}(k)$ at a general time t_k becomes

$$\mathbf{x}(k) = \mathbf{A}(k-1)\mathbf{A}(k-2)\cdots\mathbf{A}(0)\mathbf{x}(0) + \sum_{j=1}^{k}\left[\prod_{p=j}^{k-1}\mathbf{A}(p)\right]\mathbf{B}(j-1)\mathbf{u}(j-1) \quad (9.32)$$

In Eq. (9.32) the notation $\prod_{p=j}^{k-1}\mathbf{A}(p)$ indicates that the product $\mathbf{A}(k-1)\mathbf{A}(k-2)\ldots\mathbf{A}(j+1)\mathbf{A}(j)$. It is understood that if $j = k-1$, the product is just $\mathbf{A}(k-1)$ and if $j = k$, then $\prod_{p=k}^{k-1}\mathbf{A}(p) \triangleq \mathbf{I}_n$. The transition matrix for the time-varying case is given by

$$\mathbf{\Phi}(k, j) = \prod_{p=j}^{k-1}\mathbf{A}(p) \quad (9.33)$$

When this definition is used, the solution for the time-variable case, Eq. (9.32), is exactly that given in Eq. (9.29). Evaluation of the transition matrix is much more cumbersome for the time-variable case, however.

9.12 THE DISCRETE-TIME TRANSITION MATRIX

The discrete-time transition matrix has been defined and used in the previous sections. The principal properties of this important matrix are summarized here. For the most part, the same properties hold for both the continuous-time and discrete-time transition matrices. In particular, the transition matrix $\mathbf{\Phi}(k, j)$ represents the mapping of the state at time t_j into the state at time t_k provided the input sequence $\mathbf{u}$ is zero in that interval. It completely describes the unforced behavior of the state vector.

The semigroup property applies, that is, $\mathbf{\Phi}(k, m)\mathbf{\Phi}(m, j) = \mathbf{\Phi}(k, j)$ for any k, m, j satisfying $j \le m \le k$. The identity property holds, that is, $\mathbf{\Phi}(k, k) = \mathbf{I}_n$ for any time index k.

One major difference for the discrete-time transition matrix is that its inverse *need not* exist. When the inverse does exist, then the reversed time property holds:

$$\mathbf{\Phi}^{-1}(k, j) = \mathbf{\Phi}(j, k)$$

The inverse will exist if the discrete system is correctly derived as an approximation to a continuous system, since then $\mathbf{A}(k) = \mathbf{\Phi}(t_{k+1}, t_k)$ and $\mathbf{\Phi}(k, j) = \mathbf{\Phi}(t_k, t_j)$ and the continuous system transition matrix is always nonsingular.

In Problem 5.66, page 206, the formal adjoint for the discrete-time system operator was shown to be

$$\mathbf{w}(k-1) = \mathbf{A}^T(k)\mathbf{w}(k)$$

Notice the backward time indexing. If the transition matrix for this adjoint system is defined as $\mathbf{\Theta}(k, j)$, then many (but not all) of the relationships existing between $\mathbf{\Phi}$ and $\mathbf{\Theta}$ in the continuous case (Problems 9.16, 9.17, 9.19) will also be true in the discrete case. These properties are less useful in the discrete case and are not presented.

9.13 SUMMARY OF DISCRETE-TIME LINEAR SYSTEM SOLUTIONS

The most general solution for the linear discrete-time state Eq. (*9.22*) is given by Eq. (*9.29*), repeated here:

$$\mathbf{x}(k) = \mathbf{\Phi}(k, 0)\mathbf{x}(0) + \sum_{j=1}^{k} \mathbf{\Phi}(k, j)\mathbf{B}(j-1)\mathbf{u}(j-1) \tag{9.29}$$

The output is given by

$$\mathbf{y}(k) = \mathbf{C}(k)\mathbf{\Phi}(k, 0)\mathbf{x}(0) + \sum_{j=1}^{k} \mathbf{C}(k)\mathbf{\Phi}(k, j)\mathbf{B}(j-1)\mathbf{u}(j-1) + \mathbf{D}(k)\mathbf{u}(k)$$

When the discrete-time system matrix $\mathbf{A}$ is constant, then the transition matrix $\mathbf{\Phi}(k, j) = \mathbf{A}^{k-j}$ can be computed by any one of the several methods presented in Chapter 8. When $\mathbf{A}(k)$ is time-varying, no simple method exists for evaluating $\mathbf{\Phi}$ other than the direct calculation of the products indicated in Eq. (*9.33*). One hopes a digital computer would be available for this task.

REFERENCES

1. Coddington, E. A. and N. Levinson: *Theory of Ordinary Differential Equations,* McGraw-Hill, New York, 1955.
2. De Russo, P. M., R. J. Roy, and C. M. Close: *State Variables for Engineers,* John Wiley, New York, 1965.
3. Zadeh, L. A. and C. A. Desoer: *Linear System Theory, The State Space Approach,* McGraw-Hill, New York, 1963.
4. MacCamy, R. C. and V. J. Mizel: *Linear Analysis and Differential Equations,* MacMillan, New York, 1969.
5. Dorf, R. C.: *Modern Control Systems,* 5th ed., Addison-Wesley, Reading, Mass., 1989.

ILLUSTRATIVE PROBLEMS

Derivation and Verification of Solutions

9.1 Use Laplace transforms to solve $\dot{x} = ax(t) + b(t)u(t)$, with the initial condition $x(0)$, and a is constant.

Transforming gives

$$sx(s) - x(0) = ax(s) + \mathcal{L}\{b(t)u(t)\} \quad \text{or} \quad x(s) = \frac{x(0)}{s-a} + \frac{\mathcal{L}\{b(t)u(t)\}}{s-a}$$

The inverse transform gives

$$x(t) = \mathcal{L}^{-1}\{x(s)\} = x(0)e^{at} + \mathcal{L}^{-1}\left\{\frac{\mathcal{L}\{b(t)u(t)\}}{s-a}\right\}$$

Using the convolution theorem $\mathcal{L}^{-1}\{g_1(s)g_2(s)\} = \int_0^t g_1(t-\tau)g_2(\tau)\,d\tau$ on the last term gives

$$\mathcal{L}^{-1}\left\{\frac{\mathcal{L}\{b(t)u(t)\}}{s-a}\right\} = \int_0^t e^{a(t-\tau)}\,b(\tau)u(\tau)\,d\tau$$

so that

$$x(t) = e^{at}x(0) + \int_0^t e^{a(t-\tau)}\,b(\tau)u(\tau)\,d\tau$$

If b is also constant, the system is time-invariant. The solution due to any other initial condition $x(t_0)$ at time t_0 is

$$x(t) = e^{a(t-t_0)}x(t_0) + \int_{t_0}^t e^{a(t-\tau)}\,b(\tau)u(\tau)\,d\tau$$

9.2 Verify that Eq. (*9.3a*) is the solution of Eq. (*9.1*).

Verification requires showing that the postulated solution $x(t)$ satisfies the initial condition and the differential equation.

Initial condition check: With $t = t_0$,

$$x(t_0) = \{e^{\int_{t_0}^{t_0} a(\tau)\,d\tau}\}x(t_0) + \int_{t_0}^{t_0} e^{\int_\tau^{t_0} a(\zeta)\,d\zeta}\,b(\tau)u(\tau)\,d\tau$$

Since $a(t)$ is continuous, $\int_{t_0}^{t_0} a(\tau)\,d\tau = 0$ so $e^{\int_{t_0}^{t_0} a(\tau)\,d\tau} = 1$.

Likewise, $\int_{t_0}^{t_0} e^{\int a(\zeta)\,d\zeta}\,b(\tau)u(\tau)\,d\tau = 0$ provided that $b(t)$ and $u(t)$ remain finite. Therefore, $x(t_0) = x(t_0)$.

Differential equation check: Differentiating the postulated solution gives

$$\dot{x}(t) = \frac{d}{dt}\left[\int_{t_0}^t a(\tau)\,d\tau\right]e^{\int_{t_0}^t a(\tau)\,d\tau}x(t_0) + \frac{d}{dt}\left[\int_{t_0}^t e^{\int_\tau^t a(\zeta)\,d\zeta}\,b(\tau)u(\tau)\,d\tau\right]$$

Using the general formula for differentiating an integral term,

$$\frac{d}{dt}\left[\int_{f(t)}^{g(t)} h(t,\tau)\,d\tau\right] = \int_{f(t)}^{g(t)} \frac{\partial h(t,\tau)}{\partial t}\,d\tau + h(t, g(t))\frac{dg}{dt} - h(t, f(t))\frac{df}{dt}$$

gives

$$\frac{d}{dt}\left[\int_{t_0}^t a(\tau)\,d\tau\right] = a(t)$$

$$\frac{d}{dt}\left[\int_{t_0}^t e^{\int_\tau^t a(\zeta)\,d\zeta}\,b(\tau)u(\tau)\,d\tau\right] = e^{\int_t^t a(\zeta)\,d\zeta}\,b(t)u(t) + \int_{t_0}^t \frac{\partial}{\partial t}[e^{\int_\tau^t a(\zeta)\,d\zeta}]b(\tau)u(\tau)\,d\tau$$

$$= b(t)u(t) + a(t)\int_{t_0}^t e^{\int_\tau^t a(\zeta)\,d\zeta}\,b(\tau)u(\tau)\,d\tau$$

so that

$$\dot{x}(t) = a(t)\left[e^{\int_{t_0}^t a(\tau)\,d\tau}x(t_0) + \int_{t_0}^t e^{\int_\tau^t a(\zeta)\,d\zeta}\,b(\tau)u(\tau)\,d\tau\right] + b(t)u(t)$$

The term in brackets is the postulated solution for $x(t)$, so $\dot{x} = a(t)x(t) + b(t)u(t)$ and the equation is satisfied.

9.3 Solve $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}(t)\mathbf{u}(t)$ using Laplace transforms.

Let the vector $\mathbf{B}(t)\mathbf{u}(t) = \mathbf{f}(t)$ for convenience. $s\mathbf{x}(s) - \mathbf{x}(0) = \mathbf{A}\mathbf{x}(s) + \mathbf{f}(s)$ or $[s\mathbf{I} - \mathbf{A}]\mathbf{x}(s) = \mathbf{x}(0) + \mathbf{f}(s)$ so that $\mathbf{x}(s) = [s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{x}(0) + [s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{f}(s)$. Taking the inverse transform gives $\mathbf{x}(t) = \mathcal{L}^{-1}\{[s\mathbf{I} - \mathbf{A}]^{-1}\}\mathbf{x}(0) + \mathcal{L}^{-1}\{[s\mathbf{I} - \mathbf{A}]^{-1}\} * \mathbf{f}(t)$, where $g(t) * f(t)$ is used to indicate convolution. Since $\mathcal{L}^{-1}\{[s\mathbf{I} - \mathbf{A}]^{-1}\} = e^{\mathbf{A}t}$,

$$\mathbf{x}(t) = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{f}(\tau)\,d\tau = e^{\mathbf{A}t}\mathbf{x}(0) + \int_0^t e^{\mathbf{A}(t-\tau)}\mathbf{B}(\tau)\mathbf{u}(\tau)\,d\tau$$

9.4 Verify that Eq. (*9.7*) is the solution of Eq. (*9.6*).

Setting $t = t_0$ gives $\mathbf{x}(t_0) = e^{[0]}\mathbf{x}(t_0) = \mathbf{x}(t_0)$, assuming that $\mathbf{B}(t)\mathbf{u}(t)$ remains finite, i.e., contains no impulse functions.

Differentiating Eq. (*9.7*) gives

$$\dot{\mathbf{x}} = \mathbf{A}e^{\mathbf{A}(t-t_0)}\mathbf{x}(t_0) + \mathbf{B}(t)\mathbf{u}(t) + \mathbf{A}\int_{t_0}^t e^{\mathbf{A}(t-\tau)}\mathbf{B}(\tau)\mathbf{u}(\tau)\,d\tau$$

Using Eq. (*9.7*) reduces this to $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$, indicating that Eq. (*9.7*) does satisfy the differential equation.

9.5 Use a sequence of approximations for the solution of $\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x}(t)$, $\mathbf{x}(t_0)$ given, and derive an approximation for the transition matrix $\mathbf{\Phi}(t, t_0)$.

As the zeroth approximation, let $\mathbf{x}^{(0)}(t) = \mathbf{x}(t_0)$. Then use the differential equation to find the next approximation $\mathbf{x}^{(1)}(t)$ by solving $\dot{\mathbf{x}}^{(1)}(t) = \mathbf{A}(t)\mathbf{x}^{(0)}(t)$. The solution is

$$\mathbf{x}^{(1)}(t) = \mathbf{x}(t_0) + \int_{t_0}^t \dot{\mathbf{x}}^{(1)}(\tau)\,d\tau = \left[\mathbf{I} + \int_{t_0}^t \mathbf{A}(\tau_0)\,d\tau_0\right]\mathbf{x}(t_0)$$

Let $\dot{\mathbf{x}}^{(2)} = \mathbf{A}(t)\mathbf{x}^{(1)}(t)$. Then

$$\begin{aligned}\mathbf{x}^{(2)}(t) &= \mathbf{x}(t_0) + \int_{t_0}^t \dot{\mathbf{x}}^{(2)}(\tau)\,d\tau \\ &= \mathbf{x}(t_0) + \left[\int_{t_0}^t \mathbf{A}(\tau_0)\,d\tau_0 + \int_{t_0}^t \mathbf{A}(\tau_0)\int_{t_0}^{\tau_0}\mathbf{A}(\tau_1)\,d\tau_1\,d\tau_0\right]\mathbf{x}(t_0) \\ &= \left[\mathbf{I} + \int_{t_0}^t \mathbf{A}(\tau_0)\,d\tau_0 + \int_{t_0}^t \mathbf{A}(\tau_0)\int_{t_0}^{\tau_0}\mathbf{A}(\tau_1)\,d\tau_1\,d\tau_0\right]\mathbf{x}(t_0)\end{aligned}$$

Continuing this procedure with $\dot{\mathbf{x}}^{(k+1)}(t) = \mathbf{A}(t)\mathbf{x}^{(k)}(t)$ leads to

$$\begin{aligned}\mathbf{x}(t) \cong \Bigg[\mathbf{I} &+ \int_{t_0}^t \mathbf{A}(\tau_0)\,d\tau_0 + \int_{t_0}^t \mathbf{A}(\tau_0)\int_{t_0}^{\tau_0}\mathbf{A}(\tau_1)\,d\tau_1\,d\tau_0 \\ &+ \int_{t_0}^t \mathbf{A}(\tau_0)\int_{t_0}^{\tau_0}\mathbf{A}(\tau_1)\int_{t_0}^{\tau_1}\mathbf{A}(\tau_2)\,d\tau_2\,d\tau_1\,d\tau_0 + \cdots\Bigg]\mathbf{x}(t_0)\end{aligned}$$

Truncating the series in the brackets after a finite number of terms gives an approximation for $\mathbf{\Phi}(t, t_0)$.

Miscellaneous Applications

9.6 The satellite of Problem 3.16 is considered. If the two input torques are programmed to give $u_1(t) = (1/J_y)T_y(t) = C\sin\alpha t$, and $u_2(t) = (1/J_y)T_z(t) = C\cos\alpha t$, find the resultant time history of the state $\mathbf{x}(t) = [\omega_y \quad \omega_z]^T$. Use arbitrary initial conditions at time $t = 0$.

The state equations are $\dot{\mathbf{x}} = \begin{bmatrix} 0 & -\Omega \\ \Omega & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$. From Problem 8.17, the transition matrix is $\boldsymbol{\Phi}(t, 0) = \begin{bmatrix} \cos \Omega t & -\sin \Omega t \\ \sin \Omega t & \cos \Omega t \end{bmatrix}$ and

$$\mathbf{x}(t) = \boldsymbol{\Phi}(t, 0)\mathbf{x}(0) + \int_0^t \boldsymbol{\Phi}(t, \tau)\mathbf{u}(\tau)\, d\tau$$

The transition matrix properties can be used to write $\boldsymbol{\Phi}(t, \tau) = \boldsymbol{\Phi}(t, 0)\boldsymbol{\Phi}(0, \tau)$ and $\boldsymbol{\Phi}(0, \tau) = \boldsymbol{\Phi}(-\tau, 0)$, so

$$\mathbf{x}(t) = \boldsymbol{\Phi}(t, 0)\left\{\mathbf{x}(0) + C\int_0^t \begin{bmatrix} \cos \Omega\tau \sin \alpha\tau + \sin \Omega\tau \cos \alpha\tau \\ -\sin \Omega\tau \sin \alpha\tau + \cos \Omega\tau \cos \alpha\tau \end{bmatrix} d\tau\right\}$$

The trigonometric identities $\cos a \sin b + \sin a \cos b = \sin(a + b)$ and $\cos a \cos b - \sin a \sin b = \cos(a + b)$ are used inside the integral to give

$$\mathbf{x}(t) = \boldsymbol{\Phi}(t, 0)\left\{\mathbf{x}(0) + \frac{C}{\Omega + \alpha}\begin{bmatrix} 1 - \cos(\Omega + \alpha)t \\ \sin(\Omega + \alpha)t \end{bmatrix}\right\}$$

9.7 The input to the circuit of Figure 9.3 is an ideal current source $u(t)$. The output (and also the state) is the voltage across the capacitor $x(t)$. If

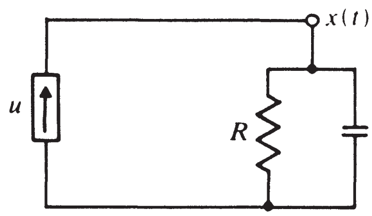


Figure 9.3

$$u(t) = e^{t/RC}\frac{10 - e^{-t_f/RC}x_0}{R \sinh(t_f/RC)}$$

and if $x(0) = x_0$, find the output $x(t_f)$ at some final time $t = t_f$.

The state equation is $\dot{x} = -x/RC + u/C$. The solution is

$$x(t) = e^{-t/RC}x_0 + \frac{1}{C}\int_0^t e^{-(t-\tau)/RC}u(\tau)\, d\tau$$

Letting $u(t) = Ke^{t/RC}$ for simplicity gives

$$x(t) = e^{-t/RC}x_0 + \frac{K}{C}e^{-t/RC}\int_0^t e^{2\tau/RC}\, d\tau$$

$$= e^{-t/RC}x_0 + \frac{K}{C}e^{-t/RC}\left\{\frac{RC}{2}[e^{2t/RC} - 1]\right\} = e^{-t/RC}x_0 + KR \sinh\left(\frac{t}{RC}\right)$$

Using $K = [10 - e^{-t_f/RC}x_0]/[R \sinh(t_f/RC)]$ and evaluating at $t = t_f$ gives $x(t_f) = 10$.

Although not proven here, the specified input $u(t)$ is the one which charges the capacitor from $x(0) = x_0$ to $x(t_f) = 10$ while minimizing the energy dissipated in R.

9.8 A system is described by $\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -2 & -2 & 0 \\ 0 & 0 & 1 \\ 0 & -3 & -4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$

(a) Find the change of variables $\mathbf{x} = \mathbf{Mq}$ which uncouples this system.
(b) If $\mathbf{x}(0) = [10 \quad 5 \quad 2]^T$ and if $\mathbf{u}(t) = [t \quad 1]^T$, find $\mathbf{x}(t)$.
(a) The modal matrix $\mathbf{M}$ must be found. $|\mathbf{A} - \mathbf{I}\lambda| = -(\lambda + 1)(\lambda + 2)(\lambda + 3)$, so the eigenvalues are $\lambda_i = -1, -2, -3$:

$$\text{Adj}[\mathbf{A}-\mathbf{I}\lambda] = \begin{bmatrix} \lambda^2+4\lambda+3 & -2(\lambda+4) & -2 \\ 0 & (\lambda+2)(\lambda+4) & 2+\lambda \\ 0 & -3(2+\lambda) & \lambda(2+\lambda) \end{bmatrix}$$

From this, the eigenvectors are $\xi_1 = [-2 \quad 1 \quad -1]^T$, $\xi_2 = [1 \quad 0 \quad 0]^T$, and $\xi_3 = [-2 \quad -1 \quad 3]^T$,

so that $\mathbf{x} = \begin{bmatrix} -2 & 1 & -2 \\ 1 & 0 & -1 \\ -1 & 0 & 3 \end{bmatrix}\mathbf{q}$ is the decoupling transformation. Using this substitution along

with $\mathbf{M}^{-1} = \begin{bmatrix} 0 & \frac{3}{2} & \frac{1}{2} \\ 1 & 4 & 2 \\ 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$ leads to

$$\dot{\mathbf{q}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -3 \end{bmatrix}\mathbf{q} + \begin{bmatrix} \frac{1}{2} & 2 \\ 3 & 6 \\ \frac{1}{2} & 1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

(b) The three uncoupled equations and their solutions are

$$\dot{q}_1 = -q_1 + \tfrac{1}{2}t + 2 \Rightarrow q_1(t) = e^{-t}q_1(0) + \tfrac{1}{2}t + \tfrac{3}{2}(1-e^{-t})$$

$$\dot{q}_2 = -2q_2 + 3t + 6 \Rightarrow q_2(t) = e^{-2t}q_2(0) + \tfrac{3}{2}t + \tfrac{9}{4}(1-e^{-2t})$$

$$\dot{q}_3 = -3q_3 + \tfrac{1}{2}t + 1 \Rightarrow q_3(t) = e^{-3t}q_3(0) + \tfrac{1}{6}t + \tfrac{5}{18}(1-e^{-3t})$$

Since $\mathbf{q}(0) = \mathbf{M}^{-1}\mathbf{x}(0) = [\frac{17}{2} \quad 34 \quad \frac{7}{2}]^T$, and since $\mathbf{x}(t) = \mathbf{M}\mathbf{q}(t)$, the solution is

$$\mathbf{x}(t) = \begin{bmatrix} -14e^{-t} + (\frac{127}{4})e^{-2t} - (\frac{58}{9})e^{-3t} + (\frac{1}{6})t - \frac{47}{36} \\ 7e^{-t} - (\frac{29}{9})e^{-3t} + (\frac{1}{3})t + \frac{11}{9} \\ -7e^{-t} + (\frac{29}{3})e^{-3t} - \frac{2}{3} \end{bmatrix}$$

9.9 The motor-generator system of Problem 3.13, page 114, has been driving the load at a constant speed $\Omega = 100$ rad/sec for some time. At time $t = 0$ the input voltage $e_f(t)$ is suddenly removed, that is, $e_f(t) = 0$ for $t \geq 0$. Find the resulting motion of the system. Assume the linear relation $e_g = K_g i_f$ and use the parameter values $b/J = 1$, $K_m/J = 2$, $K_m/(L_g + L_m) = 2.5$, $(R_g + R_m)/(L_g + L_m) = 7$, $K_g/(L_g + L_m) = 4$, $R_f/L_f = 5$, $L_f = 1$.

The state equations are $\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -1 & 2 & 0 \\ -2.5 & -7 & 4 \\ 0 & 0 & -5 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}u(t)$.

The desired solution is $\mathbf{x}(t) = \mathbf{\Phi}(t, 0)\mathbf{x}(0) = e^{\mathbf{A}t}\mathbf{x}(0)$. The initial value of $\Omega = x_1(0)$ is 100. The initial values of the other state variables can be determined from the fact that the system was initially in steady-state, $\dot{\Omega} = 0$ and $T(0) = b\Omega(0) = K_m i_m(0)$. From this, $x_2(0) = i_m(0) = (b/K_m)\Omega(0) = (b/J)(J/K_m)\Omega(0) = 50$. Also $di_m/dt = 0$ for $t \leq 0$, so $e_g - e_m = (R_g + R_m)i_m$ at $t = 0$. Therefore, $e_g(0) = e_m(0) + (R_g + R_m)i_m(0)$ and $i_f(0) = x_3(0) = e_g(0)/K_g$. $x_3(0) = [K_m\Omega(0) + (R_g + R_m)i_m(0)]/K_g = 2.5(100)/4 + 7(50)/4 = 150$ or $\mathbf{x}(0) = [100 \quad 50 \quad 150]^T$.
To find $\mathbf{\Phi}(t, 0)$, the eigenvalues of $\mathbf{A}$ are found.

$$|\mathbf{A}-\mathbf{I}\lambda| = (-\lambda-5)\begin{vmatrix} -\lambda-1 & 2 \\ -2.5 & -\lambda-7 \end{vmatrix} = (-\lambda-5)(\lambda+2)(\lambda+6);$$

$$\lambda_1 = -2, \ \lambda_2 = -5, \ \lambda_3 = -6$$

Using the Cayley-Hamilton remainder technique,

$$e^{\mathbf{A}t} = \alpha_0\mathbf{I} + \alpha_1\mathbf{A} + \alpha_2\mathbf{A}^2 = \begin{bmatrix} \alpha_0 - \alpha_1 - 4\alpha_2 & 2\alpha_1 - 16\alpha_2 & 8\alpha_2 \\ -2.5\alpha_1 + 20\alpha_2 & \alpha_0 - 7\alpha_1 + 44\alpha_2 & 4\alpha_1 - 48\alpha_2 \\ 0 & 0 & \alpha_0 - 5\alpha_1 + 25\alpha_2 \end{bmatrix}$$

where

$$\left.\begin{array}{l} e^{-2t} = \alpha_0 - 2\alpha_1 + 4\alpha_2 \\ e^{-5t} = \alpha_0 - 5\alpha_1 + 25\alpha_2 \\ e^{-6t} = \alpha_0 - 6\alpha_1 + 36\alpha_2 \end{array}\right\} \Rightarrow \left\{\begin{array}{l} \alpha_0 = (5/2)e^{-2t} - 4e^{-5t} + (5/2)e^{-6t} \\ \alpha_1 = (11/12)e^{-2t} - (8/3)e^{-5t} + (7/4)e^{-6t} \\ \alpha_2 = (1/12)e^{-2t} - (1/3)e^{-5t} + (1/4)e^{-6t} \end{array}\right.$$

Using these gives

$$\mathbf{\Phi}(t,0) = \left[\begin{array}{c|c|c} \frac{5}{4}e^{-2t} - \frac{1}{4}e^{-6t} & \frac{1}{2}e^{-2t} - \frac{1}{2}e^{-6t} & \frac{2}{3}e^{-2t} - \frac{8}{3}e^{-5t} + 2e^{-6t} \\ -\frac{5}{8}e^{-2t} + \frac{5}{8}e^{-6t} & -\frac{1}{4}e^{-2t} + \frac{5}{4}e^{-6t} & -\frac{1}{3}e^{-2t} + \frac{16}{3}e^{-5t} - 5e^{-6t} \\ 0 & 0 & e^{-5t} \end{array}\right]$$

Then $\mathbf{x}(t) = \mathbf{\Phi}(t,0)\mathbf{x}(0)$; and since $\mathbf{\Omega}(t) = y = [1 \quad 0 \quad 0]\mathbf{x}(t)$, $\mathbf{\Omega}(t) = 250e^{-2t} - 400e^{-5t} + 250e^{-6t}$.

9.10 A system is described by $\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$, with $\mathbf{A}$ and $\mathbf{B}$ constant. Develop an efficient computational procedure for finding $\mathbf{x}(T)$, assuming $\mathbf{u}(t)$ is constant over $[0, T]$.

Setting $t_{k+1} = T$ and $t_k = 0$ in Eq. (9.21) gives the form of $\mathbf{x}(T)$

$$\mathbf{x}(t) = \mathbf{\Phi}(T, 0)\mathbf{x}(0) + \int_0^T \mathbf{\Phi}(T, \tau)\, d\tau\, \mathbf{Bu}$$

It is known that $\mathbf{\Phi}(T, 0) = e^{\mathbf{A}T}$ and the series form is

$$\begin{aligned} \mathbf{\Phi}(T, 0) &= \mathbf{I} + \mathbf{A}T + (\mathbf{A}T)^2/2 + (\mathbf{A}T)^3/3! + (\mathbf{A}T)^4/4! + \cdots \\ &= \mathbf{I} + \mathbf{A}T\{\mathbf{I} + \mathbf{A}T/2 + (\mathbf{A}T)^2/3! + (\mathbf{A}T)^3/4! + \cdots\} \\ &\vdots \\ &= \mathbf{I} + \mathbf{A}T\{\mathbf{I} + \mathbf{A}T/2[\mathbf{I} + \mathbf{A}T/3(\mathbf{I} + \mathbf{A}T/4(\mathbf{I} + \cdots (\mathbf{I} + \mathbf{A}T/N)))]\} \end{aligned}$$

This nested form for $\mathbf{\Phi}(T, 0)$ does not require the direct computation of increasingly high powers of $\mathbf{A}T$ and therefore avoids many overflow and underflow problems. How many terms need to be retained depends upon $|\lambda_{max}|T$, where $|\lambda_{max}|$ is the largest magnitude eigenvalue of $\mathbf{A}$. This test is not normally used, however. On the Nth step in the nested sequence, the first neglected term would be $(\mathbf{A}T)^2/[(N)(N+1)]$, and this should be acceptably small compared with $\mathbf{A}T/N$.

The previous problem gave a result for $\mathbf{B}_1$ which depends on the existence of $\mathbf{A}^{-1}$. That is too restrictive in many cases, so another form which is better suited to machine computation is sought. Clearly,

$$\mathbf{B}_1 = \int_0^T e^{\mathbf{A}(T-\tau)}\, d\tau\, \mathbf{B} = -\int_T^0 e^{\mathbf{A}\xi}\, d\xi\, \mathbf{B} = \int_0^T e^{\mathbf{A}\xi}\, d\xi\, \mathbf{B}$$

Direct term-by-term integration of the exponential matrix gives

$$\begin{aligned} \mathbf{B}_1 &= \{\mathbf{I}T + \mathbf{A}T^2/2 + \mathbf{A}^2\mathbf{T}^3/3! + \cdots\}\mathbf{B} \\ &= T\{\mathbf{I} + \mathbf{A}T/2 + (\mathbf{A}T)^2/3! + \cdots\}\mathbf{B} \end{aligned}$$

Note that the series inside { } is the same as the one which appeared in the calculation of $\mathbf{\Phi}(T, 0)$. Therefore, the same nested form is possible. Define this part of the solution as $\mathbf{\Psi}$. That is,

$$\mathbf{\Psi} = \mathbf{I} + \mathbf{A}T/2[\mathbf{I} + \mathbf{A}T/3(\mathbf{I} + \mathbf{A}T/4(\mathbf{I} + \ldots (\mathbf{I} + \mathbf{A}T/N)))]$$

This partial result is then used to obtain the desired approximations

$$\mathbf{B}_1 = T\mathbf{\Psi}\mathbf{B}$$

$$\mathbf{\Phi} = \mathbf{I} + \mathbf{A}T\mathbf{\Psi}$$

These are widely used in obtaining discrete approximations to continuous-time systems.

9.11 A second-order system is described by $\ddot{x} + 2\dot{x} + 4x = u(t)$. Using $x = x_1$ and $x_2 = \dot{x}$ as states, find the state equations and evaluate the exact transition matrix $\mathbf{\Phi}(T, 0)$ and input matrix $\mathbf{B}_1$ using the results of Sec. 9.8 and Chapter 8 with $T = 0.2$. Then use results of Problem 9.10 to obtain approximate numerical results. Compare these.

The state equation is

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -4 & -2 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 1 \end{bmatrix}u$$

The exact state transition matrix is found to be

$$\mathbf{\Phi}(T, 0) = \begin{bmatrix} C+S & S \\ -4S & C-S \end{bmatrix}, \quad \text{where } C = e^{-T}\cos(\sqrt{3}T) \text{ and } S = e^{-T}\sin(\sqrt{3}T)/\sqrt{3}$$

Using $T = 0.2$ gives $\mathbf{\Phi} = \begin{bmatrix} 0.9306 & 0.1605 \\ -0.6420 & 0.6096 \end{bmatrix}$. Since $\mathbf{A}$ is nonsingular,

$$\mathbf{B}_1 = \mathbf{A}^{-1}[\mathbf{\Phi}(T, 0) - \mathbf{I}]\mathbf{B} = \begin{bmatrix} 0.017 \\ 0.160 \end{bmatrix}$$

The truncated series defined as $\mathbf{\Psi}$ in Problem 9.10 is now used. Note that for an Nth order approximation in T for $\mathbf{\Phi}$ and $\mathbf{B}_1$, an $(N-1)$st order approximation in $\mathbf{\Psi}$ is used.

Highest power of T	$\mathbf{\Psi}$	$\mathbf{\Phi}(T, 0)$	$\mathbf{B}_1$
1	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.2 \\ -0.8 & 0.6 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0.2 \end{bmatrix}$
2	$\begin{bmatrix} 1 & 0.1 \\ -0.4 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0.92 & 0.16 \\ -0.64 & 0.60 \end{bmatrix}$	$\begin{bmatrix} 0.02 \\ 0.16 \end{bmatrix}$
3	$\begin{bmatrix} 0.9733 & 0.0867 \\ -0.3467 & 0.8 \end{bmatrix}$	$\begin{bmatrix} 0.931 & 0.160 \\ -0.64 & 0.611 \end{bmatrix}$	$\begin{bmatrix} 0.017 \\ 0.160 \end{bmatrix}$

Depending on the application, the second- or third-order approximation may suffice. The first-order approximation probably would not, because very large differences between $(1)^k$ and $(0.93)^k$ will quickly appear in $\mathbf{\Phi}(k, 0)$ as the approximate difference equations are solved over k time steps.

9.12 Find a discrete-time approximate model for the system of Figure 9.4. Use $t_{k+1} - t_k = \Delta t = 1$ and approximate u_1 and u_2 as piecewise constant functions.

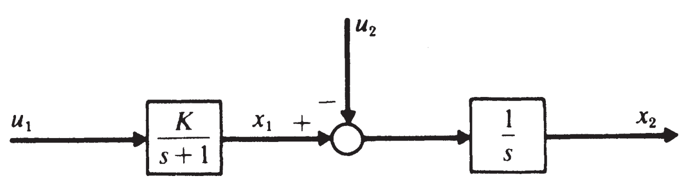


Figure 9.4

The continuous-time state equations are

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} K & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

and the transition matrix is

$$\mathbf{\Phi}(t, 0) = e^{\mathbf{A}t} = \mathscr{L}^{-1}\{[s\mathbf{I} - \mathbf{A}]^{-1}\}, \qquad \mathbf{\Phi}(s) = \frac{\begin{bmatrix} s & 0 \\ 1 & s+1 \end{bmatrix}}{s(s+1)}$$

or

$$\mathbf{\Phi}(t, 0) = \begin{bmatrix} e^{-t} & 0 \\ (1 - e^{-t}) & 1 \end{bmatrix}$$

The state at time $t_{k+1} = t_k + \Delta t$ can be written as

$$\mathbf{x}(t_{k+1}) = \mathbf{\Phi}(t_{k+1}, t_k)\mathbf{x}(t_k) + \int_{t_k}^{t_{k+1}} \mathbf{\Phi}(t_{k+1}, \tau)\, d\tau \begin{bmatrix} K & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} u_1(t_k) \\ u_2(t_k) \end{bmatrix}$$

But

$$\mathbf{\Phi}(t_{k+1}, t_k) = \mathbf{\Phi}(t_{k+1} - t_k, 0) = \begin{bmatrix} e^{-1} & 0 \\ 1 - e^{-1} & 1 \end{bmatrix} = \begin{bmatrix} 0.368 & 0 \\ 0.632 & 1 \end{bmatrix}$$

and

$$\int_{t_k}^{t_{k+1}} \mathbf{\Phi}(t_{k+1}, \tau)\, d\tau = \begin{bmatrix} 1 - e^{-1} & 0 \\ e^{-1} & 1 \end{bmatrix} = \begin{bmatrix} 0.632 & 0 \\ 0.368 & 1 \end{bmatrix}$$

The approximating difference equation is

$$\begin{bmatrix} x_1(t_{k+1}) \\ x_2(t_{k+1}) \end{bmatrix} = \begin{bmatrix} 0.368 & 0 \\ 0.632 & 1 \end{bmatrix} \begin{bmatrix} x_1(t_k) \\ x_2(t_k) \end{bmatrix} + \begin{bmatrix} 0.632K & 0 \\ 0.368K & -1 \end{bmatrix} \begin{bmatrix} u_1(t_k) \\ u_2(t_k) \end{bmatrix}$$

9.13 The system of Figure 9.4 represents a simple model of a production and inventory control system. The input $u_1(t)$ represents the scheduled production rate, $x_1(t)$ represents the actual production rate, $u_2(t)$ represents the sales rate, and $x_2(t)$ represents the current inventory level. Suppose that the production schedule is selected as $u_1(t) = c - x_2(t)$, where c is the desired inventory level. This is a feedback control policy. The system is originally in equilibrium with $x_1(0)$ equal to the sales rate and $x_2(0) = c$. At time $t = 0$ the sales rate suddenly increases by 10%. That is, $u_2(t) = 1.1x_1(0)$ for $t \geq 0$. Find the resulting system response. Use $K = \frac{3}{16}$.

The simulation diagram for the feedback system is shown in Figure 9.5.

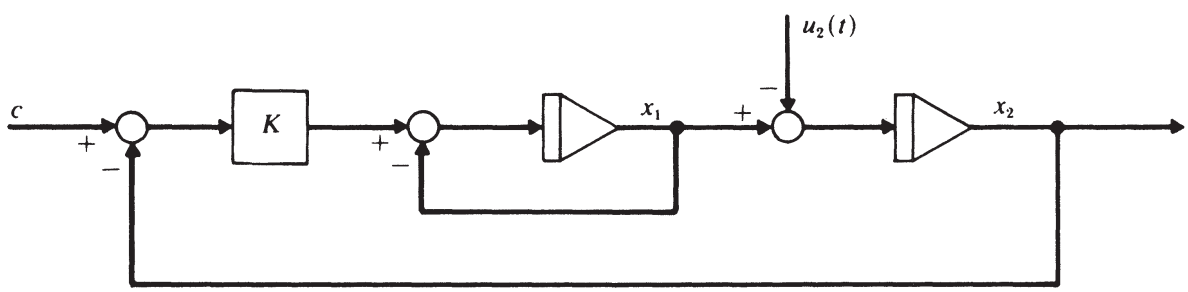


Figure 9.5

The state equations are $\dot{\mathbf{x}} = \begin{bmatrix} -1 & -K \\ 1 & 0 \end{bmatrix}\mathbf{x} + \begin{bmatrix} K & 0 \\ 0 & -1 \end{bmatrix}\begin{bmatrix} c \\ u_2 \end{bmatrix}$. The eigenvalues are determined from $|\mathbf{A} - \mathbf{I}\lambda| = \lambda^2 + \lambda + K = 0$ so that, with $K = \frac{3}{16}$, $\lambda_1 = -\frac{1}{4}$ and $\lambda_2 = -\frac{3}{4}$. The transition matrix is

$$\mathbf{\Phi}(t, 0) = e^{\mathbf{A}t} = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} = \begin{bmatrix} \alpha_0 - \alpha_1 & -3\alpha_1/16 \\ \alpha_1 & \alpha_0 \end{bmatrix}$$

Using the eigenvalues to solve for α_0 and α_1 gives

$$\mathbf{\Phi}(t, 0) = \begin{bmatrix} -\frac{1}{2}e^{-t/4} + \frac{3}{2}e^{-3t/4} & -\frac{3}{8}(e^{-t/4} - e^{-3t/4}) \\ 2(e^{-t/4} - e^{-3t/4}) & \frac{3}{2}e^{-t/4} - \frac{1}{2}e^{-3t/4} \end{bmatrix}$$

The solution is

$$\mathbf{x}(t) = \mathbf{\Phi}(t, 0)\begin{bmatrix} x_1(0) \\ c \end{bmatrix} + \int_0^t \mathbf{\Phi}(t, \tau)\, d\tau\, \mathbf{B}\begin{bmatrix} c \\ 1.1x_1(0) \end{bmatrix}$$

Using $\mathbf{\Phi}(t, \tau) = \mathbf{\Phi}(t - \tau, 0)$, carrying out the integration, and simplifying give

$$x_1(t) = x_1(0)\{1.1 - (4.3/2)e^{-t/4} + (4.1/2)e^{-3t/4}\}$$

$$x_2(t) = c + x_1(0)\{-17.6/3 + 8.6e^{-t/4} - (8.2/3)e^{-3t/4}\}$$

Additional Properties of the Transition Matrix

9.14 Assume that the eigenvectors of the constant system matrix **A** form a basis and show that $\mathbf{\Phi}(t, t_0) = \sum_{i=1}^{n} e^{\lambda_i(t-t_0)}\, \boldsymbol{\xi}_i\rangle\langle \mathbf{r}_i$, where λ_i, $\boldsymbol{\xi}_i$, and $\mathbf{r}_i$ are eigenvalues, eigenvectors, and reciprocal basis vectors of **A**, respectively.

The modal decomposition developed in Sec. 9.4 led to the expression

$$\mathbf{x}(t) = q_1(t)\boldsymbol{\xi}_1 + q_2(t)\boldsymbol{\xi}_2 + \cdots + q_n(t)\boldsymbol{\xi}_n = \sum_{i=1}^{n} q_i(t)\boldsymbol{\xi}_i$$

For the homogeneous system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$,

$$q_i(t) = e^{\lambda_i(t-t_0)}\, q_i(t_0) = e^{\lambda_i(t-t_0)}\langle \mathbf{r}_i, \mathbf{x}(t_0)\rangle$$

so that

$$\mathbf{x}(t) = \sum_{i=1}^{n} e^{\lambda_i(t-t_0)}\langle \mathbf{r}_i, \mathbf{x}(t_0)\rangle\boldsymbol{\xi}_i = \left[\sum_{i=1}^{n} e^{\lambda_i(t-t_0)}\, \boldsymbol{\xi}_i\rangle\langle \mathbf{r}_i\right]\mathbf{x}(t_0)$$

Comparing this with the known solution $\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\mathbf{x}(t_0)$ gives the desired result.

9.15 For fixed times t_0 and t, the transition matrix is a transformation of the state space Σ onto itself. A linear transformation which possesses a full set of n linearly independent eigenvectors has a spectral representation

$$\mathbf{\Phi}(t, t_0) = \sum_{i=1}^{n} \gamma_i\, \boldsymbol{\eta}_i\rangle\langle \mathbf{v}_i$$

where γ_i, $\boldsymbol{\eta}_i$ are the eigenvalues and eigenvectors of $\mathbf{\Phi}(t, t_0)$ and $\mathbf{v}_i$ are reciprocal to $\boldsymbol{\eta}_i$. This is the result of Eq. (7.5), page 263, with notational changes. Comparing this with the previous problem, draw conclusions about the relationships between eigenvalues and eigenvectors of **A** and of $\mathbf{\Phi}(t, t_0)$.

The indicated comparison suggests the following relationship between eigenvalues, a result known as Frobenius' theorem. If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are eigenvalues of the $n \times n$ matrix **A**, and if $f(x)$ is a function which is analytic inside a circle in the complex plane which contains all the λ_i, then $f(\lambda_1), f(\lambda_2), \ldots, f(\lambda_n)$ are the eigenvalues of the matrix function $f(\mathbf{A})$. In the present case the eigenvalues of $\mathbf{\Phi}(t, t_0) = e^{\mathbf{A}(t-t_0)}$ are $\gamma_i = e^{\lambda_i(t-t_0)}$. Furthermore, it can be verified that the eigenvectors of **A** and of $\mathbf{\Phi}(t, t_0)$ are the same, that is, $\boldsymbol{\xi}_i = \boldsymbol{\eta}_i$.

The Adjoint Equations

9.16 The formal adjoint of the differential equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is $\dot{\mathbf{y}} = -\mathbf{A}^T\mathbf{y}$ (see Problem 5.30). Let the transition matrix for the adjoint equation be $\mathbf{\Theta}(t, \tau)$. Show that $\mathbf{\Theta}(t, \tau) = \mathbf{\Phi}^T(\tau, t)$, where $\mathbf{\Phi}(t, \tau)$ is the transition matrix for the equation in **x**.

Since $\boldsymbol{\Theta}(t, \tau)$ is the transition matrix, it satisfies

$$\frac{d}{dt}[\boldsymbol{\Theta}(t, \tau)] = -\mathbf{A}^T\boldsymbol{\Theta}(t, \tau), \qquad \boldsymbol{\Theta}(\tau, \tau) = \mathbf{I}$$

The desired result is established by showing that $\boldsymbol{\Phi}^T(\tau, t)$ satisfies the same differential equation and initial conditions, since these equations define a unique solution. Since $\boldsymbol{\Phi}(\tau, t) = \boldsymbol{\Phi}^{-1}(t, \tau)$,

$$\frac{d}{dt}[\boldsymbol{\Phi}(\tau, t)] = \frac{d}{dt}[\boldsymbol{\Phi}^{-1}(t, \tau)] = -\boldsymbol{\Phi}^{-1}(t, \tau)\frac{d}{dt}[\boldsymbol{\Phi}(t, \tau)]\boldsymbol{\Phi}^{-1}(t, \tau)$$

But $\frac{d}{dt}[\boldsymbol{\Phi}(t, \tau)] = \mathbf{A}\boldsymbol{\Phi}(t, \tau)$. Therefore, $\frac{d}{dt}[\boldsymbol{\Phi}(\tau, t) = -\boldsymbol{\Phi}^{-1}(t, \tau)\mathbf{A} = -\boldsymbol{\Phi}(\tau, t)\mathbf{A}$. Transposing shows that $\frac{d}{dt}[\boldsymbol{\Phi}(\tau, t)]^T = -\mathbf{A}^T[\boldsymbol{\Phi}(\tau, t)]^T$. This, plus the fact that $\boldsymbol{\Phi}(\tau, \tau) = \mathbf{I}$, establishes that

$$\boldsymbol{\Theta}(t, \tau) = \boldsymbol{\Phi}^T(\tau, t)$$

It follows that the adjoint transition matrix can be expressed in terms of the fundamental matrix $\mathbf{U}(t)$ as $\boldsymbol{\Theta}^T(t, \tau) = \mathbf{U}(\tau)\mathbf{U}^{-1}(t)$.

9.17 Suppose a simulation of the system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ is available, as well as a simulation of the adjoint system $\dot{\mathbf{y}} = -\mathbf{A}^T\mathbf{y}$. Interpret the meaning of the vector functions $\mathbf{x}(t)$ and $\mathbf{y}(t)$ which are obtained if the initial conditions used are $\mathbf{x}(t_0) = [1 \quad 0 \quad 0 \quad \cdots \quad 0]^T$ and $\mathbf{y}(t_0) = [1 \quad 0 \quad 0 \quad \cdots \quad 0]^T$.

Since the solutions are $\mathbf{x}(t) = \boldsymbol{\Phi}(t, t_0)\mathbf{x}(t_0)$ and $\mathbf{y}(t) = \boldsymbol{\Theta}(t, t_0)\mathbf{y}(t_0)$, the first simulation generates the first column of $\boldsymbol{\Phi}(t, t_0)$ for all $t \geq t_0$ and the second generates the first column of $\boldsymbol{\Theta}(t, t_0)$ for all $t \geq t_0$. But the first column of $\boldsymbol{\Theta}(t, t_0)$ equals the first *row* of $\boldsymbol{\Phi}(t_0, t)$.

The adjoint system simulation provides a means of reversing the roles of t_0 and t. In a sense, a reversed time impulse response for the original system can be generated. The complete matrix $\boldsymbol{\Phi}(t_0, t)$ can be obtained, one row at a time, by modifying the initial conditions for the adjoint simulation. This property has several uses (see pages 379–394 of Reference 2).

State Equations as Linear Transformations

9.18 Consider the operator $\mathcal{A}(\mathbf{x}) \triangleq [\mathbf{I}(d/dt) - \mathbf{A}]\mathbf{x}$ as a transformation on infinite dimensional function spaces. Discuss the form of the solution for $\mathcal{A}(\mathbf{x}) = \mathbf{B}\mathbf{u}(t)$ given by Eq. (*9.20*) in terms of the results of Problem 6.22, page 240.

Problem 6.22 indicates that if $\mathbf{x}_1$ is a nonzero solution of the homogeneous equation $\mathcal{A}(\mathbf{x}) = \mathbf{0}$, then the most general solution of $\mathcal{A}(\mathbf{x}) = \mathbf{B}\mathbf{u}$ takes the form $\mathbf{x} + \mathbf{x}_1$, where $\mathbf{x}$ is a solution to the nonhomogeneous equation. This is precisely the form of Eq. (*9.20*), with $\mathbf{x}_1 = \boldsymbol{\Phi}(t, t_0)\mathbf{x}(t_0)$ being the homogeneous solution. A unique solution is not possible without specifying initial conditions.

9.19 Discuss the implications of Problem 6.23, page 240, in the context of linear state equations.

Problem 6.23 indicates that $\mathcal{A}(\mathbf{x}) = \mathbf{B}\mathbf{u}$ will have a solution for all $\mathbf{B}\mathbf{u}(t)$ if and only if the only solution to $\mathcal{A}^*(\mathbf{y}) = \mathbf{0}$ is the trivial solution. This poses an *apparent* contradiction, since nontrivial solutions to the adjoint equation have been discussed in Problem 9.17 and since solutions to $\mathcal{A}(\mathbf{x}) = \mathbf{B}\mathbf{u}(t)$ have been explicitly displayed in Eq. (*9.20*). The difficulty arises because of the differences between the adjoint transformation of Chapters 5 and 6 and the *formal* adjoint as used in this chapter.

A heuristic reconciliation is provided in a nonrigorous, formal manner. In Problem 5.30, page 199, it is shown that

$$\langle \mathbf{y}(t), \mathcal{A}(x(t))\rangle = \mathbf{y}^T(t_f)\mathbf{x}(t_f) - \mathbf{y}^T(t_0)\mathbf{x}(t_0) - \left\langle \frac{d\mathbf{y}}{dt} + \mathbf{A}^T\mathbf{y}, \mathbf{x}(t)\right\rangle$$

The *formal* adjoint was defined by dropping the two boundary terms. These two terms automatically cancel if $\mathbf{u}(t) = \mathbf{0}$ (see Problem 5.31). Generally, they are nonzero and can be included as follows:

$$\mathbf{y}^T(t_f)\mathbf{x}(t_f) - \mathbf{y}^T(t_0)\mathbf{x}(t_0) - \int_{t_0}^{t_f}\left[\frac{d\mathbf{y}^T(\tau)}{dt} + \mathbf{y}^T(\tau)\mathbf{A}\right]\mathbf{x}(\tau)\,d\tau$$

$$= \int_{t_0}^{t_f}\left\{-\frac{d\mathbf{y}^T(\tau)}{dt} - \mathbf{y}^T(\tau)\mathbf{A} + \mathbf{y}^T(t_f)\delta(t_f-\tau) - \mathbf{y}^T(t_0)\delta(\tau - t_0)\right\}\mathbf{x}(\tau)\,d\tau$$

The term in brackets is the transpose of the adjoint transformation, that is,

$$\mathscr{A}^*(\mathbf{y}) = -\frac{d\mathbf{y}}{dt} - \mathbf{A}^T\mathbf{y}(t) + \mathbf{y}(t_f)\delta(t_f - t) - \mathbf{y}(t_0)\delta(t - t_0) \tag{1}$$

Then treating the two impulse terms as forcing terms, the solution of Eq. (*1*) is

$$\mathbf{y}(t) = \mathbf{\Theta}(t, t_0)\mathbf{y}(t_0) + \int_{t_0}^{t}\mathbf{\Theta}(t, \tau)[\mathbf{y}(t_f)\delta(t_f-\tau) - \mathbf{y}(t_0)\delta(\tau - t_0)]\,d\tau$$

But

$$\int_{t_0}^{t}\mathbf{\Theta}(t, \tau)\mathbf{y}(t_f)\delta(t_f - \tau)\,d\tau = \mathbf{0} \qquad \text{for all } t < t_f$$

and

$$\int_{t_0}^{t}\mathbf{\Theta}(t, \tau)\mathbf{y}(t_0)\delta(\tau - t_0)\,d\tau = \mathbf{\Theta}(t, t_0)\mathbf{y}(t_0)$$

because of the sifting property of the impulse function. Therefore, the solution is $\mathbf{y}(t) = \mathbf{0}$ for all $t < t_f$. At the final time an identity is obtained, $\mathbf{y}(t_f) = \mathbf{y}(t_f)$. The solution $\mathbf{y}(t)$ is therefore zero for all t except possibly at the single time $t = t_f$. Such a function will be considered **0**, since its norm is zero (the Hilbert inner product norm, for example). Thus the only solution to $\mathscr{A}^*(\mathbf{y}) = \mathbf{0}$ is the trivial solution, and the results of Problem 6.23 are still true and do not lead to a contradiction.

Solution of Linear Discrete-Time State Equations

9.20 Solve for $\mathbf{x}(k)$ if

$$x_1(k+1) = \tfrac{1}{2}x_1(k) - \tfrac{1}{2}x_2(k) + x_3(k) \qquad x_3(k+1) = \tfrac{1}{2}x_3(k)$$

$$x_2(k+1) = \tfrac{1}{2}x_2(k) + 2x_3(k) \qquad x(0) = [2 \quad 4 \quad 6]^T$$

When put in matrix form $\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k)$, the system matrix is $\mathbf{A} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 1 \\ 0 & \frac{1}{2} & 2 \\ 0 & 0 & \frac{1}{2}\end{bmatrix}$. The solution for this homogeneous system is $\mathbf{x}(k) = \mathbf{A}^k\mathbf{x}(0)$. The matrix $\mathbf{A}^k$ was found in Problem 8.31, page 306. Using that result,

$$x_1(k) = 2(\tfrac{1}{2})^k - 4k(\tfrac{1}{2})^k + 6k(2-k)(\tfrac{1}{2})^{k-1}$$

$$x_2(k) = 4(\tfrac{1}{2})^k + 6k(\tfrac{1}{2})^{k-2}$$

$$x_3(k) = 6(\tfrac{1}{2})^k$$

9.21 Write the solution for the homogeneous discrete-time system

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1)\end{bmatrix} = \frac{1}{12}\begin{bmatrix} 5 & 1 \\ 1 & 5\end{bmatrix}\begin{bmatrix} x_1(k) \\ x_2(k)\end{bmatrix}, \qquad \mathbf{x}(0) = \begin{bmatrix} 2 \\ 1\end{bmatrix}$$

in the modal expansion form $\mathbf{x}(k) = \sum_{i=1}^{2}\langle \mathbf{r}_i, \mathbf{x}(0)\rangle\lambda_i^k\,\xi_i$.

The matrix $\mathbf{A} = \frac{1}{12}\begin{bmatrix} 5 & 1 \\ 1 & 5 \end{bmatrix}$ has as its eigenvalues $\lambda_1 = \frac{1}{2}, \lambda_2 = \frac{1}{3}$. The eigenvectors are $\xi_1 = [1 \quad 1]^T$ and $\xi_2 = [-1 \quad 1]^T$. Thus $\mathbf{M} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$ and $\mathbf{M}^{-1} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$.

The rows of $\mathbf{M}^{-1}$ give the reciprocal basis vectors $\mathbf{r}_1 = [\frac{1}{2} \quad \frac{1}{2}]^T$, $\mathbf{r}_2 = [-\frac{1}{2} \quad \frac{1}{2}]^T$. Since $\langle \mathbf{r}_1, \mathbf{x}(0)\rangle = \frac{3}{2}$, $\langle \mathbf{r}_2, \mathbf{x}(0)\rangle = -\frac{1}{2}$, the solution is

$$\mathbf{x}(k) = \frac{3}{2}\left(\frac{1}{2}\right)^k\begin{bmatrix} 1 \\ 1 \end{bmatrix} - \frac{1}{2}\left(\frac{1}{3}\right)^k\begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

9.22 Consider the discrete-time system

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{8} \\ \frac{1}{8} & \frac{1}{2} \end{bmatrix}\begin{bmatrix} x_1(k) \\ x_2(k) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix}$$

$$y(k) = x_1(k) + 2x_2(k)$$

Find $y(k)$ if $x_1(0) = -1, x_2(0) = 3$. The input $u_1(k)$ is obtained by sampling the ramp function t at times $t_0 = 0, t_1 = 1, \ldots, t_k = k$, and $u_2(k)$ is obtained by sampling e^{-t} at the same set of discrete times.

The transition matrix is first found:

$$\mathbf{A}^k = \alpha_0 \mathbf{I} + \alpha_1 \mathbf{A} = \begin{bmatrix} \alpha_0 + \frac{1}{2}\alpha_1 & \frac{1}{8}\alpha_1 \\ \frac{1}{8}\alpha_1 & \alpha_0 + \frac{1}{2}\alpha_1 \end{bmatrix}$$

The eigenvalues are required. $|\mathbf{A} - \mathbf{I}\lambda| = \lambda^2 - \lambda + \frac{15}{64}$ so $\lambda_1 = \frac{3}{8}, \lambda_2 = \frac{5}{8}$. Solving for α_0 and α_1,

$$\left.\begin{aligned} (\tfrac{3}{8})^k &= \alpha_0 + (\tfrac{3}{8})\alpha_1 \\ (\tfrac{5}{8})^k &= \alpha_0 + (\tfrac{5}{8})\alpha_1 \end{aligned}\right\} \Rightarrow \left\{\begin{aligned} 4[(\tfrac{5}{8})^k - (\tfrac{3}{8})^k] &= \alpha_1 \\ (\tfrac{5}{2})(\tfrac{3}{8})^k - (\tfrac{3}{2})(\tfrac{5}{8})^k &= \alpha_0 \end{aligned}\right.$$

Thus

$$\mathbf{A}^k = \mathbf{\Phi}(k, 0) = \begin{bmatrix} \frac{1}{2}[(\frac{5}{8})^k + (\frac{3}{8})^k] & \frac{1}{2}[(\frac{5}{8})^k - (\frac{3}{8})^k] \\ \frac{1}{2}[(\frac{5}{8})^k - (\frac{3}{8})^k] & \frac{1}{2}[(\frac{5}{8})^k + (\frac{3}{8})^k] \end{bmatrix}$$

Using this and the fact that $\mathbf{\Phi}(k, j) = \mathbf{A}^{k-j}$ gives

$$\mathbf{x}(k) = \begin{bmatrix} (\frac{5}{8})^k - 2(\frac{3}{8})^k \\ (\frac{5}{8})^k + 2(\frac{3}{8})^k \end{bmatrix} + \sum_{j=1}^{k}\begin{bmatrix} \frac{1}{2}[(\frac{5}{8})^{k-j} + (\frac{3}{8})^{k-j}] & \frac{1}{2}[(\frac{5}{8})^{k-j} - (\frac{3}{8})^{k-j}] \\ \frac{1}{2}[(\frac{5}{8})^{k-j} - (\frac{3}{8})^{k-j}] & \frac{1}{2}[(\frac{5}{8})^{k-j} + (\frac{3}{8})^{k-j}] \end{bmatrix}\begin{bmatrix} j-1 \\ e^{1-j} \end{bmatrix}$$

The output is $y(k) = [1 \quad 2]\mathbf{x}(k)$.

9.23 Express the following discrete-time state equations in normal form using the modal matrix in a change of basis:

$$\begin{bmatrix} x_1(k+1) \\ x_2(k+1) \\ x_3(k+1) \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ \frac{5}{6} & -\frac{13}{6} & -\frac{1}{3} \end{bmatrix}\begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix} + \begin{bmatrix} 3 & 1 \\ 2 & 0 \\ -1 & 1 \end{bmatrix}\begin{bmatrix} u_1(k) \\ u_2(k) \end{bmatrix}$$

$$\begin{bmatrix} y_1(k) \\ y_2(k) \end{bmatrix} = \begin{bmatrix} -1 & 3 & 1 \\ 0 & 1 & 1 \end{bmatrix}\begin{bmatrix} x_1(k) \\ x_2(k) \\ x_3(k) \end{bmatrix}$$

When the equations are expressed as in Eqs. (*9.30*) and (*9.31*), they are said to be in normal form. To put the equations in this form, the eigenvalues and eigenvectors must be determined:

$$|\mathbf{A} - \mathbf{I}\lambda| = (1 - \lambda)(\tfrac{1}{2} - \lambda)(-\tfrac{1}{3} - \lambda)$$

The eigenvectors are $\xi_1 = [1 \quad 1 \quad -1]^T$, $\xi_2 = [1 \quad 0 \quad 1]^T$, $\xi_3 = [0 \quad 0 \quad 1]^T$, so that

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ -1 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{M}^{-1} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -1 & 0 \\ -1 & 2 & 1 \end{bmatrix}$$

$$\mathbf{\Lambda} = \mathbf{M}^{-1}\mathbf{A}\mathbf{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & -\frac{1}{3} \end{bmatrix}, \qquad \mathbf{B}_n = \mathbf{M}^{-1}\mathbf{B} = \begin{bmatrix} 2 & 0 \\ 1 & 1 \\ 0 & 0 \end{bmatrix},$$

$$\mathbf{C}_n = \mathbf{C}\mathbf{M} = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

Collecting and using these results in $\mathbf{q}(k+1) = \mathbf{\Lambda}\mathbf{q}(k) + \mathbf{B}_n\mathbf{u}(k)$ and $\mathbf{y}(k) = \mathbf{C}_n\mathbf{q}(k)$ gives the normal form equations.

9.24 Find an expression for $\mathbf{y}(k)$, valid for all time t_k, for the system of Problem 9.23. Use $\mathbf{x}(0) = [1 \quad 2 \quad 1]^T$, $\mathbf{u}(k) = [k \quad -1-k]^T$.

Using the normal form equations,

$$\mathbf{q}(0) = \mathbf{M}^{-1}\mathbf{x}(0) = [2 \quad -1 \quad 4]^T$$

$$q_1(k) = q_1(0) + \sum_{j=1}^{k} 2u_1(j-1) = 2 + k(k-1)$$

$$q_2(k) = (\tfrac{1}{2})^k q_2(0) + \sum_{j=1}^{k} (\tfrac{1}{2})^{k-j}[u_1(j-1) + u_2(j-1)] = -(\tfrac{1}{2})^k\left[1 + \sum_{j=1}^{k} 2^j\right]$$

$$q_3(k) = (-\tfrac{1}{3})^k q_3(0) = 4(-\tfrac{1}{3})^k$$

Using the output equation $\mathbf{y}(k) = \mathbf{C}_n\mathbf{q}(k)$ yields

$$y_1(k) = q_1(k) + q_3(k) = 2 + k(k-1) + 4(-\tfrac{1}{3})^k$$

$$y_2(k) = q_2(k) + q_3(k) = (-\tfrac{1}{2})^k\left[1 + \sum_{j=1}^{k} 2^j\right] + 4(-\tfrac{1}{3})^k$$

9.25 Consider a homogeneous discrete-time system described by $\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k)$.

(a) Show that if a nontrivial steady-state (constant) solution is to exist, the matrix $\mathbf{A}$ must have unity as an eigenvalue.

(b) Construct a 2×2 nondiagonal, symmetric matrix with this property and find the steady-state solution.

(a) A constant steady-state solution implies that for k sufficiently large, $\mathbf{x}(k+1) = \mathbf{x}(k)$. Call this solution $\mathbf{x}_e$. Then the difference equation requires that $\mathbf{x}_e = \mathbf{A}\mathbf{x}_e$. But this is just the eigenvalue equation $\mathbf{A}\mathbf{x}_e = \lambda\mathbf{x}_e$, with $\lambda = 1$.

(b) Let $\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \end{bmatrix}$. Then

$$|\mathbf{A} - \mathbf{I}\lambda| = \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}^2) = 0$$

The roots are

$$\lambda_{1,2} = \tfrac{1}{2}(a_{11} + a_{22}) \pm \sqrt{[\tfrac{1}{2}(a_{11} + a_{22})]^2 - (a_{11}a_{22} - a_{12}^2)}$$

There are many possible solutions. One is obtained by arbitrarily setting $a_{11} = a_{22} = 2$. Then $\lambda_{1,2} = 2 \pm \sqrt{4 - 4 + a_{12}^2}$. If a root is to be $\lambda = 1$, then $a_{12} = 1$. Using $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, the steady-state solution $\mathbf{x}_e$ is just the eigenvector associated with the root $\lambda = 1$:

$$\text{Adj}[\mathbf{A} - \mathbf{I}\lambda]|_{\lambda=1} = \begin{bmatrix} 2-\lambda & -1 \\ -1 & 2-\lambda \end{bmatrix}\Bigg|_{\lambda=1} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

The steady-state solution will be proportional to $\mathbf{x}_e = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

9.26 Assume that a system is described by Eqs. (*3.13*) and (*3.14*) with **A, B, C,** and **D** constant. Use Z-transforms to find the transforms of $\mathbf{x}(k)$ and $\mathbf{y}(k)$. Then use the inverse Z-transform to find $\mathbf{x}(k)$. The initial condition $\mathbf{x}(0)$ is known and the input $\mathbf{u}(k)$ is zero for $k < 0$.

The Z-transform introduced in Chapters 1 and 2 applies to vector-matrix equations on an obvious component-by-component basis. If $Z\{\mathbf{x}(k)\} \triangleq \mathbf{X}(z)$, then $Z\{\mathbf{x}(k+1)\} = z\mathbf{X}(z) - z\mathbf{x}(0)$. Note the z multiplier on $\mathbf{x}(0)$, which deviates from the analogy expected from the Laplace transform of a derivative term. With this result and the linearity of the Z-transform operator, Eq. (*3.13*) gives

$$z\mathbf{X}(z) - z\mathbf{x}(0) = \mathbf{AX}(z) + \mathbf{BU}(z)$$

or

$$\mathbf{X}(z) = [z\mathbf{I} - \mathbf{A}]^{-1}\{\mathbf{BU}(z) + z\mathbf{x}(0)\} \tag{1}$$

Since the form of the initial condition response is known from earlier time-domain analysis, it is noted that

$$\mathbf{\Phi}(k, 0) \triangleq Z^{-1}\{[z\mathbf{I} - \mathbf{A}]^{-1} z\}$$

Using this definition allows the forcing function term to be written as

$$[z\mathbf{I} - \mathbf{A}]^{-1}\mathbf{BU}(z) = [z\mathbf{I} - \mathbf{A}]^{-1} zz^{-1}\mathbf{BU}(z) = Z\{\mathbf{\Phi}(k, 0)\}Z\{\mathbf{Bu}(k-1)\}$$

Then the convolution theorem of Z-transforms gives the inverse as

$$Z^{-1}\{[z\mathbf{I} - \mathbf{A}]^{-1}\mathbf{BU}(z)\} = \sum_{j=0}^{k} \mathbf{\Phi}(k, j)\mathbf{Bu}(j-1)$$

Since $\mathbf{u}(j-1) = \mathbf{0}$ for $j \le 0$, the lower summation limit is changed to $j = 1$. The solution for $\mathbf{x}(k)$ is then exactly Eq. (*9.29*). Note that the assumption that **B** was constant is unnecessary. From the transform of Eq. (*3.14*),

$$\mathbf{Y}(z) = \mathbf{CX}(z) + \mathbf{DU}(z) \tag{2}$$

Combined with (*1*) this gives

$$\mathbf{Y}(z) = \{\mathbf{C}[z\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B} + \mathbf{D}\}\mathbf{U}(z) + [z\mathbf{I} - \mathbf{A}]^{-1} z\mathbf{x}(0) \tag{3}$$

Note that a useful formula for computing input-output transfer functions has been found, namely,

$$T(z) = \mathbf{C}[z\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B} + \mathbf{D}$$

Finally, the sequence $\mathbf{y}(k)$ could be computed by inverse transforming (*3*) or by first finding $\mathbf{x}(k)$ as the inverse transform of (*1*) and then using that result in Eq. (*3.14*).

Approximation of a Continuous-Time System

9.27 A simple scalar system is described by $\dot{x} = -x + u$, $x(0) = 10$.

(a) Solve for $x(t)$ if $u(t) = e^t$.

(b) Derive a discrete approximation for the above system, using $t_{k+1} - t_k = 1$. Solve this discrete system and compare the results with the continuous solution.

(a) By inspection, $\phi(t, \tau) = e^{-(t-\tau)}$, so

$$x(t) = 10e^{-t} + \int_0^t e^{-(t-\tau)} e^{\tau}\, d\tau = 10e^{-t} + \sinh t$$

(b) Using the scalar transition matrix, the relation between $x(k+1)$ and $x(k)$ is

$$x(k+1) = e^{-1}x(k) + \int_{t_k}^{t_{k+1}} e^{-(t_{k+1}-\tau)}u(\tau)\,d\tau$$

Assuming $u(\tau)$ is constant over the interval $[t_k, t_{k+1}]$ and carrying out the integration gives

$$x(k+1) = e^{-1}x(k) + [1 - e^{-1}]u(k) \cong 0.368x(k) + 0.632u(k)$$

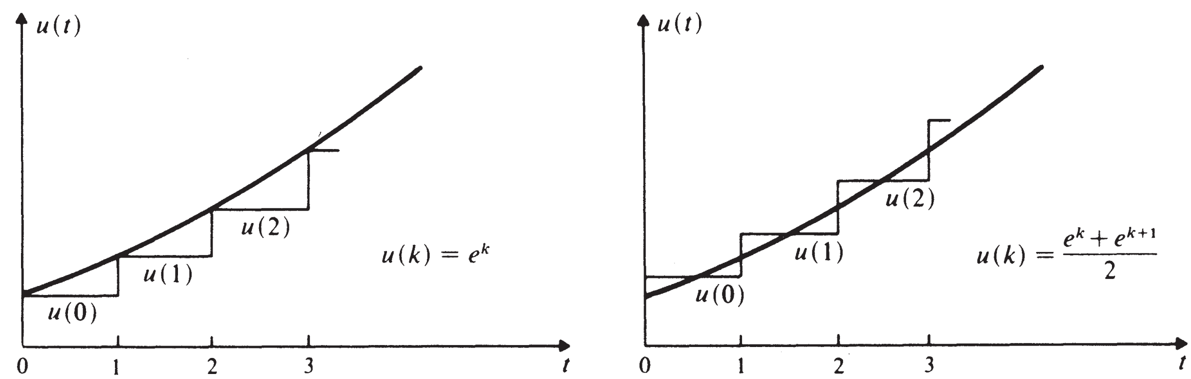
The solution for the discrete system is

$$x(k) = 10(0.368)^k + \sum_{j=1}^{k} (0.368)^{k-j}[0.632u(j-1)]$$

The initial condition term is the same for both the continuous and the discrete cases. A comparison of the forced response for the first five sampling periods is given in Table 9.1. The two separate approximations of Figure 9.6 are used for $u(k)$.

TABLE 9.1

		$t=k=0$	$t=k=1$	$t=k=2$	$t=k=3$	$t=k=4$
Continuous result: $\sinh t$		0	1.1752	3.6269	10.018	27.290
Discrete result: $\sum_{j=1}^{k} (0.368)^{k-j}[0.632u(j-1)]$	input (a)	0	0.632	1.951	5.388	14.677
	input (b)	0	1.175	3.626	10.016	27.286



Approximation (a) Approximation (b)

Figure 9.6

The accuracy of the discrete approximation depends on how closely the piecewise constant discrete input matches the continuous input. The accuracy can be good or poor as demonstrated in the example, but improves as the discrete step-size is decreased.

9.28 Apply the method of Problem 9.10 to the observable canonical form of the continuous-time system in Example 3.8 to obtain discrete state equations.

To facilitate comparisons, $T = 0.2$ is again used. Keeping fifth-order terms leads to

$$\mathbf{\Phi}(T, 0) = \begin{bmatrix} 0.0056 & 0.0793 & 0.1100 \\ -2.1224 & 0.7194 & 0.1779 \\ -1.5861 & -0.2191 & 0.9823 \end{bmatrix}, \qquad \mathbf{B}_1 = \begin{bmatrix} 0.0136 \\ 0.2347 \\ 0.5797 \end{bmatrix}$$

The output matrix is still $\mathbf{C} = [1 \quad 0 \quad 0]$, and $\mathbf{D} = 0$.

PROBLEMS

9.29 A system has two inputs $\mathbf{u} = [u_1 \quad u_2]^T$, and two outputs $\mathbf{y} = [y_1 \quad y_2]^T$. The input-output equations are $\dot{y}_1 + 3(y_1 + y_2) = u_1$ and $\ddot{y}_2 + 4\dot{y}_2 + 3y_2 = u_2$. Find $\mathbf{y}(t)$ if $y_1(0) = 1$, $y_2(0) = 2$, $\dot{y}_2(0) = 1$, and $\mathbf{u}(t) = \mathbf{0}$.

9.30 A system is described by the coupled input-output equations $\dot{y}_1 + 2(y_1 + y_2) = u_1$ and $\ddot{y}_2 + 4\dot{y}_2 + 3y_2 = u_2$. Find the output $\mathbf{y}(t) = [y_1(t) \quad y_2(t)]^T$ if $y_1(0) = 1, y_2(0) = 2, \dot{y}_2(0) = 0, u_1(t) = 0$, $u_2(t) = \delta(t)$ (i.e., an impulse at $t = 0$).

9.31 A system is described by $\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} -2 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 3 \\ 1 \end{bmatrix} u(t)$. If $\mathbf{x}(0) = [10 \quad 1]^T$ and if $u(t) = 0$, find $\mathbf{x}(t)$.

9.32 If the input to the system of the previous problem is $u(t) = e^{2t}$, what is $\mathbf{x}(t)$?

9.33 The wobbling satellite of Problems 3.16, page 117, and 9.6, page 327, has the initial state $\mathbf{x}(0) = [\omega_y(0) \quad \omega_z(0)]^T$. If the input torques are programmed as

$$u_1(t) = -\frac{1}{t_f}[\omega_y(0) \cos \Omega t - \omega_z(0) \sin \Omega t]$$

and

$$u_2(t) = -\frac{1}{t_f}[\omega_y(0) \sin \Omega t + \omega_z(0) \cos \Omega t]$$

find the state (wobble) $\mathbf{x}$ at time $t = t_f$.

9.34 Show that the approximate numerical solution of

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \tag{1}$$

at time T, expressed as

$$\mathbf{x}(T) = \mathbf{x}(0) + \dot{\mathbf{x}}(0)T + \ddot{\mathbf{x}}(0)T^2/2 + \dddot{\mathbf{x}}(0)T^3/3! + \cdots$$

leads to exactly the same series representation for $\mathbf{\Phi}(T, 0)$ and $\mathbf{B}_1$ as found in Problem 9.10. *Hint:* Repeatedly use Eq. (*1*) and its derivatives to express all derivatives of $\mathbf{x}$ in terms of $\mathbf{x}$ and $\mathbf{u}$. Treat $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{u}$ as constants. Notice that the first-order approximation is just rectangular integration of $\dot{\mathbf{x}}$, the second-order approximation is trapezoidal integration of $\dot{\mathbf{x}}$, etc.

9.35 Find the transition matrix $\mathbf{\Phi}(t, 0)$ for the feedback system of Problem 9.13 if K is increased to 2.5

9.36 Let $\mathbf{A}$ be a constant $n \times n$ matrix with n linearly independent eigenvectors. Use the Cayley-Hamilton remainder form for $\mathbf{\Phi}(t, t_0) = e^{\mathbf{A}(t - t_0)}$ to verify the results stated in Problem 9.15. That is, show that $\mathbf{\Phi}(t, t_0)\xi_i = e^{\lambda_i(t - t_0)}\xi_i$, where λ_i and ξ_i are eigenvalues and eigenvectors of $\mathbf{A}$.

9.37 Solve the following homogeneous difference equations:

$$x_1(k + 1) = x_1(k) - x_2(k) + x_3(k)$$

$$x_2(k + 1) = x_2(k) + x_3(k)$$

$$x_3(k + 1) = x_3(k)$$

with $x_1(0) = 2, x_2(0) = 5, x_3(0) = 10$.

9.38 Find the time response of the discrete model developed in Problem 9.12 if $x_1(0) = (0), x_2(0) = 10$, $u_1(k) = 1/K$, and $u_2(k) = 1$.

9.39 A single-input, single-output system is described by

$$\mathbf{x}(k+1) = \begin{bmatrix} 1 & 0 \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} 1 \\ -1 \end{bmatrix}u(k) \quad \text{and} \quad y(k) = [5 \quad 1]\mathbf{x}(k)$$

Use a change of basis to determine the normal form equations.

9.40 If a system is described by $\mathbf{x}(k+1) = \begin{bmatrix} 3 & 2 & 3 \\ 2 & 1 & 1 \\ 1 & 1 & 2 \end{bmatrix}\mathbf{x}(k)$, is it true that $\mathbf{\Phi}(j, k) = \mathbf{\Phi}^{-1}(k, j)$?

9.41 A simplified model of a motor is given by the transfer function $\theta(s)/u(s) = K/[s(\tau s + 1)]$. Let $x_1 = \theta, x_2 = \dot{\theta}$ and develop the continuous state equations. Then determine the approximate discrete-time state equations, using time points separated by $t_{k+1} - t_k = \Delta t$.

9.42 The motor of Problem 9.41 is used in a sampled-data feedback system as shown in Figure 9.7. The signal $u(k)$ is $e(t_k) = r(t_k) - \theta(k)$. Write the discrete state equations, using $r(t_k)$ as the input and $\theta(t_k)$ as the output.

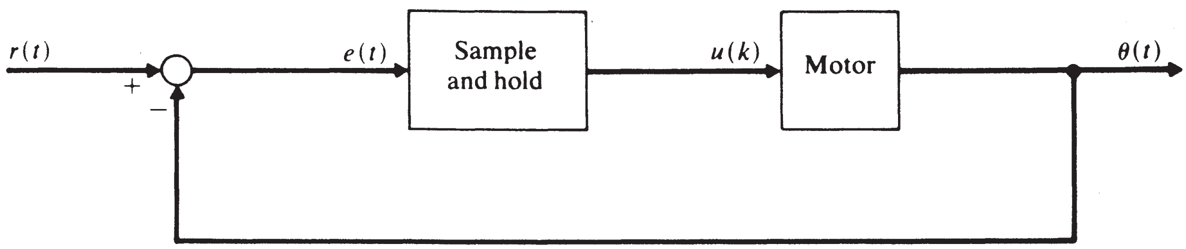


Figure 9.7

9.43 Apply the method of Problem 9.10 to the cascade realization of the continuous-time system in Example 3.8. Use $T = 0.2$ and keep terms through fifth order. After finding the approximate discrete models for **A** and **B**, find the transfer function by using

$$T(z) = \mathbf{C}[z\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}$$

9.44 Find a discrete-time state variable model for a system with transfer function

$$T(z) = \frac{0.006745(z + 0.0672)(z + 1.2416)}{(z - 0.04979)(z - 0.22313)(z - 0.60653)}$$

10 Stability

10.1 INTRODUCTION

Stability of single-input, single-output linear time-invariant systems was discussed from the transfer function point of view in Chapter 2. There the conditions for stability were given in terms of pole locations. The left half of the complex s-plane was found to be the stable region for continuous-time systems. The interior of the unit circle, centered at the origin of the Z-plane, was the stable region for discrete-time systems. Classical methods of stability analysis, including those of Nyquist, Bode, and root-locus, were presented in Chapter 2.

The goal of this chapter is to extend the previous stability concepts to multivariable systems described by state variable models. Although this chapter is primarily concerned with linear system stability, much of the machinery needed for nonlinear systems is also established here. Chapter 15 is devoted to several aspects of nonlinear control system analysis, including additional applications of stability theory.

In earlier discussions a system was either said to be stable or unstable, with perhaps some uncertainty about how to label systems which fall on the dividing line. Actually there are many different definitions of stability. A few of the more common ones are given here, along with methods of investigating them. Furthermore, a given system can exhibit behavior that is considered stable in some region of state space and unstable in other regions. Thus the question of stability should properly be addressed to the various *equilibrium points* (sometimes called critical points) of a system rather than to the system itself. This distinction is largely unnecessary for linear systems, as will be seen, but it is stressed here in preparation for nonlinear systems.

A sampling of the many treatments of stability from various points of view may be found in References 1 through 6.

10.2 EQUILIBRIUM POINTS AND STABILITY CONCEPTS

A heuristic discussion of stability is first given to help make the later mathematical treatment more intuitive. Consider the ball which is free to roll on the surface shown in Figure 10.1. The ball could be made to rest at points A, E, F, and G and anywhere between points B and D, such as at C. Each of these points is an equilibrium point of the system.

In state space, an equilibrium point for a continuous-time system is a point at which $\dot{\mathbf{x}}$ is zero in the absence of all inputs and disruptive disturbances. Thus if the system is placed in that state, it will remain there. For discrete-time systems, an equilibrium point is one for which $\mathbf{x}(k+1) = \mathbf{x}(k)$ in the absence of all control inputs or disturbances.

An infinitesimal perturbation away from points A or F will cause the ball to diverge from these points. This behavior intuitively justifies labeling A and F as *unstable* equilibrium points. After small perturbations away from E or G, the ball will eventually return to rest at these points. Thus E and G are labeled as stable equilibrium points. If the ball is displaced slightly from point C, in the absence of an initial velocity it will stay at the new position. Points like C are sometimes said to be *neutrally* stable.

Assume that the shape of the surface in Figure 10.1 changes with time. Specifically, assume that point E moves vertically so that the slope at that point is always zero, but the surface is sometimes concave upward (as shown) and sometimes concave downward. Point E is still an equilibrium point, but whether it is stable or not now depends upon time.

Thus far only *local* stability has been considered, since the perturbations were assumed to be small. If the ball were displaced sufficiently far from point G, it would not return to that point. Stability therefore depends on the size of the original perturbation and on the nature of any disturbances which may be acting. These intuitive notions are now developed for dynamical systems.

A particular point $\mathbf{x}_e \in \Sigma$ is an equilibrium point of a dynamical system if the system's state at t_0 is $\mathbf{x}_e$ and $\mathbf{x}(t) = \mathbf{x}_e$ for all $t \geq t_0$ in the absence of inputs or disturbances. For the continuous-time system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t)$, this means that $\mathbf{f}(\mathbf{x}_e, \mathbf{0}, t) = \mathbf{0}$ for $t \geq t_0$. For the discrete-time system $\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k), k)$, this means that $\mathbf{f}(\mathbf{x}_e, \mathbf{0}, k) = \mathbf{x}_e$ for all $k > 0$.

The origin of the state space is always an equilibrium point for linear systems, although it need not be the only one. In the continuous-time case, if the system matrix

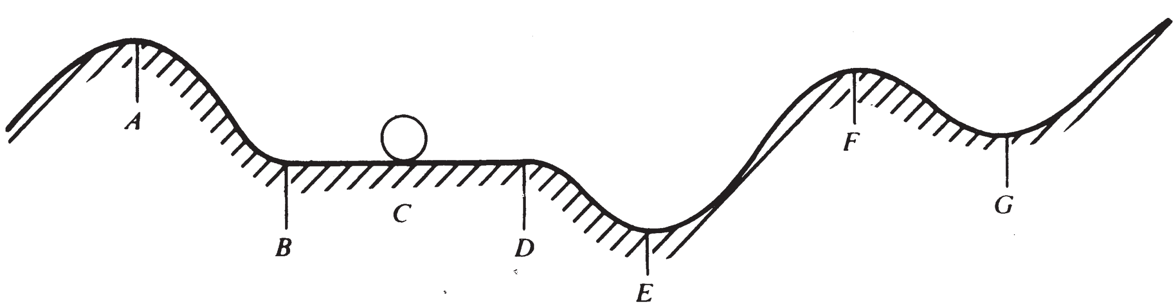


Figure 10.1

A has a zero eigenvalue, then there is an infinity of vectors (eigenvectors) satisfying $\mathbf{A}\mathbf{x}_e = \mathbf{0}$. In the discrete-time case a unity eigenvalue of **A** means there is an infinity of vectors satisfying $\mathbf{A}\mathbf{x}_e = \mathbf{x}_e$. These points loosely correspond to points between B and D of Figure 10.1. Only *isolated equilibrium points* will be considered in this text, and for linear systems the only isolated equilibrium point is the origin.

Any isolated singular point can be transferred to the origin by a change of variables, $\mathbf{x}' = \mathbf{x} - \mathbf{x}_e$. For this reason it is often assumed in the sequel that $\mathbf{x}_e = \mathbf{0}$.

Stability deals with the following questions. If at time t_0 the state is perturbed from its equilibrium point, does the state return to $\mathbf{x}_e$, or remain close to $\mathbf{x}_e$, or diverge from it? Similar questions could be raised if system inputs or disturbances are allowed. Another class of stability questions deals with the state trajectories of an unperturbed system and of a perturbed system. Let the solution to $\dot{\mathbf{x}}_1 = \mathbf{f}(\mathbf{x}_1(t), \mathbf{u}(t), t)$, with $\mathbf{x}_1(t_0)$ given, define the unperturbed trajectory $\mathbf{x}_1(t)$. Let the perturbed trajectory $\mathbf{x}_2(t)$ be defined by $\dot{\mathbf{x}}_2 = \mathbf{f}(\mathbf{x}_2(t), \mathbf{u}(t) + \mathbf{v}(t), t)$, where $\mathbf{x}_2(t_0) = \mathbf{x}_1(t_0) + \mathbf{e}(t_0)$. The initial state and control perturbations are $\mathbf{e}(t_0)$ and $\mathbf{v}(t)$, respectively. Does $\mathbf{x}_2(t)$ return to $\mathbf{x}_1(t)$, or remain close to it, or diverge from it? These questions can be studied by considering the difference $\mathbf{e}(t) = \mathbf{x}_2(t) - \mathbf{x}_1(t)$, which satisfies $\dot{\mathbf{e}} = \mathbf{f}(\mathbf{x}_1(t) + \mathbf{e}(t),\ \mathbf{u}(t) + \mathbf{v}(t), t) - \mathbf{f}(\mathbf{x}_1(t), \mathbf{u}(t), t)$ or simply $\dot{\mathbf{e}} = \mathbf{f}'(\mathbf{e}(t), \mathbf{v}(t), t)$ with $\mathbf{e}(t_0)$, $\mathbf{x}_1(t)$, and $\mathbf{u}(t)$ given. Now $\mathbf{e} = \mathbf{0}$ is an equilibrium point and the questions regarding the perturbed motion can be studied in terms of perturbations about the origin, as before.

Whether an equilibrium point is stable or not depends upon what is meant by remaining close, the magnitude of state or input disturbances, and their time of application. These qualifying conditions are the reasons for the existence of a variety of stability definitions.

10.3 STABILITY DEFINITIONS

Consider the continuous-time system with the input set to zero,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{0}, t), \qquad \mathbf{x}(t_0) = \mathbf{x}_0 \tag{10.1}$$

As was pointed out in Sec. 3.3, page 78, there will *exist* a *unique* solution to these differential equations provided that the function $\mathbf{f}(\mathbf{x}, 0, t)$ satisfies a Lipschitz condition [7] with respect to **x** and is at least piecewise continuous with respect to t throughout some region of the product space $\Sigma \times \tau$, which contains $\mathbf{x}_0, t_0$. Furthermore, the solution depends on its arguments in a continuous fashion. The solution is frequently written as $\mathbf{\Phi}(t; x_0, t_0)$ to show its arguments explicitly, but we often refer to it simply as $\mathbf{x}(t)$. Note that in the linear case discussed in Chapter 9, $\mathbf{\Phi}(t; x_0, t_0) = \mathbf{\Phi}(t, t_0)\mathbf{x}_0$.

It is assumed that an equilibrium point for system (*10.1*) is at or has been transferred to the origin. Then the following definitions apply. For the continuous-time case with zero input

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{0}, t), \qquad \mathbf{x}(t_0) = \mathbf{x}_0$$

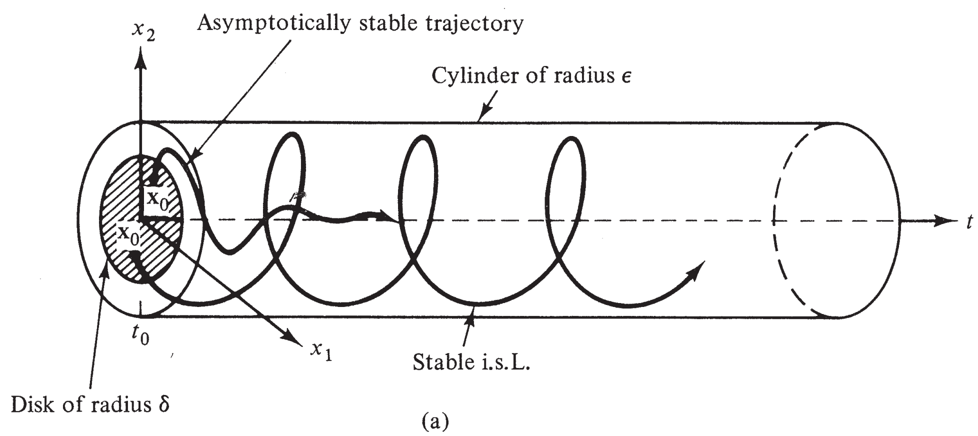
with the origin an equilibrium point, the following apply.

Definition 10.1. The origin is a *stable* equilibrium point if for any given value $\epsilon > 0$ there exists a number $\delta(\epsilon, t_0) > 0$ such that if $\|\mathbf{x}(t_0)\| < \delta$, then the resultant motion $\mathbf{x}(t)$ satisfies $\|\mathbf{x}(t)\| < \epsilon$ for all $t > t_0$.

This definition of stability is sometimes called *stability in the sense of Lyapunov,* abbreviated as stable i.s.L. If a system possesses this type of stability, then it is ensured that the state can be kept within ϵ, in norm, of the origin by restricting the initial perturbation to be less than δ, in norm. Note that it is necessarily true that $\delta \leq \epsilon$.

Definition 10.2 The origin is an *asymptotically stable* equilibrium point if (*a*) it is stable, and if in addition, (*b*) there exists a number $\delta'(t_0) > 0$ such that whenever $\|\mathbf{x}(t_0)\| < \delta'(t_0)$ the resultant motion satisfies $\lim_{t\to\infty} \|\mathbf{x}(t)\| = 0$.

Figure 10.2*a* illustrates these definitions for the two-dimensional state space Σ_2. The same notions conceptually apply in higher dimensions. Two examples of possible trajectories are shown in Figure 10.2*a,* one for a system which is stable i.s.L. and the other for an asymptotically stable system. A projection onto the state space Σ_2 is shown in Figure 10.2*b*. An *unstable* trajectory—i.e., one which is not stable—has been added to the original two.



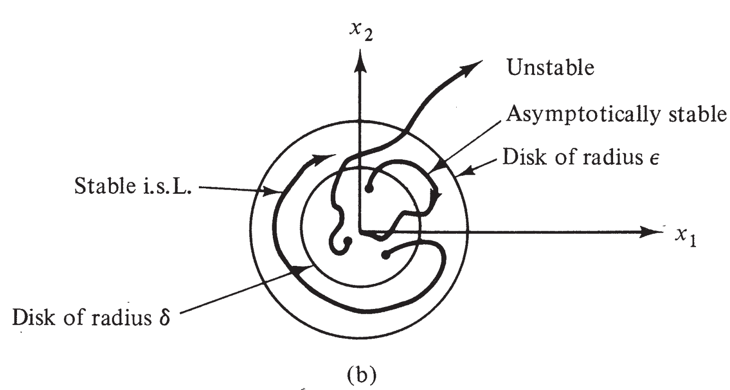


Figure 10.2*a* Illustrations of stable trajectories in $\Sigma_2 \times \tau$.
Figure 10.2*b* Illustrations of possible trajectories in Σ_2.

The two most basic definitions of stability have been given for unforced continuous-time systems. Variations upon these are defined by adding additional qualifying adjectives [2]. If δ and δ' are not functions of t_0, then the origin is said to be *uniformly stable* and *uniformly asymptotically stable,* respectively. If $\delta'(t_0)$ in Definition 10.2 can be made arbitrarily large—i.e., if all $\mathbf{x}(t_0)$ converge to $\mathbf{0}$—then the origin is said to be *globally* asymptotically stable or asymptotically stable *in the large.*

Stability definitions for the discrete-time system with zero input

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{0}, k), \qquad \mathbf{x}(0) = \mathbf{x}_0 \tag{10.2}$$

are identical to those given earlier, provided the discrete-time index k is used in place of t. As before, it is assumed that the coordinates have been chosen so that the origin is an equilibrium state.

When nonzero inputs $\mathbf{u}(t)$ or $\mathbf{u}(k)$ are considered, two additional types of stability are often used.

Definition 10.3. (Bounded input, bounded state stability.) If there is a fixed, finite constant K such that $\|\mathbf{u}\| \leq K$ for every t (or k), then the input is said to be bounded. If for every bounded input, and for arbitrary initial conditions $\mathbf{x}(t_0)$, there exists a scalar $0 < \delta(K, t_0, \mathbf{x}(t_0))$ such that the resultant state satisfies $\|\mathbf{x}\| \leq \delta$, then the system is *bounded input, bounded state* stable, abbreviated as BIBS stable.

All the previous definitions of stability deal with the behavior of the state vector relative to an equilibrium state. Frequently, the main interest is in the system output behavior. This motivates the final stability definition.

Definition 10.4. (Bounded input, bounded output stability.) Let $\mathbf{u}$ be a bounded input with K_m as the least upper bound. If there exists a scalar α such that for every t (or k), the output satisfies $\|\mathbf{y}\| \leq \alpha K_m$, then the system is *bounded input, bounded ouptut* stable, abbreviated as BIBO stable.

10.4 LINEAR SYSTEM STABILITY

The following linear continuous-time system is considered:

$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t) \\ \mathbf{y}(t) &= \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \end{aligned} \tag{10.3}$$

The unforced case is treated first. With $\mathbf{u}(t) = \mathbf{0}$, the state vector is given by

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) \tag{10.4}$$

The norm of $\mathbf{x}(t)$ is a measure of the distance of the state from the origin:

$$\|\mathbf{x}(t)\| = \|\mathbf{\Phi}(t, t_0)\mathbf{x}(t_0)\| \leq \|\mathbf{\Phi}(t, t_0)\| \, \|\mathbf{x}(t_0)\| \tag{10.5}$$

Suppose there exists a number $N(t_0)$, possibly depending on t_0, such that

$$\|\mathbf{\Phi}(t, t_0)\| \leq N(t_0) \quad \text{for all } t \geq t_0 \tag{10.6}$$

Then the conditions of Definition 10.1 can be satisfied for any $\epsilon > 0$ by letting $\delta(t_0, \epsilon) = \epsilon / N(t_0)$. It follows from Eq. (*10.5*) that Eq. (*10.6*) is *sufficient* to ensure that the origin is stable in the sense of Lyapunov. It is easy to show that this condition is also *necessary*. The origin is asymptotically stable if and only if Eq. (*10.6*) holds, and if in addition, $\|\mathbf{\Phi}(t, t_0)\| \to 0$ for $t \to \infty$. Note that for a linear system, asymptotic stability does not depend on $\mathbf{x}(t_0)$. If a linear system is asymptotically stable, it is globally asymptotically stable.

The stability types which depend upon the input $\mathbf{u}(t)$ are now considered. For the linear continuous-time system, the state vector is given by

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\, d\tau \tag{10.7}$$

BIBS stability requires that $\mathbf{x}(t)$ remain bounded for all bounded inputs. Since $\mathbf{u}(t) = \mathbf{0}$ is bounded, it is clear that stability i.s.L. is a necessary condition for BIBS stability. By taking the norm of both sides of Eq. (*10.7*) and using well-known properties of the norm, it is found that $\|\mathbf{x}(t)\|$ remains bounded, and thus the origin is BIBS stable, if Eq. (*10.6*) holds and if in addition there exists a number $N_1(t_0)$ such that

$$\int_{t_0}^{t} \|\mathbf{\Phi}(t, \tau)\mathbf{B}(\tau)\|\, d\tau \le N_1(t_0) \quad \text{for all } t \ge t_0 \tag{10.8}$$

Similar arguments show that a linear discrete-time system is BIBS stable if the discrete transition matrix satisfies Eq. (*10.6*) and if $\sum_{k=0}^{k_1} \|\mathbf{\Phi}(k_1, k)\mathbf{B}(k-1)\| \le N_1$.

BIBO stability is investigated by considering the output of a linear system

$$\mathbf{y}(t) = \mathbf{C}(t)\mathbf{x}(t) + \mathbf{D}(t)\mathbf{u}(t) \tag{10.9}$$

Substitution of Eq. (*10.7*) into Eq. (*10.9*) and viewing the initial state $\mathbf{x}(t_0)$ as having arisen because of a bounded input over the interval $(-\infty, t_0)$ gives

$$\mathbf{y}(t) = \int_{-\infty}^{t} \mathbf{W}(t, \tau)\mathbf{u}(\tau)\, d\tau \tag{10.10}$$

Only bounded inputs are considered, that is,

$$\|\mathbf{u}(\tau)\| \le K \quad \text{for all } \tau \tag{10.11}$$

The output remains bounded in norm if there exists a constant $M > 0$ such that the impulse response or weighting matrix $\mathbf{W}(t, \tau)$ satisfies

$$\int_{-\infty}^{t} \|\mathbf{W}(t, \tau)\|\, d\tau \le M \quad \text{for all } t \tag{10.12}$$

Equation (*10.12*) is the necessary and sufficient condition for BIBO stability of continuous-time systems. The analogous result, with a summation replacing the integration, holds for discrete-time systems.

The matrix norm $\|\mathbf{\Phi}(t, t_0)\|$ plays a central role in the stability conditions. This norm can be defined in various ways, including the inner product definition

$$\|\mathbf{\Phi}(t, t_0)\|^2 = \max_{\mathbf{x}} \{\langle \mathbf{\Phi}(t, t_0)\mathbf{x}, \mathbf{\Phi}(t, t_0)\mathbf{x}\rangle \,|\, \langle \mathbf{x}, \mathbf{x}\rangle = 1\} \tag{10.13}$$

By introducing the adjoint of $\Phi(t, t_0)$, it is found that Eq. (*10.13*) leads to

$$\|\Phi(t, t_0)\|^2 = \text{max eigenvalue of } \overline{\Phi}^T(t, t_0)\Phi(t, t_0) \qquad (10.14)$$

If $\Phi(t, t_0)$ is normal, then $\overline{\Phi}^T(t, t_0)\Phi(t, t_0) = \Phi(t, t_0)\overline{\Phi}^T(t, t_0)$ and then

$$\|\Phi(t, t_0)\| = \max_i |\alpha_i| \qquad (10.15)$$

where α_i is an eigenvalue of $\Phi(t, t_0)$. In all cases a useful lower bound on the norm is given by

$$\|\Phi(t, t_0)\|^2 \geq |\alpha_i|^2 \quad \text{for any eigenvalue } \alpha_i \text{ of } \Phi(t, t_0) \qquad (10.16)$$

10.5 LINEAR CONSTANT SYSTEMS

Whenever the system under consideration has a constant system matrix **A**, the following results hold:

$$\Phi(t, t_0) = \mathbf{e}^{\mathbf{A}(t - t_0)} \qquad \text{(continuous-time)} \qquad (10.17)$$

$$\Phi(k, 0) = \mathbf{A}^k \qquad \text{(discrete-time)} \qquad (10.18)$$

By virtue of the Cayley-Hamilton theorem, Chapter 8, both of these results can be expressed as polynomials in **A**. Then by Frobenius' theorem (Problem 9.15, page 333) the eigenvalues α_i of Φ are related to the eigenvalues λ_i of **A** by

$$\alpha_i = e^{\lambda_i(t - t_0)} \quad \text{or} \quad \alpha_i = \lambda_i^k$$

for the continuous-time and discrete-time cases, respectively. It is relatively simple to express the previous stability conditions in terms of the eigenvalues of the system matrix **A**. Letting these eigenvalues be $\lambda_i = \beta_i \pm j\omega_i$, the resulting conditions are summarized in Table 10.1.

The results of Table 10.1 again show the left-half plane and the interior of the unit circle as stability regions, this time in terms of locations of the eigenvalues of the **A** matrix instead of transfer function poles. The consistency between the continuous-

TABLE 10-1 STABILITY CRITERIA FOR LINEAR CONSTANT SYSTEMS

(Eigenvalues of **A** are $\lambda_i = \beta_i \pm j\omega_i$)

	Continuous Time $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$	Discrete Time $\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k)$
Unstable	If $\beta_i > 0$ for any simple root or $\beta_i \geq 0$ for any repeated root	If $\|\lambda_i\| > 1$ for any simple root or $\|\lambda_i\| \geq 1$ for any repeated root
Stable i.s.L.	If $\beta_i \leq 0$ for all simple roots and $\beta_i < 0$ for all repeated roots	If $\|\lambda_i\| \leq 1$ for all simple roots and $\|\lambda_i\| < 1$ for all repeated roots
Asymptotically Stable	If $\beta_i < 0$ for all roots	$\|\lambda_i\| < 1$ for all roots

time and discrete-time columns of Table 10.1 is noted. Recall from Problem 2.21 that any s-plane pole with a negative real part maps into a complex number in the Z-plane with a magnitude less than unity.

Classification of Equilibrium Points. Equilibrium points can be classified into several types, beyond just labeling them stable or unstable. For this purpose it is useful to expand the initial condition response, using the eigenvectors $\boldsymbol{\xi}_i$ of $\mathbf{A}$ as basis vectors:

$$\mathbf{x}(t) = w_1(0)\exp\{\lambda_1 t\}\boldsymbol{\xi}_1 + w_2(0)\exp\{\lambda_2 t\}\boldsymbol{\xi}_2 + \cdots + w_n(0)\exp\{\lambda_n t\}\boldsymbol{\xi}_n \tag{10.19}$$

where $\mathbf{w}(0) = \mathbf{M}^{-1}\mathbf{x}_0$ and $\mathbf{M}$ is the modal matrix.

Equation (*10.19*) is used to discuss the initial condition response of a second-order system. The two eigenvalues could both be real or a complex conjugate pair. In the real case, if both eigenvalues are stable, the equilibrium point is called a *stable node,* and the phase portrait for various initial conditions is shown in Figure 10.3*a*. If one eigenvalue is stable and the other is unstable, the behavior is as shown in Figure 10.3*b*, and this is called a *saddle point.* When both eigenvalues are unstable, trajectories such as those in Figure 10.3*c* result, and this is called an *unstable node.* When the eigenvalues are complex, Eq. (*10.19*) can be rewritten as

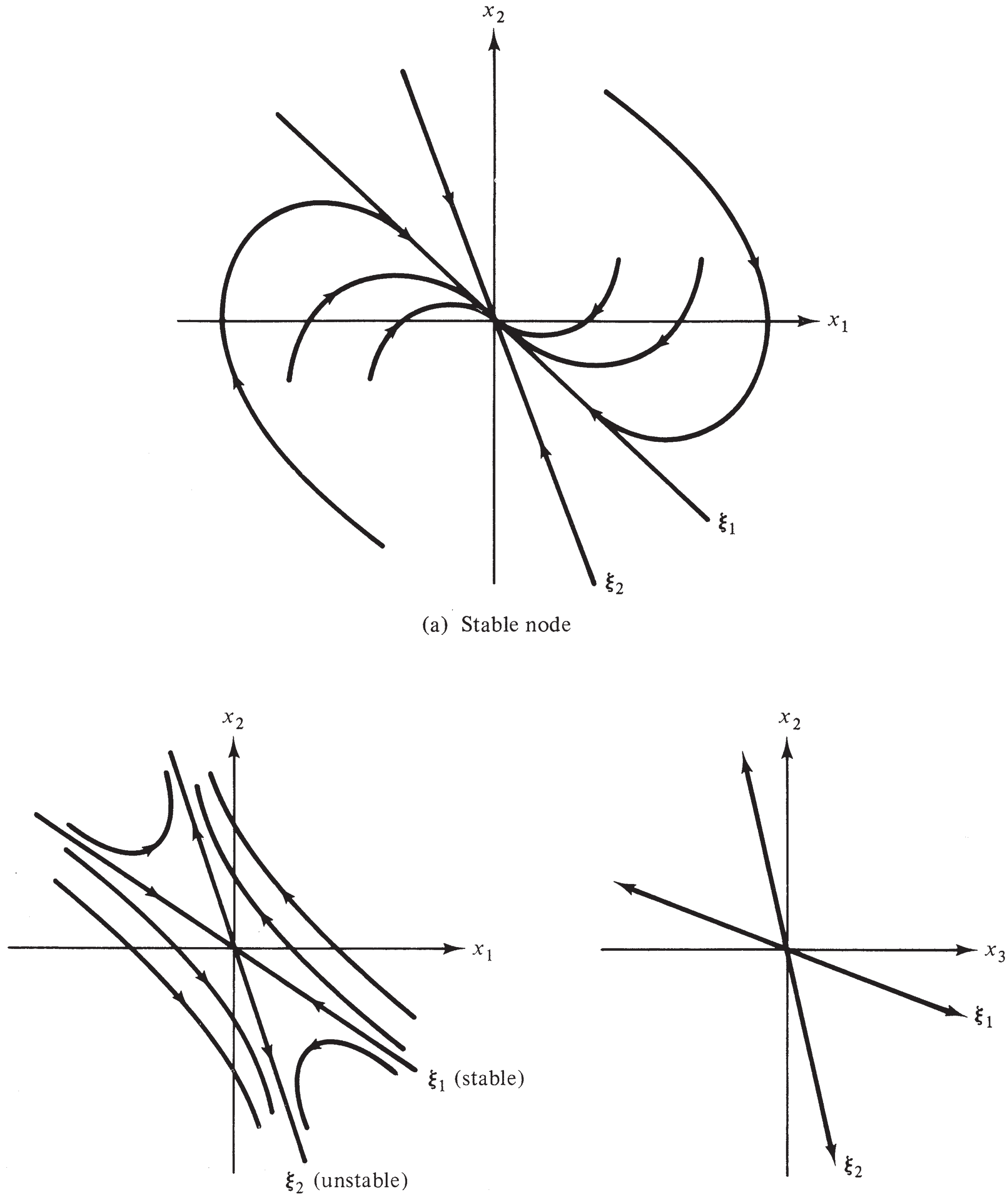
$$\mathbf{x}(t) = \exp\{\beta t\}\{[a\cos(\omega t) + b\sin(\omega t)]\boldsymbol{\xi}_R + [b\cos(\omega t) - a\sin(\omega t)]\boldsymbol{\xi}_I\}$$

where $\boldsymbol{\xi}_R$ and $\boldsymbol{\xi}_I$ are the real and imaginary parts of $\boldsymbol{\xi}$ and a, b are real constants dependent upon the initial conditions. If the real part of the eigenvalue, β, is negative the phase portraits of Figure 10.4*a* result, and this is called a *stable focus.* If $\beta > 0$, the *unstable focus* of Figure 10.4*b* results. If $\beta = 0$, the equilibrium point is called a *center,* and the phase portraits are as shown in Figure 10.4*c*. Generally, graphical display of trajectories becomes difficult or impossible in higher dimensions. However, Eq. (*10.19*) is still useful for conceptualizing the response. For example, if a third-order system has two stable complex eigenvalues and one real, unstable eigenvalue with eigenvector $\boldsymbol{\xi}_3$, a three-dimensional phase trajectory would look somewhat like Figure 10.4*a* but with the center stretched out along the vector $\boldsymbol{\xi}_3$, as sketched in Figure 10.5.

Phase-space methods have proven very useful in analyzing linear and nonlinear second-order systems. A method of extending many of phase-plane advantages to higher-order systems is called the *center manifold theory* [4]. This approach essentially selects the two modes that are providing the dominant action, those with eigenvalues nearest the stability boundary. The full development of classical phase plane methods and their more recent extensions are left for the references.

10.6 THE DIRECT METHOD OF LYAPUNOV

The general stability results for linear systems have been presented in Sec. 10.4 in terms of the properties of the transition matrix $\Phi(t, t_0)$. In effect this means that the solution of the system's differential or difference equations needs to be known before stability conclusions can be drawn. In the case of linear *constant* systems, this does not



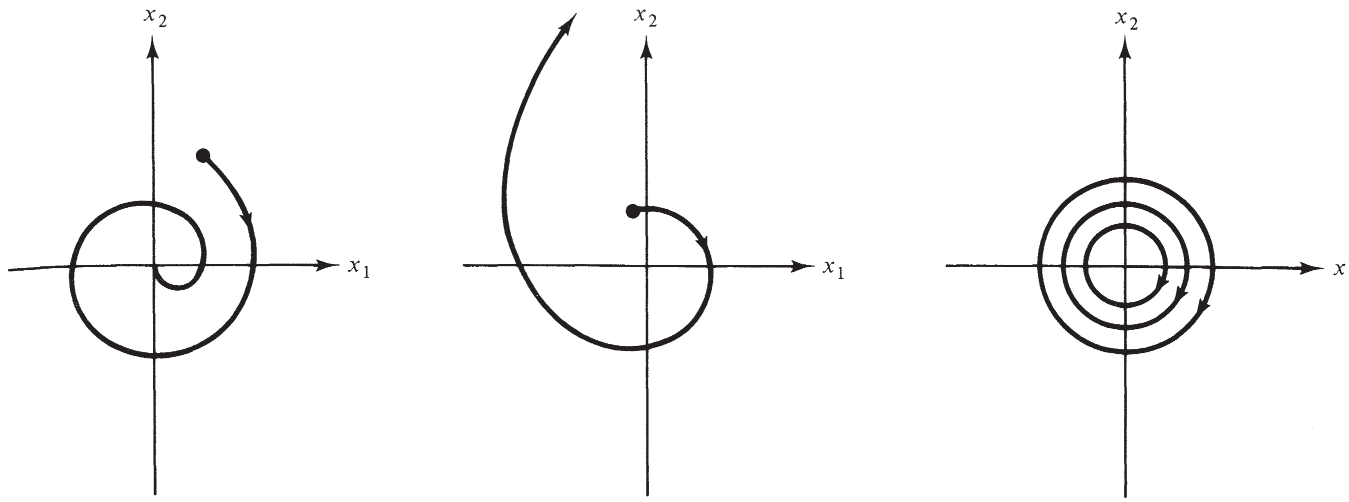
(a) Stable node

(b) Saddle point

(c) Unstable node

Figure 10.3

appear to be so because Sec. 10.5 expresses the stability conditions in terms of the eigenvalues of the system matrix **A.** Late in the nineteenth century the Russian mathematician A. M. Lyapunov developed an approach to stability analysis, now known as the direct method (or second method) of Lyapunov. The unique thing about this method is that only the form of the differential or difference equations need be known, not their solutions. Lyapunov's direct method is now widely used for stability analysis of linear and nonlinear systems, both time-invariant and time-varying. Although this chapter is primarily concerned with linear systems, Lyapunov's method is presented in



(a) Stable focus (b) Unstable focus (c) Center

Figure 10.4

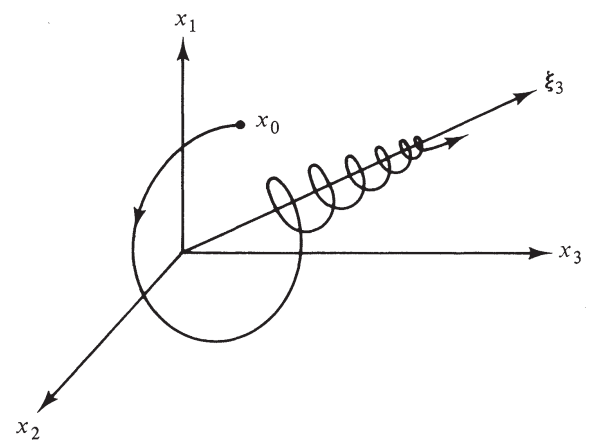


Figure 10.5

sufficient detail so that it can be later applied to nonlinear systems as well. The approach will be to present an intuitive discussion of the method first. Then a few of the main and most useful theorems will be presented without proof. Uses of the theorems are illustrated in the problems and in Chapter 15. The mathematical overhead may seem excessive for dealing with linear systems. However, for *time-varying* linear systems, reliable answers to stability questions, which might otherwise be difficult to obtain, can be found. Even for constant linear systems, the Lyapunov method provides an informative alternative approach.

Energy concepts are widely used and easily understood by engineers. Lyapunov's direct method can be viewed as a generalized energy method. Consider a second-order system, such as the unforced LC circuit of Figure 10.6*a* or the mass-spring system of Figure 10.6*b*.

In the first case the capacitor voltage v and inductor current i can be used as state variables $\mathbf{x}$, and the total energy (magnetic plus electric) in the system at any time is $Li^2/2 + Cv^2/2$. In the second case the position y (measured from the free length of the spring) and the velocity $\dot{y}$ can be used as state variables $\mathbf{x}$, and the total energy (kinetic

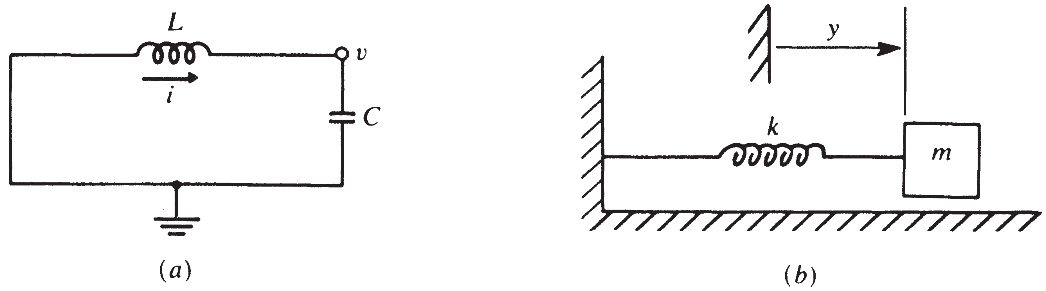


Figure 10.6

plus potential) at any time is $m\dot{y}^2/2 + ky^2/2$. In both cases the energy $\mathscr{E}$ is a quadratic function of the state variables. Thus $\mathscr{E}(\mathbf{x}) > 0$ if $\mathbf{x} \neq \mathbf{0}$, and $\mathscr{E} = 0$ if and only if $\mathbf{x} = \mathbf{0}$. If the time rate of change of the energy $\dot{\mathscr{E}}$ is always negative, except at $\mathbf{x} = \mathbf{0}$, then $\mathscr{E}$ will be continually decreasing and will eventually approach zero. Because of the nature of the energy function, $\mathscr{E} = 0$ implies $\mathbf{x} = \mathbf{0}$. Therefore, if $\dot{\mathscr{E}} < 0$ for all t, except when $\mathbf{x} = \mathbf{0}$, it is concluded that $\mathbf{x}(t) \to \mathbf{0}$ for sufficiently large t. The close relationship of this conclusion to asymptotic stability, Definition 10.2, is obvious.

If the time rate of change of energy is never positive, that is, $\dot{\mathscr{E}} \leq 0$, then $\mathscr{E}$ can never increase, but it need not approach zero either. It can then be concluded that $\mathscr{E}$, and hence $\mathbf{x}$, remain bounded in some sense. This situation is related in an obvious way to stability i.s.L., Definition 10.1.

For both systems of Figure 10.6, the energy can be expressed in the form $\mathscr{E}(\mathbf{x}) = \frac{1}{2}a_1 x_1^2 + \frac{1}{2}a_2 x_2^2$. Assuming for the present that the coefficients a_1 and a_2 are constant, the time rate of change is given by

$$\dot{\mathscr{E}} = a_1 x_1 \dot{x}_1 + a_2 x_2 \dot{x}_2 \tag{10.20}$$

Knowledge of the *form* of the differential equations $\dot{x}_1 = f_1(\mathbf{x})$ and $\dot{x}_2 = f_2(\mathbf{x})$ allows both $\mathscr{E}$ and $\dot{\mathscr{E}}$ to be expressed as a function of the state $\mathbf{x}$. No knowledge of the *solutions* of the differential equations is required in order to draw conclusions regarding stability. Lyapunov's direct method is a generalization of these ideas.

Notice that in Figure 10.1 those points E and G that were described as stable in the earlier discussion are points of relative minimum total energy. If the ball is at rest at one of these points, its kinetic energy is zero and its potential energy is at a relative minimum. A small displacement would increase the potential energy slightly. In the absence of all surface and air friction or any other dissipative phenomenon, the total energy would remain constant after release, being passed back and forth between potential and kinetic energy. The ball would oscillate forever with constant amplitude. This is an example of stability i.s.L. In the real-world case there will always be some energy dissipation, and the ball would eventually return to the point of (relative) minimum total energy. This is an example of asymptotic stability.

EXAMPLE 10.1 For the system of Figure 10.6*a*, with $x_1 = i$ and $x_2 = v$, $\dot{x}_1 = -x_2/L$, $\dot{x}_2 = x_1/C$, so that $\dot{\mathscr{E}} = Lx_1(-x_2/L) + Cx_2(x_1/C) = 0$. Thus $\dot{\mathscr{E}} \equiv 0$ for all t and $\mathscr{E}$ is constant. This conservative system is stable i.s.L. but not asymptotically stable. Of course, this is a well-known result for this undamped oscillator. If the system is modified to include a positive resistance, then the system is dissipative and $\dot{\mathscr{E}}$ would be always negative, except when $\mathbf{x} = \mathbf{0}$. The system is then asymptotically stable. ■

EXAMPLE 10.2 The system of Figure 10.6*b* is assumed to have a nonlinear frictional force d acting between the mass and the supporting surface. This does not change the definition of the energy $\mathscr{E}$. Letting $x_1 = y$ and $x_2 = \dot{y}$, the differential equations are $\dot{x}_1 = x_2$, $\dot{x}_2 = (-kx_1 + d)/m$. Then

$$\dot{\mathscr{E}} = kx_1\dot{x}_1 + mx_2\dot{x}_2 = kx_1x_2 - kx_1x_2 + x_2 d = x_2 d$$

$\dot{\mathscr{E}}$ is nonpositive if the friction force d is always opposing the direction of the velocity x_2. In this case $\mathscr{E}$ can never increase, and the system is clearly stable i.s.L. It may, in fact, be asymptotically stable, as will become clear later. On the other hand, if d and x_2 have the same sign, then $\dot{\mathscr{E}} > 0$ and $\mathscr{E}$ may increase. Neither stability nor instability can be concluded without further analysis. ■

Lyapunov's direct method makes use of a *Lyapunov function* $V(\mathbf{x})$. This scalar function of the state may be thought of as a generalized energy. In many problems the energy function can serve as a Lyapunov function. In cases where a system model is described mathematically, it may not be clear what "energy" means. The conditions which $V(\mathbf{x})$ must satisfy in order to be a Lyapunov function are therefore based on mathematical rather than physical considerations.

A single-valued function $V(\mathbf{x})$ which is continuous and has continuous partial derivatives is said to be *positive definite* in some region Ω about the origin of the state space if (1) $V(\mathbf{0}) = 0$ and (2) $V(\mathbf{x}) > 0$ for all nonzero $\mathbf{x}$ in Ω. A special case of a positive definite function was the quadratic form discussed in Sec. 7.8, page 264. If condition (2) is relaxed to $V(\mathbf{x}) \geq 0$ for all $\mathbf{x} \in \Omega$, then $V(\mathbf{x})$ is said to be *positive semidefinite*. Reversing the inequalities leads to corresponding definitions of *negative definite* and *negative semidefinite* functions.

Consider the autonomous (i.e., unforced, no explicit time dependence) system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) \tag{10.21}$$

The origin is assumed to be an equilibrium point, that is, $\mathbf{f}(\mathbf{0}) = \mathbf{0}$. The stability of this equilibrium point can be investigated by means of the following theorems.

Theorem 10.1. If a positive definite function $V(\mathbf{x})$ can be determined such that $\dot{V}(\mathbf{x}) \leq 0$ (negative semidefinite), then the origin is stable i.s.L.

A function $V(\mathbf{x})$ satisfying these requirements is called a Lyapunov function. The Lyapunov function is not unique; rather, many different Lyapunov functions may be found for a given system. Likewise, the inability to find a satisfactory Lyapunov function does not mean that the system is unstable.

Theorem 10.2. If a positive definite function $V(\mathbf{x})$ can be found such that $\dot{V}(\mathbf{x})$ is negative definite, then the origin is asymptotically stable.

Both of the previous theorems relate to local stability in the neighborhood Ω of the origin. Global asymptotic stability is considered next.

Theorem 10.3. The origin is a globally asymptotically stable equilibrium point for the system of Eq. (*10.21*) if a Lyapunov function $V(\mathbf{x})$ can be found such that (1) $V(\mathbf{x}) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ and $V(\mathbf{0}) = 0$, (2) $\dot{V}(\mathbf{x}) < 0$ for all $\mathbf{x} \neq 0$, and (3) $V(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$.

In both the preceding theorems the negative definite requirement on $\dot{V}$ can be relaxed to negative semidefinite, and asymptotic stability can still be concluded, if it can be shown that $\dot{V}$ is never zero along a solution trajectory.

As an aid in understanding these three theorems and their differences, a family of contours of $V(\mathbf{x}) =$ constant is shown in Figure 10.7 for a two-dimensional state space. These contours can never intersect because $V(\mathbf{x})$ is single-valued. They are smooth curves because $V(\mathbf{x})$ and all its partial derivatives are required to be continuous.

For the situation shown, condition (3), Theorem 10.3 is not met. It is possible for $\|\mathbf{x}\| \to \infty$ while $V(\mathbf{x})$ remains finite; for example, along the contour $V(\mathbf{x}) = C_4$. It could happen that $V(\mathbf{x})$ continually decreases, causing $V(\mathbf{x})$ to decrease from C_6 through C_5 and asymptotically approach C_4, while $\|\mathbf{x}(t)\|$ continues to grow without bound. The trajectory starting at $\mathbf{x}_3(t_0)$ illustrates this. Condition (3) of theorem 10.3 is necessary to rule out this possibility.

Referring to Figure 10.7, if $\dot{V}(\mathbf{x}) < 0$, then the origin is asymptotically stable for any initial state sufficiently close to the origin, specifically for any $\mathbf{x}(t_0)$ inside contour C_3. Any initial state inside contour C_3, such as $\mathbf{x}_2(t_0)$, must eventually approach the origin if $\dot{V}(\mathbf{x})$ is always negative for $\mathbf{x} \neq \mathbf{0}$. This is the essence of Theorem 10.2. If the restriction is relaxed so that $\dot{V}(\mathbf{x}) \leq 0$, as in Theorem 10.1, then $V(\mathbf{x})$ can never

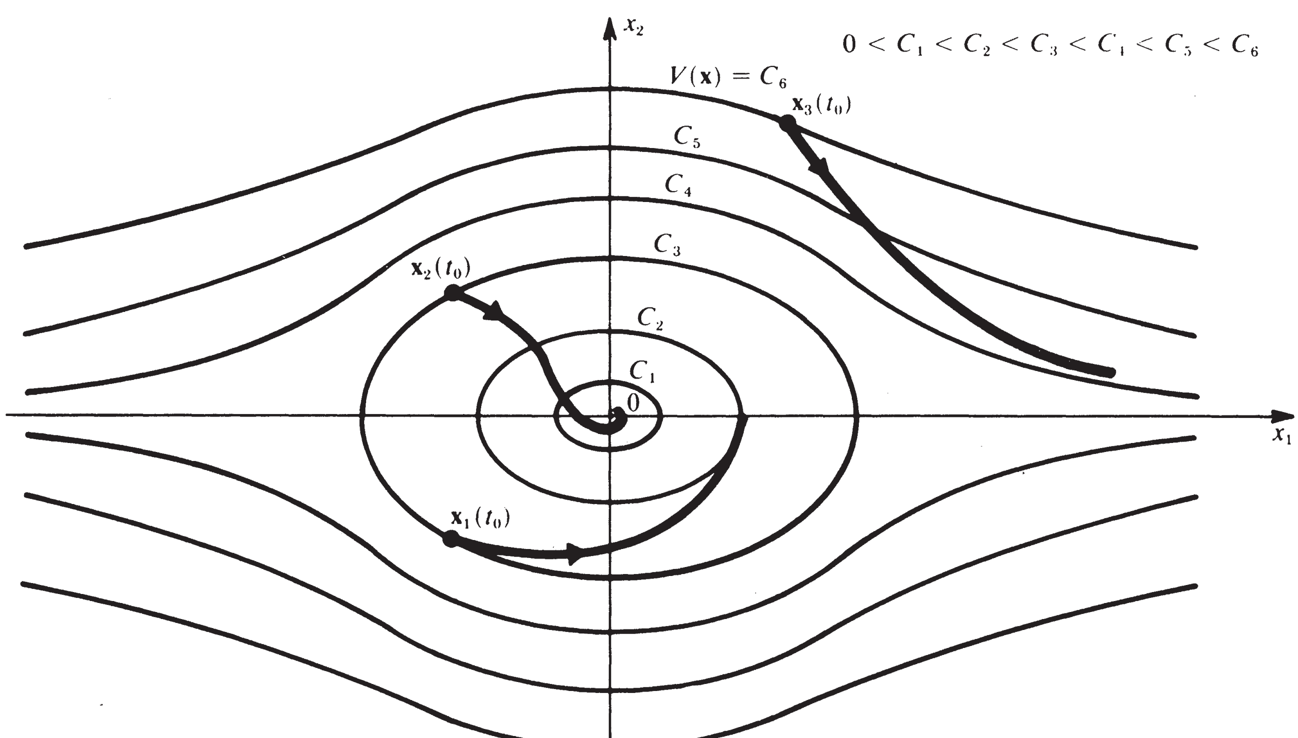


Figure 10.7

increase, but might approach a nonzero constant. If this is true and if the initial state is inside contour C_3, the state will remain bounded but need not approach zero. Initial state $\mathbf{x}_1(t_0)$ illustrates this case and indicates that the origin is stable i.s.L. but not necessarily asymptotically stable. If $\dot{V}$ is only negative semidefinite but is never zero on a solution trajectory (except at the origin), then V will always be decreasing. This rules out the behavior shown with $\mathbf{x}_1(t_0)$, and asymptotic stability is again the conclusion.

Unfortunately, the Lyapunov theorems give no indication of how a Lyapunov function might be found. There is no one unique Lyapunov function for a given system. Some are better than others. It might happen that a $V_1(\mathbf{x})$ can be found which indicates stability i.s.L., $V_2(\mathbf{x})$ might indicate asymptotic stability for initial states quite close to the origin, and $V_3(\mathbf{x})$ might indicate asymptotic stability for a much larger region, or even global asymptotic stability. The inability to find a suitable Lyapunov function does not prove instability. If a system is stable in one of the senses mentioned, it is ensured that an appropriate Lyapunov function does exist. Much ingenuity may be required to find it, however.

There is no universally best method of searching for Lyapunov functions. A form for $V(\mathbf{x})$ can be assumed, either as a pure guess or tempered by physical insight and energy-like considerations. Then $\dot{V}(\mathbf{x})$ can be tested, bringing in the system equations $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$ in the process. Another approach is to assume a form for the derivatives of $V(\mathbf{x})$, either $\dot{V}(\mathbf{x})$ or $\nabla_{\mathbf{x}} V(\mathbf{x})$. Then $V(\mathbf{x})$ can be determined by integration and tested to see if it meets the required conditions. Examples of these techniques are found in the problems for this chapter and, for nonlinear systems, in Chapter 15.

The task of finding suitable Lyapunov functions for linear systems is much more routine. A quadratic form can be used,

$$V(\mathbf{x}) = \mathbf{x}^T \mathbf{P} \mathbf{x}$$

where $\mathbf{P}$ is a real, symmetric, positive definite matrix. Then the time rate-of-change is $\dot{V}(\mathbf{x}) = \dot{\mathbf{x}}^T \mathbf{P} \mathbf{x} + \mathbf{x}^T \mathbf{P} \dot{\mathbf{x}} + \mathbf{x}^T \dot{\mathbf{P}} \mathbf{x}$. Using the linear system state equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$, this becomes

$$\dot{V}(\mathbf{x}) = \mathbf{x}^T[\mathbf{A}^T \mathbf{P} + \mathbf{P}\mathbf{A} + \dot{\mathbf{P}}]\mathbf{x}$$

If the system is asymptotically stable, $\dot{V}$ must be negative definite, and this requires that the matrix equation

$$\mathbf{A}^T \mathbf{P} + \mathbf{P}\mathbf{A} + \dot{\mathbf{P}} = -\mathbf{Q}$$

must be true for some positive definite matrix $\mathbf{Q}$. If $\mathbf{A}$ is constant, then a constant $\mathbf{P}$ will suffice, so $\dot{\mathbf{P}} = \mathbf{0}$. The remaining equation is often called the Lyapunov equation,

$$\mathbf{A}^T \mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{Q} \tag{10.22}$$

Numerical solution of equations of this type were discussed at length in Sec. 6.10. From those results it is recalled that a unique solution $\mathbf{P}$ will exist for each $\mathbf{Q}$ provided that no two eigenvalues of $\mathbf{A}$ satisfy $\lambda_i + \lambda_j = 0$. Therefore, for any arbitrary $\mathbf{Q}$, $\mathbf{P}$ can be found by solving n^2 linear equations (because $\mathbf{P} = \mathbf{P}^T$, only $(n^2 + n)/2$ are really needed):

$$[\mathbf{I}_n \otimes \mathbf{A}^T + \mathbf{A}^T \otimes \mathbf{I}_n](\mathbf{P}) = -(\mathbf{Q}) \tag{10.23}$$

where (**P**) and (**Q**) are the column vectorized versions of **P** and **Q**. One approach to using Eq. (*10.22*) would be to guess a trial positive definite symmetric matrix **P** and then check to see whether the calculated **Q** is positive definite. If it is, asymptotic stability is established. If **Q** is only positive semidefinite, then stability i.s.L. is established. If **Q** is negative definite, instability is concluded. If **Q** is indefinite, *nothing* can be concluded. A more powerful result is obtained by working in the reverse direction, that is, by starting with a positive definite **Q** matrix and using Eq. (*10.22*) to determine **P**. It turns out that any symmetric positive definite **Q** will suffice, and typically $\mathbf{Q} = \mathbf{I}$ is used. The result is formalized as follows.

Theorem 10.4. The constant matrix **A** is asymptotically stable—i.e., has all its eigenvalues strictly in the left-half plane—if and only if the solution **P** of Eq. (*10.22*) is positive definite when **Q** is positive definite.

Note that this is not just a special case of Theorem 10.2 because both necessary and sufficient conditions are given here. In view of the one-to-one relationship between **P** and **Q** established by Eq. (*10.22*), it may seem strange that starting with an arbitrary positive definite **Q** gives a definite answer, whereas starting with an arbitrary positive definite **P** may not. The Venn diagram of Figure 10.8 provides the explanation.

EXAMPLE 10.3 A linear system with $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}$ is known to be asymptotically stable. If $\mathbf{P} = \mathbf{I}_2$ is arbitrarily selected, then Eq. (*10.22*) gives $\mathbf{Q} = -\begin{bmatrix} 0 & -1 \\ -1 & -6 \end{bmatrix}$. **Q** is *not* positive definite. If, instead, $\mathbf{Q} = \mathbf{I}_2$ is selected, then solving Eq. (*10.22*) for **P** gives $\mathbf{P} = \begin{bmatrix} 1.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$, which *is* positive definite. ■

EXAMPLE 10.4 A third-order system expressed in controllable canonical form is

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -a_0 & -a_1 & -a_2 \end{bmatrix} \mathbf{x}$$

The Lyapunov equation can be reduced to

$$\begin{bmatrix} -2a_0P_{13} & P_{11} - a_0P_{23} - a_1P_{13} & P_{12} - a_0P_{33} - a_2P_{13} \\ & 2P_{12} - 2a_1P_{23} & P_{22} + P_{13} - a_1P_{33} - a_2P_{23} \\ & & 2P_{23} - 2a_2P_{33} \end{bmatrix} = -\mathbf{I}_3$$

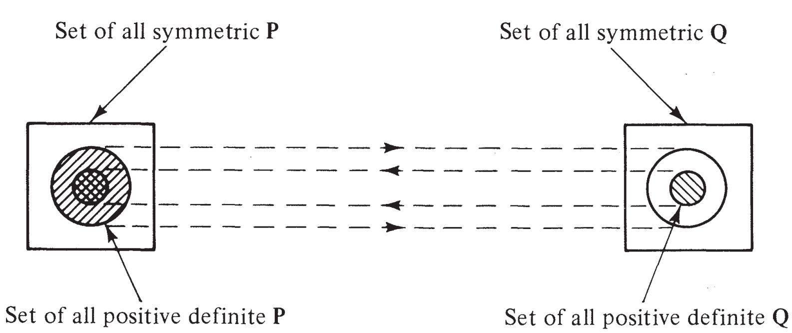


Figure 10.8 The mapping of Eq. (*10.22*) with stable **A**.

(Only the upper triangle of these symmetric equations has been shown, and the symmetry of **P** has been used explicitly.) From the 1, 1 element it is immediate that $P_{13} = 1/(2a_0)$. The five remaining unknowns are related by

$$\begin{bmatrix} 1 & 0 & 0 & -a_0 & 0 \\ 0 & 1 & 0 & 0 & -a_0 \\ 0 & 0 & 1 & -a_2 & -a_1 \\ 0 & 2 & 0 & -2a_1 & 0 \\ 0 & 0 & 0 & 2 & -2a_2 \end{bmatrix} \begin{bmatrix} P_{11} \\ P_{12} \\ P_{22} \\ P_{23} \\ P_{33} \end{bmatrix} = \begin{bmatrix} a_1/(2a_0) \\ a_2/(2a_0) \\ -1/(2a_0) \\ -1 \\ -1 \end{bmatrix}$$

These are solved using Gaussian elimination to give

$$P_{33} = [a_0 a_1 + a_0 + a_2]/[2a_0 a_1 a_2 - 2a_0^2]$$

$$P_{23} = 1/(2a_1) + a_2/(2a_0 a_1) + a_0 P_{33}/a_1$$

$$P_{22} = -1/(2a_0) + a_1 P_{33} + a_2 P_{23}$$

$$P_{12} = a_2/(2a_0) + a_0 P_{33}$$

$$P_{11} = a_1/(2a_0) + a_0 P_{23}$$

The principal minors of **P** are

$$\Delta_1 = P_{11}, \qquad \Delta_2 = P_{22}\Delta_1 - P_{12}^2$$

$$\Delta_3 = P_{33}\Delta_2 - P_{23}(P_{11}P_{23} - P_{12}P_{13}) + P_{13}(P_{12}P_{23} - P_{22}P_{13})$$

The requirement for asymptotic stability is that each of these be positive. This gives expressions somewhat like those of Routh's criterion for checking stability. This example demonstrates that the procedure can become quite messy if carried out analytically. Numerical solution, for a specific system, is quite straightforward. This is demonstrated in the problems. ■

A more general, time-varying system is now considered:

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) \tag{10.24}$$

It is assumed that the origin is an equilibrium point, $\mathbf{f}(\mathbf{0}, t) = \mathbf{0}$ for all t. The previous stability theorems are still basically what is needed to conclude stability. However, now the Lyapunov function and its time derivative may be explicit functions of time as well as the state. The definitions of positive and negative definite must be modified to reflect this added generality. Only Theorem 10.3 is generalized.

Theorem 10.5. If a single-valued scalar function $V(\mathbf{x}, t)$ exists, which is continuous and has continuous first partial derivatives and for which

(1) $V(\mathbf{0}, t) = 0$ for all t;
(2) $V(\mathbf{x}, t) \geq \sigma(\|\mathbf{x}\|) > 0$ for all $\mathbf{x} \neq \mathbf{0}$ and for all t, where $\sigma(\cdot)$ is a continuous, nondecreasing scalar function with $\sigma(0) = 0$;
(3) $\dot{V}(\mathbf{x}, t) \leq -\kappa(\|\mathbf{x}\|) < 0$ for all $\mathbf{x} \neq \mathbf{0}$ and for all t, where $\kappa(\cdot)$ is a continuous nondecreasing scalar function with $\kappa(0) = 0$;
(4) $V(\mathbf{x}, t) \leq \nu(\|\mathbf{x}\|)$ for all $\mathbf{x}$ and t, where $\nu(\cdot)$ is a continuous nondecreasing scalar function with $\nu(0) = 0$;
(5) $\sigma(\|\mathbf{x}\|) \rightarrow \infty$ as $\|\mathbf{x}\| \rightarrow \infty$;

then $\mathbf{x} = \mathbf{0}$ is uniformly globally asymptotically stable.

The functions σ, κ, and ν are all positive definite in the earlier sense, where time did not explicitly appear. $V(\mathbf{x}, t)$ is said to be positive definite if $V(\mathbf{0}, t) = 0$ and if $V(\mathbf{x}, t)$ is always greater than or equal to a time-invariant positive definite function such as σ. Conditions (1) and (2) simply require $V(\mathbf{x}, t)$ to be positive definite. Similarly, condition (3) requires $\dot{V}(\mathbf{x}, t)$ to be negative definite. Condition (5) requires $V(\mathbf{x}, t)$ to become infinite as $\|\mathbf{x}\| \to \infty$, as in Theorem 10.3, and condition (4) prevents $V(\mathbf{x}, t)$ from becoming infinite when $\|\mathbf{x}\|$ is finite. The "uniformly" in the conclusion indicates that the asymptotic stability does not depend on any particular initial time t_0.

There are a number of instability theorems [1, 8] which are useful in avoiding fruitless searches for Lyapunov functions for unstable systems. One typical theorem of this sort follows.

Theorem 10.6. If a scalar function $V(\mathbf{x}, t)$ is continuous, single valued, and has continuous first partial derivatives, and if

(1) $\dot{V}(\mathbf{x}, t)$ is positive definite in some region Ω containing the origin;
(2) $V(\mathbf{0}, t) = 0$ for all t;
(3) $V(\mathbf{x}, t) > 0$ at some point in Ω arbitrarily near the origin;

then the origin is an *unstable* equilibrium point of the system of Eq. (*10.24*).

Lyapunov's stability theorems can also be used to investigate the stability of discrete-time systems:

$$\mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k)) \qquad (10.25)$$

The origin is assumed to be an equilibrium point, so that $\mathbf{f}(\mathbf{0}) = \mathbf{0}$. The preceding theorems for stability, asymptotic stability, and global asymptotic stability apply provided the time derivative $\dot{V}(\mathbf{x})$ is replaced by the first difference, $\Delta V = V(\mathbf{x}(k+1)) - V(\mathbf{x}(k))$.

Lyapunov's direct method of stability analysis is a general method of approach, but its successful use requires considerable ingenuity. Beginning with the very general Lyapunov philosophy, various methods of generating Lyapunov functions and easy-to-use results have been developed by restricting the system under consideration in various ways. The methods of Zubov, Lure, Popov, and others fall into this category [8, 9]. The circle criterion is one popular result of this type, which is in essence a generalized frequency domain criterion similar to the Nyquist criterion for linear systems [10]. The circle criterion is discussed in Chapter 15 in conjunction with nonlinear systems.

10.7 A CAUTIONARY NOTE ON TIME-VARYING SYSTEMS

Many control systems can be modeled as linear systems with time-varying coefficients. Examples include aircraft or spacecraft whose mass decreases as fuel is burned; machines for processing paper, wire, or other materials which cause the driven inertia to change as material is wound on the spool; robotic manipulators with unknown or varying loads; and linearized approximations to nonlinear systems whose coefficients change as they are evaluated at different points along the nominal trajectory.

A commonly used stability analysis technique is the so-called frozen coefficient method, in which all time-varying coefficients are frozen and then the system stability is analyzed as if it were a constant coefficient system. It is almost a folk theorem that if the eigenvalues (poles) are safely within the stability region at all time points (or at least a representative sampling of them) and if the coefficients and eigenvalues are not changing "too rapidly," then the time-varying system can be presumed stable. Empirical evidence suggests that when used with caution, this approach will *usually* give correct results. However, the theoretical basis for the folk theorem and a precise definition of "too rapidly" are not well understood. Total dependence on eigenvalue or pole location can be misleading for time-varying systems. The direct method of Lyapunov can be used to advantage in making stability decisions in these cases. Two carefully selected examples are given to reinforce these words of caution and to give examples of using the Lyapunov techniques.

EXAMPLE 10.5 [6] Consider the unforced time-varying linear system

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} -1 + \alpha\cos^2(t) & 1 - \alpha\sin(t)\cos(t) \\ -1 - \alpha\sin(t)\cos(t) & -1 + \alpha\sin^2(t) \end{bmatrix}\mathbf{x}(t)$$

Find the eigenvalues of the 2×2 $\mathbf{A}(t)$ matrix and then use Lyapunov's direct method to draw conclusions about system stability as a function of the scalar parameter α. The characteristic equation is $|\lambda\mathbf{I} - \mathbf{A}(t)| = 0$ or $\lambda^2 + (2-\alpha)\lambda + (2-\alpha) = 0$, from which the eigenvalues are

$$\lambda_i = -(2-\alpha)/2 \pm \sqrt{(2-\alpha)^2/4 - (2-\alpha)}$$

If α is any real number less than 2, the system has strictly left-half-plane eigenvalues. The eigenvalues are not functions of time, so the question of whether they are moving too rapidly does not apply here. It is not always possible to find the transition matrix for time-varying systems, but in this case it is known to be [6]

$$\mathbf{\Phi}(t, 0) = \begin{bmatrix} e^{(\alpha-1)t}\cos(t) & e^{-t}\sin(t) \\ -e^{(\alpha-1)t}\sin(t) & e^{-t}\cos(t) \end{bmatrix}$$

This can be verified by showing that both $\mathbf{\Phi}(0, 0) = \mathbf{I}$ and $\dot{\mathbf{\Phi}} = \mathbf{A}(t)\mathbf{\Phi}$ are satisfied. From knowledge of $\mathbf{\Phi}$, it is clear that for any real value of $\alpha > 1$, exponential growth occurs and the system is unstable by any definition, even though both eigenvalues are in the left-half plane for $\alpha < 2$.

A quadratic Lyapunov function is adequate for analyzing linear systems, and in this case a very simple one suffices. let $V = \frac{1}{2}\mathbf{x}^T\mathbf{x}$. Then $\dot{V} = \frac{1}{2}\mathbf{x}^T[\mathbf{A}^T(t) + \mathbf{A}(t)]\mathbf{x}$. The quadratic form $\dot{V}$ is negative definite if the matrix $\mathbf{A}_s = \frac{1}{2}[\mathbf{A}^T(t) + \mathbf{A}(t)]$ is negative definite. Since this is a symmetric matrix, its sign definiteness can be checked by using its principal minors or its eigenvalues, as presented in Sec. 7.8. It is worth emphasizing that the system stability can be assured if *all* eigenvalues of $\mathbf{A}_s$—i.e., the symmetric part of $\mathbf{A}(t)$—are in the left-half plane. (The factor $\frac{1}{2}$ is superfluous. It does not change the sign of the eigenvalues.) For this example, the eigenvalues of the symmetric part of $\mathbf{A}(t)$ are $\lambda = -2$ and $-2(1-\alpha)$. Therefore $\dot{V}$ is negative definite for all real $\alpha < 1$. Lyapunov's second method correctly indicates that the system is stable for $\alpha < 1$. If the transition matrix were not known, the question of whether the system might still be stable for some values of α larger than 1 could legitimately be raised because Lyapunov's method does not always give a tight bound. To answer this question one could search for a better Lyapunov function or attempt to use the instability Theorem 10.5. ■

It is not true that the system must be unstable just because **A** (or its symmetric part, $\mathbf{A}_s$) has an eigenvalue in the right-half plane. It is true that if *all* eigenvalues of $\mathbf{A}_s$ are in the right-half plane, the system is unstable. These results are proven as follows. From Problem 7.37 it is seen that

$$\lambda_{\min} \|\mathbf{x}(t)\|^2 \leq \mathbf{x}^T \mathbf{A}_s \mathbf{x} \equiv \dot{V} \leq \lambda_{\max} \|\mathbf{x}(t)\|^2$$

and this remains true for all t if the eigenvalues are time-varying. Dividing by $V \equiv \|\mathbf{x}(t)\|^2$, this inequality can be written

$$\lambda_{\min}\, dt \leq dV/V \leq \lambda_{\max}\, dt$$

Integrating gives $\int \lambda_{\min}\, dt \leq \ln\{V(t)/V(t_0)\} \leq \int \lambda_{\max}\, dt$. When raised to the exponential, this inequality gives

$$\exp\{\textstyle\int \lambda_{\min}\, dt\} \leq V(t)/V(t_0) \leq \exp\{\textstyle\int \lambda_{\max}\, dt\}$$

or

$$\|\mathbf{x}(t_0)\|^2 \exp\{\textstyle\int \lambda_{\min}\, dt\} \leq \|\mathbf{x}(t)\|^2 \leq \|\mathbf{x}(t_0)\|^2 \exp\{\textstyle\int \lambda_{\max}\, dt\}$$

Since $\mathbf{A}_s$ is symmetric, its eigenvalues are all real (Problem 7.25). If $\lambda_{\max}$ of $\mathbf{A}_s$ is negative (then they all are), the integral in the exponent will approach $-\infty$ as $t \to \infty$. Thus $\|\mathbf{x}(t)\| \to 0$, and asymptotic stability is assured. Likewise if $\lambda_{\min}$ is positive (then so are all of the eigenvalues of $\mathbf{A}_s$), the leftmost inequality shows that $\|\mathbf{x}(t)\| \to \infty$, and the system is unstable. Although it is the sign of the time integral of the eigenvalues and not the eigenvalues themselves that is important here, constant bounds on the eigenvalue excursions are easier to work with. These kinds of results are occasionally useful, but they are usually ultraconservative. A system is now demonstrated whose matrix **A** has a right-half plane eigenvalue for all time, and yet the system is asymptotically stable for a certain range of the parameter ω.

EXAMPLE 10.6 Consider the linear, time-varying system

$$\dot{\mathbf{x}} = \begin{bmatrix} -4 - \sqrt{50}\sin(\omega t) & 1 \\ 25\cos(2\omega t) & -4 + \sqrt{50}\sin(\omega t) \end{bmatrix} \mathbf{x}$$

The eigenvalues of $\mathbf{A}(t)$ are $\lambda_1 = 1$ and $\lambda_2 = -9$ for all t. For this system we have

$$\mathbf{A}_s = \begin{bmatrix} -8 + 2\sqrt{50}\sin(\omega t) & 1 + 25\cos(2\omega t) \\ 1 + 25\cos(2\omega t) & -8 + 2\sqrt{50}\sin(\omega t) \end{bmatrix}$$

The eigenvalues of $\mathbf{A}_s$ are given by

$$\lambda = -8 \pm \{200\sin^2(\omega t) + [1 + 25\cos(2\omega t)]^2\}^{1/2}$$

The minimum eigenvalue is clearly always negative. The maximum eigenvalue is approximately bounded by $7.179 < \lambda_{\max} < 19.856$. Yet this system is asymptotically stable for $\omega > 6.14$. This is only an approximate bound, found by numerical integration. Simulation shows that if ω is reduced to 5.9, for instance, the system is definitely unstable—but not in the simple exponential way the eigenvalue of **A** at $+1$ might suggest. When ω is further decreased to about 1, then the response looks more like the expected exponential growth. Even there, the rate of growth does not appear as simple as e^t. When ω is increased to about 6.2, the oscillations are clearly seen to be decaying. Figures 10.9 and 10.10 show the initial condition response of the two state components of **x** for $\omega = 5$ and $\omega = 6.8$, respectively. ■

Even though the previous system has a right-half plane eigenvalue, it is still stable if the oscillations are fast enough. Note that the folk theorem states that the frozen coefficient analysis based on eigenvalue locations will usually give the correct stability result if the system parameters are not changing too rapidly. If the oscillations in this system are slowed down by reducing ω, then the eigenvalue locations (one unstable) do correctly predict an unstable system. This example gives a quantitative meaning to *too rapidly* for this problem. Also notice that the simple bound using the integral of the maximum eigenvalue was *not* evaluated explicitly. Numerical integration indicates that the bounds will not be satisfied and therefore cannot be used to establish the stability regime. The simple use of the constant upper limit for λ_{max} is even more obviously of no help here. The existence of a suitable Lyapunov function is guaranteed. Finding it will require looking beyond the simple unit diagonal form used earlier.

10.8 USE OF LYAPUNOV'S METHOD IN FEEDBACK DESIGN

Stability is an overriding requirement in almost all control system designs. It is possible to use the Lyapunov method directly in the design of the feedback controller in some situations. Since response time or settling time is frequently involved in the system-design specifications, it is first noted that the Lyapunov function and its derivative can be used to estimate the system's speed of response, i.e., its dominant time constant [2, 3, 9]. Assume that the origin is a stable equilibrium point. Let $V(\mathbf{x}, t)$ be a Lyapu-

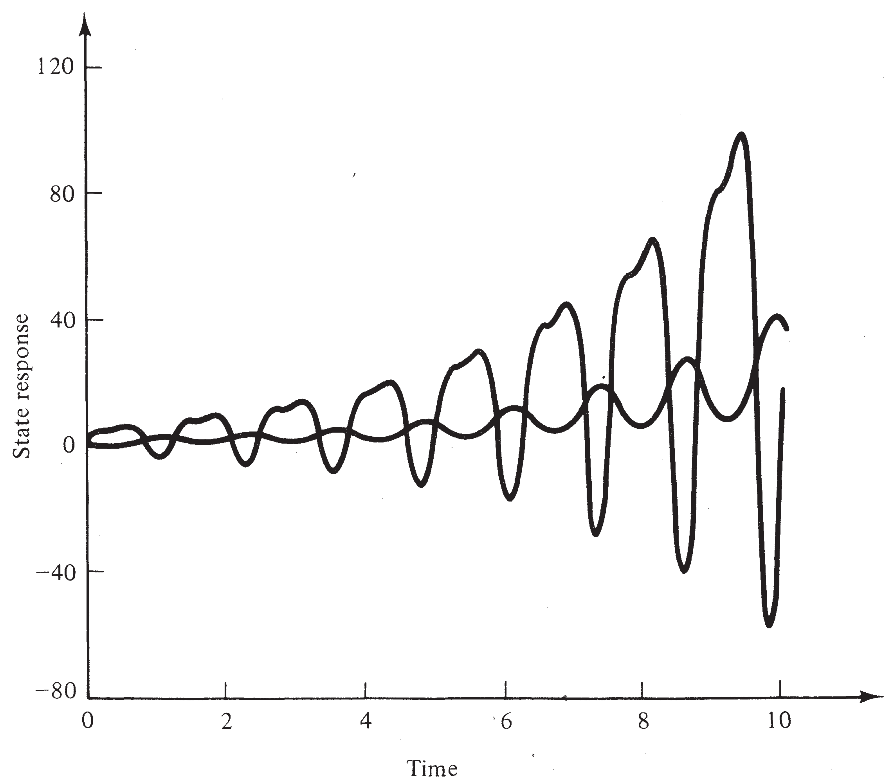


Figure 10.9 Response with $\omega = 5$.

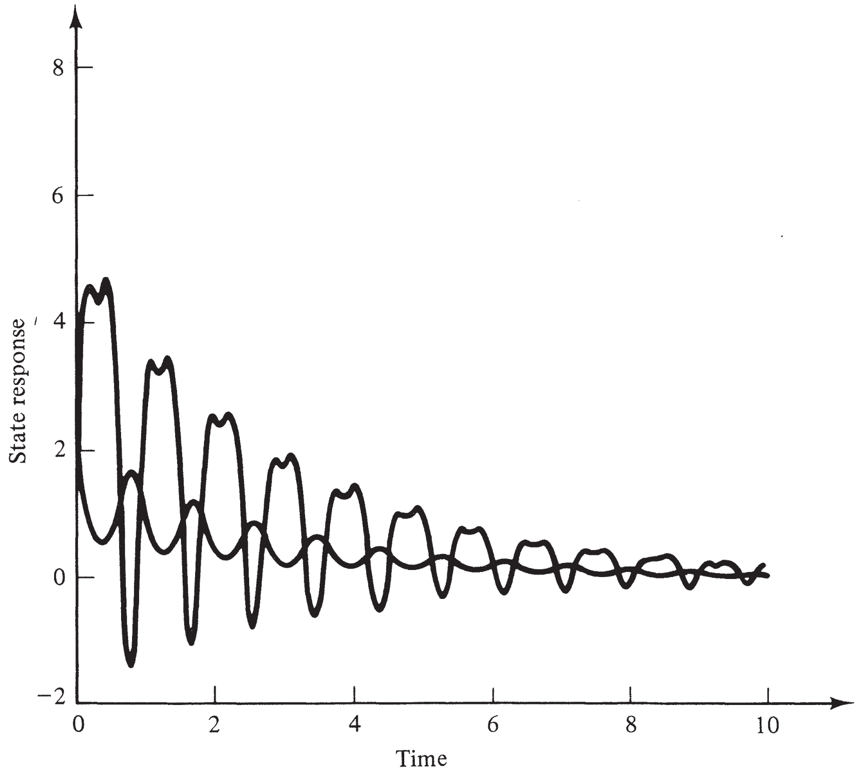


Figure 10.10 Response with $\omega = 6.8$.

nov function, and define $\eta = -\dot{V}(\mathbf{x}, t)/V(\mathbf{x}, t)$. Integrating both sides of this definition gives

$$\int_{t_0}^{t} \eta\, dt = -\int_{V(\mathbf{x}(t_0), t_0)}^{V(\mathbf{x}(t), t)} \frac{dV}{V}$$

from which

$$\ln\left[\frac{V(\mathbf{x}(t), t)}{V(\mathbf{x}(t_0), t_0)}\right] = -\int_{t_0}^{t} \eta\, dt \quad \text{or} \quad V(\mathbf{x}(t), t) = V(\mathbf{x}(t_0), t_0)e^{-\int_{t_0}^{t} \eta\, dt}$$

Although η is not generally constant, if $\eta_{\min}$ is defined as the minimum value of η, then

$$V(\mathbf{x}(t), t) \le V(\mathbf{x}(t_0), t_0)e^{-\eta_{\min}(t - t_0)}$$

For an asymptotically stable system, $\mathbf{x}(t) \to \mathbf{0}$ as $V(\mathbf{x}, t) \to 0$. It is seen that $1/\eta_{\min}$ is an approximate bound on the energy decay rate time constant. This is sometimes taken as a bound on the system's dominant time constant. However, since energy (and frequently the generalized energy V) is quadratic in $\mathbf{x}$, $V \approx V_0 e^{-\eta t}$ suggests that $\mathbf{x}(t) \approx \mathbf{x}(0)e^{-\eta t/2}$ might be a better bound on the state behavior. From this it is suggested that $2/\eta_{\min}$ be used as an approximation of the system's dominant time constant. This will be used later in designing a controller with a settling time specification. With or without the factor of 2, when several candidate Lyapunov functions are under consideration, it is seen that the one with the larger ratio $-\dot{V}/V$ will give the faster time response.

Since $V(\mathbf{x}, t)$ does not depend on $\dot{\mathbf{x}}$ and hence on $\mathbf{u}$—but $\dot{V}$ does—the following controller design procedure can sometimes be used to advantage. First select a Lyapunov function for the homogeneous system equations. This V should be positive definite, and $\dot{V}$ should be at least negative semidefinite. This assumes that the uncontrolled system is at least stable i.s.L. Then allow the control variable **u** to be nonzero and add these control-dependent terms to form a modified $\dot{V}$ term. V itself will not change. With V known from the unforced analysis, the only unknowns in the modified $\dot{V}$ are terms containing **u**. These can be selected to force $\dot{V}$ to be as negative as possible within available control limits. This would yield a control system which approximates a minimum time response system. It typically leads to a nonlinear control law, even when the original system is linear.

EXAMPLE 10.7 Consider the system

$$\dot{\mathbf{x}} = \begin{bmatrix} -3 & 2 \\ -1 & -1 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}\mathbf{u}(t)$$

The input control vector is limited by $\mathbf{u}^T(t)\mathbf{u}(t) \leq 1$ for all time t. Use the quadratic Lyapunov function found for this system in Problem 10.10. Specify the feedback control law for computing $\mathbf{u}(t)$ as a function of $\mathbf{x}(t)$ which will make the modified $\dot{V}$ as negative as possible. This controller will rapidly drive the state back to $\mathbf{x}(t) = \mathbf{0}$ if it is perturbed by initial conditions or disturbances.

Using the Lyapunov function of Problem 10.10, $V(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x}$ and then $\dot{V}(\mathbf{x}) = \dot{\mathbf{x}}^T\mathbf{P}\mathbf{x} + \mathbf{x}^T\mathbf{P}\dot{\mathbf{x}} = -\mathbf{x}^T\mathbf{x} + 2\mathbf{u}^T\mathbf{B}^T\mathbf{P}\mathbf{x}$. The control $\mathbf{u}(t)$ is selected to minimize $\mathbf{u}^T\mathbf{B}^T\mathbf{P}\mathbf{x}$ subject to the restriction placed upon $\mathbf{u}(t)$. Obviously $\mathbf{u}(t)$ should be parallel to $\mathbf{B}^T\mathbf{P}\mathbf{x}$, but with the opposite sign, and have its largest possible magnitude. That is,

$$\mathbf{u}(t) = \frac{-\mathbf{B}^T\mathbf{P}\mathbf{x}(t)}{\|\mathbf{B}^T\mathbf{P}\mathbf{x}(t)\|}$$

Using the given form for the matrices **B** and **P**, this gives

$$\mathbf{u}(t) = \frac{\begin{bmatrix} -7x_1(t) + x_2(t) \\ 8x_1(t) - 19x_2(t) \end{bmatrix}}{(113x_1^2 - 318x_1x_2 + 362x_2^2)^{1/2}}$$

■

Other control system designs might require a specified response time rather than the maximal effort, minimum time response just demonstrated. One such example follows.

EXAMPLE 10.8 Consider the linearized equations of motion for the spin-stabilized satellite of Problem 3.16. Assume that the value of Ω is 8 rad/min. Design a linear controller which will drive any wobbling errors to zero with a time constant of 0.5 min.

The equations of motion are rewritten here as

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & -\Omega \\ \Omega & 0 \end{bmatrix}\mathbf{x} + \mathbf{u}(t)$$

A quadratic Lyapunov function $V = \mathbf{x}^T\mathbf{P}\mathbf{x}$ is selected, and $\mathbf{P} = \mathbf{I}_2$ is adequate for present purposes. Then $\dot{V} = \mathbf{x}^T[\mathbf{A}^T + \mathbf{A}]\mathbf{x} + \mathbf{u}^T\mathbf{x} + \mathbf{x}^T\mathbf{u}$. With $\mathbf{u} = \mathbf{0}$, $\dot{V}$ is identically zero, due to the skew-symmetric matrix **A**. The conclusion to be drawn from this particular Lyapunov function is that the uncontrolled system is stable i.s.L. but not necessarily asymptotically stable. Actually, this was known at the outset, since no energy dissipation has been included in the model. If the linear controller $\mathbf{u} = -k\mathbf{x}$ is selected, then the **u**-dependent terms give the modified $\dot{V} = -2k\mathbf{x}^T\mathbf{x}$. This is negative definite for any positive real scalar k. The ratio $\eta = -\dot{V}/V = 2k$ can be used to select k in order to meet the settling time-constant specification $\tau = 0.5 \approx 2/(2k)$. Therefore, $k = 2$ is selected, meaning that the feedback control law is $\mathbf{u}(t) = -2\mathbf{x}(t)$. Figure 10.11 shows the time response due to a unit initial condition error. The $\pm e^{-2t}$ decay envelope is also included.

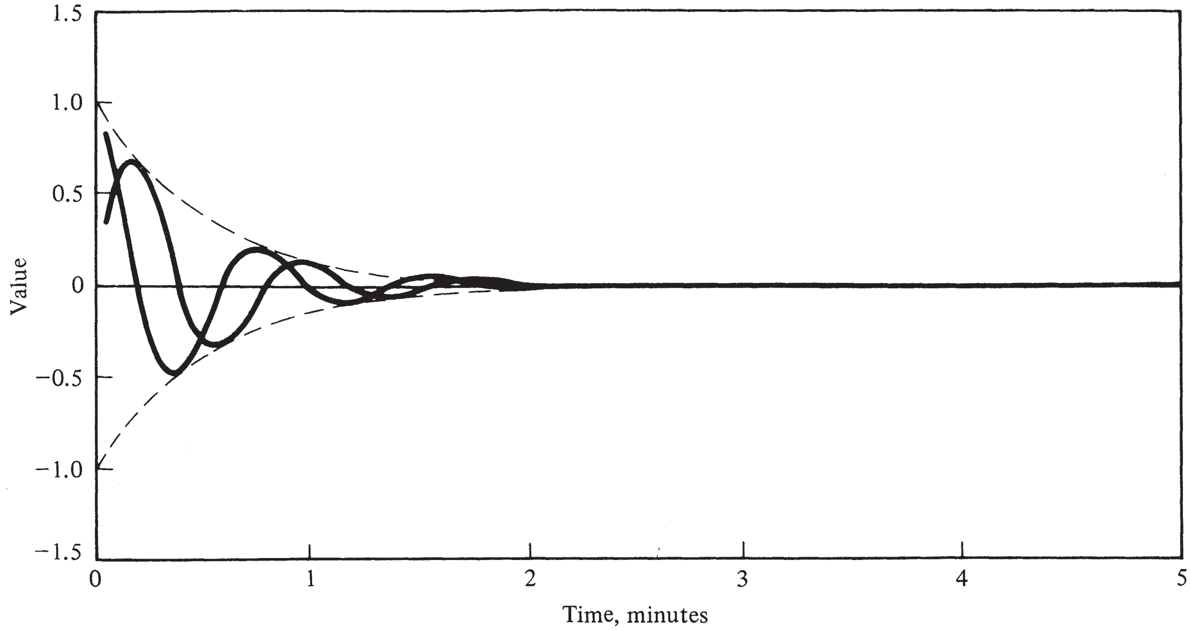


Figure 10.11 Satellite wobble control example, Example 10.7, Linear Feedback $u = -2x$.

REFERENCES

1. LaSalle, J. and S. Lefschetz: *Stability by Liapunov's Direct Method, with Applications,* Academic Press, New York, 1961.
2. Kalman, R. E. and J. E. Bertram: "Control System Analysis and Design Via the Second Method of Lyapunov I. Continuous-Time Systems," *Trans. ASME Journal of Basic Engineering,* Vol. 82D, June 1960, pp. 371–393.
3. Kalman, R. E. and J. E. Bertram: "Control System Analysis and Design Via the Second Method of Lyapunov II. Discrete-Time Systems," *Trans. ASME Journal of Basic Engineering,* Vol. 82D, June 1960, pp. 394–400.
4. Guckenheimer, J. and P. Holmes: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields,* Springer-Verlag, New York, 1983.
5. Schultz, D. G. and J. E. Gibson: "The Variable Gradient Method for Generating Liapunov Functions," *AIEE Trans.,* Part II, Vol. 81, Sept. 1962, pp. 203–210.
6. Desoer, C. A.: *Notes for a Second Course on Linear Systems,* Van Nostrand Reinhold, New York, 1970.
7. Coddington, E. A. and N. Levinson: *Theory of Ordinary Differential Equations,* McGraw-Hill, New York, 1955.
8. Willems, J. L.: *Stability Theory of Dynamical Systems,* John Wiley, New York, 1970.
9. Ogata, K.: *State Space Analysis of Control Systems,* Prentice Hall, Englewood Cliffs, N.J., 1967.
10. Atherton, D. P.: *Nonlinear Control Engineering,* Van Nostrand Reinhold, New York, 1975.
11. Reiss, R. and G. Geiss: "The Construction of Liapunov Functions," *IEEE Transactions on Automatic Control,* Vol. AC-8, No. 4, Oct. 1963, pp. 382–383.
12. Puri, N. N. and C. N. Weygandt: "Second Method of Liapunov and Routh's Canonical Form," *Journal of The Franklin Institute,* Vol. 276, No. 5, 1963, pp. 365–384.
13. Ku, Y. H. and N. N. Puri: "On Liapunov Functions of High Order Nonlinear Systems," *Journal of The Franklin Institute,* Vol. 276, No. 5, 1963, pp. 349–364.

ILLUSTRATIVE PROBLEMS

Constant, Linear System Stability

10.1 Comment on the stability of the four constant coefficient systems of Problem 3.1, page 104.

(a) The eigenvalue is $\lambda = -\alpha$ and the system is asymptotically stable if $\lambda < 0$. If $\lambda = 0$, the system is stable in the sense of Lyapunov. If $\lambda > 0$, it is unstable.

(b) The eigenvalue is again $\lambda = -\alpha$ and the results of part (a) still apply.

(c) The eigenvalues of the 2×2 **A** matrix satisfy $\lambda^2 + 2\zeta\omega\lambda + \omega^2 = 0$. They are $\lambda = -\zeta\omega \pm \omega\sqrt{\zeta^2 - 1}$. Both eigenvalues are negative for any positive value of ζ and then the system is asymptotically stable. This assumes that ω is a positive constant. If $\zeta = 0$, then $\lambda = \pm j\omega$ and the system is stable i.s.L. For $\zeta < 0$, the system is unstable.

(d) The results of part (c) again apply.

These examples illustrate that the poles of a transfer function are eigenvalues of the matrix **A** in the state space representation. This will always be true, but the converse need not be true.

10.2 Is the system of Problem 3.2, page 105, stable? Is it asymptotically stable?

Evaluating $|\mathbf{A} - \mathbf{I}\lambda| = 0$ gives $\lambda^4 + 3\lambda^3 + 3\lambda^2 + \lambda = 0$. Thus one eigenvalue is $\lambda = 0$. The remaining cubic $\lambda^3 + 3\lambda^2 + 3\lambda + 1 = 0$ can be investigated using Routh's criterion. This indicates that there are no right-half-plane eigenvalues. Thus the system is stable, but not asymptotically stable because of the root $\lambda = 0$.

10.3 **(a)** Is the system of Problem 3.3, page 106, asymptotically stable?

(b) Is the system of Problem 3.4, page 107, asymptotically stable?

(a) All three of the forms given for **A** have the same eigenvalues, $\lambda = -3, -1$, and -6. (These are also poles of the original transfer function.) This system is asymptotically stable.

(b) The 6×6 **A** matrix is given in Jordan form, from which it is apparent that the eigenvalues are $0, 0, -3, -3, -3, -1$. This system is unstable because of the double root $\lambda = 0$. With $\mathbf{u} = \mathbf{0}$, $\dot{x}_2 = 0$ or $x_2 =$ constant. Since $\dot{x}_1 = x_2$, $x_1(t)$ is a linear function of time. Thus unless $x_2(0) = 0$, $x_1(t) \to \pm\infty$ and as a result $\|\mathbf{x}(t)\| \to \infty$.

10.4 Investigate the stability of the continuous-time system and its discrete-time approximation, as described in Problem 9.12, page 331.

The continuous-time system matrix **A** has eigenvalues $\lambda = 0, -1$. Therefore, the system is stable in the sense of Lyapunov but not asymptotically stable. The discrete-time system matrix **A** has eigenvalues $\lambda_1 = 0.368$ and $\lambda_2 = 1$. Since $|\lambda_1| < 1$, but $|\lambda_2| = 1$, the discrete system is stable in the sense of Lyapunov, but not asymptotically stable.

10.5 Comment on the stability of the linear discrete-time systems described in

(a) Problem 9.20, page 335;

(b) Problem 9.21, page 335;

(c) Problem 9.23, page 336.

(a) Since the **A** matrix is triangular, its eigenvalues are found by inspection as $\lambda_1 = \lambda_2 = \lambda_3 = \frac{1}{2}$. Since all $|\lambda_i| < 1$, the system is asymptotically stable. This is verified by the explicit solution found in Problem 9.20, which shows that $\mathbf{x}(k) \to \mathbf{0}$ as $k \to \infty$.

(b) The eigenvalues were given earlier as $\lambda_1 = \frac{1}{2}, \lambda_2 = \frac{1}{3}$. Since $|\lambda_i| < 1$, the system is asymptotically stable.

(c) The eigenvalues are $\lambda_1 = 1, \lambda_2 = \frac{1}{2}, \lambda_3 = -\frac{1}{3}$. This system is stable i.s.L., but not asymptotically stable, since $\lambda_1 = 1$.

10.6 Show that if a continuous-time, linear constant system is asymptotically stable, then its adjoint is unstable.

Asymptotic stability of $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ implies that all eigenvalues of **A** have negative real parts. The adjoint equation is $\dot{\mathbf{z}} = -\mathbf{A}^T\mathbf{z}$. The eigenvalues of $-\mathbf{A}^T$ are the roots of $|-\mathbf{A}^T - \mathbf{I}\gamma| = 0$ or $|\mathbf{A}^T - \mathbf{I}(-\gamma)| = \mathbf{0}$. Since $\mathbf{A}^T$ and **A** have the same eigenvalues, the roots γ are the negative of the eigenvalues of **A**. Then **A** having all left half plane eigenvalues implies that $-\mathbf{A}^T$ has all right half plane eigenvalues, and the adjoint system is unstable.

10.7 Show that if a continuous-time, linear constant system is asymptotically stable, it is also BIBS stable.

The conditions for BIBS stability are given by Eqs. (*10.6*) and (*10.8*). If the system is asymptotically stable, then $\|\mathbf{\Phi}(t, t_0)\| \leq N(t_0)$ for all $t \geq t_0$. Furthermore, using norm inequalities gives

$$\int_{t_0}^{t} \|\mathbf{\Phi}(t, \tau)\mathbf{B}\|\, d\tau \leq \int_{t_0}^{t} \|\mathbf{\Phi}(t, \tau)\| \cdot \|\mathbf{B}\|\, d\tau$$

Since $\mathbf{B}$ and hence $\|\mathbf{B}\|$ are constant, this term can be taken outside the integral. Also, expressing $\mathbf{\Phi}$ in terms of the Jordan form and modal matrix of $\mathbf{A}$ (see page 317) leads to

$$\int_{t_0}^{t} \|\mathbf{\Phi}(t, \tau)\mathbf{B}\|\, d\tau \leq \|\mathbf{M}^{-1}\| \cdot \|\mathbf{M}\| \int_{t_0}^{t} \|e^{\mathbf{J}(t-\tau)}\|\, d\tau \|\mathbf{B}\|$$

If all the eigenvalues of $\mathbf{A}$ are distinct, then $\|e^{\mathbf{J}(t-\tau)}\| = e^{\beta_i(t-\tau)}$, where β_i is the largest real part of the eigenvalues of $\mathbf{A}$. But β_i is negative since the system is asymptotically stable. In this case the integral is bounded and thus BIBS stability follows. If $\mathbf{A}$ has repeated eigenvalues, then $\|e^{\mathbf{J}(t-\tau)}\|$ can be bounded by $\sqrt{p((t)}\, e^{\beta_i(t-\tau)}$, where $p(t)$ is a polynomial in t. Asymptotic stability ensures that $\beta_i < 0$, and again the integral is bounded. Therefore, for linear constant systems, asymptotic stability implies BIBS stability.

10.8 Show that if a continuous-time, linear constant system is asymptotically stable, it is also BIBO stable.

The system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$, $\mathbf{y} = \mathbf{C}\mathbf{x} + \mathbf{D}\mathbf{u}$ has the output solution

$$\mathbf{y}(t) = \mathbf{C}\mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{C}\mathbf{\Phi}(t, \tau)\mathbf{B}\mathbf{u}(\tau)\, d\tau + \mathbf{D}\mathbf{u}(t)$$

Since the norm of any constant matrix is bounded, straightforward application of norm inequalities leads to

$$\|\mathbf{y}(t)\| \leq \|\mathbf{C}\|\, \|\mathbf{\Phi}(t, t_0)\|\, \|\mathbf{x}(t_0)\| + \|\mathbf{C}\| \cdot \|\mathbf{B}\| K \int_{t_0}^{t} \|\mathbf{\Phi}(t, \tau)\|\, d\tau + \|\mathbf{D}\| K$$

where $\|\mathbf{u}(t)\| \leq K$ for all t. Asymptotic stability ensures that $\|\mathbf{\Phi}(t, t_0)\|$ is bounded by a decaying exponential and thus $\int_{t_0}^{t} \|\mathbf{\Phi}(t, \tau)\|\, d\tau$ is also bounded for all $t \geq t_0$. It follows that the output is bounded in norm also.

It is also known [8] that *if the system is completely controllable and completely observable,* then the implication can also be reversed. In that case asymptotic stability $\Leftrightarrow$ BIBO stability.

Lyapunov Methods

10.9 Verify the stability of the third-order system of Problem 3.3. This is of the same form as the problem treated in Example 10.3 with parameter values $a_0 = 18$, $a_1 = 27$, and $a_2 = 10$. Numerical solution of the Lyapunov equation (*10.22*) with $\mathbf{Q} = \mathbf{I}_3$, using the full 9×9 Kronecker product form of Eq. (*10.23*), yields

$$\mathbf{P} = \begin{bmatrix} 1.94841 & 1.297619 & 0.027777 \\ 1.297619 & 2.167769 & 0.066578 \\ 0.027777 & 0.066578 & 0.056657 \end{bmatrix}$$

From this the principal minors are $\Delta_1 = 1.94841$, $\Delta_2 = 2.53989$, and $\Delta_3 = 0.138395$. Since they are all positive, $\mathbf{P}$ is positive definite and the system is asymptotically stable. This verifies a known result, since the eigenvalues of $\mathbf{A}$ were given in Problem 3.3 as $\lambda = -1, -3$, and -6.

10.10 Investigate the stability of the system described by

$$\dot{\mathbf{x}} = \begin{bmatrix} -3 & 2 \\ -1 & -1 \end{bmatrix}\mathbf{x} \qquad (1)$$

Let $\mathbf{P} = [P_{ij}]$ with $P_{12} = P_{21}$. Then

$$\mathbf{A}^T\mathbf{P} + \mathbf{PA} = \begin{bmatrix} -6P_{11} - 2P_{12} & -4P_{12} - P_{22} + 2P_{11} \\ -4P_{12} - P_{22} + 2P_{11} & 4P_{12} - 2P_{22} \end{bmatrix}$$

Using $\mathbf{Q} = \mathbf{I}_2$, the unit matrix, and solving $\mathbf{A}^T\mathbf{P} + \mathbf{PA} = -\mathbf{Q}$ gives $\mathbf{P} = \begin{bmatrix} \frac{7}{40} & -\frac{1}{40} \\ -\frac{1}{40} & \frac{18}{40} \end{bmatrix}$. The principal minors of $\mathbf{P}$ are $\Delta_1 = \frac{7}{40} > 0$ and $\Delta_2 = |\mathbf{P}| = \frac{5}{64} > 0$. Therefore, $\mathbf{P}$ is positive definite and the system is asymptotically stable.

10.11 Derive conditions for asymptotic stability of the origin for the following systems. Use Lyapunov's direct method and proceed by assuming a suitable form for $\dot{V}(\mathbf{x})$. $V(\mathbf{x})$ is then found by integration.

(a) $\dot{x} = ax$

(b) $\dot{x}_1 = x_2, \dot{x}_2 = -ax_1 - bx_2$

(c) $\dot{x}_1 = x_2, \dot{x}_2 = x_3, \dot{x}_3 = -ax_1 - bx_2 - cx_3$

A form for $\dot{V}$ is assumed, which is at least negative semidefinite. A choice which is often successful is $\dot{V} = -x_n^2$ [11]. Then $V(\mathbf{x}(t)) - V(\mathbf{x}(t_1)) = \int_{t_1}^{t} \dot{V}(\mathbf{x})\,dt$. The lower limit is selected so that $\mathbf{x}(t_1) = \mathbf{0}$ and $V(\mathbf{0}) = 0$. If the $V(\mathbf{x})$ found in this manner is positive definite, then the method is successful.

(a) Try $\dot{V} = -x^2$. Assuming $a \neq 0$, this gives $\dot{V} = -x\dot{x}/a$. Then

$$V(x) = -\frac{1}{a}\int_{t_1}^{t} x\dot{x}\,dt = -\frac{1}{a}\int_0^x x\,dx = -\frac{x^2}{2a}$$

$V(x)$ is positive definite if $a < 0$. Since $\dot{V}$ is negative definite, asymptotic stability results if $a < 0$.

(b) Try $\dot{V} = -x_2^2 = (-x_2)x_2$. If $a \neq 0$, then $-x_2 = (\dot{x}_2 + ax_1)/b$ so that $\dot{V} = \dot{x}_2 x_2/b + ax_1x_2/b$. Using $x_2 = \dot{x}_1$ gives

$$V(\mathbf{x}) = (a/b)\int_0^{x_1} x_1\,dx_1 + (1/b)\int_0^{x_2} x_2\,dx_2 = \frac{ax_1^2}{2b} + \frac{x_2^2}{2b}$$

If $a > 0$ and $b > 0$, $V(\mathbf{x})$ is positive definite. $\dot{V}$ is negative semidefinite, but is never zero on any trajectory of this system except at $\mathbf{x} = \mathbf{0}$. Therefore, $a > 0$ and $b > 0$ ensure asymptotic stability.

(c) Try $\dot{V} = -x_3^2$. Then $V(\mathbf{x}) = -\int_{t_1}^{t} x_3^2\,dt = -\int_{t_1}^{t} x_3\dot{x}_2\,dt$. Using integration by parts, $V(\mathbf{x}) = -x_3x_2 + \int x_2\dot{x}_3\,dt$. Using the differential equation to replace $\dot{x}_3$,

$$V(\mathbf{x}) = -x_3x_2 - \int x_2(ax_1 + bx_2 + cx_3)\,dt = -x_3x_2 - \frac{ax_1^2}{2} - b\int x_2^2\,dt - \frac{cx_2^2}{2}$$

The integral term is evaluated, using $\dot{x}_1 = x_2$ and integration by parts:

$$b\int x_2^2\,dt = bx_1x_2 - b\int x_1\dot{x}_2\,dt = bx_1x_2 - b\int x_1x_3\,dt$$

From the differential equation, $-x_1 = (1/a)(\dot{x}_3 + bx_2 + cx_3)$. Hence

$$b\int x_2^2\,dt = bx_1x_2 + \frac{b}{2a}x_3^2 + \frac{b^2}{2a}x_2^2 + (bc/a)\int_{t_2}^{t} x_3^2\,dt$$

which gives

$$V(\mathbf{x}) = -\frac{a}{2}\left(x_1 + \frac{b}{a}x_2\right)^2 - \frac{b}{2a}\left(x_3 + \frac{a}{b}x_2\right)^2 - \frac{c - a/b}{2}x_2^2 - \frac{bc}{a}\int x_3^2\,dt$$

This expression is *not* positive definite. In fact, it can be made negative definite. A new trial function is selected as $V'(\mathbf{x}) = -V(\mathbf{x}) - (bc/a)\int_{t_1}^{t} x_3^2\,dt$. Then

$$\dot{V}'(\mathbf{x}) = -\dot{V}(x) - (bc/a)x_3^2 = -(bc/a - 1)x_3^2$$

$\dot{V}'(\mathbf{x})$ is negative semidefinite if $bc/a > 1$. $V'(\mathbf{x})$ is positive definite if, in addition, $a > 0, b > 0, c > 0$. Since $\dot{V}'(\mathbf{x})$ can never vanish on a trajectory of this system, the conditions for asymptotic stability are $a > 0$, $b > 0$, $c > 0$, and, $bc - a > 0$.

In all three cases the well-known results of the Routh stability criterion have been determined. The procedure illustrated by these examples can be extended to nonlinear problems [11].

10.12 Assume that a system is described by a matrix $\mathbf{A}(t)$ that is composed of a sign-definite matrix $\mathbf{A}_1(t)$ plus a skew-symmetric term $\mathbf{S}(t)$. Show that the system is asymptotically stable if $\mathbf{A}_1$ is negative definite and unstable if $\mathbf{A}_1$ is positive definite.

Use the quadratic Lyapunov function $V = \mathbf{x}^T\mathbf{P}\mathbf{x}$, with $\mathbf{P} = \mathbf{I}$. Then $\dot{V} = \mathbf{x}^T[\mathbf{A}^T + \mathbf{A}]\mathbf{x} \equiv \mathbf{x}^T[\mathbf{A}_1^T + \mathbf{A}_1]\mathbf{x}$, since $\mathbf{S}^T + \mathbf{S} = \mathbf{0}$. Thus by Theorem 10.2 (or Theorem 10.4), the system is asymptotically stable if $\mathbf{A}_1$ is negative definite. By Theorem 10.6 the system is unstable if $\mathbf{A}_1$ is positive definite. This result also shows that many (but not all) of the surprises regarding stability and eigenvalue locations in the linear time-varying case of Sec. 10.7 can be resolved by examining the eigenvalues of $\mathbf{A}_s = \mathbf{A}^T + \mathbf{A}$ rather than the eigenvalues of $\mathbf{A}$.

10.13 Find a constant feedback gain matrix which will ensure asymptotic stability for the system of Example 10.6, regardless of the frequency ω.

If a feedback term $\mathbf{u} = -K\mathbf{x} = -k\mathbf{I}_2\mathbf{x}$ is added to the system equation in Example 10.6, the $\mathbf{A}_s$ matrix is modified to

$$\mathbf{A}_s = \begin{bmatrix} -8 - 2k + 2\sqrt{50}\,\sin(\omega t) & 1 + 25\cos(2\omega t) \\ 1 + 25\cos(2\omega t) & -8 - 2k + 2\sqrt{50}\,\sin(\omega t) \end{bmatrix}$$

This means that the eigenvalues are both shifted to the left by $2k$. From the bound previously given for λ_{max}, it is clear that both eigenvalues will always remain in the left-hand plane for $k > 19.856/2 = 9.928 \approx 10$. This will ensure asymptotic stability.

10.14 Describe a method of generating Lyapunov functions, beginning with an assumed form for the gradient $\nabla_x V$. This method is known as the variable gradient method [5, 8, 9].

Consider a general class of nonlinear, autonomous systems, $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, for which $\mathbf{f}(\mathbf{0}) = \mathbf{0}$. Then for any candidate Lyapunov function $V(\mathbf{x})$,

$$\frac{dV(\mathbf{x})}{dt} = [\nabla_{\mathbf{x}} V]^T\dot{\mathbf{x}} = [\nabla_{\mathbf{x}} V]^T\mathbf{f}(\mathbf{x}) \tag{1}$$

A general form for the gradient is assumed, such as

$$\nabla_{\mathbf{x}} V = \begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n \\ \vdots \qquad\qquad \vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n \end{bmatrix} \tag{2}$$

The coefficients a_{ij} need not be constants, but may be functions of the components of $\mathbf{x}$. The function $V(\mathbf{x})$ can be determined by evaluating the line integral in state space from $\mathbf{0}$ to a general point $\mathbf{x}$, $V(\mathbf{x}) = \int_0^{\mathbf{x}} (\nabla_{\mathbf{x}} V)^T\,d\mathbf{x}$. The line integral can be made independent of the path of integration if a set of generalized curl conditions is imposed.

Let the ith component of $\nabla_{\mathbf{x}} V$ be called ∇V_i. Then the curl requirements are that $\partial \nabla V_i/\partial x_j = \partial \nabla V_j/\partial x_i$ for $i, j = 1, 2, \ldots, n$. That is, the matrix of second partials $\nabla_{\mathbf{x}}(\nabla_{\mathbf{x}} V)^T = [\partial^2 V/\partial x_i\,\partial x_j]$ must be symmetric. Satisfying this requirement allows the line integral to be written as the sum of n simple scalar integrals:

$$V(\mathbf{x}) = \int_0^{x_1} \nabla V_1\,dx_1\Big|_{x_2 = x_3 = \cdots = x_n = 0} + \int_0^{x_2} \nabla V_2\,dx_2\Big|_{\substack{x_1 = x_1 \\ x_3 = x_4 = \cdots = x_n = 0}} + \cdots$$

$$+ \int_0^{x_n} \nabla V_n\,dx_n\Big|_{\substack{x_1 = x_1 \\ x_2 = x_2 \\ \cdots\cdots}} \tag{3}$$

The procedure then consists of assuming $\nabla_{\mathbf{x}} V$ as in Eq. (*2*), specifying the coefficients a_{ij} such that (1) $\dot{V}(\mathbf{x})$, Eq. (*1*), is at least negative semidefinite and (2) the generalized curl equations are satisfied. Then $V(\mathbf{x})$ is found as in Eq. (*3*) and checked for positive definiteness. Usually it will be possible to let many a_{ij} terms equal zero. Generally $a_{nn}=1$ is chosen at the outset to ensure that $V(\mathbf{x})$ is quadratic in x_n. This technique is illustrated in Chapter 15 on nonlinear systems. It can also be used on linear systems, either constant or time-varying.

10.15 Use the variable gradient method to determine sufficient conditions for asymptotic stability of the time-varying second-order system

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = a_1(t)x_1 + a_2(t)x_2 \tag{1}$$

Assume

$$\nabla V = \begin{bmatrix} \alpha_{11}x_1 + \alpha_{12}x_2 \\ \alpha_{21}x_1 + \alpha_{22}x_2 \end{bmatrix} \tag{2}$$

Select $\alpha_{12} = \alpha_{21}$ so that the curl equations will be satisfied. This allows determination of V by a path-independent line integral of ∇V from $\mathbf{0}$ to $\mathbf{x}$. That is,

$$\begin{aligned} V(\mathbf{x}, t) &= \int_0^{x_1} (\alpha_{11}x_1 + \alpha_{12}x_2)\,dx_1 \bigg|_{x_2=0} \\ &\quad + \int_0^{x_2} (\alpha_{12}x_1 + \alpha_{22}x_2)\,dx_2 \bigg|_{x_1 \text{ fixed}} \\ &= \tfrac{1}{2}\mathbf{x}^T \begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{12} & \alpha_{22} \end{bmatrix} \mathbf{x} \end{aligned} \tag{3}$$

In order to make V positive definite, it is required that $\alpha_{11}(t) > 0$ and $\alpha_{11}\alpha_{22} - \alpha_{12}^2 > 0$ for all t. The time rate-of-change is given by

$$\dot{V} = (\nabla V)^T \dot{\mathbf{x}} + \partial \mathbf{V}/\partial t = \mathbf{x}^T \begin{bmatrix} \alpha_{12}a_1 + \dot{\alpha}_{11}/2 & \alpha_{11} + \alpha_{12}a_2 + \dot{\alpha}_{12}/2 \\ \alpha_{22}a_1 + \dot{\alpha}_{12}/2 & \alpha_{12} + \alpha_{22}a_2 + \dot{\alpha}_{22}/2 \end{bmatrix} \mathbf{x} \tag{4}$$

This must be forced to be negative definite for all t by proper choice of α_{ij}. It is suggested [5] that the coefficient of the highest state component can often be set to 1, so $\alpha_{22} = 1$ is selected. Also, some of the α_{ij} terms can usually be set to zero. Here we tentatively try $\alpha_{12} = 0$. Then Eq. (*4*) reduces to

$$\dot{V} = \mathbf{x}^T \begin{bmatrix} \dot{\alpha}_{11}/2 & \alpha_{11} \\ a_1 & a_2 \end{bmatrix} \mathbf{x}$$

and will be negative definite if $\dot{\alpha}_{11} < 0$ and $\dot{\alpha}_{11}a_2 - 2\alpha_{11}a_1 > 0$. By selecting $\alpha_{11} = -a_1$, these requirements become $\dot{a}_1 > 0$ and $-\dot{a}_1 a_2 + 2a_1^2 > 0$. V is assured of being positive definite if $a_1 < 0$. In order to satisfy all requirements, it is sufficient that

$$a_1 < 0, \qquad a_2 < 0, \quad \text{and} \quad \dot{a}_1 > 0 \quad \text{for all } t \tag{5}$$

(Actually a_2 need only be less than $2a_1^2/\dot{a}_1$ for $\dot{a}_1$ strictly positive). This same problem has been considered [9] using a slightly different Lyapunov function, and a different set of sufficient conditions were found:

$$a_1 < 0, \qquad a_2 < 0, \quad \text{and} \quad \dot{a}_1 < 2a_1a_2 \quad \text{for all } t. \tag{6}$$

The major difference is in the $\dot{a}_1$ condition. Equation (*5*) allows all positive $\dot{a}_1$ and Eq. (*6*) allows all negative $\dot{a}_1$ (plus some positive). Taken together, it is clear that asymptotic stability is assured for *any* $\dot{a}_1$ rate as long as a_1 and a_2 remain negative for all t. Note that these conditions are the same as obtained from applying Routh's criterion to the constant-coefficient case.

PROBLEMS

10.16 Comment on the stability of the systems described in
(a) Example 9.1, page 314;
(b) Problem 9.8, page 328;
(c) Figure 9.5, page 332, with $K = \frac{3}{16}$.

10.17 Investigate the stability of the systems described in conjunction with controllability and observability in
(a) Problem 11.2, page 388;
(b) Problem 11.21, page 397;
(c) Problem 11.26, page 401.

10.18 If the input of Example 10.7 has constraints on the individual components, $|u_i(t)| \leq M_i$, what should the control law be if it is desired to rapidly correct for initial state perturbations?

10.19 Verify by simulation that the time-varying system

$$\dot{\mathbf{x}} = \begin{bmatrix} -2 & g(t) \\ -g(t) & -3 \end{bmatrix}\mathbf{x}$$

is asymptotically stable for any reasonable $g(t)$ function. Note that numerical overflow may occur on the computer if a $g(t)$ which grows without bound is selected. One suggestion is $g(t) = \alpha e^{bt}\sin(\gamma t)$, with $b \leq 0$.

10.20 Modify the matrix $\mathbf{A}(t)$ used in Problem 10.19 by changing the sign of A_{21}. Analyze the stability of the new system.

10.21 Modify the matrix $\mathbf{A}(t)$ used in Problem 10.19 by changing the amplitude coefficient on the A_{21} term. Do not change the A_{12} term. Analyze the stability of the new system.

10.22 Consider the second-order system of Problem 10.15. Form the symmetric matrix $\mathbf{A}_s = \mathbf{A} + \mathbf{A}^T$. Show that the stability results of Sec. 10.7, which are expressed in terms of the eigenvalues of $\mathbf{A}_s$, are not particularly useful because λ_{max} will always be positive or zero.

10.23 The system

$$\ddot{x} + (\gamma - \sin\omega t)\dot{x} + (2 + \beta e^{-\alpha t})x = 0$$

is assured of being asymptotically stable by the results of Problem 10.15 if $\beta > 0$, $\alpha > 0$, and $\gamma > 1$. Use $\alpha = 0.01$, $\beta = 1$, $\gamma = 2$ and $\omega = 15$ and verify by simulation that the origin is a stable focus, as shown in Figure 10.12. Note that $\dot{a}_1 = \beta\alpha e^{-\alpha t}$ is positive.

10.24 Repeat Problem 10.23, but let $\beta = -1$ so that $\dot{a}_1 = \beta\alpha e^{-\alpha t}$ is now negative, thus satisfying Eq. (*6*) of Problem 10.15.

10.25 Demonstrate by simulation that the sufficient conditions of Eq. (*6*) in Problem 10.15 are *not* necessary conditions. This can be demonstrated using the system of Problem 10.23 with $\gamma < 1$, so that a_2 is positive part of the time. With $\alpha = 0.01$, $\beta = 1$, $\gamma = 0.25$, and $\omega = 10$, the phase portrait of a stable focus, shown in Figure 10.13, should be obtained.

10.26 Repeat the variable gradient steps used in Problem 10.15 to develop sufficient conditions for asymptotic stability of a general third-order time-varying system of the form

$$\dddot{y} + a_2(t)\ddot{y} + a_1(t)\dot{y} + a_0(t)y = 0$$

10.27 Consider the motion of the mass-spring-damper system of Figure 10.14 acting under the influence of gravity, g.
(a) Measure the position y of the mass from the unloaded free length of the spring and write the equation of motion

$$m\ddot{y} + b\dot{y} + ky = mg \tag{1}$$

Show that the equilibrium point is $y_e = mg/k$, $\dot{y}_e = 0$. Then, by referencing the position of the

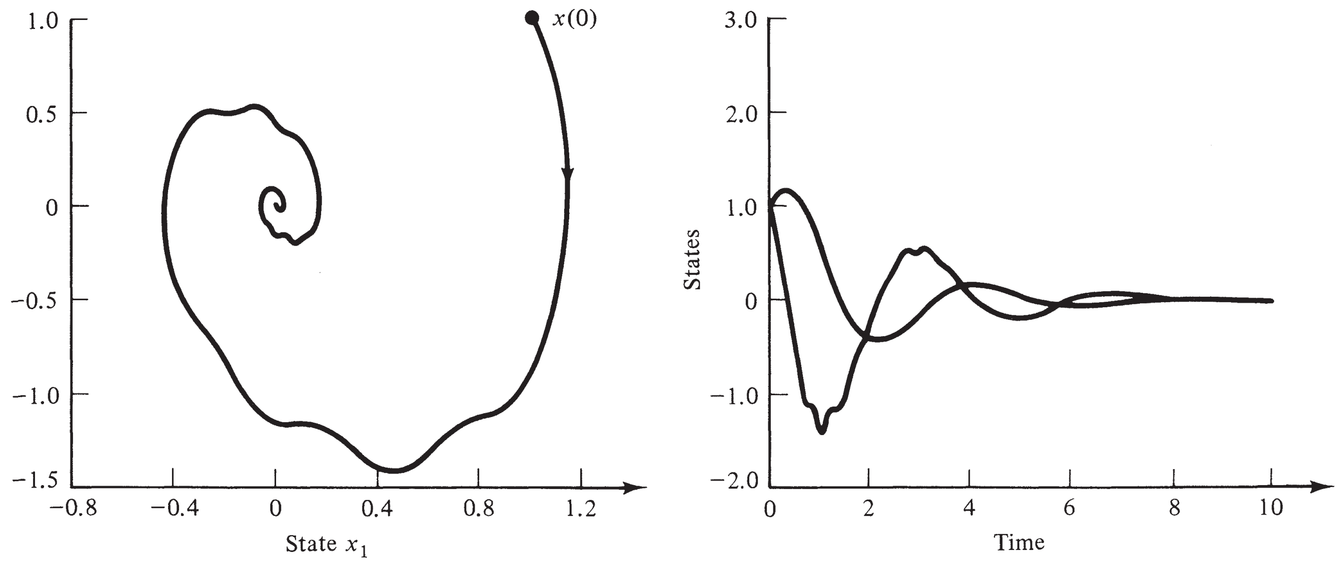


Figure 10.12

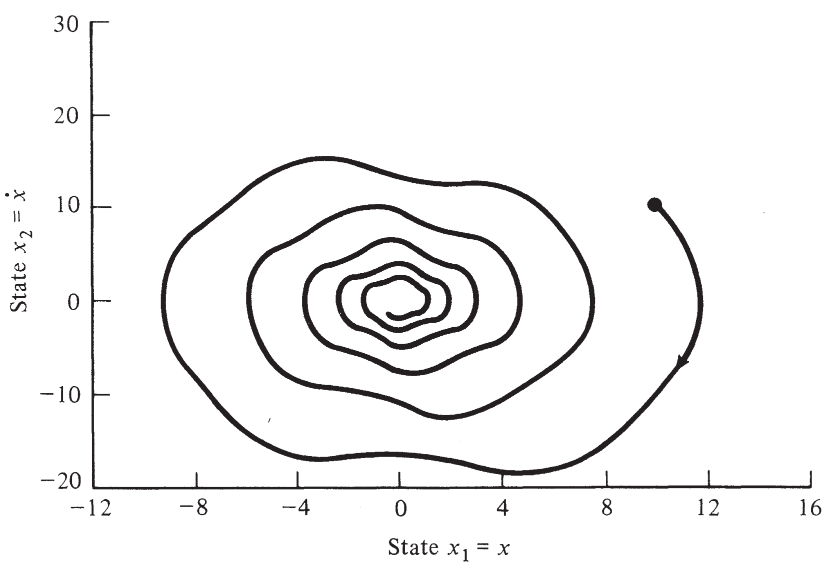


Figure 10.13

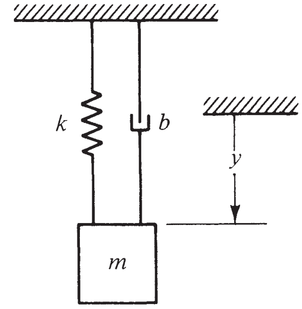


Figure 10.14

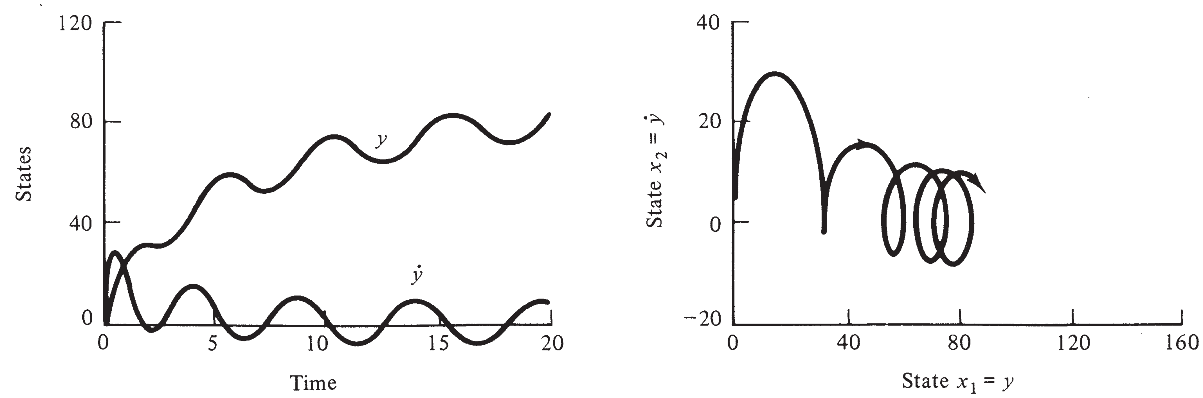


Figure 10.15

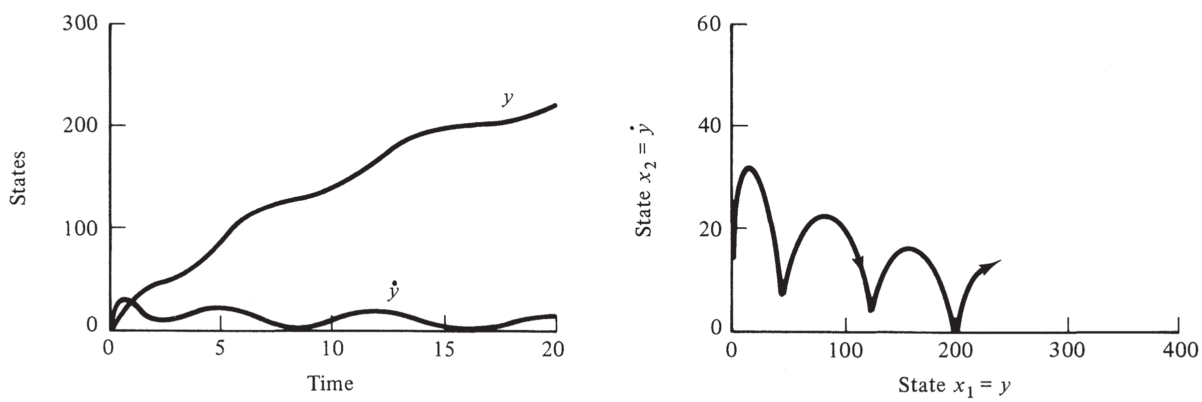


Figure 10.16

mass with respect to the loaded free length of the spring, transfer the equilibrium point to the origin of the new coordinates, so that

$$m\ddot{y}_1 + b\dot{y}_1 + ky_1 = 0 \tag{2}$$

(b) Assume that the damping coefficient b is constant, but that the spring coefficient k is time variable. Show that the frozen coefficient approach mentioned in Sec. 10.7 indicates asymptotic stability for all $k > 0$ and $b > 0$, no matter how small.

(c) Let $b = 1$ lb/ft/s, and $mg = 100$ lb, and assume that the spring becomes softer with time according to

$$k(t) = 50/(1 + 10t)$$

Investigate the behavior of y and $\dot{y}$ in Eq. (*1*). Is the system stable? Response curves for $\mathbf{x}(0) = \mathbf{0}$ are given in Figure 10.15.

(d) Repeat part (c) with $b = 0.2$ and verify the response curves of Figure 10.16.