14

An Introduction to Optimal Control Theory

14.1 INTRODUCTION

The word *optimal* intuitively means doing a job in the best possible way. Before beginning a search for such an optimal solution, the *job* must be defined, a mathematical scale must be established for quantifying what *best* means, and the *possible* alternatives must be spelled out. Unless there is agreement on these qualifiers, a claim that a system is optimal is really meaningless. A crude, inaccurate system might be considered optimal because it is inexpensive, is easy to fabricate, and gives adequate performance. Conversely, a very precise and elegant system could be rejected as nonoptimal because it is too expensive or is too heavy or would take too long to develop.

An introductory account of a dynamic programming [1] approach to optimal control problems is given in this chapter. The emphasis is on linear system, quadratic cost problems, both discrete-time and continuous-time. Other optimization techniques are discussed in many references, including References 2, 3, 4, and 5. A brief look at one of thse techniques, the minimum principle, is included in Section 14.5 and related problems. An elementary example of a minimum norm problem was given in Problem 5.36. Methods of generalizing this approach are suggested by Problem 14.1.

14.2 STATEMENT OF THE OPTIMAL CONTROL PROBLEM

A mathematical statement of the optimal control problem consists of

1. A description of the system to be controlled.
2. A description of system constraints and possible alternatives.
3. A description of the task to be accomplished.
4. A statement of the criterion for judging optimal performance.

The dynamic systems to be considered are described in state variable form by one of the forms in Eq. (*14.1*). It is assumed that all states are available as output measurements, i.e., $\mathbf{y}(t) = \mathbf{x}(t)$. If this is not the case, the state of an *observable* linear system can be estimated by using an observer or a Kalman filter.

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t) \quad \text{or} \quad \mathbf{x}(k+1) = \mathbf{f}(\mathbf{x}(k), \mathbf{u}(k)) \qquad (14.1)$$

Constraints will sometimes exist on allowable values of the state variables. However, in this chapter only control variable constraints are considered. The set of admissible controls U is a subset of the r-dimensional input space $\mathcal{U}^r$ of Chapter 3, $U \subseteq \mathcal{U}^r$. For example, U could be defined as the set of all piecewise continuous vectors $\mathbf{u}(t) \in \mathcal{U}^r$ satisfying $|\mathbf{u}_i(t)| \le M$ or $\|\mathbf{u}(t)\| \le M$ for all t, and for some positive constant M. If there are no constraints, $U = \mathcal{U}^r$.

The task to be performed often takes the form of additional boundary conditions on Eq. (*14.1*). An example might be to transfer the state from a known initial state $\mathbf{x}(t_0)$ to a specified final state $\mathbf{x}(t_f) = \mathbf{x}_d$ at a specified time t_f, or the minimum possible t_f. The task might be to transfer the state to a specified region of state space, called a *target set,* rather than to a specified point. Often, the task to be performed is implicitly specified by the performance criterion.

The most general continuous or discrete general performance criteria to be considered are

$$J = S(\mathbf{x}(t_f), t_f) + \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t)\, dt$$

or

$$J = S(\mathbf{x}(N)) + \sum_{k=0}^{N-1} L(\mathbf{x}(k), \mathbf{u}(k)) \qquad (14.2)$$

S and L are real, scalar-valued functions of the indicated arguments. S is the cost or penalty associated with the error in the stopping or terminal state at time t_f (or N). L is the cost or loss function associated with the transient state errors and control effort. The functions S and L must be selected by the system designer to put more or less emphasis on terminal accuracy, transient behavior, and the expended control effort in the total cost function J.

EXAMPLE 14.1 Set $S = 0, L = 1$. Then $J = \int_{t_0}^{t_f} dt = t_f - t_0$. This is the minimum time problem. ■

EXAMPLE 14.2 Set $S = 0, L = \mathbf{u}^T\mathbf{u}$. Then $J = \int_{t_0}^{t_f} \mathbf{u}^T\mathbf{u}\, dt$ is a measure of the control effort expended. In many cases this term can be interpreted as control energy. This is called the least-effort problem. ■

EXAMPLE 14.3 Set $S = [\mathbf{x}(t_f) - \mathbf{x}_d]^T[\mathbf{x}(t_f) - x_d]$ and $L = 0$. Then minimizing J is equivalent to minimizing the square of the norm of the error between the final state $\mathbf{x}(t_f)$ and a desired final state $\mathbf{x}_d$. This is the minimum terminal error problem. ■

EXAMPLE 14.4 Set $S = 0, L = [\mathbf{x}(t) - \eta(t)]^T[\mathbf{x}(t) - \eta(t)]$. Then minimizing J is equivalent to minimizing the integral of the norm squared of the transient error between the actual state trajectory $\mathbf{x}(t)$ and a desired trajectory $\eta(t)$. ■

EXAMPLE 14.5 A general quadratic criterion which gives a weighted trade-off between the previous three criteria uses $S = [\mathbf{x}(t_f) - \mathbf{x}_d]^T\mathbf{M}[x(t_f) - \mathbf{x}_d]$ and $L = [\mathbf{x}(t) - \eta(t)]^T\mathbf{Q}[\mathbf{x}(t) - \eta(t)] + \mathbf{u}(t)^T\mathbf{R}\mathbf{u}(t)$. The $n \times n$ weighting matrices **M** and **Q** are assumed to be positive semidefinite to ensure a well-defined finite minimum for J. The $r \times r$ matrix **R** is assumed to be positive definite because $\mathbf{R}^{-1}$ will be required in future manipulations. Values for **M, Q,** and **R** should be selected to give the desired trade-offs among terminal error, transient error, and control effort. ■

The discrete-time versions of the above examples require obvious minor modifications. Other performance criteria would be appropriate in specific cases.

The *optimal control problem* is now stated as:

From among all admissible control functions (or sequences) $\mathbf{u} \in U$, find that one which minimizes J of Eq. (*14.2*) subject to the dynamic system constraints of Eq. (*14.1*) and all initial and terminal boundary conditions that may be specified.

If the control is determined as a function of the initial state and other given system parameters, the control is said to be open-loop. If the control is determined as a function of the current state, then it is a closed-loop or feedback control law. Examples of both types will be given in the sequel.

The importance of the property of controllability should be evident. If the system is completely controllable, there is at least one control which will transfer any initial state to any desired final state. If the system is not controllable, it is not meaningful to search for the optimal control. However, controllability does not guarantee that a solution exists for every optimal control problem. Whenever the admissible controls are restricted to the set U, certain final states may not be attainable. Even though the system is completely controllable, the required control may not belong to U (see Problem 14.42).

14.3 DYNAMIC PROGRAMMING

14.3.1 General Introduction to the Principle of Optimality

Dynamic programming provides an efficient means for sequential decision-making. Its basis is R. Bellman's principle of optimality, "An optimal policy has the property that whatever the initial state and the initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision." As used here, a "decision" is a choice of control at a particular time and the "policy" is the entire control sequence (or function).

Consider the nodes in Figure 14.1 as states, in a general sense. A decision is the choice of alternative paths leaving a given node. The goal is to move from state a to state l with minimum cost. A cost is associated with each segment of the line graph. Define J_{ab} as the cost between a and b. J_{bd} is the cost between b and d, etc. For path a, b, d, l the total cost is $J = J_{ab} + J_{bd} + J_{dl}$, and the optimal path (policy) is defined by

$$\min J = \min[J_{ab} + J_{bd} + J_{dl}, J_{ab} + J_{be} + J_{el}, J_{ac} + J_{ch} + J_{hl}, J_{ac} + J_{ck} + J_{kl}] \tag{14.3}$$

If the initial state is a and if the initial decision is to go to b, then the path from b to l

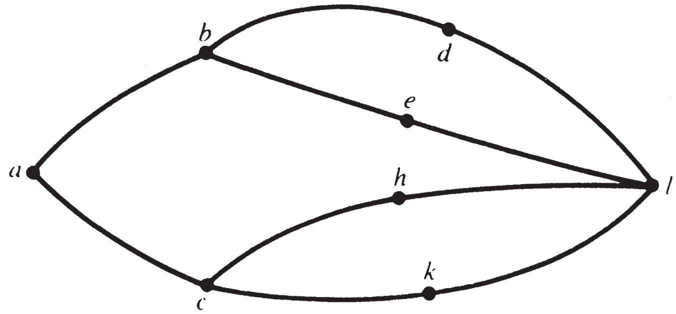


Figure 14.1

must certainly be selected optimally if the overall path from a to l is to be optimum. If the final decision is to go to c, then the path from c to l must then be selected optimally.

Let g_b and g_c be the minimum costs from b and c, respectively, to l. Then $g_b = \min[J_{bd} + J_{dl}, J_{be} + J_{el}]$ and $g_c = \min[J_{ch} + J_{hl}, J_{ck} + J_{kl}]$. The principle of optimality allows equation (*14.3*) to be written as

$$g_a \triangleq \min J = \min[J_{ab} + g_b, J_{ac} + g_c] \tag{14.4}$$

The key feature is that the quantity to be minimized consists of two parts:

1. The part directly attributable to the current decision, such as costs J_{ab} and J_{ac}.
2. The part representing the minimum value of all future costs, starting with the state which results from the first decision.

The principle of optimality replaces a choice between all alternatives (Eq. (*14.3*)) by a sequence of decisions between fewer alternatives (find g_b, g_c, and then g_a from Eq. (*14.4*)). Dynamic programming allows us to concentrate on a sequence of current decisions rather than being concerned about all decisions simultaneously.

Division of cost into the two parts, current and future, is typical, but these parts do not necessarily appear as a sum. A simple example illustrates the sequential nature of the method.

EXAMPLE 14.6 Given N numbers $x_1, x_2, \ldots, x_N$, find the smallest one.

Rather than consider all N numbers simultaneously, define g_k as the minimum of x_k through x_N. Then $g_N = x_N$, $g_{N-1} = \min\{x_{N-1}, g_N\}$, $g_{N-2} = \min\{x_{N-2}, g_{N-1}\}, \ldots, g_k = \min\{x_k, g_{k+1}\}$. Continuing to choose between two alternatives eventually leads to $g_1 = \min\{x_1, g_2\} = \min\{x_1, x_2, \ldots, x_N\}$.

The desired result g_1 need not be unique, since more than one number may have the same smallest value. The recursive nature of the formula $g_k = \min\{x_k, g_{k+1}\}$ is typical of all discrete dynamic programming solutions. ■

EXAMPLE 14.7 Use dynamic programming to find the minimum cost route between node a and node l of Figure 14.2. The line segments can only be traversed from left to right, and the travel costs are shown beside each segment.

Define g_k as the minimum cost from a general node k to l. Obviously, $g_l = 0$. Since there is only one admissible path from h to l, $g_h = 4$. Similarly, $g_j = 5$, $g_d = 2 + g_h = 6$, and $g_f = 6 + g_j = 11$. The first point where a decision must be made is at node e. $g_e = \min\{3 + g_h, 4 + g_j\} = \min\{3 + 4, 4 + 5\} = 7$. The best route from e to l passes through h. Continuing, it is found that $g_b = \min\{8 + g_d, 5 + g_e\} = 12$, $g_c = \min\{4 + g_e, 6 + g_f\} = 11$, and $g_a = \min\{3 + g_b, 5 + g_c\} = 15$. The minimum cost is 15, and the best path is a, b, e, h, l.

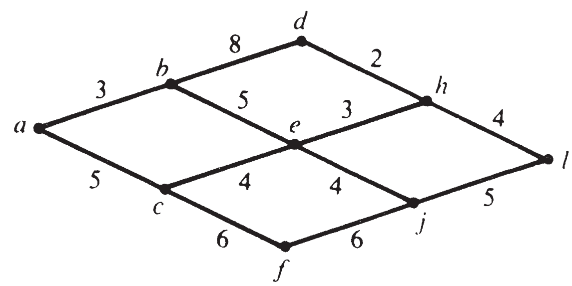


Figure 14.2 Simple routine problem.

The preceding procedure has actually answered the sequence of questions: If past decisions cause the route to be at point k, what is the best decision to make, starting at that point? The answers to these questions must be stored for future use and are shown by the arrows in Figure 14.3. The minimum cost path from any node to l is now obvious. For example, if the path starts at c, then c, e, h, l is cheapest. From f, the cheapest path is f, j, l. ■

14.3.2 Application to Discrete-Time Optimal Control

To make the transition from the simple graph to the more general control problems, the following analogies are made. The graph of Figure 14.2 or 14.3 is a plot of possible states (nodes) at discrete-time points t_k. Point a is at t_0, b and c at $t_1, \ldots, l$ at t_4. The choice of possible directions, say up or down from e, is analogous to the set of admissible controls. Selecting a control $\mathbf{u}(k)$ is analogous to selecting a direction of departure from a given node $\mathbf{x}(k)$. The line segments connecting nodes play the same role as the difference Equation (*14.1*), since both determine the next node $\mathbf{x}(k+1)$ to be encountered.

EXAMPLE 14.8 Consider the scalar system $x(k+1) = x(k) + u(k)$ with boundary conditions $x(0) = 0$ and $x(3) = 3$. Find the controls $u(0)$, $u(1)$, and $u(2)$ which minimize $J = \sum_{k=0}^{2} \{u(k)^2 + \Delta t_k^2\}$. This performance criterion is the sum of the squares of three hypotenuses in the tx plane. This is a form of a minimum distance problem and the optimal sequence of points $x(k)$ will lie on a straight line in the tx plane. Verifying this obvious result will serve to illustrate the dynamic programming procedure.

Let $g(x(k))$ be the minimum cost from $x(k)$ to the terminal point. Since $x(3)$ is the terminal point, $g(x(3)) = 0$. Then

$$g(x(2)) = \min_{u(2)} \{\text{cost from } x(2) \text{ to the terminal point}\} = \min_{u(2)} \{u(2)^2 + \Delta t_2^2 + g(x(3))\}$$

This minimization is not without restriction since $x(3)$ must equal 3. Using $x(3) = x(2) + u(2)$, it is found that $u(2) = 3 - x(2)$. For any $x(2)$, there is a uniquely required $u(2)$. This is analogous to the decisions at points h and j of Example 14.7.

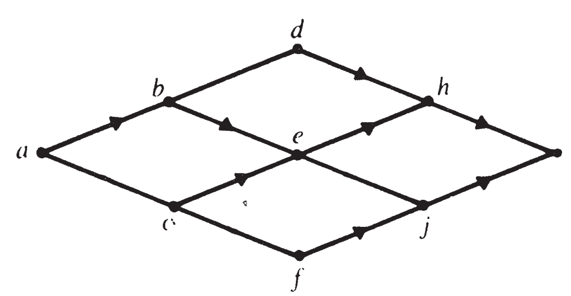


Figure 14.3 The optimal paths.

For convenience, let $\Delta t_k = 1$. Then $g(x(2)) = [3 - x(2)]^2 + 1$. Moving back one stage, $g(x(1)) = \min_{u(1)} \{u(1)^2 + 1 + g(x(2))\}$. Using $x(2) = x(1) + u(1)$ leads to

$$g(x(1)) = \min_{u(1)} \{u(1)^2 + 2 + [3 - x(1) - u(1)]^2\}$$

There are no restrictions on $u(1)$, so the minimization can be accomplished by setting $\frac{\partial}{\partial u(1)}\{u(1)^2 + 2 + [3 - x(1) - u(1)]^2\} = 0$. This gives $u(1) = [3 - x(1)]/2$, and then

$$g(x(1)) = \left[3 - x(1) - \frac{3 - x(1)}{2}\right]^2 + 2 + \left[\frac{3 - x(1)}{2}\right]^2$$

This is used along with $x(1) = x(0) + u(0)$ in $g(x(0)) = \min_{u(0)} \{u(0)^2 + 1 + g(x(1))\}$. Minimizing again gives $u(0) = [3 - x(0)]/3$. Since $x(0) = 0$ is given, $u(0) = 1$. The difference equation gives $x(1) = 1$. This is used in the previously computed expression to give $u(1) = 1$. Then $x(2) = x(1) + u(1) = 2$ and $u(2) = 3 - x(2) = 1$. Finally, $x(3) = x(2) + u(2) = 3$ as required. Note that the sequence $x(0), x(1), x(2), x(3)$ lies on a straight line in tx space, as expected. ■

Discrete dynamic programming solutions of optimal control problems usually consist of two stage-by-stage passes through the time stages. First, a *backward* pass answers the questions, "What is the minimum cost if the problem is started at time t_k with state $\mathbf{x}(k)$, and what is the optimal control as a function of $\mathbf{x}(k)$?" If $\mathbf{u}(k)$ is unrestricted and the criterion function is simple enough, the optimal $\mathbf{u}(k)$ can be obtained in explicit equation form in terms of $\mathbf{x}(k)$. This can be done by setting the gradient with respect to $\mathbf{u}(k)$ equal to zero as in Example 14.8. When this is not possible, a discrete set of grid points for the state and control variables would be used. A computer search routine would be used to find the optimal $\mathbf{u}(k)$, in tabular form, for each discrete value of $\mathbf{x}(k)$. This is a generalization of the method in Example 14.7.

The backward pass is completed when time t_0 is reached. Since $\mathbf{x}(0)$ is known, $\mathbf{u}(0)$ can be found in terms of that specific state. The second pass is in the *forward* direction. The system difference equation uses $\mathbf{x}(0)$ and $\mathbf{u}(0)$ to obtain $\mathbf{x}(1)$. The previously computed function or table is used to determine the value of $\mathbf{u}(1)$ associated with $\mathbf{x}(1)$. Then, in turn, $\mathbf{u}(1)$ gives $\mathbf{x}(2)$ and $\mathbf{x}(2)$ gives $\mathbf{u}(2)$, and so on.

The preceding verbal description is now expressed in equation form for the discrete versions of Eqs. (*14.1*) and (*14.2*). Define $g(\mathbf{x}(k))$ as the minimum cost of the process, starting at $t_k, \mathbf{x}(k)$. Obviously, $g(\mathbf{x}(N)) = S(\mathbf{x}(N))$. Then the principle of optimality gives

$$g(\mathbf{x}(N-1)) = \min_{\mathbf{u}(N-1)} \{L(\mathbf{x}(N-1), \mathbf{u}(N-1)) + g(\mathbf{x}(N))\}$$

Equation (*14.1*) is used to eliminate $\mathbf{x}(N)$. The minimization with respect to $\mathbf{u}(N-1)$ is carried out by setting the gradient with respect to $\mathbf{u}(N-1)$ equal to zero (*if* there are no restrictions on $\mathbf{u}$), or by a computer search routine. In either case the optimal $\mathbf{u}(N-1)$ must be stored for each possible $\mathbf{x}(N-1)$. Also, $g(\mathbf{x}(N-1))$ must be stored for use in the next stage. This continues, stage-by-stage, with a typical step being

$$g(\mathbf{x}(k)) = \min_{\mathbf{u}(k)} \{L(\mathbf{x}(k), \mathbf{u}(k)) + g(\mathbf{x}(k+1))\}$$

$$= \min_{\mathbf{u}(k)} \{L(\mathbf{x}(k), \mathbf{u}(k)) + g(\mathbf{f}(x(k), \mathbf{u}(k)))\} \qquad (14.5)$$

Equation (*14.5*) is an extremely powerful and general result. It is the key to the solution of many discrete-time optimal control problems. This nonlinear difference equation has the boundary condition $g(\mathbf{x}(N)) = S(\mathbf{x}(N))$. The solution of Eq. (*14.5*) yields the optimal control $\mathbf{u}^*(k)$‡ at each time step, and also the optimal trajectory $\mathbf{x}^*(k)$.

The most general case of Eq. (*14.5*) can be solved (in principle at least) by using a tabular computational approach. That is, a discretized grid of possible $\mathbf{x}(k)$ and $\mathbf{u}(k)$ values is determined at each time point. The results of the backward-in-time pass through this grid will consist of a table of optimal $\mathbf{u}^*(k)$ values for each possible $\mathbf{x}(k)$ value. The storage requirements quickly become excessive for all but the lowest order systems. The emphasis here will be on the case of linear systems with quadratic cost functions.

14.3.3 The Discrete-Time Linear Quadratic Problem

The linear system, quadratic cost function (LQ) optimal control problem has received special attention in the literature and in applications. This is because it can be solved analytically, and the resulting optimal controller is expressed in easy-to-implement state feedback form.

Consider the system $\mathbf{x}(k+1) = \mathbf{A}\mathbf{x}(k) + \mathbf{B}\mathbf{u}(k)$ with $\mathbf{x}(0)$ known. The goal is to find the control sequence $\mathbf{u}(k)$ which minimizes the quadratic cost function for the finite-time regulator problem,

$$J = \tfrac{1}{2}\mathbf{x}(N)^T\mathbf{M}\mathbf{x}(N) + \tfrac{1}{2}\sum_{k=0}^{N-1} \{\mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k)\} \qquad (14.6)$$

$\mathbf{M}$ and $\mathbf{Q}$ are symmetric positive semidefinite $n \times n$ matrices and $\mathbf{R}$ is a symmetric positive definite $r \times r$ matrix. No restrictions are placed on $\mathbf{u}(k)$.

Let $g[\mathbf{x}(k)]$ = min cost from k, $\mathbf{x}(k)$ to N, $\mathbf{x}(N)$. Equation (*14.5*) gives

$$g[\mathbf{x}(k)] = \min_{\mathbf{u}(k)} \{\tfrac{1}{2}\mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \tfrac{1}{2}\mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k) + g[\mathbf{x}(k+1)]\} \qquad (14.7)$$

This is a difference equation and the boundary condition is $g[\mathbf{x}(N)] = \tfrac{1}{2}\mathbf{x}(N)^T\mathbf{M}\mathbf{x}(N)$. This equation is solved by assuming a solution

$$g[\mathbf{x}(k)] = \tfrac{1}{2}\mathbf{x}(k)^T\mathbf{W}(N-k)\mathbf{x}(k) + \mathbf{x}(k)^T\mathbf{V}(N-k) + Z(N-k) \qquad (14.8)$$

where $\mathbf{W}$, $\mathbf{V}$, and Z are an $n \times n$ matrix, an $n \times 1$ vector, and a scalar, respectively. They will be selected so as to force Eq. (*14.8*) to satisfy Eq. (*14.7*). Equation (*14.7*) becomes

$$\begin{aligned}
&\tfrac{1}{2}\mathbf{x}(k)^T\mathbf{W}(N-k)\mathbf{x}(k) + \mathbf{x}(k)^T\mathbf{V}(N-k) + Z(N-k) \\
&= \min_{\mathbf{u}(k)} \{\tfrac{1}{2}\mathbf{x}^T(k)\mathbf{Q}\mathbf{x}(k) + \tfrac{1}{2}\mathbf{u}^T(k)\mathbf{R}\mathbf{u}(k) + \tfrac{1}{2}\mathbf{x}(k+1)^T\mathbf{W}(N-k-1)\mathbf{x}(k+1) \qquad (14.9) \\
&+ \mathbf{x}(k+1)^T\mathbf{V}(N-k-1) + Z(N-k-1)\}
\end{aligned}$$

‡ The * on a vector, such as $\mathbf{u}^*(k)$, indicates the optimal vector. It should not be confused with the notation for an adjoint transformation, such as $\mathcal{A}^*$.

Substituting $\mathbf{x}(k+1) = \mathbf{Ax}(k) + \mathbf{Bu}(k)$ and regrouping, the right-hand side of Eq. (*14.9*) becomes

$$\begin{aligned}\text{R.H.S.} = \min_{\mathbf{u}(k)} \{&\tfrac{1}{2}\mathbf{x}(k)^T[\mathbf{Q} + \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{A}]\mathbf{x}(k) \\ &+ \tfrac{1}{2}\mathbf{u}(k)^T[\mathbf{R} + \mathbf{B}^T\mathbf{W}(N-k-1)\mathbf{B}]\mathbf{u}(k) \\ &+ \mathbf{u}(k)^T[\mathbf{B}^T\mathbf{V}(N-k-1) + \mathbf{B}^T\mathbf{W}(N-k-1)\mathbf{Ax}(k)] \\ &+ \mathbf{x}(k)^T\mathbf{A}^T\mathbf{V}(N-k-1) + Z(N-k-1)\}\end{aligned} \tag{14.10}$$

Since there are no restrictions on $\mathbf{u}(k)$, the minimizing $\mathbf{u}(k)$ is found by setting $\partial\{\ \}/\partial\mathbf{u}(k) = 0$. This yields

$$\begin{aligned}\mathbf{u}^*(k) = -&[\mathbf{R} + \mathbf{B}^T\mathbf{W}(N-k-1)\mathbf{B}]^{-1}\mathbf{B}^T[\mathbf{V}(N-k-1) \\ &+ \mathbf{W}(N-k-1)\mathbf{Ax}(k)] = \mathbf{F}_c(k)\mathbf{V}(N-k-1) - \mathbf{G}(k)\mathbf{x}(k)\end{aligned} \tag{14.11}$$

The problem remains to determine the feed-forward and feedback gain matrices $\mathbf{F}_c(k)$ and $\mathbf{G}(k)$ and the external input $\mathbf{V}(k)$. Rearranging and simplifying Eq. (*14.10*) gives

$$\begin{aligned}\text{R.H.S.}& \\ = &\tfrac{1}{2}\mathbf{x}(k)^T[\mathbf{Q} + \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{A} - \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{BUB}^T\mathbf{W}(N-k-1)\mathbf{A}]\mathbf{x}(k) \\ &+ \mathbf{x}(k)^T[\mathbf{A}^T\mathbf{V}(N-k-1) - \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{BUB}^T\mathbf{V}(N-k-1)] \\ &+ [Z(N-k-1) - \tfrac{1}{2}\mathbf{V}(N-k-1)^T\mathbf{BUB}^T\mathbf{V}(N-k-1)]\end{aligned}$$

where $\mathbf{U} \triangleq [\mathbf{R} + \mathbf{B}^T\mathbf{W}(N-k-1)\mathbf{B}]^{-1}$. Equating the left-hand and right-hand sides, the assumed form for $g[\mathbf{x}(k)]$ can be forced to be a solution *for all* $\mathbf{x}(k)$ by requiring that the quadratic terms, the linear terms, and the terms not involving $\mathbf{x}$ all balance individually. This requires

$$\begin{aligned}\mathbf{W}(N-k) = \mathbf{Q} &+ \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{A} \\ &- \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{BUB}^T\mathbf{W}(N-k-1)\mathbf{A}\end{aligned} \tag{14.12}$$

$$\mathbf{V}(N-k) = \mathbf{A}^T\mathbf{V}(N-k-1) - \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{BUB}^T\mathbf{V}(N-k-1) \tag{14.13}$$

$$Z(N-k) = Z(N-k-1) - \tfrac{1}{2}\mathbf{V}(N-k-1)^T\mathbf{BUB}^T\mathbf{V}(N-k-1) \tag{14.14}$$

The boundary conditions are $\mathbf{W}(N-N) = \mathbf{M}$, $\mathbf{V}(N-N) = 0$, and $Z(N-N) = 0$. A computer solution, backward in time, easily gives $\mathbf{W}(N-k)$. Normally (*14.12*) is solved first. Then its solution $\mathbf{W}(N-k)$ is used as a known coefficient matrix while solving (*14.13*) for $\mathbf{V}(N-k)$. Then $\mathbf{V}$ acts as a known forcing function in (*14.14*). Actually, (*14.14*) never needs to be solved if the only interest is in finding the optimal control. $Z(N-k)$ is only needed if $J_{\min}$ must be calculated. For the cost function considered here (the so-called regulator problem), (*14.13*) is a homogeneous equation with zero initial conditions, so $\mathbf{V}(N-k)$ is zero for all stages. Therefore, (*14.13*) is not needed either.

Equation (*14.12*) remains as the principal result. It is interesting to rewrite (*14.12*) as follows: Let $k' = N - k$ be a backward running time index and let

$\mathbf{W}(N-k) = \mathbf{M}(k')$. Introducing two new intermediate variables $\mathbf{K}$ and $\mathbf{P}$ allows Eq. (*14.12*) to be replaced by

$$\mathbf{M}(k') = \mathbf{A}^T\mathbf{P}(k'-1)\mathbf{A} + \mathbf{Q} \tag{14.15}$$

$$\mathbf{K}(k') = \mathbf{M}(\mathbf{k}')\mathbf{B}[\mathbf{B}^T\mathbf{M}(k')\mathbf{B} + \mathbf{R}]^{-1} \tag{14.16}$$

$$\mathbf{P}(k') = [\mathbf{I} - \mathbf{K}(k')\mathbf{B}^T]\mathbf{M}(k') \tag{14.17}$$

These are exactly the same as (*1*), (*2*), and (*3*) of Problem 6.18 (the Kalman filter algorithm), except that $\mathbf{B}^T$ replaces $\mathbf{C}$, $\mathbf{A}^T$ replaces $\mathbf{\Phi}$, and k' replaces k. The optimal mean square estimator problem and the optimal regulator problem are duals of each other. The initial condition on (*14.15*) is $\mathbf{M}(0) = \mathbf{M}$. The optimal feedback control is given by

$$\mathbf{u}^*(k) = -\mathbf{K}^T(k'-1)\mathbf{A}\mathbf{x}(k)$$

so that the feedback gain matrix is $\mathbf{G}(k) = \mathbf{K}^T(k'-1)\mathbf{A}$. This duality has a practical significance. If a computer program is available to compute the Kalman gain matrix $\mathbf{K}(k)$ (Eq. (*2*) of Problem 6.18), then the same algorithm can be used to find the control gains $\mathbf{G}(k)$. Just make the interchanges with $\mathbf{B}$, $\mathbf{C}$, $\mathbf{A}$, and $\mathbf{\Phi}$ as mentioned above and remember that time is reversed.

The optimal control sequence can next be found for the same linear system, but with the more general cost function for the tracking problem,

$$J = \tfrac{1}{2}[\mathbf{x}(N) - \mathbf{x}_d]^T\mathbf{M}[\mathbf{x}(N) - \mathbf{x}_d] + \tfrac{1}{2}\sum_{k=0}^{N-1}\{[\mathbf{x}(k) - \boldsymbol{\eta}(k)]^T\mathbf{Q}[\mathbf{x}(k) - \boldsymbol{\eta}(k)] + \mathbf{u}(k)^T\mathbf{R}\mathbf{u}(k)\} \tag{14.18}$$

This cost function attempts to make $\mathbf{x}(k)$ follow the specified sequence $\boldsymbol{\eta}(k)$.

Most of the solution details are the same as for the regulator problem and will not be repeated. Expanding the quadratic terms in J shows that there are four additional terms to deal with, due to $\mathbf{x}_d(k)$ and $\boldsymbol{\eta}(k)$. Two of these terms become forcing functions on Eqs. (*14.13*) and (*14.14*), modified here as

$$\mathbf{V}(N-k) = \mathbf{A}^T\mathbf{V}(N-k-1) - \mathbf{A}^T\mathbf{W}(N-k-1)\mathbf{B}\mathbf{U}\mathbf{B}^T\mathbf{V}(N-k-1) - \mathbf{Q}\boldsymbol{\eta}(k) \tag{14.19}$$

$$Z(N-k) = Z(N-k-1) - \tfrac{1}{2}\mathbf{V}^T(N-k-1)\mathbf{B}\mathbf{U}\mathbf{B}^T\mathbf{V}(N-k-1) + \tfrac{1}{2}\boldsymbol{\eta}(k)^T\mathbf{Q}\boldsymbol{\eta}(k) \tag{14.20}$$

Equation (*14.12*)—or its expanded counterparts (*14.15*), (*14.16*) and (*14.17*)—remains unchanged. The other two additional terms come into the boundary conditions, which are

$$\mathbf{W}(N-N) = \mathbf{M} \quad \text{(unchanged)}$$

$$\mathbf{V}(N-N) = -\mathbf{M}\mathbf{x}_d$$

$$Z(N-N) = \tfrac{1}{2}\mathbf{x}_d^T\mathbf{M}\mathbf{x}_d$$

A form of Eq. (*14.19*) that is consistent with the notation introduced in (*14.15*), (*14.16*), and (*14.17*) is

$$\mathbf{V}(k') = \mathbf{A}^T[\mathbf{I} - \mathbf{K}^T(k'-1)\mathbf{B}^T]\mathbf{V}(k'-1) - \mathbf{Q}\boldsymbol{\eta}(k) \tag{14.21}$$

The control law is

$$\mathbf{u}^*(k) = -\mathbf{K}^T(k'-1)\mathbf{A}\mathbf{x}(k) - \mathbf{U}(k'-1)\mathbf{B}^T\mathbf{V}(k'-1) \tag{14.22}$$

In both the regulator and tracking problems, the feedback gain

$$\mathbf{G}(k) = [\mathbf{R} + \mathbf{B}^T\mathbf{W}(N-k-1)\mathbf{B}]^{-1}\mathbf{B}^T\mathbf{W}(N-k-1)\mathbf{A}$$

and the feed-forward gain

$$\mathbf{F}_c(k) = -[\mathbf{R} + \mathbf{B}^T\mathbf{W}(N-k-1)\mathbf{B}]^{-1}\mathbf{B}^T$$

depend on the positive definite symmetric matrix **W,** which must be obtained by solving Eq. (*14.12*) or the equivalent triple, Eqs. (*14.15*) through (*14.17*). This is called the discrete-time Riccati equation. In terms of the backward running time index k', this discrete Riccati equation is

$$\begin{aligned}\mathbf{W}(k') = \mathbf{Q} &+ \mathbf{A}^T\mathbf{W}(k'-1)\mathbf{A}\\ &- \mathbf{A}^T\mathbf{W}(k'-1)\mathbf{B}[\mathbf{R} + \mathbf{B}^T\mathbf{W}(k'-1)\mathbf{B}]^{-1}\mathbf{B}^T\mathbf{W}(k'-1)\mathbf{A}\end{aligned} \tag{14.23}$$

Assume that **A** is nonsingular. Then this nonlinear difference equation can be reduced to a pair of coupled linear equations by replacing the $n \times n$ nonsingular matrix **W** by $\mathbf{W}(k') = \mathbf{E}(k')\mathbf{F}(k')^{-1}$. Making this substitution in Eq. (*14.23*) and then post-multiplying by $\mathbf{F}(k')$ and premultiplying by $\mathbf{A}^{-T}$ (the transpose of $\mathbf{A}^{-1}$) gives

$$\begin{aligned}\mathbf{A}^{-T}\mathbf{E}(k') &= \mathbf{A}^{-T}\mathbf{Q}\mathbf{F}(k') + \mathbf{W}(k'-1)\mathbf{A}\mathbf{F}(k')\\ &\quad - \mathbf{W}(k'-1)\mathbf{B}[\mathbf{R} + \mathbf{B}^T\mathbf{W}(k'-1)\mathbf{B}]^{-1}\mathbf{B}^T\mathbf{W}(k'-1)\mathbf{A}\mathbf{F}(k')\end{aligned} \tag{14.24}$$

If the judicious choice

$$\mathbf{E}(k'-1) = \mathbf{A}^{-T}\mathbf{E}(k') - \mathbf{A}^{-T}\mathbf{Q}\mathbf{F}(k') \tag{14.25}$$

is made, then Eq. (*14.24*) becomes

$$\begin{aligned}\mathbf{E}(k'-1) &= \mathbf{W}(k'-1)\mathbf{A}\mathbf{F}(k')\\ &\quad - \mathbf{W}(k'-1)\mathbf{B}[\mathbf{R} + \mathbf{B}^T\mathbf{W}(k'-1)\mathbf{B}]^{-1}\mathbf{B}^T\mathbf{W}(k'-1)\mathbf{A}\mathbf{F}(k')\end{aligned}$$

Premultiplication by $\mathbf{W}(k'-1)^{-1} = \{\mathbf{E}(k'-1)\mathbf{F}(k'-1)^{-1}\}^{-1}$ gives

$$\mathbf{F}(k'-1) = \mathbf{A}\mathbf{F}(k') - \mathbf{B}[\mathbf{R} + \mathbf{B}^T\mathbf{W}(k'-1)\mathbf{B}]^{-1}\mathbf{B}^T\mathbf{W}(k'-1)\mathbf{A}\mathbf{F}(k')$$

The matrix inversion lemma of Section 4.9, $[\mathbf{R} + \mathbf{B}^T\mathbf{W}\mathbf{B}]^{-1} = \mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{B}^T[\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T + \mathbf{W}^{-1}]^{-1}\mathbf{B}\mathbf{R}^{-1}$, is used to rewrite this as

$$\begin{aligned}\mathbf{F}(k'-1) &= \mathbf{A}\mathbf{F}(k')\\ &\quad - \{\mathbf{I} - \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T[\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T + \mathbf{W}(k'-1)^{-1}]^{-1}\}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{W}(k'-1)\mathbf{A}\mathbf{F}(k')\end{aligned}$$

When the unit matrix is written

$$\mathbf{I} = [\mathbf{BR}^{-1}\mathbf{B}^T + \mathbf{W}(k'-1)^{-1}][\mathbf{BR}^{-1}\mathbf{B}^T + \mathbf{W}(k'-1)^{-1}]^{-1}$$

the preceding equation can be rearranged to

$$\mathbf{F}(k'-1) - \mathbf{AF}(k') = -\mathbf{W}(k'-1)^{-1}[\mathbf{BR}^{-1}\mathbf{B}^T + \mathbf{W}(k'-1)^{-1}]^{-1}\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{W}(k'-1)\mathbf{AF}(k')$$

Using $\mathbf{W}^{-1}[\mathbf{BR}^{-1}\mathbf{B}^T + \mathbf{W}^{-1}]^{-1} = [\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{W} + \mathbf{I}]^{-1}$ and then premultiplying by $[\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{W} + \mathbf{I}]$ and replacing the remaining $\mathbf{W}$ terms by $\mathbf{EF}^{-1}$ allows the linear equation to be obtained:

$$\mathbf{F}(k'-1) = \mathbf{AF}(k') - \mathbf{BR}^{-1}\mathbf{B}^T\mathbf{E}(k'-1)$$

Using Eq. (*14.25*) to eliminate $\mathbf{E}(k'-1)$ gives the final form

$$\mathbf{F}(k'-1) = [\mathbf{A} + \mathbf{BR}^{-1}\mathbf{B}^T\mathbf{A}^{-T}\mathbf{Q}]\mathbf{F}(k') - \mathbf{BR}^{-1}\mathbf{B}^T\mathbf{A}^{-T}\mathbf{E}(k') \tag{14.26}$$

Equations (*14.25*) and (*14.26*) can be combined into

$$\begin{bmatrix}\mathbf{F}(k'-1)\\ \mathbf{E}(k'-1)\end{bmatrix} = \begin{bmatrix}\mathbf{A} + \mathbf{BR}^{-1}\mathbf{B}^T\mathbf{A}^{-T}\mathbf{Q} & -\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{A}^{-T}\\ -\mathbf{A}^{-T}\mathbf{Q} & \mathbf{A}^{-T}\end{bmatrix}\begin{bmatrix}\mathbf{F}(k')\\ \mathbf{E}(k')\end{bmatrix} \tag{14.27}$$

The $2n \times 2n$ coefficient matrix on the right side of Eq. (*14.27*) is often called the Hamiltonian matrix $\mathbf{H}$. When $\mathbf{A}$, $\mathbf{B}$, $\mathbf{Q}$, and $\mathbf{R}$ (and hence $\mathbf{H}$) are all constant, the solution to Eq. (*14.27*) can be written as

$$\begin{bmatrix}\mathbf{F}(k')\\ \mathbf{E}(k')\end{bmatrix} = \mathbf{H}^{-k'}\begin{bmatrix}\mathbf{I}\\ \mathbf{M}\end{bmatrix} \tag{14.28}$$

where the initial condition $\mathbf{W}(k'=0) = \mathbf{M}$ was used to select $\mathbf{E}(0) = \mathbf{M}$ and $\mathbf{F}(0) = \mathbf{I}$.

In order to bring Eq. (*14.27*) into the standard form, it was premultiplied by $\mathbf{H}^{-1}$ before solving. This is what causes the negative power on the exponent of $\mathbf{H}$ in Eq. (*14.28*). It is known [2] that if λ is an eigenvalue of $\mathbf{H}$, then so is $1/\lambda$. That is, n of the eigenvalues are stable (inside the unit circle) and n are unstable (outside the unit circle). The Jordan form of $\mathbf{H}$ is $\mathbf{J} = \text{Diag}[\mathbf{J}_s \quad \mathbf{J}_u]$, where $\mathbf{J}_s$ and $\mathbf{J}_u$ are the $n \times n$ blocks associated with the stable and unstable eigenvalues, respectively. In the case of distinct eigenvalues, these will be diagonal blocks and $\mathbf{J}_u = \mathbf{J}_s^{-1}$ because the eigenvalues occur in reciprocal pairs. Let the modal matrix of eigenvectors for $\mathbf{H}$ be written as $\mathbf{T} = \begin{bmatrix}\mathbf{T}_{11} & \mathbf{T}_{12}\\ \mathbf{T}_{21} & \mathbf{T}_{22}\end{bmatrix}$ and let $\mathbf{T}^{-1} = \begin{bmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{bmatrix}$. The columns of $\begin{bmatrix}\mathbf{T}_{11}\\ \mathbf{T}_{21}\end{bmatrix}$ represent the eigenvectors associated with the stable eigenvalues. Then

$$\mathbf{H}^{-k'} = \mathbf{TJ}^{-k'}\mathbf{T}^{-1} = \begin{bmatrix}\mathbf{T}_{11}\mathbf{J}_s^{-k'}\mathbf{V}_{11} + \mathbf{T}_{12}\mathbf{J}_u^{-k'}\mathbf{V}_{21} & \mathbf{T}_{11}\mathbf{J}_s^{-k'}\mathbf{V}_{12} + \mathbf{T}_{12}\mathbf{J}_u^{-k'}\mathbf{V}_{22}\\ \mathbf{T}_{21}\mathbf{J}_s^{-k'}\mathbf{V}_{11} + \mathbf{T}_{22}\mathbf{J}_u^{-k'}\mathbf{V}_{21} & \mathbf{T}_{21}\mathbf{J}_s^{-k'}\mathbf{V}_{12} + \mathbf{T}_{22}\mathbf{J}_u^{-k'}\mathbf{V}_{22}\end{bmatrix}$$

Using this result in Eq. (*14.28*) and then using $\mathbf{W} = \mathbf{EF}^{-1}$ gives

$$\begin{aligned}\mathbf{W}(k') = {} & [\mathbf{T}_{21}\mathbf{J}_s^{-k'}\mathbf{V}_{11} + \mathbf{T}_{22}\mathbf{J}_u^{-k'}\mathbf{V}_{21} + (\mathbf{T}_{21}\mathbf{J}_s^{-k'}\mathbf{V}_{12} + \mathbf{T}_{22}\mathbf{J}_u^{-k'}\mathbf{V}_{22})M] * \\ & [\mathbf{T}_{11}\mathbf{J}_s^{-k'}\mathbf{V}_{11} + \mathbf{T}_{12}\mathbf{J}_u^{-k'}\mathbf{V}_{21} + (\mathbf{T}_{11}\mathbf{J}_s^{-k'}\mathbf{V}_{12} + \mathbf{T}_{12}\mathbf{J}_u^{-k'}\mathbf{V}_{22})M]^{-1}\end{aligned} \tag{14.29}$$

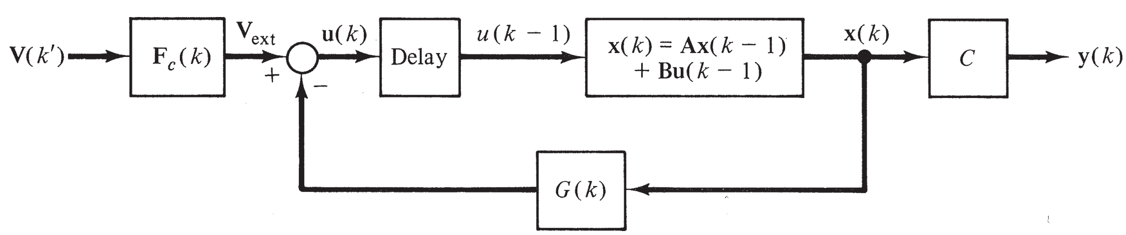
The solution for $\mathbf{W}(k') = \mathbf{W}(N - k - 1)$ can then be used in the equations given previously for the feedback matrix $\mathbf{G}(k)$ and the feed-forward matrix $\mathbf{F}_c(k)$, leading to the implementations shown in Figure 14.4. Figure 14.4*b* is obtained by placing the factor common to both $\mathbf{G}$ and $\mathbf{F}_c$ in the forward path.

14.3.4 The Infinite Horizon, Constant Gain Solution

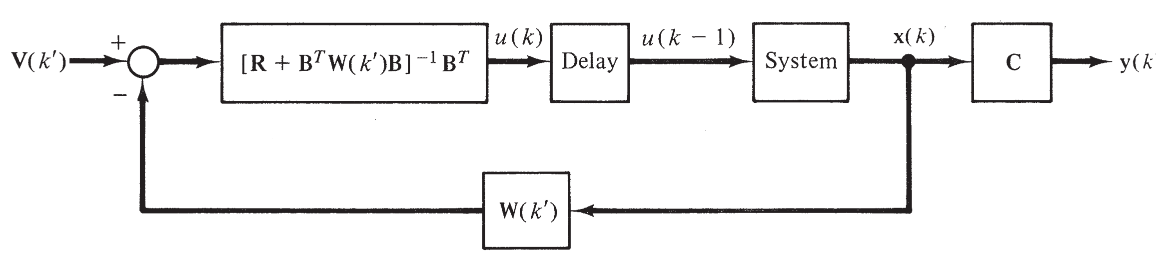
For constant coefficient systems whose operating time is very long compared with the system time constants, it is often justifiable to assume that the terminal time is infinitely far in the future. This so-called infinite horizon case leads to a constant feedback gain matrix, with attendant implementation advantages. This approximation may cause little or no degradation in optimality because the optimal time-varying gains approach constant values in a few time constants (backward from the final time). Thus the optimal gains are constant for most of the operating period. Furthermore, in the regulator problem the states are driven to zero in a few time constants after the initial time t_0. Therefore, the control $\mathbf{u} = -\mathbf{Gx}$ will be essentially zero during the final part of the operation, regardless of whether a constant or time-varying $\mathbf{G}$ is used.

As the time remaining until the end of the problem approaches infinity, $k' \to \infty$, the general solution for the time-varying matrix $\mathbf{W}(k')$ simplifies to a constant. One approach to finding the infinite-time-to-go solution is to set $\mathbf{W}(k') = \mathbf{W}(k' - 1)$ in Eq. (*14.23*) and solve the so-called discrete algebraic Riccati equation (DARE). This solution is easily obtained from the general solution of Sec. 14.3.3, Eq. (*14.29*). Note that $\mathbf{J}_u^{-k'} \to \mathbf{0}$ and $\mathbf{J}_s^{-k'} \to \infty$ as $k' \to \infty$. The $\mathbf{J}_u$ terms are dropped and the $\mathbf{J}_s$ terms are retained (although they too will be found to drop out), giving

$$\begin{aligned}\mathbf{W}_\infty &= [\mathbf{T}_{21}\mathbf{J}_s^{-k'}\mathbf{V}_{11} + \mathbf{T}_{21}\mathbf{J}_s^{-k'}\mathbf{V}_{12}\mathbf{M}][\mathbf{T}_{11}\mathbf{J}_s^{-k'}\mathbf{V}_{11} + \mathbf{T}_{11}\mathbf{J}_s^{-k'}\mathbf{V}_{12}\mathbf{M}]^{-1} \\ &= \mathbf{T}_{21}\mathbf{T}_{11}^{-1}\end{aligned} \qquad (14.30)$$



(a)



(b)

Figure 14.4

Thus the desired answer is found by determining the eigenvectors associated with the stable eigenvalues of **H** and partitioning them into the two $n \times n$ blocks required in Eq. (*14.30*). Another obvious way of calculating $\mathbf{W}_\infty$ without solving an eigenvalue problem is to cycle through Eq. (*14.23*) or (*14.27*) until an unchanging result is obtained. In either case, the constant feedback gain matrix is then given by

$$\mathbf{G}_\infty = [\mathbf{B}^T\mathbf{W}_\infty\mathbf{B} + \mathbf{R}]^{-1}\mathbf{B}^T\mathbf{W}_\infty\mathbf{A}$$

and the control law is

$$\mathbf{u}(k) = -\mathbf{G}_\infty\mathbf{x}(k) + \mathbf{F}_c\mathbf{V}(k)$$

In the regulator problem the input term $\mathbf{V}(k)$ is zero. In tracking problems this extra term must be evaluated from Eq. (*14.13*) or (*14.21*). When needed, the feed-forward control matrix $\mathbf{F}_c$ is given by

$$\mathbf{F}_c = -[\mathbf{B}^T\mathbf{W}_\infty\mathbf{B} + \mathbf{R}]^{-1}\mathbf{B}^T$$

EXAMPLE 14.9 Consider the second-order continuous-time plant shown in Figure 14.5. A sample and zero-order hold is placed in each input path, so that a discrete-time controller can be implemented. The continuous-time state equations are selected as

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 0 & -0.5 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}\mathbf{u} \quad \text{and} \quad \mathbf{y} = [1 \quad 0]\mathbf{x}$$

The results of Sec. 9.8 are used to derive discrete-time state equations $\mathbf{x}(k+1) = \mathbf{A}_1\mathbf{x}(k) + \mathbf{B}_1\mathbf{u}(k)$ for three different sampling times $T = 1, \frac{1}{3}$, and $\frac{1}{10}$. Note that the system time constant is $\tau = 2$, so these will provide 2, 6, and 20 samples per time constant, respectively.

T	$\mathbf{A}_1$	$\mathbf{B}_1$
1	$\begin{bmatrix} 1 & 0.786939 \\ 0 & 0.606531 \end{bmatrix}$	$\begin{bmatrix} 0.426123 & 1 \\ 0.786939 & 0 \end{bmatrix}$
$\frac{1}{3}$	$\begin{bmatrix} 1 & 0.307036 \\ 0 & 0.846482 \end{bmatrix}$	$\begin{bmatrix} 0.052593 & 0.333333 \\ 0.307036 & 0 \end{bmatrix}$
$\frac{1}{10}$	$\begin{bmatrix} 1 & 0.097541 \\ 0 & 0.951229 \end{bmatrix}$	$\begin{bmatrix} 0.004918 & 0.1 \\ 0.097541 & 0 \end{bmatrix}$

To simplify notation, the subscripts on $\mathbf{A}_1$ and $\mathbf{B}_1$ will be omitted in the following. Figure 14.6 shows the transient behavior of the gain components derived from the discrete Riccati equation plotted versus time remaining. For this example the parameters were $T = \frac{1}{3}$, $\mathbf{Q} = \mathbf{R} = \mathbf{I}$, with the boundary condition for $\mathbf{W}(k' = 0) = \mathbf{M} = [\mathbf{0}]$. Note that the final constant values are essentially achieved by $t_r = 4$—that is, within just two system time constants. Similar results are found for

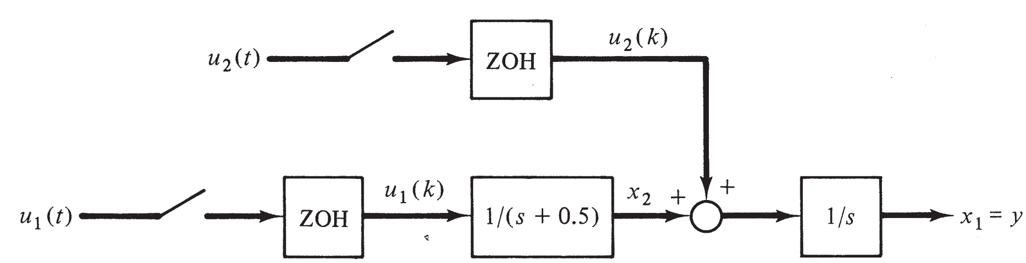


Figure 14.5

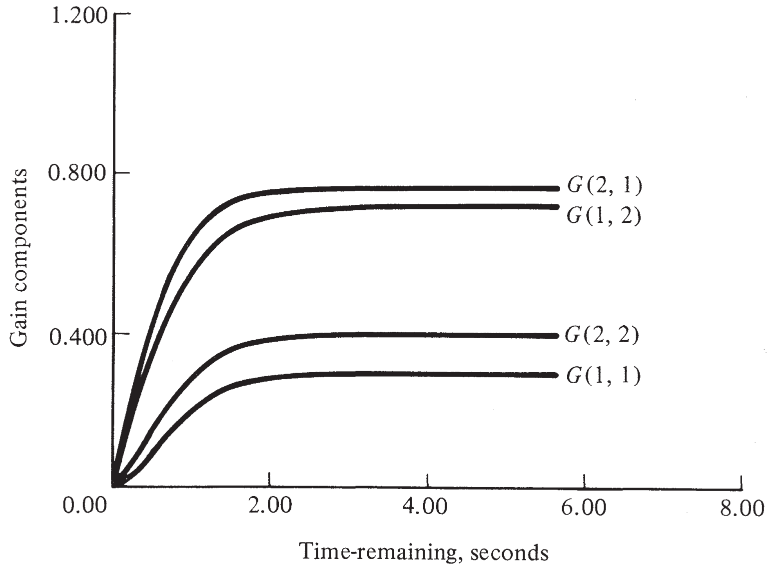


Figure 14.6 Transient gain behavior: $T = 0.3333, Q = R = I$.

the other sampling times and for other choices of **Q, R,** and **M.** For most problems four times the dominant time constant is a good conservative estimate of the time to settle. Note, however, that the number of discrete *cycles* of Eq. (*14.23*) required to reach these constant levels is a function of T.

The constant gain versions of the LQ state feedback controller are now used. To show the effects of the weights **Q** and **R,** three choices for **Q** are used for each sample time T. $\mathbf{R} = I$ is used in all nine cases. For the comparisons, a simple tracking problem is defined with $\boldsymbol{\eta} = [1 \quad 0]^T$, which asks for a unit step in x_1 while maintaining x_2 at or near zero. This requires solution for the infinite time-remaining version of $\mathbf{V}(k')$ and $\mathbf{F}_c$. By setting $\mathbf{V}(k') = \mathbf{V}(k'-1) = \mathbf{V}_\infty$ in Eq. (*14.21*), it is found that

$$\begin{aligned}\mathbf{V}_\infty &= -\{\mathbf{I} - \mathbf{A}^T + \mathbf{A}^T\mathbf{W}_\infty\mathbf{B}[\mathbf{R} + \mathbf{B}^T\mathbf{W}_\infty\mathbf{B}]^{-1}\mathbf{B}^T\}^{-1}\mathbf{Q}\boldsymbol{\eta} \\ &= -\{\mathbf{I} - \mathbf{A}^T[\mathbf{I} + \mathbf{W}_\infty\mathbf{B}\mathbf{F}_{c\infty}]\}^{-1}\mathbf{Q}\boldsymbol{\eta}\end{aligned} \tag{14.31}$$

The external input vector $\mathbf{v}_{ext}$ of Figure 14.4*a* is given by $\mathbf{F}_{c\infty}\mathbf{V}_\infty$. Table 14.1 gives the constant feedback gains, the required external input, and the resulting closed-loop eigenvalues for each case.

Note that in each case $\mathbf{v}_{ext}$ is just column one of $\mathbf{G}_\infty$, that is, $\mathbf{G}_\infty\boldsymbol{\eta}$. This is true here because the selected $\boldsymbol{\eta}$ happens to be an equilibrium point of the system; that is, it can be maintained with zero control, as seen from $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\eta}$. In this special case, Figure 14.4 shows that $\mathbf{F}_{c\infty}\mathbf{V}_\infty = \mathbf{G}_\infty\boldsymbol{\eta}$ if $\mathbf{u}$ is to be zero. To see that this is not generally true, the reader should rework a case from Table 14.1 using $\boldsymbol{\eta} = [0 \quad 1]^T$. With $T = 1$ and $\mathbf{Q} = \mathbf{R} = \mathbf{I}$, it will be found that $\mathbf{G}_\infty$ is unchanged but that $\mathbf{v}_{ext} = [0.4405 \quad -0.27403]^T$, the state vector approaches $[0 \quad 0.4444]^T$ rather than $\boldsymbol{\eta}$, and the control does not go to zero but rather to $\mathbf{u} = [0.2222 \quad -0.4444]^T$.

The transient responses for the cases of Table 14.1 are given in Figure 14.7*a, b,* and *c.* Notice that as **Q** increases, which is equivalent to making **R** relatively smaller, the response gets faster. This is also borne out by the magnitude of the eigenvalues becoming smaller in the tabulated data. At the same time the gains generally get larger. This is most evident in the off-diagonal components. The corresponding transient control signals would also get larger as **Q** increases. Regarding variations in sampling time, it is seen that all three cases give about the same results, except that the slower sampling rate gives a less smooth response due to the larger but less frequent changes at sampling times. As a general rule of thumb, sampling rates in the range of six to ten per dominant time constant are frequently used. ■

TABLE 14.1

T	$\mathbf{Q}$	$\mathbf{G}_\infty$	$\mathbf{v}_{ext}$	λ_{cl}
1	10**I**	$\begin{bmatrix} 0.053625 & 0.69693 \\ 0.89339 & 0.45096 \end{bmatrix}$	$\begin{bmatrix} 0.053625 \\ 0.89339 \end{bmatrix}$	$0.07091 \pm j0.03848$
	I	$\begin{bmatrix} 0.17640 & 0.49114 \\ 0.53973 & 0.38343 \end{bmatrix}$	$\begin{bmatrix} 0.17640 \\ 0.53973 \end{bmatrix}$	$0.30257 \pm j0.14195$
	0.1**I**	$\begin{bmatrix} 0.16088 & 0.28932 \\ 0.18717 & 0.20982 \end{bmatrix}$	$\begin{bmatrix} 0.16088 \\ 0.18717 \end{bmatrix}$	$0.56156 \pm j0.15516$
$\frac{1}{3}$	10**I**	$\begin{bmatrix} 0.19094 & 1.6476 \\ 1.8740 & 0.47914 \end{bmatrix}$	$\begin{bmatrix} 0.19094 \\ 1.8740 \end{bmatrix}$	$0.35295 \pm j0.05835$
	I	$\begin{bmatrix} 0.29835 & 0.72404 \\ 0.76431 & 0.39803 \end{bmatrix}$	$\begin{bmatrix} 0.29835 \\ 0.76431 \end{bmatrix}$	$0.67685 \pm j0.09853$
	0.1**I**	$\begin{bmatrix} 0.19409 & 0.33294 \\ 0.21669 & 0.21348 \end{bmatrix}$	$\begin{bmatrix} 0.19409 \\ 0.21669 \end{bmatrix}$	$0.83091 \pm j0.07419$
$\frac{1}{10}$	10**I**	$\begin{bmatrix} 0.36152 & 2.3921 \\ 2.6645 & 0.48467 \end{bmatrix}$	$\begin{bmatrix} 0.36152 \\ 2.6645 \end{bmatrix}$	$0.72484 \pm j0.03560$
	I	$\begin{bmatrix} 0.36586 & 0.82744 \\ 0.86783 & 0.39982 \end{bmatrix}$	$\begin{bmatrix} 0.36586 \\ 0.86783 \end{bmatrix}$	$0.89097 \pm j0.03861$
	0.1**I**	$\begin{bmatrix} 0.20775 & 0.34882 \\ 0.22792 & 0.21390 \end{bmatrix}$	$\begin{bmatrix} 0.20775 \\ 0.22792 \end{bmatrix}$	$0.94670 \pm j0.02527$

14.4 DYNAMIC PROGRAMMING APPROACH TO CONTINUOUS-TIME OPTIMAL CONTROL

Dynamic programming applies in a similar way to continuous-time systems. The cost of operating the system from a general time and state t, $\mathbf{x}(t)$ to the terminal time and state t_f, $\mathbf{x}(t_f)$ is defined as

$$g(\mathbf{x}(t), t_f - t) \triangleq \min_{\mathbf{u}(t)} \left\{ S(\mathbf{x}(t_f), t_f) + \int_t^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t)\, dt \right\}$$

Breaking the integral into two segments gives

$$g(\mathbf{x}(t), t_f - t) = \min_{\mathbf{u}(t)} \left\{ S(\mathbf{x}(t_f), t_f) + \int_{t+\delta t}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t)\, dt + \int_t^{t+\delta t} L(\mathbf{x}(t), \mathbf{u}(t), t)\, dt \right\} \tag{14.32}$$

The principle of optimality states that if the total cost is to be minimum, then the cost

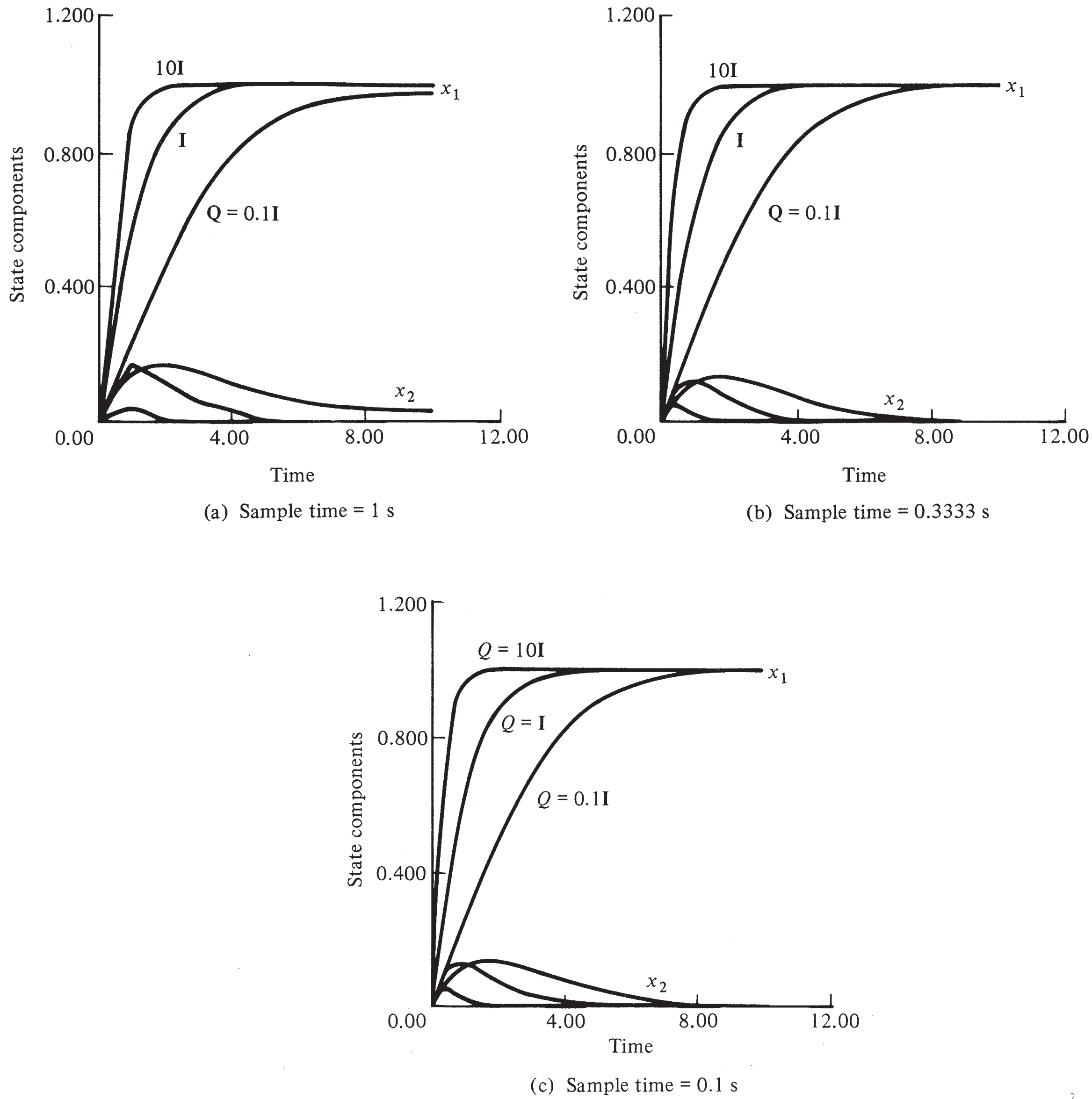


Figure 14.7 Transient response, constant LQ gains

from $t + \delta t$, $\mathbf{x}(t + \delta t)$, must also be minimum. The first two terms on the right side of Eq. (*14.32*) must therefore be equal to $g(\mathbf{x}(t + \delta t), t_f - t - \delta t)$, so that

$$g(\mathbf{x}(t), t_f - t) = \min_{\mathbf{u}(t)} \left\{ g(\mathbf{x}(t + \delta t), t_f - t - \delta t) + \int_t^{t+\delta t} L(\mathbf{x}(t), \mathbf{u}(t), t)\, dt \right\} \qquad (14.33)$$

For δt sufficiently small,

$$\int_t^{t+\delta t} L(\mathbf{x}(t), \mathbf{u}(t), t)\, dt \cong L(\mathbf{x}(t), \mathbf{u}(t), t)\, \delta t \quad \text{and} \quad \mathbf{x}(t + \delta t) \cong \mathbf{x}(t) + \dot{\mathbf{x}}\delta t$$

Taylor series expansion gives

$$g(\mathbf{x}(t + \delta t), t_f - t - \delta t) \cong g(\mathbf{x}(t), t_f - t) + [\nabla_{\mathbf{x}} g]^T \dot{\mathbf{x}}\, \delta t - \frac{\partial g}{\partial t_r} \delta t$$

where $t_r \triangleq t_f - t$ is the time remaining. Using these in Eq. (*14.33*) gives

$$g(\mathbf{x}(t), t_r) = \min_{\mathbf{u}(t)} \left\{ g(\mathbf{x}(t), t_r) + [\nabla_\mathbf{x} g]^T \dot{\mathbf{x}}\, \delta t - \frac{\partial g}{\partial t_r} \delta t + L\, \delta t \right\} \tag{14.34}$$

By definition, $g(\mathbf{x}(t), t_r)$ is a function only of the current state and the time remaining (and not a function of $\mathbf{u}(t)$). Therefore, Eq. (*14.34*) leads to the *Hamilton-Jacobi-Bellman* (H.J.B.) partial differential equation [3]:

$$\frac{\partial g}{\partial t_r} = \min_{\mathbf{u}(t)} \{ L(\mathbf{x}(t), \mathbf{u}(t), t) + [\nabla_\mathbf{x} g]^T \dot{\mathbf{x}} \} \tag{14.35}$$

The boundary condition is $g(\mathbf{x}(t), t_r)|_{t_r=0} = S(\mathbf{x}(t_f), t_f)$. Equation (*14.35*) is the continuous-time counterpart of Eq. (*14.5*). It is the major key to solving continuous-time optimal control problems. Whether it can be solved or not, and with what difficulty level, depends upon the class of systems, cost functions, admissible controls, and state constraints. The optimal control is the one which minimizes the right side of Eq. (*14.35*). If there are no restrictions on $\mathbf{u}(t)$, i.e., the admissible control set $\mathbf{U}$ is the entire space $\mathcal{U}^r$, then the minimum can be found by differentiating with respect to $\mathbf{u}(t)$ and setting the resultant gradient vector to zero. This gives the necessary condition for optimality.

$$\frac{\partial L}{\partial \mathbf{u}} + \frac{\partial \mathbf{f}}{\partial \mathbf{u}} \nabla_\mathbf{x} g = 0$$

If the system is linear, $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$ and $\partial\mathbf{f}/\partial\mathbf{u} = \mathbf{B}^T(t)$. If the loss function L is quadratic, i.e., L of Example 14.5, then $\partial L/\partial\mathbf{u} = 2\mathbf{R}\mathbf{u}(t)$ so that the optimal $\mathbf{u}(t)$ for the continuous version of the *LQ* problem is given by

$$\mathbf{u}^*(t) = -\tfrac{1}{2}\mathbf{R}^{-1}\mathbf{B}^T \nabla_\mathbf{x} g(\mathbf{x}(t), t_r) \tag{14.36}$$

The *LQ* problem is examined in detail next.

14.4.1 Linear-Quadratic (LQ) Problem; the Continuous Riccati Equation

Consider the linear system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}$. The optimal control is sought to minimize the quadratic performance criterion of Example 14.5, i.e., the tracking problem. There are no restrictions on $\mathbf{u}(t)$, and t_f is fixed.

Equation (*14.35*) specializes to

$$\frac{\partial g}{\partial t_r} = \min_{\mathbf{u}(t)} \{ [\mathbf{x}(t) - \boldsymbol{\eta}(t)]^T \mathbf{Q} [\mathbf{x}(t) - \boldsymbol{\eta}(t)] + \mathbf{u}(t)^T \mathbf{R}\mathbf{u}(t) + (\nabla_\mathbf{x} g)^T [\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u}] \}$$

with boundary conditions

$$g[\mathbf{x}(t), t_r]|_{t_r=0} = [\mathbf{x}(t_f) - \mathbf{x}_d]^T \mathbf{M} [\mathbf{x}(t_f) - \mathbf{x}_d]$$

Since $\mathbf{u}(t)$ is unrestricted, taking the derivative with respect to $\mathbf{u}(t)$ and setting it equal to zero gives $\mathbf{u}^*(t) = -\tfrac{1}{2}\mathbf{R}^{-1}\mathbf{B}^T \nabla_\mathbf{x} g[\mathbf{x}(t), t_r]$. The H.J.B. equation then reduces to

$$\frac{\partial g}{\partial t_r} = [\mathbf{x}(t) - \boldsymbol{\eta}(t)]^T \mathbf{Q} [\mathbf{x}(t) - \boldsymbol{\eta}(t)] + (\nabla_\mathbf{x} g)^T \mathbf{A}\mathbf{x}(t) - \tfrac{1}{4} (\nabla_\mathbf{x} g)^T \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T \nabla_\mathbf{x} g$$

This nonlinear partial differential equation can be solved by assuming a solution $g[\mathbf{x}(t), t_r] = \mathbf{x}^T(t)\mathbf{W}(t_r)\mathbf{x}(t) + \mathbf{x}^T(t)\mathbf{V}(t_r) + Z(t_r)$, where $\mathbf{W}$ is an unknown symmetric $n \times n$ matrix, $\mathbf{V}$ is an unknown $n \times 1$ vector, and Z is an unknown scalar. Differentiating the assumed answer, treating $\mathbf{x}(t)$ and t as independent variables, gives

$$\frac{\partial g}{\partial t_r} = \mathbf{x}^T(t)\frac{d\mathbf{W}}{dt_r}\mathbf{x}(t) + \mathbf{x}^T(t)\frac{d\mathbf{V}}{dt_r} + \frac{dZ}{dt_r} \quad \text{and} \quad \nabla_{\mathbf{x}} g = 2\mathbf{W}(t_r)\mathbf{x}(t) + \mathbf{V}(t_r)$$

Using these, the H.J.B. equation becomes

$$\begin{aligned}\mathbf{x}^T(t)\frac{dW}{dt_r}\mathbf{x}(t) + \mathbf{x}^T(t)\frac{d\mathbf{V}}{dt_r} + \frac{dZ}{dt_r} &= \mathbf{x}^T\{\mathbf{Q} + 2\mathbf{WA} - \mathbf{WBR}^{-1}\mathbf{B}^T\mathbf{W}\}\mathbf{x} \\ &+ \mathbf{x}^T\{-2\mathbf{Q}\boldsymbol{\eta} + \mathbf{A}^T\mathbf{V} - \mathbf{WBR}^{-1}\mathbf{B}^T\mathbf{V}\} \\ &+ \{\boldsymbol{\eta}^T\mathbf{Q}\boldsymbol{\eta} - \tfrac{1}{4}\mathbf{V}^T\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{V}\}\end{aligned}$$

In order for the assumed form to actually be a solution for all $\mathbf{x}(t)$, the quadratic terms in $\mathbf{x}$, the linear terms, and the terms not involving $\mathbf{x}$ must balance individually. Therefore,

$$\frac{dZ}{dt_r} = \boldsymbol{\eta}(t)^T\mathbf{Q}\boldsymbol{\eta}(t) - \tfrac{1}{4}\mathbf{V}^T(t_r)\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{V}(t_r) \qquad \text{(terms not involving } \mathbf{x}(t)) \qquad (14.37)$$

The linear terms in $\mathbf{x}(t)$ require

$$\frac{d\mathbf{V}}{dt_r} = -2\mathbf{Q}\boldsymbol{\eta}(t) + \mathbf{A}^T\mathbf{V}(t_r) - \mathbf{W}(t_r)\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{V}(t_r) \qquad (14.38)$$

Each matrix involved in the quadratic terms is symmetric except $\mathbf{x}^T\{2\mathbf{WA}\}\mathbf{x}$, which is rewritten as $\mathbf{x}^T\{\mathbf{WA} + \mathbf{A}^T\mathbf{W}\}\mathbf{x} + \mathbf{x}^T\{\mathbf{WA} - \mathbf{A}^T\mathbf{W}\}\mathbf{x}$. The second term is always zero since the matrix is skew-symmetric. All quadratic terms in the H.J.B. equation can be combined into the form $\mathbf{x}^T\mathbf{Px} = 0$. Since every matrix term in $\mathbf{P}$ is now symmetric, it can be concluded that $\mathbf{P} = \mathbf{0}$, or

$$\frac{d\mathbf{W}}{dt_r} = \mathbf{Q} + \mathbf{WA} + \mathbf{A}^T\mathbf{W} - \mathbf{WBR}^{-1}\mathbf{B}^T\mathbf{W} \qquad (14.39)$$

The boundary conditions for the three sets of differential equations are $\mathbf{W}(t_r = 0) = \mathbf{M}$, $\mathbf{V}(t_r = 0) = -2\mathbf{Mx}_d$, and $Z(t_r = 0) = \mathbf{x}_d^T\mathbf{Mx}_d$. Equation (*14.39*) is known as the matrix Riccati differential equation. It can be solved first, and the result can then be used in solving Eq. (*14.38*), after which Eq. (*14.37*) can be integrated. The optimal feedback control system can be represented by either Figure 14.8*a* or *b*. The equivalence is established by using the relations $\mathbf{G}(t_r) = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{W}(t_r)$ and $\mathbf{F}_c(t_r) = -\tfrac{1}{2}\mathbf{R}^{-1}\mathbf{B}^T$. Since the major effort involves solving the Riccati equation, it is considered in detail next. The form of the matrix Riccati equation that is found most frequently in the controls literature is

$$-\dot{\mathbf{W}}(t) = \mathbf{W}(t)\mathbf{A} + \mathbf{A}^T\mathbf{W}(t) - \mathbf{W}(t)\mathbf{BR}^{-1}\mathbf{B}^T\mathbf{W}(t) + \mathbf{Q} \qquad (14.40)$$

The alternate form expressed in terms of time remaining, t_r, has a sign change on the derivative term, since $d\{\ \}/dt_r = -d\{\ \}/dt$. This nonlinear differential equation can be

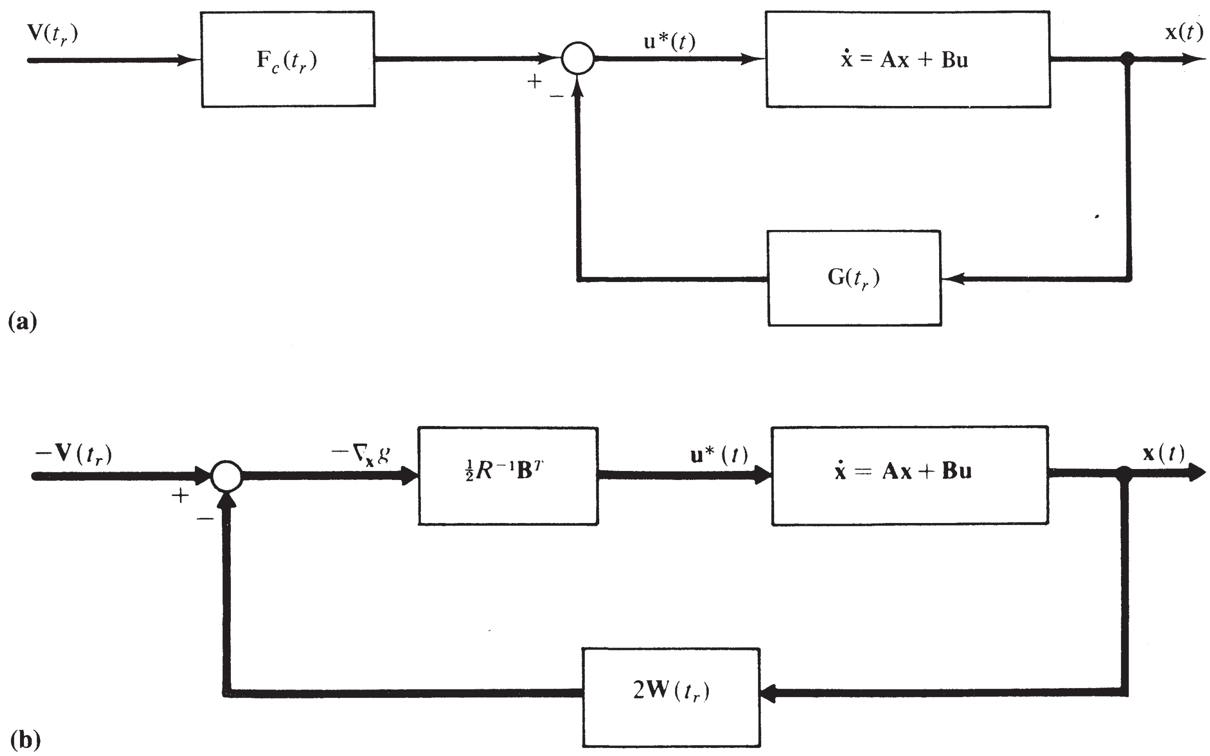


Figure 14.8

transformed into a pair of linear differential equations, in an analogous fashion to the treatment of the discrete-time Riccati equation in Sec. 14.3.3. Let $\mathbf{W} = \mathbf{E}\mathbf{F}^{-1}$. Since $\mathbf{F}\mathbf{F}^{-1} = \mathbf{I}$, $d\{\mathbf{F}\mathbf{F}^{-1}\}/dt = [\mathbf{0}]$, or

$$\dot{\mathbf{F}}\mathbf{F}^{-1} + \mathbf{F}d\{\mathbf{F}^{-1}\}/dt = [\mathbf{0}] \quad \text{or} \quad d\{\mathbf{F}^{-1}\}/dt = -\mathbf{F}^{-1}\dot{\mathbf{F}}\mathbf{F}^{-1}$$

From this, $\dot{\mathbf{W}} = \dot{\mathbf{E}}\mathbf{F}^{-1} - \mathbf{E}\mathbf{F}^{-1}\dot{\mathbf{F}}\mathbf{F}^{-1}$. Using this in Eq. (*14.40*) and then postmultiplying by $\mathbf{F}$ gives

$$-(\dot{\mathbf{E}} - \mathbf{E}\mathbf{F}^{-1}\dot{\mathbf{F}}) = \mathbf{E}\mathbf{F}^{-1}\mathbf{A}\mathbf{F} + \mathbf{A}^T\mathbf{E} - \mathbf{E}\mathbf{F}^{-1}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{E} + \mathbf{Q}\mathbf{F}$$

If the terms linear in $\mathbf{E}$ and $\mathbf{F}$ are equated, that is,

$$-\dot{\mathbf{E}} = \mathbf{A}^T\mathbf{E} + \mathbf{Q}\mathbf{F} \tag{14.41}$$

then the remaining nonlinear terms give

$$\mathbf{E}\mathbf{F}^{-1}\dot{\mathbf{F}} = \mathbf{E}\mathbf{F}^{-1}\mathbf{A}\mathbf{F} - \mathbf{E}\mathbf{F}^{-1}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{E}$$

When this is premultiplied by $[\mathbf{E}\mathbf{F}^{-1}]^{-1}$, the second linear equation is found to be

$$\dot{\mathbf{F}} = \mathbf{A}\mathbf{F} - \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{E} \tag{14.42}$$

Equations (*14.41*) and (*14.42*) can be stacked into one homogeneous linear system of coupled equations:

$$\begin{bmatrix} \dot{\mathbf{F}} \\ \dot{\mathbf{E}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T \\ -\mathbf{Q} & -\mathbf{A}^T \end{bmatrix} \begin{bmatrix} \mathbf{F} \\ \mathbf{E} \end{bmatrix} \tag{14.43}$$

Note that the superior dots indicate $d\{\ \}/dt$ rather than $d\{\ \}/dt_r$. The $2n \times 2n$ coefficient matrix on the right side of Eq. (*14.43*) is generally referred to as the Hamiltonian matrix **H.** Its form is different from—and should not be confused with—the discrete-time version of Sec. 14.3. However, similar to the discrete case, this **H** also has its eigenvalues occurring in stable-unstable pairs. That is, if λ is an eigenvalue of **H,** then so is $-\lambda$. Equation (*14.43*) has a solution in terms of the exponential matrix, that is,

$$\begin{bmatrix}\mathbf{F}(t)\\ \mathbf{E}(t)\end{bmatrix} = \exp\{(t - t_0)\mathbf{H}\}\begin{bmatrix}\mathbf{F}(t_0)\\ \mathbf{E}(t_0)\end{bmatrix} \tag{14.44}$$

for any t, t_0 pair. Since the "initial" conditions on **W**—and hence **E** and **F**—are given at time t_f and not t_0, Eq. (*14.44*) does not appear ready for use. Two possible modification methods exist, which show that it is possible to replace t_0 by t_f. The first method reverts to the time-remaining variable t_r and gives a sign change on the derivative terms. Since $t = t_f$ implies $t_r = 0$, this means that

$$\begin{bmatrix}\mathbf{F}(t_r)\\ \mathbf{E}(t_r)\end{bmatrix} = \exp\{-t_r\mathbf{H}\}\begin{bmatrix}\mathbf{F}(t_r = 0)\\ \mathbf{E}(t_r = 0)\end{bmatrix} \tag{14.45}$$

The other method evaluates Eq. (*14.44*) with the general time t replaced by the final time t_f and the initial time t_0 replaced by the current time t. This gives

$$\begin{bmatrix}\mathbf{F}(t_f)\\ \mathbf{E}(t_f)\end{bmatrix} = \exp\{(t_f - t)\mathbf{H}\}\begin{bmatrix}\mathbf{F}(t)\\ \mathbf{E}(t)\end{bmatrix} \tag{14.46}$$

Inverting the exponential matrix, which just changes the sign in its exponent, to solve for current values gives Eq. (*14.44*) with t_0 replaced by t_f. Equation (*14.45*) gives the same expression, provided it is recognized that the different methods of indexing time arguments both refer to the same physical time instant. Since $\mathbf{W}(t = t_f) = \mathbf{M}$, we set $\mathbf{E}(t = t_f) \equiv \mathbf{E}(t_r = 0) = \mathbf{M}$ and $\mathbf{F}(t_r = 0) = \mathbf{I}$. The similarity transformation relation between **H** and its Jordan form $\mathbf{J} = \text{Diag}[\mathbf{J}_s \quad \mathbf{J}_u]$ is $\mathbf{H} = \mathbf{TJT}^{-1}$, where, as before, $\mathbf{J}_s$ and $\mathbf{J}_u$ contain the stable and unstable blocks and where $\mathbf{T} = \begin{bmatrix}\mathbf{T}_{11} & \mathbf{T}_{12}\\ \mathbf{T}_{21} & \mathbf{T}_{22}\end{bmatrix}$ is the modal matrix of eigenvectors of **H,** with the first n columns being associated with the stable eigenvectors. Let $\mathbf{T}^{-1} = \begin{bmatrix}\mathbf{V}_{11} & \mathbf{V}_{12}\\ \mathbf{V}_{21} & \mathbf{V}_{22}\end{bmatrix}$. Since $\exp\{-t_r\mathbf{H}\} = \mathbf{T}\ \text{Diag}[\exp\{-t_r\mathbf{J}_s\}\ \exp\{-t_r\mathbf{J}_u\}]\mathbf{T}^{-1}$, the solution for **W** at a given time t (or the same corresponding time-remaining value t_r) is found using essentially the same steps which lead to Eq. (*14.29*)

$$\begin{aligned}\mathbf{W}(t_r) = [&\mathbf{T}_{21}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{11} + \mathbf{T}_{22}\exp\{-t_r\mathbf{J}_u\}\mathbf{V}_{21}\\ &+ (\mathbf{T}_{21}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{12} + \mathbf{T}_{22}\exp\{-t_r\mathbf{J}_u\}\mathbf{V}_{22})\mathbf{M}] \ast\\ &[\mathbf{T}_{11}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{11} + \mathbf{T}_{12}\exp\{-t_r\mathbf{J}_u\}\mathbf{V}_{21}\\ &+ (\mathbf{T}_{11}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{12} + \mathbf{T}_{12}\exp\{-t_r\mathbf{J}_u\}\mathbf{V}_{22})\mathbf{M}]^{-1}\end{aligned} \tag{14.47}$$

Once $\mathbf{W}(t_r)$ is determined, the control law is given by

$$\mathbf{u}^*(t) = -\mathbf{R}^{-1}\mathbf{B}^T\{\mathbf{W}(t_r)\mathbf{x}(t) + \tfrac{1}{2}\mathbf{V}(t_r)\} \tag{14.48}$$

The external input command $\mathbf{V}(t_r)$ is zero for the regulator problem. In the tracking problem it must be determined by solving Eq. (*14.38*) using the now-known matrix $\mathbf{W}(t_r)$ and the specified state trajectory $\boldsymbol{\eta}(t)$ which is to be tracked.

14.4.2 Infinite Time-to-Go Problem; The Algebraic Riccati Equation

A commonly used simplification of the previous LQ solution is to let $t_r \to \infty$. This is equivalent to letting the derivative $\dot{\mathbf{W}}$ go to zero, leaving the so-called algebraic Riccati equation (ARE)

$$\mathbf{A}^T\mathbf{W} + \mathbf{W}\mathbf{A} - \mathbf{W}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{W} + \mathbf{Q} = \mathbf{0} \qquad (14.49)$$

A number of ways of solving the ARE exist. For low-order problems, it may be feasible to write out the components explicitly. Using the known symmetry of $\mathbf{W}$ will yield $n(n+1)/2$ coupled quadratic equations. Since quadratics have multiple solutions, a question arises about which, if any, of these solutions is the correct one for the problem at hand. Under the assumptions that $\mathbf{R}$ is positive definite and that $\mathbf{Q}$ is at least positive semidefinite, it is known that one unique positive definite solution for $\mathbf{W}$ exists provided that the system $\{\mathbf{A}, \mathbf{B}\}$ is stabilizable and $\{\mathbf{A}, \mathbf{C}\}$ is detectible, where $\mathbf{C}^T\mathbf{C} = \mathbf{Q}$. A stronger and numerically safer set of requirements is sometimes given, namely: Either $\mathbf{A}$ is asymptotically stable or $\{\mathbf{A}, \mathbf{B}\}$ is controllable and $\{\mathbf{A}, \mathbf{C}\}$ is observable.

If a positive definite solution is known to exist because of satisfaction of these requirements, it could be found by numerical integration of Eq. (*14.39*), backward in time, until $\mathbf{W}$ approaches its constant final value. This approach is discussed in Problem 14.11, along with some suggestions for integration step-size selection. The general solution found in the last section can be used to develop another approach to finding the infinite-time-remaining solution $\mathbf{W}_\infty$. When $t_r \to \infty$, $\exp\{-t_r\mathbf{J}_s\} \to \infty$ and $\exp\{-t_r\mathbf{J}_u\} \to \mathbf{0}$. The general results then reduce to

$$\mathbf{W}_\infty = [\mathbf{T}_{21}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{11} + \mathbf{T}_{21}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{12}\mathbf{M}] *$$
$$[\mathbf{T}_{11}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{11} + \mathbf{T}_{11}\exp\{-t_r\mathbf{J}_s\}\mathbf{V}_{12}\mathbf{M}]^{-1}$$

which reduces to

$$\mathbf{W}_\infty = \mathbf{T}_{21}\mathbf{T}_{11}^{-1} \qquad (14.50)$$

It is of interest to note that if Eq. (*14.44*) is treated in a similar fashion but with $t \to \infty$, it is found that a "steady-state" solution for $\mathbf{W}$ is given by $\mathbf{W}_{ss} = \mathbf{T}_{22}\mathbf{T}_{12}^{-1}$. This is *not* the correct solution and explains why the commonly used terminology *steady-state solution* must be used with care.

In calculating Eq. (*14.50*) and/or Eq. (*14.30*), the eigenvalues and eigenvectors of $\mathbf{H}$ will often be complex. However, Problem 4.23 shows that only real arithmetic is needed to calculate $\mathbf{W}_\infty$. It is also pointed out that the n stable eigenvalues of $\mathbf{H}$ are exactly the closed-loop eigenvalues of the system when the constant feedback gain control law $\mathbf{u}(t) = -\mathbf{R}^{-1}\mathbf{B}^T\mathbf{W}_\infty\mathbf{x}(t) = -\mathbf{G}_\infty\mathbf{x}(t)$ is used. The following example uses simple second-order systems to illustrate the determination of the optimal feedback gains and several potential pitfalls.

EXAMPLE 14.10 Analyze the ARE for each of the following systems, with $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = 1$.

(a) $\mathbf{A} = \begin{bmatrix} -3 & 1 \\ 0 & 2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. This system is in the Kalman canonical form, which shows it to be unstable, and the unstable mode is uncontrollable; hence it is not stabilizable. Ignoring this fact, it is found that the eigenvalues of $\mathbf{H}$ are ± 2 and ± 3.16227. The eigenvectors associated with the stable eigenvalues are found to be

$$\begin{bmatrix} \mathbf{T}_{11} \\ \hline \mathbf{T}_{21} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 0 & 0 \\ \hline 0 & -0.16228 \\ -1 & -0.13962 \end{bmatrix}$$

Clearly $\mathbf{T}_{11}$ is singular, and the method of Eq. (*14.50*) fails, as well it should. If the numerical integration method is attempted, it also fails because the solution never settles to a constant matrix.

(b) $\mathbf{A} = \begin{bmatrix} -3 & 1 \\ 0 & -2 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. The uncontrollable mode is now stable; hence this system is stabilizable. Using the eigenvalues-eigenvectors of $\mathbf{H}$, it is found that $\mathbf{W}_\infty = \begin{bmatrix} 0.16228 & 0.031435 \\ 0.031435 & 0.26547 \end{bmatrix}$, $\mathbf{G}_\infty = [0.16228 \quad 0.031435]$, and the closed loop eigenvalues are $\lambda = -2, -3.16228$. Note that the uncontrollable mode has its eigenvalue unchanged from the open-loop value of $\lambda = -2$. This solution is also verified by using fourth-order Runge-Kutta integration on Eq. (*14.39*).

(c) $\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & -3 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. The same two eigenvalues occur here as in part (a), but now the unstable mode is controllable; hence the system is stabilizable. Routine application of both the eigenvector method and numerical integration give $\mathbf{W}_\infty = \begin{bmatrix} 4.23607 & 0.80902 \\ 0.80902 & 0.32725 \end{bmatrix}$, $\mathbf{G}_\infty = [4.2361 \quad 0.80902]$, and the closed-loop eigenvalues are at $\lambda = -3, -2.23607$. Again the uncontrollable mode has its eigenvalue unchanged.

(d) $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$. The Hamiltonian matrix is $\mathbf{H} = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & -3 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 3 \end{bmatrix}$. The eigenvalues are found to be at $\lambda = -1, 1, -3$, and 3. The eigenvectors for -1 and -3 are $[-1 \quad 0 \quad -1 \quad -0.25]^T$ and $[0.375 \quad -1 \quad 0.125 \quad -0.145833]^T$, respectively. These could be used in Eq. (*14.50*) to find $\mathbf{W}_\infty$. Since this system is stabilizable, it must have a unique positive definite solution for $\mathbf{W}_\infty$. The direct solution of the ARE will be used here to find the result. $\mathbf{A}^T\mathbf{W} + \mathbf{W}\mathbf{A} - \mathbf{W}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{W} + \mathbf{Q} = \mathbf{0}$ expands to

$$\begin{bmatrix} 0 & 0 \\ 1 & -3 \end{bmatrix}\begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix} + \begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & -3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$- \begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} w_{11} & w_{12} \\ w_{12} & w_{22} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

From the 1, 1 term, $w_{11}^2 = 1$. $\mathbf{W}$ must be positive definite, so $w_{11} = 1$. From the 1, 2 or 2, 1 terms, $w_{11} - 3w_{12} - w_{12}w_{11} = 0$, so $w_{12} = 0.25$. From the 2, 2 term, $2w_{12} - 6w_{22} - w_{12}^2 + 1 = 0$ gives $w_{22} = 0.23958$. From this, the constant gain matrix is $\mathbf{G}_\infty = [1 \quad 0.25]$, and the closed-loop eigenvalues are at $\lambda = -1$ and -3. Attempts to solve this problem with numerical integration gave mixed results. The correct answer was sometimes found and other times not, depending upon the integration step size and the stopping criteria.

(e) A similar example, with pure integrators in the open-loop system, has $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Note that **B** has been changed to make this system controllable. The Hamiltonian matrix is $\mathbf{H} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & -1 & 0 \end{bmatrix}$. The eigenvalues are found to be at $\lambda = -0.86615 \pm 0.5j$ and $0.86615 \pm 0.5j$. Again, the clean separation into mirror-image stable and unstable eigenvalues is noted. It is easy to use the expanded components of the ARE, as in part (d), to find $\mathbf{W}_\infty = \begin{bmatrix} \sqrt{3} & 1 \\ 1 & \sqrt{3} \end{bmatrix}$, $\mathbf{G}_\infty = [1 \quad \sqrt{3}]$, and the closed loop eigenvalues are at $\lambda = -0.866 \pm 0.5j$. The same results are obtained by using the two stable eigenvectors $\boldsymbol{\xi}_1 = [-0.5 - 0.866j \quad 0.866 + 0.5j \quad -j \quad 1]^T$ and its complex conjugate in Eq. (*14.50*). To avoid inverting a complex matrix, the result of Problem 4.23 can be used to write $\mathbf{W}_\infty = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} -0.5 & -0.866 \\ 0.866 & 0.5 \end{bmatrix}^{-1}$. This problem can also be solved using numerical integration, if appropriate step-size and stopping criterion are selected. The suggestions given in Problem 14.11 are not useful when all the open-loop poles are at the origin. ■

For simple low-order problems, a variety of solution methods can be used, which become impractical in realistic problems. The eigenvector method of Eq. (*14.50*) is capable of solving most ARE problems for which unique solutions are known to exist. There may be occasions where the $2n \times 2n$ Hamiltonian matrix has eigenvalues on the $j\omega$ axis. The clean separation into stable and unstable values thus breaks down. In these cases, a small change to **Q, R,** or even **A** might be used to allow the algorithm to proceed successfully. The system of Example 14.10(e), with $\mathbf{Q} = [\mathbf{0}]$, is one such case. Essentially the same answer would be obtained by changing to a very small nonzero matrix **Q** or by setting the diagonal terms of **A** to a small nonzero value ϵ. Other solution methods are also known [2].

Note that when the cost of control is very expensive—i.e., when $\mathbf{R} \to \infty$—the quadratic terms become vanishingly small and Eq. (*14.49*) reduces to a Lyapunov equation. Solutions of the Lyapunov equation were examined in detail in Sec. 6.10. This limiting result does not seem to be especially useful because it leads to a vanishingly small feedback gain $\mathbf{G}_\infty = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{W}$ as well, so the system has no feedback control.

14.5 PONTRYAGIN'S MINIMUM PRINCIPLE

Optimal control problems can be analyzed from a number of alternative viewpoints. The continuous-time versions of Eqs. (*14.1*) and (*14.2*) are considered again. If the set of admissible controls is unrestricted, the calculus of variations [6] can be used to derive necessary conditions which characterize the optimal solution. When the admissible control set is bounded, unrestricted variations in $\mathbf{u}(t)$ are not allowed. This situation is analogous to the problem of finding the minimum of a function on a closed and bounded interval. If the minimum occurs at a boundary point, then it is not necessarily true that the first variation (analogous to the slope) vanishes at that minimal point. Pontryagin's minimum principle [5] is an extension of the methods of variational calculus to problems with bounded control and/or state variables. The simple version of the minimum principle presented next provides a set of necessary conditions for optimality (see Problems 14.12 and 14.13). This brief introduction to the theory is incomplete in that bounded states are not considered, sufficiency conditions are not treated, etc. A heuristic relation to dynamic programming results is given.

The pre-Hamiltonian is defined as the scalar function

$$\mathcal{H}(\mathbf{x}, \mathbf{u}, \mathbf{p}, t) \triangleq L(\mathbf{x}, \mathbf{u}, t) + \mathbf{p}^T(t)\mathbf{f}(\mathbf{x}, \mathbf{u}, t) \qquad (14.51)$$

where $\mathbf{p}(t)$ is the $n \times 1$ *costate vector* and satisfies

$$\dot{\mathbf{p}} = -\left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]^T \mathbf{p} - \nabla_{\mathbf{x}} L(\mathbf{x}, \mathbf{u}, t) \qquad (14.52)$$

The minimum principle states that the optimal control $\mathbf{u}^*(t)$ is that member of the admissible control set U which minimizes $\mathcal{H}$ at every time. If $\mathbf{u}(t)$ has r components, then minimizing $\mathcal{H}$ gives r algebraic equations which allow the determination of $\mathbf{u}^*(t)$ in terms of the still unknown $\mathbf{p}(t)$ and $\mathbf{x}(t)$. Then $\mathbf{u}(t)$ can be eliminated from equations (*14.1*) and (*14.52*). These equations can also be written in canonical form as

$$\dot{\mathbf{x}} = \frac{\partial \mathcal{H}}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial \mathcal{H}}{\partial \mathbf{x}} \qquad (14.53)$$

Equation (*14.53*) consists of $2n$ first-order differential equations, so $2n$ boundary conditions are required for solution. The initial conditions $\mathbf{x}(t_0) = \mathbf{x}_0$ give n of them. The remaining n conditions will apply at the final time t_f. Their exact nature depends on the particular problem. $\mathbf{x}(t_f)$ will be specified directly in some cases. If $\mathbf{x}(t_f)$ is free, then $\mathbf{p}(t_f) = \nabla_{\mathbf{x}} S(\mathbf{x}(t), t)|t_f$. Combinations of these two kinds of conditions, as well as others, can arise. If the final time t_f is not specified, an additional equation is required to determine it (see Problem 14.13).

Equation (*14.53*), along with n boundary conditions at t_0 and n more at t_f, constitutes a two-point boundary value problem. Linear two-point boundary value problems are easily solved, at least in principle. Nonlinear two-point boundary value problems are generally difficult to solve, even numerically. Much of the effort in optimal control theory has been devoted to the development of efficient algorithms for computer solution of these problems (Problem 14.14).

Note that if $\mathbf{p}$ is defined as $\nabla_x g$, then Eq. (*14.35*) of dynamic programming also indicates that the pre-Hamiltonian must be minimized at each instant by proper choice of $\mathbf{u}(t)$. When the pre-Hamiltonian is evaluated with all its arguments optimally selected, it is called simply the Hamiltonian $\mathcal{H}^*$. (Often the * is omitted.) Equation (*14.35*) gives physical meaning to $\mathcal{H}^*$. It is the rate of change of the cost g as time-remaining changes, $\partial g/\partial t_r = \mathcal{H}^*$. If sufficient continuity properties are assumed for g, the gradient with respect to the state $\mathbf{x}$ gives

$$\nabla_x(\partial g/\partial t_r) = \partial\{\nabla_x g\}/\partial t_r = \nabla_x \mathcal{H}^*$$

By reverting to t instead of t_r, using the definition $\mathbf{p} = \nabla_x g$, and using the alternate notation $\partial\{\ \}/\partial\mathbf{x}$ for the gradient, the second of Eq. (*14.53*) is obtained. The first of this canonical pair is automatically satisfied because of the definition of $\mathcal{H}$. Thus the essential features of the minimum principle follow directly from the dynamic programming approach presented earlier. In some problems, especially those involving the exact satisfaction of terminal boundary conditions, the minimum principle formulation seems more convenient. This is explored in the examples and problems.

EXAMPLE 14.11 A simplified model of the linear motion of an automobile is $\dot{x} = u$, where $x(t)$ is the vehicle velocity and $u(t)$ is the acceleration or deceleration. The car is initially moving at x_0 ft/sec. Find the optimal $u(t)$ which brings the velocity $x(t_f)$ to zero in *minimum time* t_f. Assume that acceleration and braking limitations require $|u(t)| \leq M$ for all t.

The minimum time performance criterion is $J = \int_0^{t_f} 1\ dt$, so the Hamiltonian is $\mathcal{H} = 1 + p(t)u(t)$. In order to minimize $\mathcal{H}$, it is obvious that $u(t) = -M$ if $p(t) > 0$ and $u(t) = M$ if $p(t) < 0$. That is, $u^*(t) = -M \text{ sign } (p(t))$. The optimal control has been found as a function of the unknown $p(t)$. The differential equation for p is $\dot{p} = -\partial\mathcal{H}/\partial x = 0$, so $p(t)$ is constant. The value of this constant must be determined from boundary conditions. (Actually only the sign of the constant is needed for this problem.) The available boundary conditions are $x(0) = x_0$, $x(t_f) = 0$. The form of the solution for $x(t)$ is $x(t) = x_0 + \int_0^t u(\tau)\, d\tau$, but $u(t)$ is a constant, either $+M$ or $-M$. Therefore, $x(t_f) = x_0 \pm Mt_f = 0$. Clearly, if $x_0 > 0$, then $u = -M$, maximum deceleration. If $x_0 < 0$, then $u = M$, maximum acceleration. That is, $u^*(t) = -M \text{ sign } (x_0)$. The minimum stopping time is $t_f = |x_0|/M$. Note that $u^*(t)$ is expressed in terms of $x(t_0)$, so this represents an open-loop control law. This is a typical result of using the minimum principle. ■

EXAMPLE 14.12 Consider the linear, constant system $\dot{\mathbf{x}} = \mathbf{Ax} + \mathbf{Bu}$ with $\mathbf{u}(t)$ unrestricted. Find $\mathbf{u}(t)$ which minimizes a trade-off between terminal error and control effort,

$$J = [\mathbf{x}(t_f) - \mathbf{x}_d]^T[\mathbf{x}(t_f) - \mathbf{x}_d] + \int_0^{t_f} \mathbf{u}^T(t)\mathbf{u}(t)\, dt$$

Time t_f is fixed.

The Hamiltonian is $\mathcal{H} = \mathbf{u}^T\mathbf{u} + \mathbf{p}^T[\mathbf{Ax} + \mathbf{Bu}]$. $\mathcal{H}$ is minimized by setting

$$\frac{\partial\mathcal{H}}{\partial\mathbf{u}} = \mathbf{0} = 2\mathbf{u} + \mathbf{B}^T\mathbf{p} \quad \text{or} \quad \mathbf{u}^*(t) = -\tfrac{1}{2}\mathbf{B}^T\mathbf{p}(t)$$

This is not yet a useful answer, since $\mathbf{p}(t)$ is unknown. However, $\dot{\mathbf{p}} = -\partial\mathcal{H}/\partial\mathbf{x} = -\mathbf{A}^T\mathbf{p}$. Combining this with the original system equation gives, after eliminating $\mathbf{u}(t)$,

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{p}} \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{A} & -\tfrac{1}{2}\mathbf{BB}^T \\ \hline \mathbf{0} & -\mathbf{A}^T \end{array}\right] \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix} \tag{14.54}$$

Note that the $2n \times 2n$ coefficient matrix is exactly the Hamiltonian matrix arising out of our solution to the Riccati equation of Sec. 14.4, in the special case of $\mathbf{R} = \mathbf{I}$ and $\mathbf{Q} = [\mathbf{0}]$. The $2n$ boundary conditions are $\mathbf{x}(0) = \mathbf{x}_0$ and $\mathbf{p}(t_f) = \nabla_\mathbf{x} S|_{t_f} = 2[\mathbf{x}(t_f) - \mathbf{x}_d]$. This linear two-point boundary value problem is treated in Problem 14.15. ■

14.6 THE SEPARATION THEOREM

In previous sections of this chapter it has been assumed that the entire state vector $\mathbf{x}$ can be measured and used in forming the feedback control signal. In Chapter 13 it was shown that if the system is *linear* and if the control law is *linear,* then a *linear* observer can be used to estimate $\mathbf{x}$ from the available outputs $\mathbf{y}$ without changing the closed-loop poles which the controller is designed to give.

Another version of the separation theorem is more commonly stated in conjunction with optimal control problems. It states that *if*

a. the system models are linear (both the dynamics of Eq. (*3.11*) or (*3.13*) and the output equation of Eq. (*3.12*) or (*3.14*)) and
b. all measurement errors and disturbances have Gaussian probability density functions, and
c. the cost function of Eq. (*14.2*) is quadratic,

then the expected value of the cost function J is minimized by

1. designing an optimal controller based on the assumption that all states are available,
2. designing an optimal estimator to provide an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$,
3. using $\hat{\mathbf{x}}$ in place of $\mathbf{x}$ in the control law of step 1.

These two versions of the separation principle are closely related because of the following facts: The optimal control law for a linear system with a quadratic cost function is a linear control law. The optimal estimator for a linear system subjected to Gaussian noise is a linear estimator of the same form as the linear observer of Chapter 13. While version one is true for *any* linear controller and linear estimator, regardless of how they were designed, and without any Gaussian assumptions, all that is guaranteed is that the desired closed-loop poles are still achieved. Version two guarantees the optimality of J in a stochastic sense, but demands somewhat stricter assumptions. The class of problems that meet these restrictions is referred to as *LQG* problems (linear, quadratic, Gaussian). Most successful applications of optimal control theory so far have been members of this class.

The implication of the separation theorem is that the assumption that the full state is available can be relaxed. The optimal controllers in this chapter can be cascaded with an observer (as in Chapter 13) or a Kalman filter (Problem 6.18) to obtain the overall design. This practice is generally not valid outside the *LQG* class, although it is still sometimes used as a rather effective suboptimal scheme.

Intuitively, one might expect that if control inputs could be tailored in some way, certain system modes would be more strongly excited, making their states easier to estimate. This idea is the basis for "probing"-type control signals. Likewise, it seems intuitive that if one or more states are only known to within some level of uncertainty, but the control system behaves as if it were exactly known, then overly bold or decisive control actions might incorrectly be taken at times. A more cautious controller which somehow takes into account its knowledge of the uncertainty in the state estimates might give better overall performance. The notions of probing and caution come into play in those cases where the separation theorem does not apply. This is still a subject of active research. To give any more than the above intuitive definitions would go beyond the scope of this book.

14.7 ROBUSTNESS ISSUES

Kalman [7] first showed that a single input control system designed with the LQ theory has a scalar return difference (first introduced in Sec. 2.4) which satisfies $|F_d(j\omega)| \geq 1$ for all ω. Using Nyquist frequency response methods, it can be shown [7] that this guarantees a phase margin $PM \geq 60°$ and an upward gain margin $GM = \infty$. The downward gain reduction margin is 0.5. This means that the LQ system will remain stable for any arbitrary gain increase and for all gains at least half the nominal design value.

These reassuring stability margins have also been shown to apply to the multiple-input case, at least for a diagonal $\mathbf{R}$ weighting matrix [8]. This was shown by using the return difference matrix $\mathbf{R}_d$ and by using its smallest singular value as the measure of how close it approaches singularity. The gain margin portion of this result is easily verified in the time-domain, state space domain by using Lyapunov stability theory of Sec. 10.6. Suppose the positive definite matrix $\mathbf{W}_\infty$ from the ARE is used to define a Lyapunov function

$$V = \mathbf{x}^T \mathbf{W}_\infty \mathbf{x}$$

for the closed-loop LQ regulator described by

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{B}\mathbf{G}_\infty)\mathbf{x}$$

(Since the presence of an external input for the LQ tracking problem does not alter the loop stability characteristics, only the regulator problem is considered for simplicity.) Then

$$\begin{aligned} \dot{V} &= \dot{\mathbf{x}}^T \mathbf{W}_\infty \mathbf{x} + \mathbf{x}^T \mathbf{W}_\infty \dot{\mathbf{x}} \\ &= \mathbf{x}^T \{\mathbf{A}^T \mathbf{W}_\infty + \mathbf{W}_\infty \mathbf{A} - \mathbf{G}_\infty^T \mathbf{B}^T \mathbf{W}_\infty - \mathbf{W}_\infty \mathbf{B} \mathbf{G}_\infty\} \mathbf{x} \end{aligned}$$

Assume that the gain actually used in the feedback loop is $\mathbf{G}_\infty = \beta \mathbf{R}^{-1} \mathbf{B}^T \mathbf{W}_\infty$, that is, it equals the optimal LQ gain only when $\beta = 1$. Then

$$\dot{V} = \mathbf{x}^T \{\mathbf{A}^T \mathbf{W}_\infty + \mathbf{W}_\infty \mathbf{A} - 2\beta \mathbf{W}_\infty \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{W}_\infty\} \mathbf{x}$$

which can be regrouped into

$$\dot{V} = \mathbf{x}^T \{\mathbf{A}^T \mathbf{W}_\infty + \mathbf{W}_\infty \mathbf{A} - \mathbf{W}_\infty \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{W}_\infty + \mathbf{Q}\} \mathbf{x} - \mathbf{x}^T \{\mathbf{Q} + (2\beta - 1) \mathbf{W}_\infty \mathbf{B} \mathbf{R}^{-1} \mathbf{B}^T \mathbf{W}_\infty\} \mathbf{x}$$

Since $\mathbf{W}_\infty$ satisfies the ARE, the first quadratic form is zero. Since $\mathbf{Q}$ is at least positive semidefinite, the second quadratic form will be at least positive semidefinite, provided that $\beta \geq 0.5$. Lyapunov's theorem therefore guarantees that the gain-perturbed LQ system remains stable provided that $\frac{1}{2} \leq \beta < \infty$.

While this simple verification gives insight into the robustness issue, most thorough analyses of the subject are based on singular value analysis of the return difference matrix. It should be pointed out that singular value analysis can sometimes give overly conservative estimates of system robustness. The minimum singular value is a measure of how close the matrix is to the nearest singular matrix. Actual perturbations in a physical model may restrict the kinds of changes that can occur in a system matrix in such a way as to preclude the "nearest singular matrix" from ever occurring. For example, if a resistor value is the major unknown parameter, these model perturbations may cause only a single matrix element to vary. This points up the difference between unstructured model perturbations and structured perturbations [9, 10]. Obviously, an analysis that uses more information about the form which model perturbations might take is likely to give sharper answers.

A useful theorem [11, 12], which is valid for any arbitrary unstructured model perturbations, gives the following bounds on the stability margins. Let $\boldsymbol{\sigma}(\omega)$ be the minimum singular value of a return difference matrix at any given frequency ω. If there exists a bound $\alpha \le 1$ such that $\boldsymbol{\sigma}(\omega) \ge \alpha$ for all ω, then

$$PM \ge \pm 2 \sin^{-1}(\alpha/2) \quad \text{and} \quad GM = 1/(1 \pm \alpha)$$

These margins apply for gain or phase perturbations introduced in the control loop at the point for which the return difference matrix applies. Figure 14.9 shows the multivariable LQ regulator in block diagram form. The return difference of interest is computed at the input to the plant as $\mathbf{R}_d(s) = \mathbf{I} + \mathbf{G}_\infty(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}$. The bound $\alpha = 1$ has been shown to apply to this return difference matrix. The preceding theorem then guarantees the previously stated stability margins for arbitrary gain-phase perturbations introduced at the plant input.

When perfect full state feedback is not available for use in the LQ control loop, the separation theorem allows estimates of the states to be used in forming the feedback control signals. Figure 14.10*a* through *d* shows block diagram representations of the controller-observer system. These are obtained, one from another, by standard block diagram manipulations. Figure 14.10 is the multivariable equivalent of Figure 13.15, except here the control gain $\mathbf{G}_\infty$ is explicitly identified.

If perfect models were available, the resulting system would perform as designed, except perhaps for a brief transient due to initial condition mismatches between the true and estimated states. In fact, the transfer functions from $\mathbf{v}$ to $\mathbf{y}$ are the same for Figures 14.9 and 14.10 for any arbitrary control gain $\mathbf{G}$. For Figure 14.9, $\mathbf{x}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{u}(s) = (s\mathbf{I} - \mathbf{A})^{-1}\{\mathbf{v} - \mathbf{G}\mathbf{x}\}$, so that $\mathbf{x} = [\mathbf{I} + (s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{G}]^{-1}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{v} = [s\mathbf{I} - \mathbf{A} + \mathbf{B}\mathbf{G}]^{-1}\mathbf{B}\mathbf{v}$. Since $\mathbf{y} = \mathbf{C}\mathbf{x}$, the full state transfer function is $\mathbf{H}_f = \mathbf{C}[s\mathbf{I} - \mathbf{A} + \mathbf{B}\mathbf{G}]^{-1}\mathbf{B}$. From Figure 14.10, $\mathbf{y} = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\mathbf{u} = \mathbf{W}_1(s)\mathbf{u}$. But $\mathbf{u} = \mathbf{v} - \mathbf{G}(s\mathbf{I} - \mathbf{A}_c)^{-1}\mathbf{B}\mathbf{u} - \mathbf{G}(s\mathbf{I} - \mathbf{A}_c)^{-1}\mathbf{L}\mathbf{y} = \mathbf{v} - \mathbf{G}\mathbf{W}_3\mathbf{B}\mathbf{u} - \mathbf{G}\mathbf{W}_2\mathbf{y}$. Combining and solving for $\mathbf{y}$ leads to the input-output transfer function

$$\mathbf{H}_o = \{\mathbf{I} + \mathbf{W}_1[\mathbf{I} + \mathbf{G}\mathbf{W}_3]^{-1}\mathbf{G}\mathbf{W}_2\}^{-1}\mathbf{W}_1[\mathbf{I} + \mathbf{G}\mathbf{W}_3]^{-1}$$

The claim is that $\mathbf{H}_f = \mathbf{H}_o$ when exact models are used. The proof uses a result from Problem 4.4, which is restated here and is referred to as a gain rearrangement identity. This identity is true for any conformable matrices $\mathbf{F}$ and $\mathbf{G}$ for which the indicated inverses exist:

$$[\mathbf{I} + \mathbf{F}\mathbf{G}]^{-1}\mathbf{F} = \mathbf{F}[\mathbf{I} + \mathbf{G}\mathbf{F}]^{-1}$$

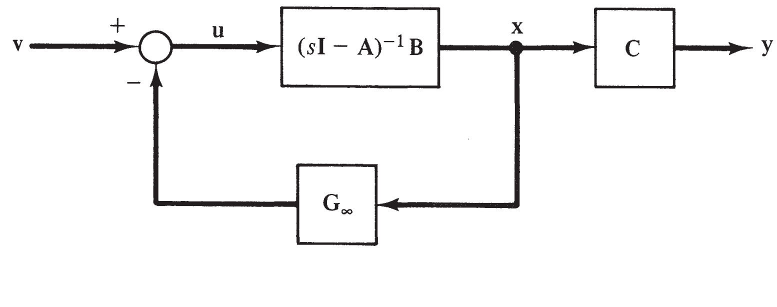


Figure 14.9

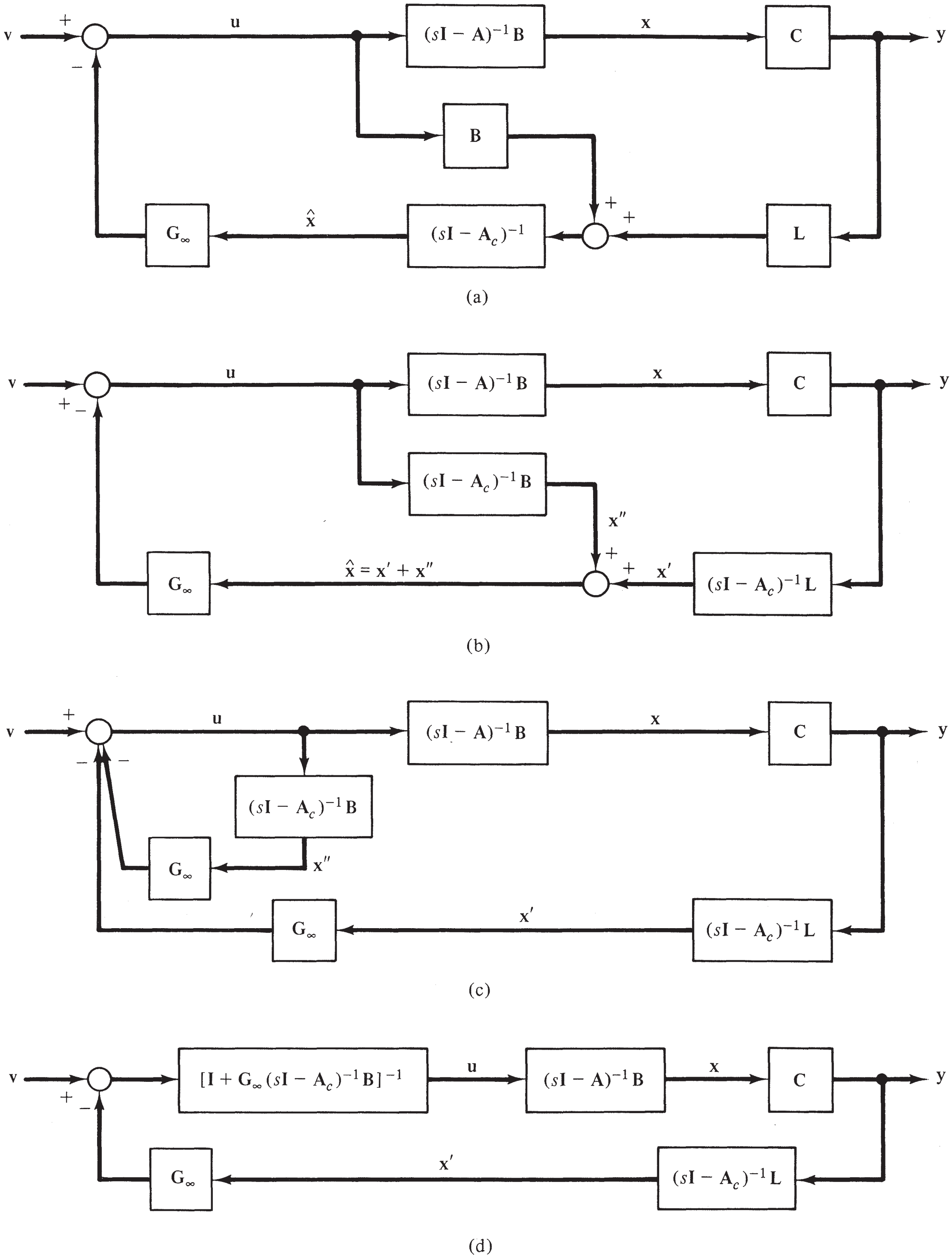

v
u
$(sI - A)^{-1}B$
x
C
y
B
$\hat{x}$
G_∞
$(sI - A_c)^{-1}$
L
(a)
$(sI - A_c)^{-1}B$
x″
$\hat{x} = x' + x''$
x′
$(sI - A_c)^{-1}L$
(b)
(c)
$[I + G_\infty(sI - A_c)^{-1}B]^{-1}$
(d)

Figure 14.10

Then a series of standard manipulations gives

$$\begin{aligned}\mathbf{H}_o &= \mathbf{W}_1\{\mathbf{I} + [\mathbf{I} + \mathbf{G}\mathbf{W}_3]^{-1}\mathbf{G}\mathbf{W}_2\mathbf{W}_1\}^{-1}[\mathbf{I} + \mathbf{G}\mathbf{W}_3]^{-1} \\ &= \mathbf{W}_1\{[\mathbf{I} + \mathbf{G}\mathbf{W}_3][\mathbf{I} + [\mathbf{I} + \mathbf{G}\mathbf{W}_3]^{-1}\mathbf{G}\mathbf{W}_2\mathbf{W}_1]\}^{-1} \\ &= \mathbf{W}_1\{\mathbf{I} + \mathbf{G}\mathbf{W}_3 + \mathbf{G}\mathbf{W}_2\mathbf{W}_1\}^{-1}.\end{aligned}$$

In terms of the original system matrices this becomes $\mathbf{H}_o = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\{\mathbf{I} + \mathbf{G}(s\mathbf{I} - \mathbf{A}_c)^{-1}[\mathbf{I} + \mathbf{L}\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}]\mathbf{B}\}^{-1}$. Replacing $\mathbf{I}$ by $(s\mathbf{I} - \mathbf{A})(s\mathbf{I} - \mathbf{A})^{-1}$ in the term in brackets allows an $(s\mathbf{I} - \mathbf{A})^{-1}$ to be factored out. Then using $\mathbf{A}_c = \mathbf{A} - \mathbf{L}\mathbf{C}$ gives $\mathbf{H}_o = \mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\{\mathbf{I} + \mathbf{G}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\}^{-1}$. Using the gain rearrangement identity to move $\mathbf{B}$ allows the two inverse terms to be combined, giving $\mathbf{H}_o = \mathbf{C}(s\mathbf{I} - \mathbf{A} + \mathbf{B}\mathbf{G})^{-1}\mathbf{B} \equiv \mathbf{H}_f$. It is clear from Figure 14.10 that the combined system is of order $2n$ (or higher with some rearrangements). Since the input-output transfer function equals that for an nth-order system, some modes must be either uncontrollable, unobservable, or both. In this case all the modes introduced by the observer are uncontrollable.

EXAMPLE 14.13 Consider the unstable system described by $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 2 & -3 \end{bmatrix}$, $\mathbf{B} = \mathbf{I}_2$, $\mathbf{C} = [1 \quad 0]$, and $\mathbf{D} = [0 \quad 0]$. Design a constant feedback *LQ* controller using $\mathbf{Q} = \mathbf{R} = \mathbf{I}_2$. Then design an observer by using the duality between controllers and observers. That is, replace $\mathbf{A}$ by $\mathbf{A}^T$ and $\mathbf{B}$ by $\mathbf{C}^T$, let $\mathbf{Q} = \mathbf{I}_2$, and $\mathbf{R} = r$, and then solve the ARE for $\mathbf{W}'_\infty$. The observer gain $\mathbf{L}$ is the transpose of the dual control gain matrix. That is, $\mathbf{L} = \mathbf{W}'_\infty\mathbf{C}^T\mathbf{R}^{-T}$. The separation theorem assures us that the final $2n$th-order system will have n poles determined by the $\mathbf{Q}$, $\mathbf{R}$ weights—i.e., the desired controller poles—and n poles determined by the observer. Select the scalar r so that the observer poles are faster—i.e., further to the left in the complex plane—than the closed-loop controller poles. Investigate the composite controller-observer system for controllability, observability, and eigenvalue locations. Then compare the step response of the full state feedback system with the controller-observer system.

Solving Eq. (*14.49*) gives $\mathbf{W}_\infty = \begin{bmatrix} 1.6012 & 0.4392 \\ 0.4392 & 0.26887 \end{bmatrix}$, and since $\mathbf{Q} = \mathbf{B} = \mathbf{I}$, $\mathbf{G}_\infty = \mathbf{W}_\infty$. Using this gain with full state feedback gives the closed-loop eigenvalues $\lambda = -1.1818$ and -3.6882. Solving the dual problem with three values of r gives:

$$r = 10 \Rightarrow \mathbf{L} = [1.1919 \quad .66033]^T \quad \text{and} \quad \lambda_0 = -0.62729, -3.5646$$

$$r = 1 \Rightarrow \mathbf{L} = [1.6350 \quad .8366]^T \quad \text{and} \quad \lambda_0 = -1.0411, -3.5939$$

$$r = 0.1 \Rightarrow \mathbf{L} = [3.5876 \quad 1.4353]^T \quad \text{and} \quad \lambda_0 = -2.487, -4.10$$

Note that the observer gains get higher and its response gets faster as the value of r is decreased. The same effect occurs as $\mathbf{Q}$ is made larger, since only the ratio $\mathbf{Q}/r$ is significant in minimizing the quadratic performance function. The value $r = 0.1$ is accepted, since it gives observer eigenvalues somewhat faster than the controller eigenvalues. The composite system is described by

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\hat{\mathbf{x}}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{B}\mathbf{G}_\infty \\ \mathbf{L}\mathbf{C} & \mathbf{A} - \mathbf{B}\mathbf{G}_\infty - \mathbf{L}\mathbf{C} \end{bmatrix}\begin{bmatrix} \mathbf{x} \\ \hat{\mathbf{x}} \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{B} \end{bmatrix}\mathbf{v}$$

where $\mathbf{v}$ is an arbitrary external input and $\mathbf{y} = [1 \quad 0 \quad 0 \quad 0]\begin{bmatrix} \mathbf{x} \\ \hat{\mathbf{x}} \end{bmatrix}$. The 4×4 composite system matrix is

$$\begin{bmatrix} 0 & 1 & -1.6012 & -0.4392 \\ 2 & -3 & -0.4392 & -0.26887 \\ 3.5876 & 0 & -5.1888 & 0.5608 \\ 1.4353 & 0 & 0.1255 & -3.26887 \end{bmatrix}$$

Its eigenvalues are verified to consist of the desired controller and observer eigenvalues, $\lambda = -1.1818, -3.6882, -2.487$, and -4.100. The composite system is observable but not controllable. In fact, the Kalman controllable canonical form for this fourth-order system has

$$\mathbf{A}' = \left[\begin{array}{cc|cc} -1.6012 & 0.5608 & -0.557862 & 5.17739 \\ 1.5608 & -3.26887 & -0.311709 & 1.86785 \\ \hline 0 & 0 & -2.9645 & -0.550433 \\ 0 & 0 & -0.985733 & -3.62310 \end{array}\right],$$

$$\mathbf{B}' = \left[\begin{array}{cc} 1.41421 & 0 \\ 0 & 1.41421 \\ \hline 0 & 0 \\ 0 & 0 \end{array}\right]$$

and $\mathbf{C}' = [0.707107 \quad 0 \quad -0.016186 \quad 0.7070]$. The 2×2 lower right corner of $\mathbf{A}'$ describes the uncontrollable modes, and the eigenvalues of this partition are $\lambda = -2.4869$ and -4.100. These are the observer modes. To test the transient performance, it is necessary to specify the inputs $v_1(t)$ and $v_2(t)$ as well as the four initial conditions. Only one case is presented, with v_1 equal to a unit step function, $v_2 = 0$, and $x_1(0) = x_2(0) = \hat{x}_2(0) = 0, \hat{x}_1(0) = 1$. Figure 14.11*a* shows the response of x_1, $\hat{x}_1$, and x_1^*, where x_1^* is the response obtained with the same inputs and initial conditions $\mathbf{x}(0)$, but with full state feedback. Figure 14.11*b* gives the same comparisons for the second state variable and its estimate. The estimate, actual, and optimal curves all agree in the limit. There is a transient difference between $\mathbf{x}(t)$ and $\hat{\mathbf{x}}(t)$ due to initial condition mismatch. This

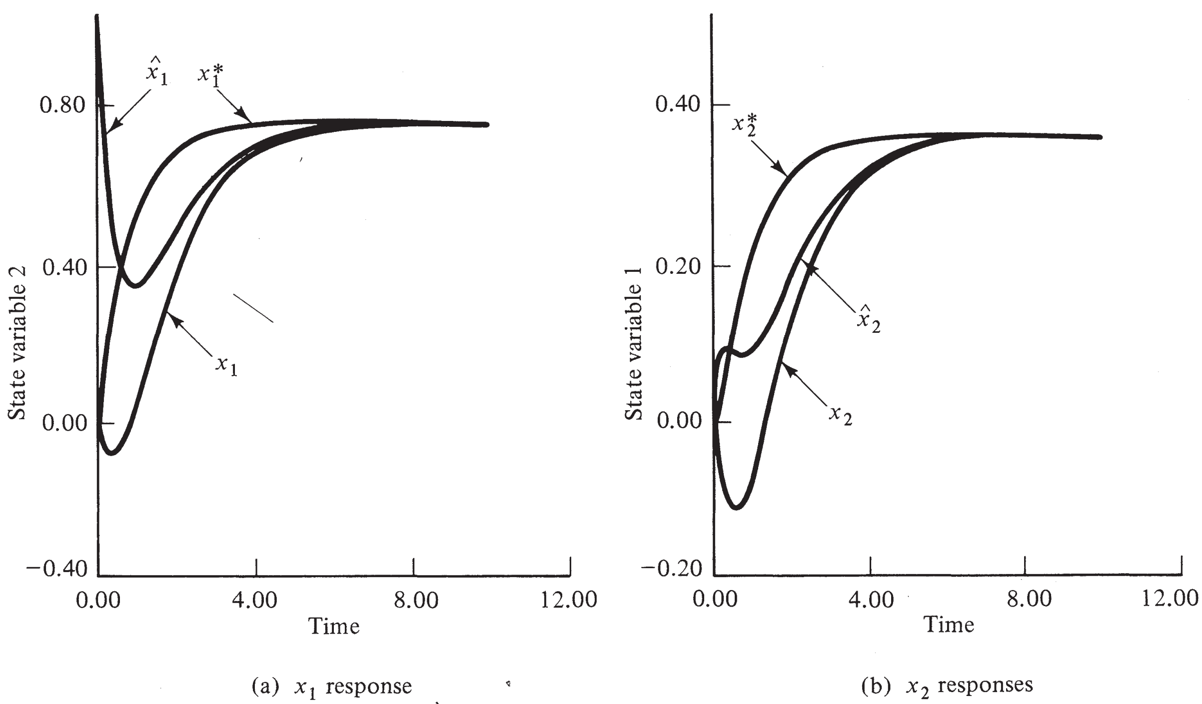


Figure 14.11 Comparison of full state step response with observer true and estimated.

difference decays according to the observer eigenvalue response. The composite system responses deviate from the optimal full state feedback case because of this transient. Even with zero input and all initial conditions zero except for $\hat{x}_1$, for example, the output y, which is just x_1, still shows a nonzero transient. This indicates that the observer modes' initial condition responses are observable in the output, but as stated earlier, they are not controllable. ■

Since all models have some degree of inaccuracy or oversimplification, the question of robustness needs to be considered here also. It has been shown [13, 14] that the introduction of a poorly designed observer can cause the comfortable stability margins of the LQ problem to be lost. The return difference matrix analysis is somewhat more complicated for the observer system shown in Figure 14.10. Consider a perturbation $\delta\mathbf{u}$ entering at the input to the plant. From Figure 14.10, it will affect $\mathbf{x}$, $\mathbf{x}'$, and $\hat{\mathbf{x}}$ but not $\mathbf{x}''$ if the loop is opened at $\mathbf{G}_\infty$, for example. A comparison of Figures 14.9 and 14.10 shows that any such $\delta\mathbf{u}$ will affect both $\mathbf{x}$ and $\mathbf{x}'$ (and hence $\hat{\mathbf{x}}$) in the same way, provided that

$$(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B} = (s\mathbf{I}-\mathbf{A}_c)^{-1}\mathbf{LC}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B} \tag{14.55}$$

This condition, after modest algebraic manipulations, leads to Eq. (*14.56*), which is called the *Doyle-Stein robust observer condition.* It has been shown [13] that if the extra freedom which exists in specifying observer poles and finding the observer gain $\mathbf{L}$ is used to satisfy, or approach, the Doyle-Stein condition, then the stability robustness associated with the full state LQ system of Figure 14.9 can be asymptotically recovered in the LQG system of Figure 14.10. The original derivation of the Doyle-Stein condition assumed that there were r inputs and $m = r$ outputs, so that $\mathbf{C}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}$ is square. Then postmultiplying by the inverse of this matrix and premultiplying by $(s\mathbf{I}-\mathbf{A})$ alters Eq. (*14.55*) to

$$\mathbf{B}[\mathbf{C}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}]^{-1} = (s\mathbf{I}-\mathbf{A})(s\mathbf{I}-\mathbf{A}+\mathbf{LC})^{-1}\mathbf{L}$$

where the definition $\mathbf{A}_c = \mathbf{A}-\mathbf{LC}$ has been used. The left side is in the desired final form. The right side can be rearranged to give

$$(s\mathbf{I}-\mathbf{A})[\{\mathbf{I}+\mathbf{LC}(s\mathbf{I}-\mathbf{A})^{-1}\}(s\mathbf{I}-\mathbf{A})]^{-1}\mathbf{L} = [(\mathbf{I}+\mathbf{LC}(s\mathbf{I}-\mathbf{A})^{-1}]^{-1}\mathbf{L}$$

Again, using the gain rearrangement identity allows the two $\mathbf{L}$ terms to be rearranged, giving

$$\mathbf{B}[\mathbf{C}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}]^{-1} = \mathbf{L}[(\mathbf{I}+\mathbf{C}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{L}]^{-1} \tag{14.56}$$

Other forms of Eq. (*14.55*) which do not require the assumption that $r = m$ can be written, such as

$$(s\mathbf{I}-\mathbf{A}_c)(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B} = \mathbf{LC}(s\mathbf{I}-\mathbf{A})^{-1}\mathbf{B}$$

which leads to

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

A full exploitation of the robust observer results is left to the references. One of the major results that comes from this analysis is that the placement of all observer poles far to the left in order to speed convergence of $\hat{\mathbf{x}}$ to $\mathbf{x}$ is not always the best strategy when the possibility of model errors exists. Rather, some of the observer poles should approach the stable plant zeros. The rest of the observer poles do migrate outward to the left along the classical Butterworth configuration.

14.8 EXTENSIONS

Lower-Order Controllers via Projective Controls [15, 16, 17]

From results of the past two chapters, it is known that the closed-loop eigenvalues of a completely controllable and observable system can be relocated as desired by using full state feedback. In most practical problems, all states are not available for use. The only signals that realistically can be assumed available for feedback are the outputs. Static output feedback can control the location only of a subset of m eigenvalues. The remaining $n - m$ eigenvalues may or may not take on acceptable values. An observer of order n, which is a dynamic feedback compensator, can be used with output feedback and total discretion about eigenvalue locations is again possible. A reduced order observer, which is a dynamic compensator of order $n - m$, where $\text{rank}(\mathbf{C}) = m$, can also be used. The order of these feedback compensators may be higher than desired in many cases. One extension to the LQ feedback theory provides for suboptimal lower-order controllers, which preserve some of the optimal eigenstructure associated with the full state LQ feedback solution. For simplicity the discussion is restricted to constant, strictly proper systems ($\mathbf{D} = [\mathbf{0}]$) which are both controllable and observable. Only the infinite control horizon is considered, so that the algebraic Riccati equation governs the optimal solution. The method of projective controls makes it possible to use output feedback through a dynamic feedback compensator of order p, with $0 \le p \le n - m$. The projective controls procedure allows for matching $p + m$ of the eigenvalues and eigenvectors of the suboptimal system with those of the optimal LQ system. The case of static output feedback is considered first. The optimal LQ closed-loop system dynamics, with optimal feedback gain $\mathbf{G}_0$, are governed by the matrix $\mathbf{A} - \mathbf{B}\mathbf{G}_0 = \mathbf{A} - \mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{W}$. Let the eigenvalues of this matrix be $\{\lambda_i, i = 1, n\}$ and the corresponding eigenvectors be $\boldsymbol{\xi}_i$. Suppose that m of these eigenvalue-eigenvector pairs, perhaps the dominant modes, must be retained in the output feedback solution. Form the $n \times m$ matrix $\mathbf{X}_m$ with columns made up of the selected eigenvectors. The $n \times n$ projection matrix $\mathbf{P} = \mathbf{X}_m(\mathbf{C}\mathbf{X}_m)^{-1}\mathbf{C}$ (note $\mathbf{P}^2 = \mathbf{P}$; see Sec. 5.13) can be shown to project any arbitrary state $\mathbf{x}$ onto the subspace spanned by the m selected eigenvectors. If the feedback control law is

$$\mathbf{u} = -\mathbf{G}_0\mathbf{P}\mathbf{x} = -\mathbf{G}_0[\mathbf{X}_m(\mathbf{C}\mathbf{X}_m)^{-1}\mathbf{C}]\mathbf{x} = -\mathbf{G}_0[\mathbf{X}_m(\mathbf{C}\mathbf{X}_m)^{-1}]\mathbf{y}$$

then the suboptimal system will have the desired m eigenvalues and eigenvectors. The balance of the eigenvalues-eigenvectors is determined by the matrix $\mathbf{A}_{22} - \mathbf{N}\mathbf{A}_{12}$, where $\mathbf{A}_{22}$ is the lower right $(n - m) \times (n - m)$ partition of the original system matrix $\mathbf{A}$, $\mathbf{A}_{12}$ is the $m \times (n - m)$ upper right partition, and $\mathbf{N} = \mathbf{X}_b\mathbf{X}_a^{-1}$. The eigenvector matrix $\mathbf{X}_m$ is partitioned into the upper m rows, $\mathbf{X}_a$, and the lower $n - m$ rows, $\mathbf{X}_b$. The obvious restriction on choice of $\mathbf{X}_m$ applies. If $\boldsymbol{\xi}_i$ is complex and is selected for inclusion, then the conjugate column must also be selected. There are many instances where the static output controller will not be satisfactory. If only one output is available and both eigenvalues of an optimal second-order system are complex, the preceding procedure cannot be used. In other cases the closed loop system will be unstable or otherwise unsatisfactory due to the uncontrolled eigenvalues. In these cases a pth-order dynamic controller can be considered with

$$\dot{\mathbf{z}} = \mathbf{H}\mathbf{z} + \mathbf{L}\mathbf{y} \quad \text{and} \quad \mathbf{u} = -\{\mathbf{G}_z\mathbf{z} + \mathbf{G}_y\mathbf{y}\} \tag{14.57}$$

The controller parameters $\mathbf{H}$ and $\mathbf{L}$ as well as the two gain matrices must be determined. A composite $(n + p)$th-order system can be written as

$$\begin{bmatrix} \dot{\mathbf{z}} \\ \dot{\mathbf{x}} \end{bmatrix} = \begin{bmatrix} \mathbf{H} & \mathbf{LC} \\ \mathbf{0} & \mathbf{A} \end{bmatrix}\begin{bmatrix} \mathbf{z} \\ \mathbf{x} \end{bmatrix} + \begin{bmatrix} \mathbf{0} \\ \mathbf{B} \end{bmatrix}\mathbf{u}$$

shortened to $\dot{\mathbf{x}}_a = \mathbf{A}_a\mathbf{x}_a + \mathbf{B}_a\mathbf{u}$, and

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{y} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix}\begin{bmatrix} \mathbf{z} \\ \mathbf{x} \end{bmatrix}$$

notationally shortened to $\mathbf{y}_a = \mathbf{C}_a\mathbf{x}_a$. An augmented $\mathbf{Q}_a = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{Q} \end{bmatrix}$ is also defined so that the augmented problem has the same cost function J as the original problem.

The standard ARE could be applied to the augmented problem if $\mathbf{H}$ and $\mathbf{L}$ were known. However, it can be shown that if $\mathbf{H}$ is asymptotically stable, the only possible solution to the ARE for the enlarged problem is $\mathbf{W}_a = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{W} \end{bmatrix}$, where $\mathbf{W}$ is the solution to the parent ARE equation. This means that the determination of $\mathbf{H}$ and $\mathbf{L}$ can be postponed until later.The previous projection results for state output feedback can be formally applied to the enlarged problem by defining the matrix of $m + p$ desired eigenvectors,

$$\mathbf{X}_{m+p} = \begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \\ \mathbf{X}_{31} & \mathbf{X}_{32} \end{bmatrix}$$

The eigenvectors desired for the original nth-order system are contained in the partitions $\begin{bmatrix} \mathbf{X}_{21} & \mathbf{X}_{22} \\ \mathbf{X}_{31} & \mathbf{X}_{32} \end{bmatrix} = \{\boldsymbol{\xi}_i,\ i = 1,\ m + p\}$ and are known. The pth-order controller requires augmentation with the still unknown blocks $\mathbf{X}_{11}$ $(p \times p)$ and $\mathbf{X}_{12}$ $(p \times m)$. It is convenient to assume that the states and measurements are arranged so that $\mathbf{C} = [\mathbf{I}_m \quad \mathbf{0}]$. Then the projection matrix for the augmented problem is

$$\mathbf{P}_a = \mathbf{X}_{m+p}[\mathbf{C}_a\mathbf{X}_{m+p}]^{-1}\mathbf{C}_a = \mathbf{X}_{m+p}\begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{I}_p & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_m & \mathbf{0} \end{bmatrix}$$

Therefore, $\mathbf{P}_a\mathbf{x}_a = \begin{bmatrix} \mathbf{I}_{m+p} \\ \mathbf{N} \end{bmatrix}\begin{bmatrix} \mathbf{z} \\ \mathbf{x} \end{bmatrix}$, where

$$\mathbf{N} = [\mathbf{X}_{31} \quad \mathbf{X}_{32}]\begin{bmatrix} \mathbf{X}_{11} & \mathbf{X}_{12} \\ \mathbf{X}_{21} & \mathbf{X}_{22} \end{bmatrix}^{-1} = [\mathbf{N}_p \quad \mathbf{N}_m] \tag{14.58}$$

The projective feedback control is $\mathbf{G}_a\mathbf{x}_a = -\mathbf{R}^{-1}\mathbf{B}_a^T\mathbf{W}_a\mathbf{P}_a\mathbf{x}_a$, which reduces to $\mathbf{G}_a\mathbf{x}_a = -\{\mathbf{G}_z\mathbf{z} + \mathbf{G}_y\mathbf{y}\}$, where

$$\mathbf{G}_z = \mathbf{G}_0\begin{bmatrix}\mathbf{0}\\ \mathbf{N}_p\end{bmatrix} = \mathbf{G}_u\mathbf{N}_p \quad \text{and} \quad \mathbf{G}_y = \mathbf{G}_0\begin{bmatrix}\mathbf{I}_m\\ \mathbf{N}_m\end{bmatrix} = \mathbf{G}_m + \mathbf{G}_u\mathbf{N}_m$$

The original full state optimal gain $\mathbf{G}_0 = [\mathbf{G}_m \quad \mathbf{G}_u]$ has been partitioned into $\mathbf{G}_m$, which multiplies measured states, and $\mathbf{G}_u$, which multiplies unmeasured states. The projective control form is now known, but the matrix $\mathbf{N}$ of Eq. (*14.58*) used in forming the gains depends upon the still-unknown eigenvector blocks $\mathbf{X}_{11}$ and $\mathbf{X}_{12}$. These are directly related to the dynamics of the controller, $\mathbf{H}$ and $\mathbf{L}$ of Eq. (*14.57*), which have not yet been specified. As long as the controller matrix $\mathbf{H}$ is asymptotically stable, any $\mathbf{X}_{11}$ and $\mathbf{X}_{12}$ could be used, and the primary objective of retaining $m + p$ eigenvalues and eigenvectors of the optimal system will be met. The remaining $n - m$ eigenvalues, called the complementary, or residual, spectrum, are affected by the choice of the remaining unknowns, since these eigenvalues are determined by the matrix

$$\mathbf{A}_r = \mathbf{A}_{22} - \mathbf{N}\mathbf{A}_{12} = \mathbf{A}_{22} - [\mathbf{X}_{31} \quad \mathbf{X}_{32}]\begin{bmatrix}\mathbf{X}_{11} & \mathbf{X}_{12}\\ \mathbf{X}_{21} & \mathbf{X}_{22}\end{bmatrix}^{-1}\mathbf{A}_{12}$$

This is similar to the problem encountered in pole placement using output feedback in Sec. 13.5. Thus the selection of $\mathbf{X}_{11}$ and $\mathbf{X}_{22}$ can be used to give acceptable eigenvalues to $\mathbf{A}_r$. As a minimum, all the residual eigenvalues must be stable. With $\mathbf{X}_{11}$ and $\mathbf{X}_{12}$ determined, the gains $\mathbf{G}_z$ and $\mathbf{G}_y$ can be evaluated. The last task is to determine the control matrices $\mathbf{H}$ and $\mathbf{L}$. The augmented closed-loop eigenvector equation is $[\mathbf{A}_a - \mathbf{B}_a\mathbf{G}_a]\mathbf{X}_{m+p} = \mathbf{X}_{m+p}\mathbf{\Lambda}_{m+p}$. It has $m + p$ known, specified eigenvectors in $\mathbf{X}_{m+p}$ and eigenvalues in $\mathbf{\Lambda}_{m+p}$. By assumption we are dealing with the simple eigenvalues cases I or II_1 of Sec. 7.4, so that $\mathbf{\Lambda}_{m+p}$ will be diagonal. However, in the case of complex eigenvalues and eigenvectors, it may be more convenient to work with purely real arithmetic. Postmultiplication by a transformation $\mathbf{T}$ (see Problem 4.23) gives

$$[\mathbf{A}_a - \mathbf{B}_a\mathbf{G}_a]\mathbf{X}_{m+p}\mathbf{T} = \mathbf{X}_{m+p}\mathbf{T}\mathbf{T}^{-1}\mathbf{\Lambda}_{m+p}\mathbf{T}, \quad \text{or} \quad [\mathbf{A}_a - \mathbf{B}_a\mathbf{G}_a]\mathbf{X}'_{m+p} = \mathbf{X}'_{m+p}\mathbf{\Lambda}'_{m+p}$$

In the primed form only real numbers are required, but $\mathbf{\Lambda}_{m+p}$ will no longer be diagonal. The prime is dropped below, and the expanded partitioned form of these equations can be solved to give

$$[\mathbf{H} \quad \mathbf{L}] = [\mathbf{X}_{11} \quad \mathbf{X}_{12}]\mathbf{\Lambda}_{m+p}\begin{bmatrix}\mathbf{X}_{11} & \mathbf{X}_{12}\\ \mathbf{X}_{21} & \mathbf{X}_{22}\end{bmatrix}^{-1} \tag{14.59}$$

EXAMPLE 14.14 The third-order system with $\mathbf{A} = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ -2 & -1 & -1\end{bmatrix}$, $\mathbf{B} = \begin{bmatrix}0\\0\\1\end{bmatrix}$, and $\mathbf{C}^T = \begin{bmatrix}1\\0\\0\end{bmatrix}$ and with $\mathbf{Q} = \mathbf{I}_3$, $R = 5$ has the optimal full state feedback gain $\mathbf{G}_0 = [0.04939 \quad 1.2591 \quad 0.92826]$ and the closed-loop eigenvalues are $\lambda_i = \{-1.37085, -0.2787 \pm 1.1905j\}$, and the corresponding eigenvectors are

$$\begin{bmatrix}-0.532106 & 0.599405 \pm 0.296929j\\ 0.729456 & 0.186430 \mp 0.79634j\\ -1 & -1\end{bmatrix}$$

If static output feedback is used, one eigenvalue-eigenvector from the optimal set can be retained, and it must be the real eigenvalue because complex pairs cannot be split. It is easy to calculate that $\mathbf{G}_y = \mathbf{G}_0\,\boldsymbol{\xi}_1[\mathbf{C}\boldsymbol{\xi}_1]^{-1} = 0.067813$ will in fact retain the first eigenvalue-eigenvector.

However, the other two eigenvalues will be in the right-half plane, so this solution is not useful. A dynamic controller of order $p = 1$ can be designed which will retain the complex conjugate eigenpair, as follows. The values $\mathbf{X}_{11} = -1$ and $\mathbf{X}_{12} = 1$ were found to give stable complementary eigenvalues. Using these, $\mathbf{N} = \begin{bmatrix} 1.76109 & 3.24908 \\ -0.981661 & -3.30605 \end{bmatrix}$, $\mathbf{G}_z = 1.30615$, $\mathbf{G}_y = 1.07144$, $\mathbf{H} = -3.806533$, and $\mathbf{L} = -7.871706$. It can be verified that these values do retain the complex eigenvalue-eigenvector pair in the suboptimal solution. The two remaining eigenvalues are at -0.23494 and -4.014. The first of these may be too close to the $j\omega$ axis. A systematic search for other values of $\mathbf{X}_{11}$ and $\mathbf{X}_{12}$ which give a more satisfactory residual spectrum can be carried out [16]. ■

Robustness Enhancement via Frequency-Weighted Cost Function [18, 19]

There are many situations where the system model used in control design is known to be inaccurate at high frequencies. The steady-state D.C. error may be particularly important in other cases. The suppression of certain vibration modes may be crucial in other circumstances. All these criteria suggest that errors at certain frequencies may be more costly than others. It is possible to extend the infinite time *LQ* optimization theory to frequency-sensitive cost functions by applying Parseval's theorem:

$$\int_0^\infty |f(t)|^2\,dt = \frac{1}{2\pi}\int_{-\infty}^\infty |F(j\omega)|^2\,d\omega \tag{14.60}$$

where $F(j\omega)$ is the Fourier transform of $f(t)$. The cost functional J can be written in a similar way by defining $\mathbf{f}^T(t) = [(\mathbf{Q}^{1/2}\mathbf{x})^T(\mathbf{R}^{1/2}\mathbf{u})^T]$, because then

$$\begin{aligned}\int_0^\infty |\mathbf{f}(t)|^2\,dt &= \int_0^\infty \{\mathbf{x}^T\mathbf{Q}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u}\}\,dt \\ &= \frac{1}{2\pi}\int_{-\infty}^\infty \{\mathbf{x}^*(j\omega)\mathbf{Q}(j\omega)\mathbf{x}(j\omega) + \mathbf{u}^*(j\omega)\mathbf{R}\mathbf{u}(j\omega)\}\,d\omega\end{aligned} \tag{14.61}$$

where ( )* indicates the complex conjugate of the transpose of the quantity inside. Up until now the weighting matrices $\mathbf{Q}$ and $\mathbf{R}$ have been constants. In order to obtain frequency-dependent weighting while still being able to utilize the standard *LQ* results, we select $\mathbf{Q}(j\omega) = \mathbf{T}^*(j\omega)\mathbf{T}(j\omega)$ and $\mathbf{R}(j\omega) = \mathbf{U}^*(j\omega)\mathbf{U}(j\omega)$ and define two new vector variables whose transforms are related to $\mathbf{x}$ and $\mathbf{u}$ according to

$$\mathbf{z}(j\omega) = \mathbf{T}(j\omega)\mathbf{x}(j\omega) \quad \text{and} \quad \mathbf{w}(j\omega) = \mathbf{U}(j\omega)\mathbf{u}(j\omega) \tag{14.62}$$

The implication is that $\mathbf{z}$ is the output of a filter or dynamic system with transfer function $\mathbf{T}(j\omega)$ and input $\mathbf{x}(j\omega)$. If $\mathbf{T}$ is a proper transfer function, then there is a corresponding time domain relation involving new states $\mathbf{z}_s$,

$$\dot{\mathbf{z}}_s = \mathbf{F}_z\mathbf{z}_s + \mathbf{G}_z\mathbf{x} \quad \text{and} \quad \mathbf{z} = \mathbf{H}_z\mathbf{z}_s + \mathbf{D}_z\mathbf{x} \tag{14.63}$$

A similar relationship can be written between $\mathbf{u}$ and $\mathbf{w}$ by introducing additional states $\mathbf{w}_s$ for the transfer function $\mathbf{U}$ using the methods of Chapters 3 and 12:

$$\dot{\mathbf{w}}_s = \mathbf{F}_u\mathbf{w}_s + \mathbf{G}_u\mathbf{u}, \qquad \mathbf{w} = \mathbf{H}_u\mathbf{w}_s + \mathbf{D}_u\mathbf{u} \tag{14.64}$$

Equations (*14.63*) and (*14.64*) can be combined with the original system state equations to give

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{z}}_s \\ \dot{\mathbf{w}}_s \end{bmatrix} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \mathbf{0} \\ \mathbf{G}_z & \mathbf{F}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F}_u \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{z}_s \\ \mathbf{w}_s \end{bmatrix} + \begin{bmatrix} \mathbf{B} \\ \mathbf{0} \\ \mathbf{G}_u \end{bmatrix} \mathbf{u} \tag{14.65}$$

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{z} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} \mathbf{C} & \mathbf{0} & \mathbf{0} \\ \mathbf{D}_z & \mathbf{H}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{H}_u \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{z}_s \\ \mathbf{w}_s \end{bmatrix} + \begin{bmatrix} \mathbf{D} \\ \mathbf{0} \\ \mathbf{D}_u \end{bmatrix} \mathbf{u} \tag{14.66}$$

The cost function is now $J = \int \{\mathbf{z}^T\mathbf{z} + \mathbf{w}^T\mathbf{w}\}\,dt$. The $\mathbf{z}^T\mathbf{z}$ term can be written as a quadratic in the new states $\mathbf{x}$ and $\mathbf{z}_s$ by using Eq. (*14.66*). If the same approach is used with the $\mathbf{w}^T\mathbf{w}$ term, cross products of state and control terms arise. It is not difficult to return to Sec. 14.4 and derive a modified Riccati equation to provide the solution in the presence of these cross terms. An alternate approach is first to solve for the control variable $\mathbf{w}$ that optimizes

$$J = \int \{\mathbf{x}_a^T \mathbf{Q}_a \mathbf{x}_a + \mathbf{w}^T \mathbf{R}_a \mathbf{w}\}\,dt \tag{14.67}$$

with the definitions $\mathbf{x}_a^T = [\mathbf{x}^T \quad \mathbf{z}_s^T \quad \mathbf{w}_s^T]$

$$\mathbf{Q}_a = \begin{bmatrix} \mathbf{D}_z^T\mathbf{D}_z & \mathbf{D}_z^T\mathbf{H}_z & \mathbf{0} \\ \mathbf{H}_z^T\mathbf{D}_z & \mathbf{H}_z^T\mathbf{H}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}, \qquad \mathbf{R}_a = \mathbf{I} \tag{14.68}$$

Equation (*14.65*) must be rewritten in terms of $\mathbf{w}$ rather than $\mathbf{u}$. This is done by using the last partition of Eq. (*14.66*), assuming that $\mathbf{D}_u$ is invertible. This gives

$$\mathbf{u} = \mathbf{D}_u^{-1}(\mathbf{w} - \mathbf{H}_u\mathbf{w}_s) \tag{14.69}$$

$$\dot{\mathbf{x}}_a = \begin{bmatrix} \mathbf{A} & \mathbf{0} & -\mathbf{B}\mathbf{D}_u^{-1}\mathbf{H}_u \\ \mathbf{G}_z & \mathbf{F}_z & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{F}_u - \mathbf{G}_u\mathbf{D}_u^{-1}\mathbf{H}_u \end{bmatrix} \mathbf{x}_a + \begin{bmatrix} \mathbf{B}\mathbf{D}_u^{-1} \\ \mathbf{0} \\ \mathbf{G}_u\mathbf{D}_u^{-1} \end{bmatrix} \mathbf{w} \tag{14.70}$$

Equations (*14.67*) and (*14.70*) are in the standard forms used in the earlier *LQ* problems, and the solution for the optimal $\mathbf{w}$ is obtained by solving an ARE for $\mathbf{W}_a$. Then $\mathbf{w} = -\mathbf{R}_a^{-1}\mathbf{B}_a^T\mathbf{W}_a\mathbf{x}_a$, and using Eq. (*14.69*) gives $\mathbf{u} = -\mathbf{K}_x\mathbf{x} - \mathbf{K}_z\mathbf{z}_s - \mathbf{K}_w\mathbf{w}_s$.

EXAMPLE 14.15 A fourth-order system with input-output transfer function $y(s)/u(s) = (s+5)/[s(s+1)(s^2/\omega_m^2 + 2\zeta/\omega_m s + 1)]$ is to be approximated by a second-order model $(s+5)/[s(s+1)]$, which is a good low-frequency approximation if $\omega_m >> 1$. Assume $\zeta = 0.5$ and $\omega_m = 10$. Then the state variable equations for the true system and the model have the following matrices.

$$\mathbf{A} = \begin{bmatrix} -11 & 1 & 0 & 0 \\ -110 & 0 & 1 & 0 \\ -100 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 100 \\ 500 \end{bmatrix} \quad \mathbf{A}_m = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \quad \mathbf{B}_m = \begin{bmatrix} 1 \\ 5 \end{bmatrix}$$

The full-state feedback optimal *LQ* solution for the fourth-order system with $\mathbf{Q} = \text{Diag}[1\ 1\ 0\ 0]$ and $\mathbf{R} = 1$ has the feedback gains $\mathbf{G}_\infty = [-4.9368 \quad 0.4406 \quad 0.09376 \quad 0.00822]$, and this gives closed-loop $\lambda_i = \{-5.583 \pm 1.3338j;\ -6.659 \pm 11.102j\}$.

When the second-order approximate model is used with a cost function containing $\mathbf{Q} = \text{Diag}[1 \quad 1]$ and $\mathbf{R} = 1$, the resulting gains are $\mathbf{G}_{m\infty} = [0.41421 \quad 1]$, and the eigenvalues are $\lambda_i = \{-1.414, -5\}$. When these two approximate gains are used in the true fourth-order system, the step response of Figure 14.12 shows a very strong ringing at the frequency of the unmodeled modes. A frequency weighted cost function is selected in an attempt to suppress the oscillation. Various approaches could be studied, such as weighting both diagonals in **Q.** This would mean at least a second-order controller. Alternatively, a weighting on **R** could be introduced, and to keep the example simple $w(s) = [(10s + 1)/(s + 1)]u(s)$ is selected. As frequency (or s) gets large, this weights the fluctuations in u^2 100 times more than in the original $\mathbf{R} = 1$ case. The dynamics of this weighting filter are described by $\dot{w}_s = -w_s - 9u$ and $w = w_s + 10u$.

Augmenting the model as in Eqs. (*14.65*) through (*14.70*) gives the third-order design system (no extra $\mathbf{z}_s$ states are needed here):

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{w}_s \end{bmatrix} = \begin{bmatrix} -1 & 1 & 0 \\ 0 & 0 & -0.5 \\ 0 & 0 & -0.1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0.1 \\ 0.5 \\ -0.9 \end{bmatrix} w \qquad \mathbf{Q}_a = \text{Diag}[1 \quad 1 \quad 0], \qquad \mathbf{R}_a = 1$$

The ARE solution shows the optimal $w = -[0.19968 \quad 1.2145 \quad -0.63393]\mathbf{x}_a$. Converting back to the actual control gives $\mathbf{u} = -[0.019968 \quad 0.12145 \quad 0.036607][x_1 \quad x_2 \quad w_s]^T$. When this controller is used with the true fourth-order model, the high-frequency oscillation is replaced by a few cycles of a much lower frequency damped response. The step responses for (1) the true system with full optimal feedback, (2) the true system with static feedback of only the first two states (gains derived from the second-order model), and (3) the true system with the first-order dynamic controller just derived are compared in Figure 14.12. ■

The preceding example illustrates that the frequency weighting concept increases the order of the design model and the resulting controller, although not necessarily to

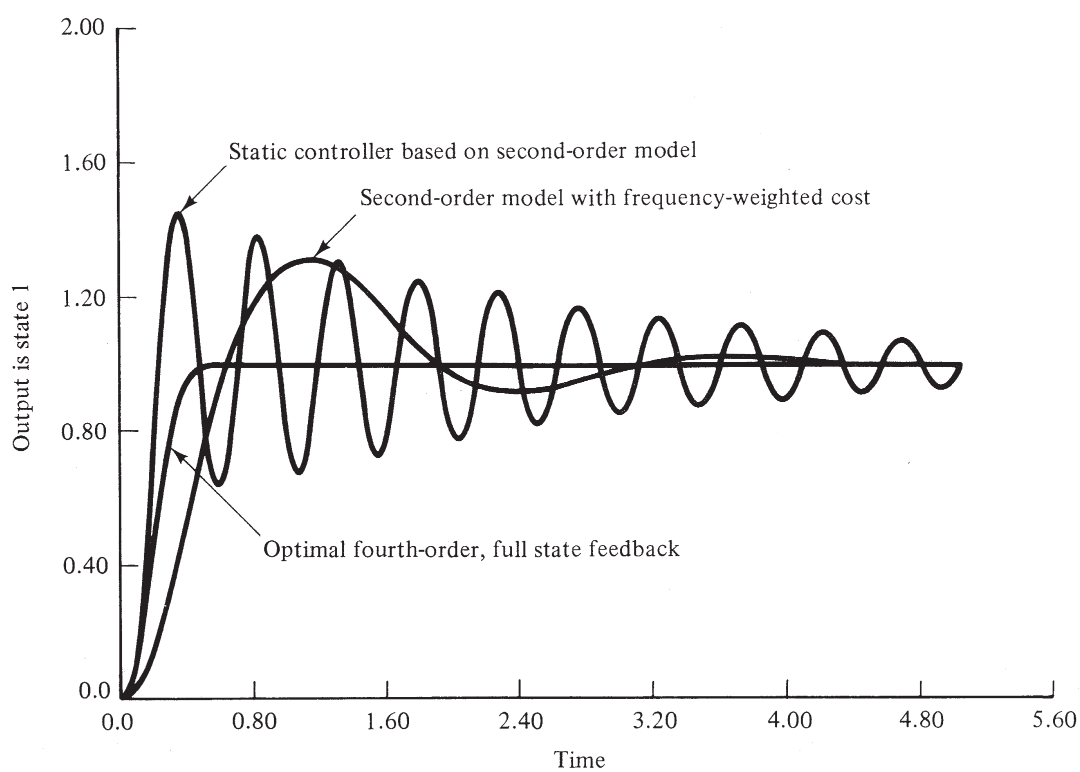


Figure 14.12

the full order of the true model, which is usually unknown. The projective controls approach reduces the order of the implemented controller. These two concepts were introduced for different fundamental reasons, one for controller order reduction and the other for enhancement of system robustness. Both concepts can be combined into a single design methodology [20].

14.9 CONCLUDING COMMENTS

Dynamic programming can be applied to a wide variety of optimization problems. An optimal assignment problem is described in Reference 21, and applications to control of systems described by partial differential equations are described in Reference 22. In principle, almost any optimal control problem can be solved numerically by using the dynamic programming approach. Let time be divided into a sufficiently small set of discrete points, and also quantize the allowable range of each state variable into a finite set of points. The set of allowable controls at each state is also quantized to a finite number of discrete possibilities. This means that the state-time space is then covered by a multidimensional mesh of node points similar to those in Figures 14.1 through 14.3 and in Problem 14.2. The solution process will normally require an interpolation step because an arbitrarily quantized set of controls at one time step will not necessarily give states at the next time step which fall on the discrete state values. Stated differently, the best control at a given state-time node will frequently fall between two of the discrete controls. Note that in this purely numerical approach, bounds on admissible controls and state variables are actually helpful, since they limit the range of values that need to be quantized and searched over. The biggest limitation to this method is what Bellman [1] called the "curse of dimensionality." If there are n state variables and if each is quantized into 100 points, there will be $(100)^n$ grid points for each time step. At each of these, the best control $\mathbf{u}[\mathbf{x}(k)]$ and the best cost $g[\mathbf{x}(k)]$ must be found (usually by direct enumeration) and stored. Computer storage and solution time become the limiting factors.

In terms of analytical solutions, dynamic programming yields the Hamilton-Jacobi-Bellman equation (*14.35*). This is a nonlinear partial differential equation, and general closed-form solutions are not known. When the set of admissible controls and states are not constrained, the minimization required in the HJB equation can be explicitly attempted. If the system is linear and if the cost is quadratic, this approach leads to the Riccati equation, which is still nonlinear but is an ordinary differential equation. The Riccati equation can be reduced to a coupled pair of linear differential equations, essentially equivalent to those obtained using the minimum principle. In the constant coefficient case, a closed-form solution can be written in terms of the exponential matrix. The infinite time-to-go solution can be found by solving the simpler algebraic Riccati equation, and the optimal control is expressed in terms of state feedback with constant gains.

Major emphasis here has been placed on the linear-quadratic problem and some interesting extensions to it. Notice that even here, computational aides are essential. Even a lowly second-order example becomes a fourth-order Hamiltonian system, which is not conducive to hand solution. In solving optimal control problems, it seems not to be a question of whether a computer will be used but how it will be used.

Tools for solving the Riccati equation are widely available. Solutions also can be obtained by using standard eigenvalue-eigenvector or numerical integration packages, which are even more widely available. Thus the mechanics of basic LQ theory can be applied by the controls practitioner in an easy fashion. Many extensions to the basic theory, such as those of Sec. 14.8, assume knowledge of and a facility with LQ theory as a starting point. Some adaptive and self-tuning control methods contain an inner loop consisting of an LQ optimal controller or a pole placement controller. The LQ theory is the most widely used aspect of optimal control theory, which is why it was stressed here.

There is a significant learning benefit derived from developing the software tools as needed rather than using canned packages. One reason for providing the many numerical examples here was that they allow the reader to calibrate his or her programs against known results.

REFERENCES

1. Bellman, R. and S. E. Dreyfus: *Applied Dynamic Programming,* Princeton University Press, Princeton, N.J., 1962.
2. Anderson, B. D. O. and J. B. Moore: *Linear Optimal Control,* Prentice Hall, Englewood Cliffs, N.J., 1971.
3. Bryson, A. E. and Y. C. Ho: *Applied Optimal Control,* Blaisdell (Xerox), Waltham, Mass., 1969.
4. Kirk, D. E.: *Optimal Control Theory,* Prentice Hall, Englewood Cliffs, N.J., 1970.
5. Pontryagin, L. S., V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko: *The Mathematical Theory of Optimal Processes,* Interscience: John Wiley, New York, 1962 (translated from Russian edition).
6. Gelfand, I. M. and S. V. Fomin: *Calculus of Variations,* Prentice Hall, Englewood Cliffs, N.J., 1963 (translated from Russian edition).
7. Kalman, R. E.: "When Is a Linear Control System Optimal?" *Trans. of the ASME, Journ. of Basic Engineering,* Vol. 86D, March 1964, pp. 51–60.
8. Safonov, M. G. and M. Athans: "Gain and Phase Margins for Multiloop LQG Regulators," *IEEE Trans. Automatic Control,* Vol. AC-22, No. 2, Apr. 1977, pp. 173–179.
9. Doyle, J.: "Analysis of Feedback Systems with Structured Uncertainties," IEE Proceedings, Part D, Vol. 129, Nov. 1982, pp. 242–250.
10. Morari, M. and E. Zafiriou: *Robust Process Control,* Prentice Hall, Englewood Cliffs, N.J., 1989.
11. Lehtomaki, N. A., N. R. Sandell, Jr., and M. Athans: "Robustness Results on LQG Based Multivariable Control System Designs," *IEEE Trans. Automatic Control,* Vol. AC-26, No. 1, Feb. 1981, pp. 66–74.
12. Friedland, B.: *Control System Design,* McGraw-Hill, New York, 1986.
13. Doyle, J. C. and G. Stein: "Robustness with Observers," *IEEE Trans. Automatic Control,* Vol. AC-24, No. 4, Aug. 1979, pp. 607–611.
14. Doyle, J. C. and G. Stein: "Multivariable Feedback Design: Concepts for a Classical/Modern Synthesis," *IEEE Trans. Automatic Control,* Vol. AC-26, No. 1, Feb. 1981, pp. 4–16.
15. Hopkins, W. E. Jr., J. V. Medanic, and W. R. Perkins: "Output Feedback Pole Placement in the Design of Suboptimal Linear Quadratic Regulator," *Int. Journ. of Control,* Vol. 34, No. 3, 1981, pp. 593–612.

16. Medanic, J. V. and Z. Uskokovic: "The Design of Optimal Output Regulators for Linear Multivariable Systems with Constant Disturbances," *Int. Journ. of Control,* Vol. 37, No. 4, 1983, pp. 809–830.
17. Meo, J. A. C., J. V. Medanic, and W. R. Perkins: "Design of Digital PI+ Dynamic Controllers Using Projective Controls," *Int. Journ. of Control,* Vol. 43, No. 2, 1986, pp. 539–559.
18. Gupta, N. K.: "Frequency-Shaped Cost Functionals: Extension of Linear-Quadratic-Gaussian Design Methods," *AIAA J. Guidance and Control,* Vol. 3, No. 6, Nov.–Dec. 1980, pp. 529–535.
19. Anderson, B. D. O. and Mingori, D. L.: "Use of Frequency Dependence in Linear Quadratic Control Problems to Frequency-Shape Robustness," *AIAA J. Guidance,* Vol. 8, No. 3, May–June 1985, pp. 397–401.
20. Tharp, H. S., J. V. Medanic, and W. R. Perkins: "Parameterization of Frequency Weighting for a Two-Stage Linear Quadratic Regulator Based Design," *Automatic,* Vol. 24, No. 3, 1988, pp. 415–418.
21. Brogan, W. L.: "Algorithm for Ranked Assignments with Applications to Multiobject Tracking," *AIAA J. Guidance, Control and Dynamics,"* Vol. 12, No. 3, May–June 1989, pp. 357–364.
22. 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.
23. Perkins, W. R. and J. B. Cruz: "Feedback Properties of Linear Regulators," *IEEE Trans. Automatic Control,* Vol. AC-16, No. 6, Dec. 1971, pp. 659–664.
24. Cruz, J. B., J. S. Freudenberg, and D. P. Looze: "A Relationship Between Sensitivity and Stability of Multivariable Feedback Systems," *IEEE Trans. Automatic Control,* Vol. AC-26, No. 1, Feb. 1981, pp. 66–74.
25. Dorato, P.: "A Historical Review of Robust Control," *IEEE Control Systems Magazine,* Vol. 7, No. 2, Apr. 1987, pp. 44–47.
26. Brogan, W. L.: "Computer Control System Design for a Radar Camouflage Production Line," *Ninth Hawaii International Conference on System Sciences,* Jan. 1976.

ILLUSTRATIVE PROBLEMS

Linear, Time-Varying Minimum Norm Problem

14.1 Consider the general linear time-varying system $\dot{\mathbf{x}}(t) = \mathbf{A}(t)\mathbf{x}(t) + \mathbf{B}(t)\mathbf{u}(t)$. Find the control which transfers the state from $\mathbf{x}(t_0)$ to $\mathbf{x}_d$ at a fixed time t_f while minimizing the control effort, $J = \int_{t_0}^{t_f} \mathbf{u}^T(t)\mathbf{u}(t)\,dt$.

Since at the final time t_f, $\mathbf{x}(t_f)$ must equal $\mathbf{x}_d$, the solution to the state equation must satisfy

$$\mathbf{x}_d - \mathbf{\Phi}(t_f, t_0)\mathbf{x}(t_0) = \int_{t_0}^{t_f} \mathbf{\Phi}(t_f, \tau)\mathbf{B}(\tau)\mathbf{u}(\tau)\,d\tau \tag{1}$$

The right-hand side of equation (*1*) defines a linear operator $\mathcal{A}(\mathbf{u})$ such that $\mathcal{A}: \mathcal{U} \rightarrow \Sigma$. The left-hand side is a fixed vector in the state space Σ. The input function $\mathbf{u}_{[t_0, t_f]}$ is an element of the infinite dimensional input function space $\mathcal{U}$. The performance criterion is the norm squared of an element in $\mathcal{U}$. The problem is therefore one of finding the minimum norm solution to a linear operator equation.

Results of Problem 6.26b are directly applicable, $\mathbf{u}^*(t) = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}[\mathbf{x}_d - \mathbf{\Phi}(t_f, t_0)\mathbf{x}(t_0)]$. The adjoint operator $\mathcal{A}^*$ is defined by $\langle \mathcal{A}(\mathbf{u}), \mathbf{v}\rangle_\Sigma = \langle \mathbf{u}, \mathcal{A}^*\mathbf{v}\rangle_{\mathcal{U}}$. Assuming $\mathbf{\Phi}$, $\mathbf{B}$, and $\mathbf{u}$ are real, the results of Sec. 11.6 give $\mathcal{A}^* = \mathbf{B}^T(t)\mathbf{\Phi}^T(t_f, t)$. The optimal control is

$$\mathbf{u}^*(t) = \mathbf{B}^T(t)\mathbf{\Phi}^T(t_f, t)\left[\int_{t_0}^{t_f} \mathbf{\Phi}(t_f, \tau)\mathbf{B}(\tau)\mathbf{B}^T(\tau)\mathbf{\Phi}^T(t_f, \tau)\, d\tau\right]^{-1}[\mathbf{x}_d - \mathbf{\Phi}(t_f, t_0)\mathbf{x}(t_0)]$$

Note that the indicated inverse exists if the system is controllable. By making use of $\mathbf{\Phi}(t_f, \tau) = \mathbf{\Phi}(t_f, t_0)\mathbf{\Phi}(t_0, \tau)$, the solution can be expressed in alternative but equivalent forms.

Other problems can be solved in the same way whenever the performance criterion J can be interpreted as some more general norm of a function $\mathbf{u}(t) \in \mathcal{U}$.

Discrete Dynamic Programming

14.2 Find the minimum cost path which starts at point a and ends at any one of the points h, j, k, l of Figure 14.13. Travel costs are shown beside each path segment and toll charges are shown by each node.

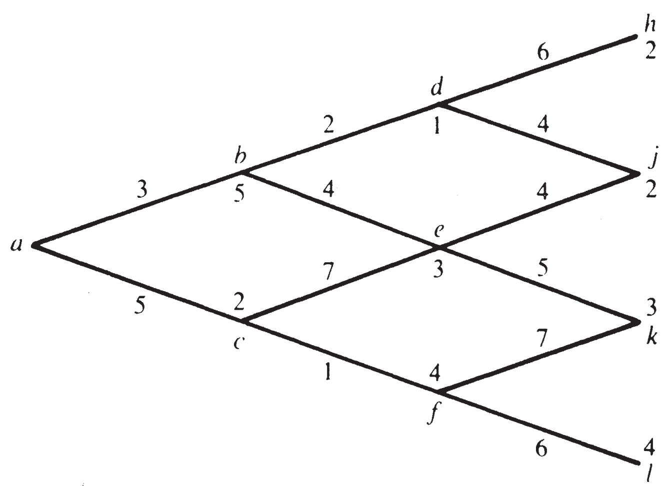


Figure 14.13

Let g_α be the minimum cost from node α to any one of the four possible termination points.

Last stage: $g_h = 2 \quad g_k = 3$
$g_j = 2 \quad g_l = 4$

Next stage: $g_d = 1 + \min[6 + g_h, 4 + g_j]$
$= 1 + \min[6 + 2, 4 + 2] = 7,$ from d to j
$g_e = 3 + \min[4 + g_j, 5 + g_k] = 9,$ from e to j
$g_f = 4 + \min[7 + g_k, 6 + g_l] = 14,$ from f to k or l

Next stage: $g_b = 5 + \min[2 + g_d, 4 + g_e] = 14,$ from b to d
$g_c = 2 + \min[7 + g_e, 1 + g_f] = 17,$ from c to f

Initial stage: $g_a = 0 + \min[3 + g_b, 5 + g_c] = 17,$ from a to b

The minimum total cost is $g_a = 17$ and is achieved on path a, b, d, j.

14.3 A two-stage discrete-time system is described by $x(k+1) = x(k) + u(k)$ with $x(0) = 10$. Use dynamic programming to find $u(0)$ and $u(1)$ which minimize $J = [x(2) - 20]^2 + \sum_{k=0}^{1} \{x^2(k) + u^2(k)\}$. There are no restrictions on $u(k)$.

$g[x(k)]$ is defined as the minimum cost from state $x(k)$ at time k to some final state $x(2)$. Then $g[x(2)] = [x(2) - 20]^2$ and

$$g[x(1)] = \min_{u(1)} \{x^2(1) + u(1)^2 + g[x(2)]\}$$

$$= \min_{u(1)} \{x^2(1) + u^2(1) + [x(1) + u(1) - 20]^2\}$$

Differentiating with respect to $u(1)$ and equating to zero gives $u(1) = [20 - x(1)]/2$, so that $g[x(1)] = x^2(1) + [20 - x(1)]^2/2$. Using this in $g[x(0)] = \min_{u(0)}\{x^2(0) + u^2(0) + g[x(1)]\}$ and setting $\partial\{\ \}/\partial u(0) = 0$ gives $u(0) = [20 - 3x(0)]/5$. This completes the backward pass. Since it is known that $x(0) = 10, u(0) = -2$. This, plus the difference equation, yields $x(1) = 8$. Using this in the expression found for $u(1)$ gives $u(1) = 6$. Finally, $x(2) = x(1) + u(1) = 14$. The minimum cost is $g[x(0)] = 240$.

14.4 Apply the general recursive algorithm of Sec. 14.3.3 to recalculate the solution of Problem 14.3.

For this case $\mathbf{A} = \mathbf{B} = \mathbf{M} = \mathbf{Q} = \mathbf{R} = 1$ and $\boldsymbol{\eta}(k) = 0$. The final time is $N = 2$. The initial conditions at $k = 2$ (or equivalently $k' = 0$) are $\mathbf{M}(0) = 1$ and $\mathbf{V}(0) = -20$. The calculations are tabulated from left to right in the order performed. The table also indicates the correct starting points and time sequencing relationships among the variables. The extra quantities $\mathbf{U}$ and $(\mathbf{I} - \mathbf{KB}^T)$ are given for convenience since they are both used more than once per line of table

k	$k'=N-k$	$\mathbf{u}(k)=-\mathbf{K}^T\mathbf{A}\mathbf{x}(k)-\mathbf{U}\mathbf{B}^T\mathbf{V}$	$\mathbf{V}$	$\mathbf{M}$	$\mathbf{U}=[\mathbf{B}^T\mathbf{M}\mathbf{B}+\mathbf{R}]^{-1}$	$\mathbf{K}=\mathbf{M}\mathbf{B}\mathbf{U}$	$(\mathbf{I}-\mathbf{K}\mathbf{B}^T)$	$\mathbf{P}$
2	0	—	−20	1	1/2	1/2	1/2	1/2
1	1	$\mathbf{u}(1)=-1/2\mathbf{x}(1)+10$	−10	3/2	2/5	3/5	2/5	3/5
0	2	$\mathbf{u}(0)=-3/5\mathbf{x}(0)+4$	−4	8/5	—	—	—	—

entries in the optimal tracking problem. Once $k = 0$ is reached on the backward pass, the knowledge that $\mathbf{x}(0) = 10$ is used to give $\mathbf{u}(0) = -\frac{3}{5}(10) + 4 = -2$. This control gives $\mathbf{x}(1) = 8$ and then $\mathbf{u}(1) = -\frac{1}{2}(8) + 10 = 6$. The final state is $\mathbf{x}(2) = \mathbf{x}(1) + \mathbf{u}(1) = 14$. These all agree with Problem 14.3.

The general mechanization of the optimal controller for the optimal tracker is given in Figure 14.3.

14.5 Figure 14.14*a* gives a schematic for an industrial system [26] typical of many in the pulp and paper industries. A linearized model of the fluidic control aspects will be treated here. The three major system elements are (1) the variable speed pump, which controls total flow volume, (2) the dilution valve, which determines the fraction of dilution water to concentrate, and (3) the head box, which acts as a reservoir and applies the mixture to a moving substrate. Each of these can be approximated by a first-order time lag. This leads to the block diagram of Figure 14.14*b*. Use the values $K = 10$, $\tau_h = 2$, $\tau_p = 1$, and $\tau_v = 0.5$ and obtain a discrete-time state model for a sample time $T = 0.2$. Design a constant-gain LQ feedback controller which attempts to maintain the state near $\eta = [4 \quad 0.6 \quad 1]^T$. Maintaining the primary output x_1 near 4 is most important. It is also important to keep u_1 small to avoid valve saturation.

The continuous-time state model is

$$\dot{\mathbf{x}} = \begin{bmatrix} -1/\tau_h & -K/\tau_h & K/\tau_h \\ 0 & -1/\tau_v & 0 \\ 0 & 0 & -1/\tau_p \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 & 0 \\ 1/\tau_v & 0 \\ 0 & 1/\tau_p \end{bmatrix}\mathbf{u}$$

When numerical values are substituted, Eq. (*9.21*) and Problem 9.10 give the discrete model

$$\mathbf{x}(k+1) = \begin{bmatrix} 0.904837 & -0.781725 & 0.861067 \\ 0 & 0.67032 & 0 \\ 0 & 0 & 0.818731 \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} -0.16990 & 0.09056 \\ 0.32968 & 0 \\ 0 & 0.18127 \end{bmatrix}\mathbf{u}(k)$$

A complete parametric study of the effect of Q and R values is not possible here. Three sets are presented. The value of Q_{22} was successively increased in an attempt to force a positive value for x_2. The DARE was solved to find the matrix $\mathbf{W}_\infty$. Then the gains $\mathbf{G}_\infty$, and $\mathbf{F}_{c\infty}$, and the input $\mathbf{V}_\infty$ is calculated from Eq. (*14.31*). The total external input $\mathbf{v}_{ext} = \mathbf{F}_{c\infty}\mathbf{V}_\infty$ is tabulated along with $\mathbf{G}_\infty$.

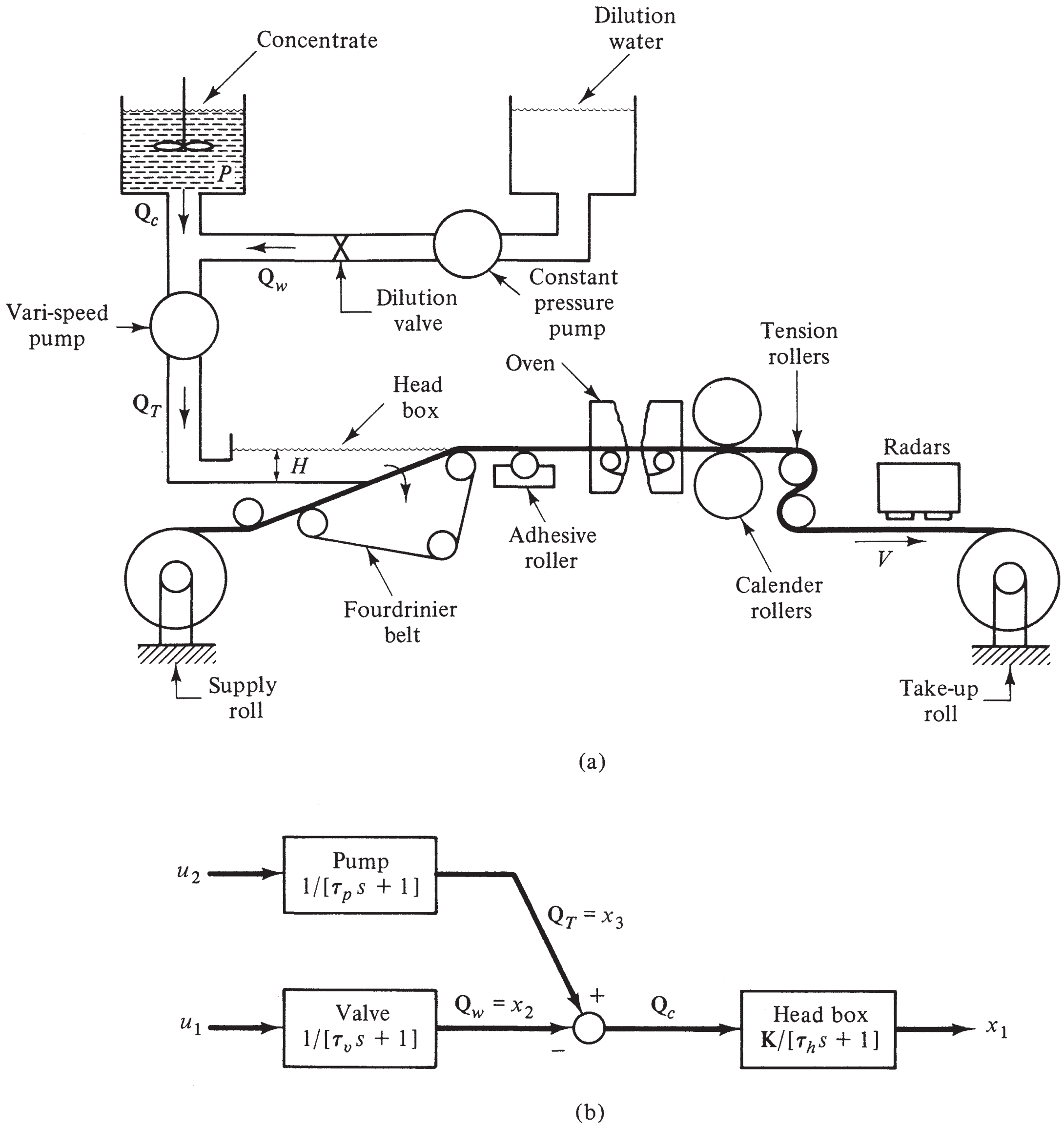


Figure 14.14

Case 1: $\mathbf{Q} = \text{Diag}[10, 1, 1]$, $\mathbf{R} = \text{Diag}[10, 4]$

$$\mathbf{v}_{\text{ext}} = \begin{bmatrix} -2.035 \\ 3.2415 \end{bmatrix} \qquad \mathbf{G}_\infty = \begin{bmatrix} -0.41905 & 0.71697 & -0.85416 \\ 0.57808 & -1.0044 & 1.4288 \end{bmatrix}$$

Case 2: $\mathbf{Q} = \text{Diag}[10, 4, 2]$, $\mathbf{R} = \text{Diag}[10, 4]$

$$\mathbf{v}_{\text{ext}} = \begin{bmatrix} -1.8476 \\ 3.5009 \end{bmatrix} \qquad \mathbf{G}_\infty = \begin{bmatrix} -0.40645 & 0.7620 & -0.80626 \\ 0.58728 & -0.96201 & 1.5118 \end{bmatrix}$$

Case 3: $\mathbf{Q} = \text{Diag}[10, 8, 2]$, $\mathbf{R} = \text{Diag}[10, 4]$

$$\mathbf{v}_{\text{ext}} = \begin{bmatrix} -1.6562 \\ 3.7437 \end{bmatrix} \qquad \mathbf{G}_\infty = \begin{bmatrix} -0.38764 & 0.80754 & -0.76125 \\ 0.60643 & -0.93136 & 1.5725 \end{bmatrix}$$

The transient responses for these cases are presented in Figure 14.15. The responses of x_1 are almost indistinguishable, but the effect of $\mathbf{Q}$ on x_2 and x_3 is clearly evident.

14.6 In order to compare quadratic optimal controllers with earlier methods, the dc motor system of Example 2.6 and Problems 2.22 through 2.24 is reconsidered. Using a sample time of $T = 1$ and

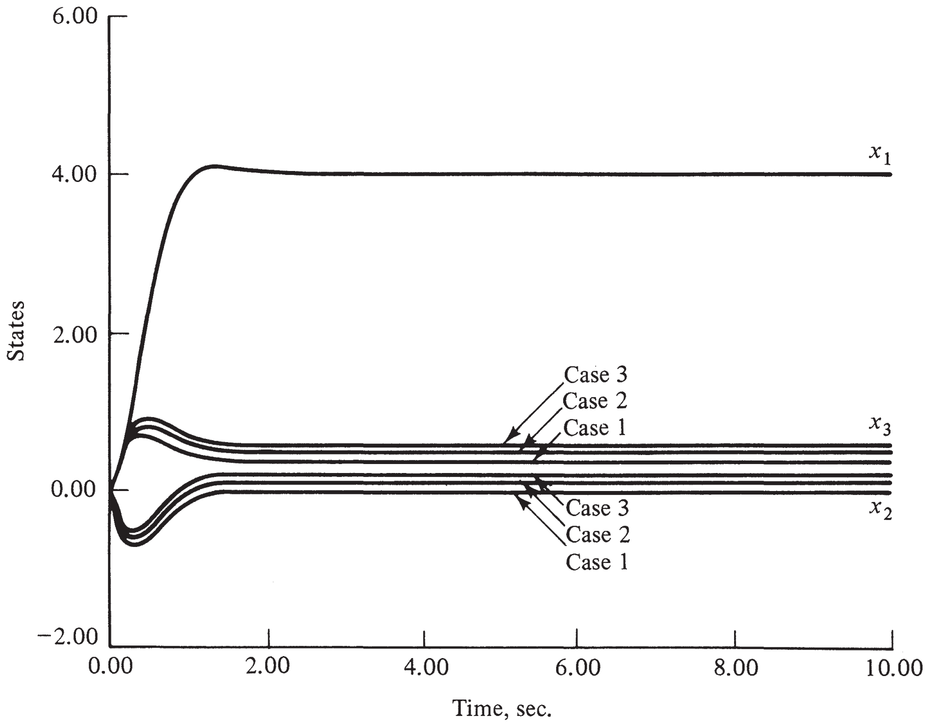


Figure 14.15

a motor gain of $K = 0.5$ allows the following observable canonical form state equations to be written for the open-loop system.

$$\mathbf{x}(k+1) = \begin{bmatrix} 1.6065 & 1 \\ -0.6065 & 0 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 0.1065 \\ 0.0902 \end{bmatrix} E(k)$$

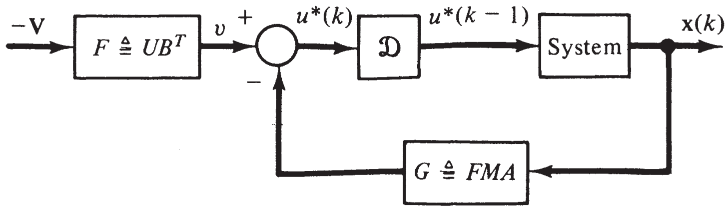
The first component $\mathbf{x}_1(k)$ is the motor shaft position angle, and its measurement is assumed available for use. It can be seen that the second state can be determined as $\mathbf{x}_2(k) = -0.6065\mathbf{x}_1(k-1) + 0.0902E(k-1)$. Therefore, it is justifiable to assume that both states are available for use. Design a controller which minimizes the cost function J of Eq. (*14.18*).

Equations (*14.15*), (*14.16*), and (*14.17*), plus equations (*14.21*) and (*14.22*) were solved with $N = 15$ steps of $T = 1$ s each. The first group of cases studied all had $\boldsymbol{\eta}(k) = \mathbf{0}$ and $\mathbf{x}_d = \mathbf{0}$ (i.e., the regulator problem). For each case the steady-state feedback gain matrix is listed, along with the closed-loop eigenvalue locations. In each case steady state was reached in about seven steps, a little over three motor time constants.

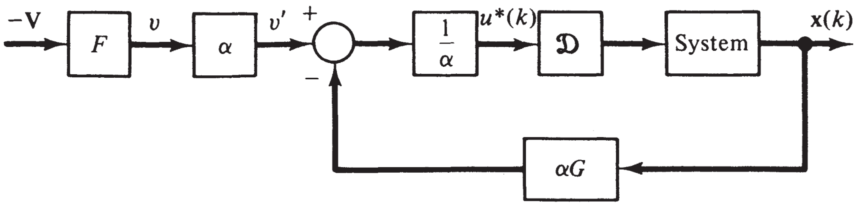
Case	M	Q	R	$\mathbf{G}_{ss} = \mathbf{K}^T \mathbf{A}$	Closed-loop Eigenvalues
1	**I**	**I**	1	[2.106 1.9160]	$0.6057 \pm j0.1756$
2	0	**I**	1	[2.1060 1.9160]	$0.6057 \pm j0.1756$
3	20**I**	**I**	1	[2.1060 1.9160]	$0.6057 \pm j0.1756$
4	**I**	**I**	10	[0.8163 0.7834]	0.6305, 0.8184 (real)
5	**I**	**I**	0.1	[4.516 3.810]	$0.3909 \pm j0.2505$

From this it is seen that the final results are independent of the initial **M.** For relatively smaller **R** values, larger gains and a faster, more responsive system is obtained. Larger **R** values give slower, smoother response and smaller gains. Smaller motor inputs will be required. The function of the **R** weighting term is to prevent large control signals from being called for. The **Q** weighting term is intended to get the state near to $\boldsymbol{\eta}(k)$.

Another group of cases with $\boldsymbol{\eta}(k) = [1. \quad -0.6065]^T$ was run, and for simplicity **M** was set to zero and $\mathbf{R} = 1$ was held fixed. In addition to the steady-state feedback gains, the steady-state



(a) Move F



(b) Pick normalizing $\alpha = \frac{1}{v_{s.s.}}$

Figure 14.16

values of the external inputs $\mathbf{v} = -\mathbf{U}\mathbf{B}^T\mathbf{V}$ are given, along with the scalar closed-loop transfer function $\mathbf{H}(z)$ from $\mathbf{v}(z)$ to $\mathbf{x}_1(z)$ (normalized as described below). The eigenvalues are, of course, the denominator roots. The increase in $\mathbf{Q}$ is equivalent to a relative decrease in $\mathbf{R}$. The previous observation remains true: as $\mathbf{Q}$ gets bigger relative to $\mathbf{R}$ the system response gets faster, but at the expense of larger control effort. The role of the external input needs to be clarified. This input $\mathbf{V}$ is a two-component vector. By shifting gains around, an equivalent normalized scalar input system can be obtained from Figure 14.4. The two steps used in shifting and normalizing are shown in Figure 14.16. The transfer functions given above are from v' to $\mathbf{x}$. The final value theorem shows that each transfer function has a steady-state gain of one, meaning that there will be no steady state error in the position $\mathbf{x}_1$ after a unit input at v'. This means that in steady state $\mathbf{Gx}$ must equal v.

Case	Q	Gain	v	Transfer Function
6	I	[same as case 1]	0.9439	$0.3971(z - 0.5324)/[(z - 0.6057 \pm j0.1756)]$
7	100I	[7.1525 5.6840]	3.7052	$1.2744(z - 0.4281)/[(z - 0.166 \pm j0.182)]$
8	10000I	[8.3299 6.4773]	4.4013	$1.4714(z - 0.4116)/[(z - 0.0068)(z - 0.1283)]$

The desired state $\eta(k)$ above happens to be a point of equilibrium where x_1 and x_2 are in balance when the input E is zero. Two more cases are solved where this is not true. x_1 is selected as unity and x_2 is chosen arbitrarily. In case 9 the errors in both states are equally weighted, and in case 10 the error in x_2 is ignored so η_2 could have been any value without changing the results. (This is not to say that the solution does not depend on Q_{22}, however.)

We see from the preceding results that the closed-loop system poles can be moved around by changing the relative importance of $\mathbf{Q}$ and $\mathbf{R}$. If any one of the above sets of eigenvalues had been selected a priori, then the pole-placement methods of Chapter 13 would give the same feedback gain. Sometimes it may be more natural to select pole locations based on their relation

Case	Q	Gain	v	Transfer Function
9	100I	[7.1525 5.684]	1.06	$1.2766(z - 0.4260)/[z - 0.1650 \pm j0.1883]$
10	Diag[100, 0]	[7.7430 5.696]	4.288	$1.3384(z - 0.3698)/[z - 0.1340 \pm j0.3060]$

to a desired time response. Pole placement would easily give the gains. Sometimes it may seem more natural to express the design goals in terms of a cost function to be minimized. In either case, if extremes are requested, the magnitudes of the resultant control signals may become excessive or even ridiculous. It is not possible to make a turbogenerator set, which normally takes minutes to come up to rated speed, respond with a time constant in the microsecond range just because "the math in Chapter 13 says you can put the poles anywhere you want." Neither can state variable feedback transform a light pleasure aircraft into a high-performance military fighter. The fallacy is due to the validity of the linear models breaking down and due to state or control limits being exceeded. The linear models are intended for use over a limited range of values of the signals. Conductors tend to vaporize when their rated loads are drastically exceeded. Wings or control surfaces tend to get ripped off under excessive loads. Similar limitations exist in other situations.

The step responses for the designs of cases 6, 7, 8, and 10 are shown in Figure 14.17 for comparison with the Chapter 2 results. Perhaps case 7 gives the best response of any considered here or in Chapter 2. Further parametric studies may uncover a better response between case 6 and case 7.

14.7 Find the optimal feedback control law for the unstable scalar system $\dot{x} = x + u$ which minimizes $J = Mx(t_f)^2 + \int_0^{t_f} u(t)^2\,dt$. There are no restrictions on $u(t)$, and t_f is fixed.

Using the dynamic programming results of Sec. 14.4.1 with $A = 1$, $B = 1$, $x_d = 0$, $Q = 0$, $R = 1$, the optimal control is $u^*(t) = \frac{1}{2}\nabla_x g[x(t), t_r]$, where $\nabla_x g = 2W(t_r)x(t) + V(t_r)$.

The Riccati equation for $W(t_r)$ is $dW/dt_r = 2W - W^2$, with $W(t_r = 0) = M$.

The equation for $V(t_r)$ is $dV/dt_r = [1 - W(t_r)]V$, but since $V(t_r = 0) = 0$, $V(t_r) \equiv 0$ for all t_r.

Using $W(t_r) = E(t_r)/F(t_r)$ for this scalar case leads to $dE/dt_r = E$, $dF/dt_r = E - F$, with $E(t_r = 0) = M$, $F(t_r = 0) = 1$. Solving gives $E(t_r) = e^{t_r}M$ and $F(t_r) = e^{-t_r} + (e^{t_r} - e^{-t_r})M/2$.

Since $W(t_r) = E(t_r)/F(t_r)$, it is seen that $W \to 2$ for large t_f. This in turn gives $u^*(t) = -2x(t)$ and the stable closed-loop system $\dot{x} = -x(t)$.

The general feedback control system is given in Figure 14.18.

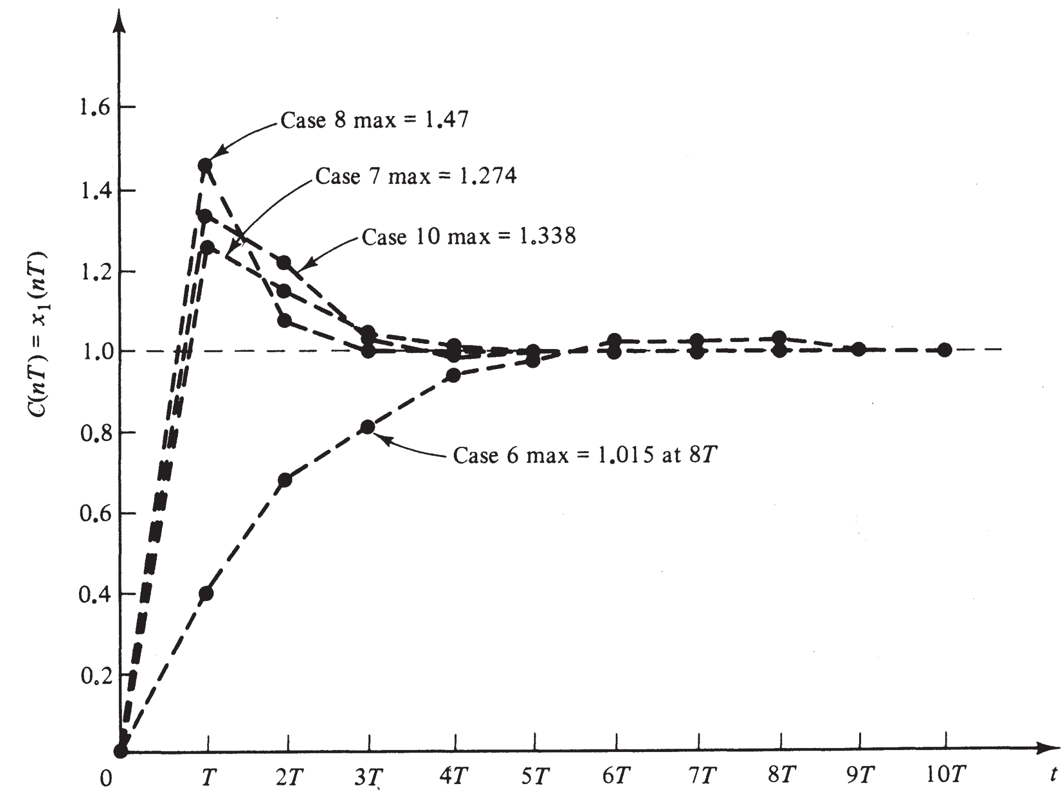


Figure 14.17 (Compare with Figures 2.29, 2.30, and 2.32.)

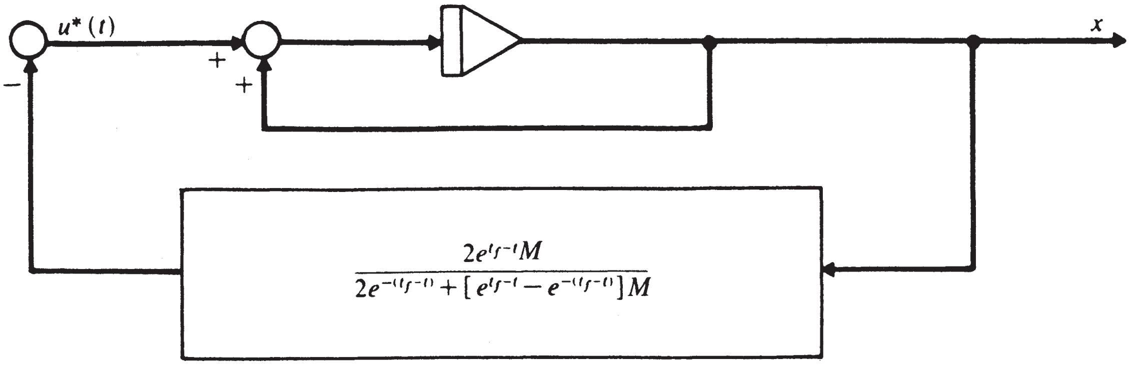


Figure 14.18

14.8 Consider the scalar control system $\dot{x} = Ax + u$. Investigate the constant state feedback control system which minimizes

$$J = x(t_f)^2 M + \int_0^{t_f} \{Qx^2 + u^2\}\, dt \qquad \text{as } t_f \to \infty$$

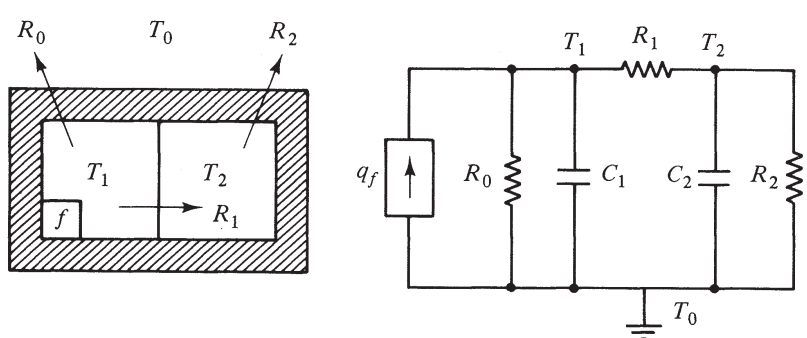
What is $u^*(t)$ if $Q = 0$? What if $Q = -A^2$?

The Riccati equation is $dW/dt_r = Q + 2WA - W^2$. In general, the steady-state solution can be found by numerically integrating until steady state is reached, or by setting the derivative equal to zero and solving a nonlinear algebraic equation for W. In this scalar case, setting $dW/dt_r = 0$ gives a simple quadratic equation for the steady-state values of W. Its solutions are $W = A \pm \sqrt{A^2 + Q}$.

Since $V(t_r) = 0$, $u^*(t) = -W(t_r)x(t)$, the steady-state feedback system satisfies $\dot{x} = Ax - [A \pm \sqrt{A^2+Q}]x = \mp\sqrt{A^2+Q}x$.

For a stable system, the plus sign in W must be selected, giving $\dot{x} = -\sqrt{A^2+Q}x$. If $Q = 0$, $\dot{x} = -|A|\mathbf{x}$. If the original system was stable, $-|A| = A$, so this implies that $u^* = 0$. But $Q = 0$ means that it does not matter what $x(t)$ is, so the optimal control strategy is to do nothing. This gives a zero value to the integral part of J. Also, since $t_f \to \infty$, $x(t_f) \to 0$ for a stable system. If the original system is unstable, the feedback system is stabilized. If $Q = -A^2$, then $W = A$ in steady state, so $u^*(t) = -Ax$. This gives $\dot{x} = 0$, and the integral term in J is once again identically zero.

14.9 A building is divided into two zones, as shown in Figure 14.19a. Figure 14.19b shows the analogous electric network, using the through variable/across variable analogies of Sec. 1.3. A furnace with heat output q_f is located in zone 1. If T_0 is the ambient temperature and T_1 and T_2 are the zone temperatures, the heat flow equations can be written as



(a) (b)

Figure 14.19

$$C_1\dot{T}_1 = q_f - (T_1 - T_0)/R_0 - (T_1 - T_2)/R_1$$
$$C_2\dot{T}_2 = (T_1 - T_2)/R_1 - (T_2 - T_0)/R_1$$

where C_i is the thermal capacitance of zone i. R_i are the lumped thermal resistances shown in Figure 14.19*a*. Let the control u be the furnace heat input q_f and let x_1 and x_2 be zone temperatures above ambient. The state equations become

$$\dot{\mathbf{x}} = \begin{bmatrix} -1/C_1[1/R_0 + 1/R_1] & 1/(C_1R_1) \\ 1/(C_2R_1) & -1/C_2[1/R_1 + 1/R_2] \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1/C_1 \\ 0 \end{bmatrix}u$$

(a) Design a constant gain full state LQ controller which minimizes

$$J = \int_0^\infty \{(\mathbf{x} - \boldsymbol{\eta})^T\mathbf{Q}(\mathbf{x} - \boldsymbol{\eta}) + \mathbf{R}u^2\}\,dt.$$

where $\boldsymbol{\eta} = [1 \quad 1]^T$ is the vector of desired zone temperatures above ambient, i.e., the set points. For simplicity, let $C_1 = C_2 = 1$, $R_0 = 10$, $R_1 = .1$, and $R_2 = 2$. Let $\mathbf{R} = 1$ and consider $\mathbf{Q} = \alpha\mathbf{I}$ for four values of α: .1, 1, 10, and 100.

(b) Show that if only one state is a measured output, an equivalent feedback transfer function $\mathbf{H}_{eq}(s)$ can be found which gives the same feedback signal as does full state feedback through the gain matrix G_∞. In particular, if the thermostat measurement is in zone 2, H_{eq} is a PD- (proportional-derivative-) type lead compensator. If the thermostat is in zone 1, then H_{eq} is a lag compensator.

(a) For this tracking problem, an external input is required in addition to the feedback gains. Setting $\dot{V} = 0$ in Eq. (*14.38*) and solving for $\mathbf{V}$ gives

$$\mathbf{V} = 2[\mathbf{A}^T - \mathbf{WBR}^{-1}\mathbf{B}^T]^{-1}\mathbf{Q}\boldsymbol{\eta}$$

and since $\mathbf{F}_\infty = -(\frac{1}{2})\mathbf{R}^{-1}\mathbf{B}^T$, the composite term is

$$v_{\text{ext}} = \mathbf{F}_\infty\mathbf{V} = -\mathbf{G}_\infty[\mathbf{A}^T\mathbf{W} - \mathbf{WBR}^{-1}\mathbf{B}^T\mathbf{W}]^{-1}\mathbf{Q}\boldsymbol{\eta}$$
$$= \mathbf{G}_\infty[\mathbf{WA} + \mathbf{Q}]^{-1}\mathbf{Q}\boldsymbol{\eta}$$

The last form used the ARE. It shows that if $\mathbf{Q}$ was large compared with $\mathbf{WA}$, then the control signal u depends on the total error $\mathbf{x} - \boldsymbol{\eta}$ as in many error-nulling servo systems; i.e., $u = \mathbf{G}_\infty(\boldsymbol{\eta} - \mathbf{x})$. This situation does not happen exactly because as $\mathbf{Q}$ increases, so does $\mathbf{W}$. The key results for the four values of $\mathbf{Q}$ are as follows:

α	$\mathbf{G}_\infty$	v_{ext}	λ_{c1}
0.1	[0.077102 0.07317]	0.27004	−0.374, −20.30
1.0	[0.48751 0.45427]	1.30478	−0.773, −20.30
10.0	[2.0883 1.8273]	4.43	−2.2645, −20.42
100.0	[7.7286 5.7925]	14.126	−6.714, −21.61

(b) Since all states may not be available as measured outputs, an observer may be needed to implement this controller. An alternative approximate method can often be used with single-input, single-output systems. The Laplace transform of the feedback signal under full state feedback is

$$F(s) = \mathbf{G}_\infty\mathbf{x}(s) = \mathbf{G}_\infty(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}u(s)$$

If instead, $y(s)$ is fed back through a transfer function $H(s)$, the feedback signal is

$$F(s) = H(s)y(s) = Hs)\mathbf{C}\mathbf{x}(s) = H(s)\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}u(s)$$

Equating gives $H(s) = \{\mathbf{G}_\infty(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\}/\{\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B}\}$. For this problem, if the thermostat is in zone 2, $\mathbf{C} = [0 \quad 1]$, and then $\mathbf{C}(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} = 10/[(s + 0.298)(s + 20.302)]$; for the case $\alpha = 10$, $\mathbf{G}_\infty(s\mathbf{I} - \mathbf{A})^{-1}\mathbf{B} = 2.088(s + 19.25)/[(s + 0.298)(s + 20.302)]$, so that $H(s) = 0.2088(s + 19.25)$. Since this is not a proper transfer function—that is, the numerator degree is higher than the denominator—it is not physically realizable. It is clearly of the PD-type lead compensator. A small time constant τ, say $\tau = 0.01$, could be used to form an approximation which is physically realizable, $H(s) \approx 0.2088(s + 19.25)/(\tau s + 1)$. This can be shown to give a response due to the v_{ext} step input, which is very close to the full state feedback response. If the thermostat is located in zone 1, then $\mathbf{C} = [1 \quad 0]$ leads to the physically realizable transfer function $H(s) = 2.088(s + 19.25)/(s + 10)$. This is recognized as a lag compensator.

14.10 The longitudinal dynamics of an aircraft cruising at constant speed can be approximated by [3, pp. 171–172] (see Figure 14.20)

$$\dot{\theta} = q$$

$$\dot{q} = -\omega^2(\alpha - S\delta)$$

$$\dot{\alpha} = -\alpha/\tau + q$$

Assume the values $\tau = 0.25$, $\omega = 2.5$, and $S = 1.6$. Then

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & -6.25 \\ 0 & 1 & -4 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ 10 \\ 0 \end{bmatrix} u$$

Design a constant feedback LQ controller which attempts to maintain the velocity vector near horizontal while keeping control effort small. That is, it is desired to minimize

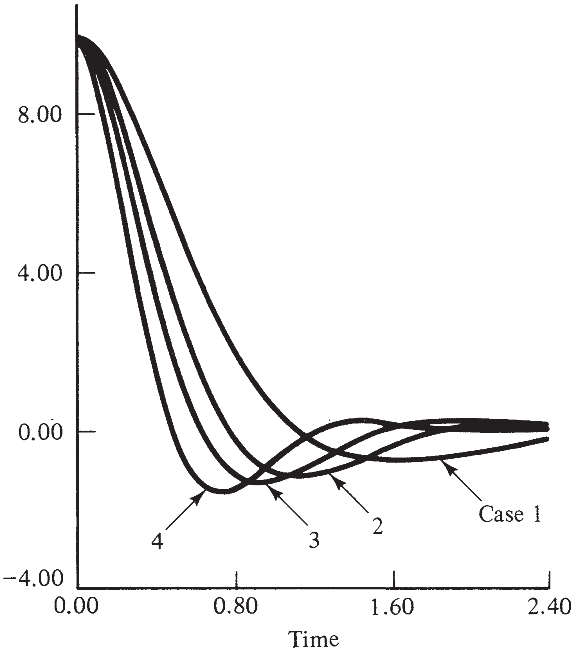
$$J = \int_0^\infty \{Q(\theta - \alpha)^2 + R\delta^2\}\,dt \qquad \textit{Note:} \text{ The matrix } \mathbf{Q} = Q\begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

Select a Q/R ratio that gives fast response to a sudden 10° change in α due to a gust or wind shear but that requires a maximum $|\delta| \le 10°$.

In this case Q and R are scalars and only the ratio is important. By trial and error, it is found that a value $Q/R = 2.8$ is the maximum permissible value—i.e., the fastest response—for which δ stays within the specified limit for an initial perturbation $\mathbf{x}(0) = [10 \quad 0 \quad 10]^T$. The feedback gains for four trial cases are tabulated, and the transient responses for α, θ, and δ are shown in Figure 14.21*a*, *b*, and *c*.

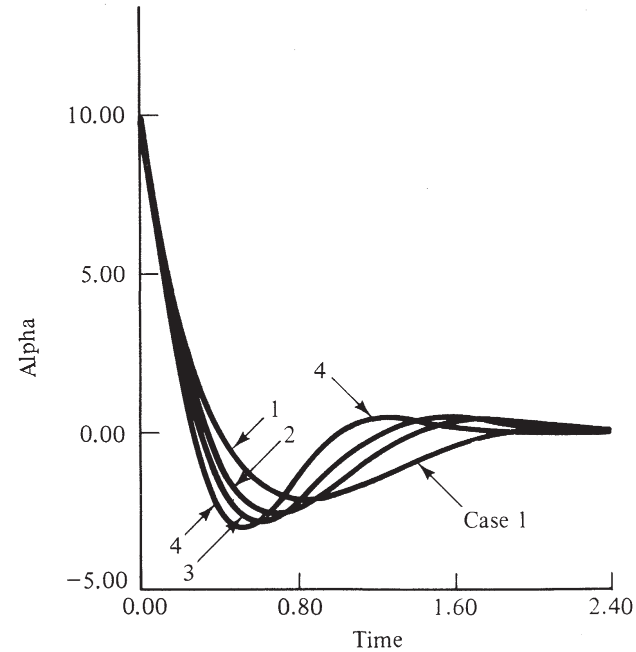
Case	Q/R	$\mathbf{G}_\infty$	λ_{c1}
1	0.1	[0.31623 0.14943 −0.20457]	−1.982, −1.756 ± *j*1.816
2	1.0	[1 0.33116 −0.45166]	−3.429, −1.942 ± *j*2.810
3	2.8	[1.6733 0.44895 −0.66554]	−4.1098, −2.1898 ± *j*3.1389
4	10.0	[3.1623 0.63169 −1.1671]	−5.0872, −2.6148 ± *j*4.2458

Figure 14.20



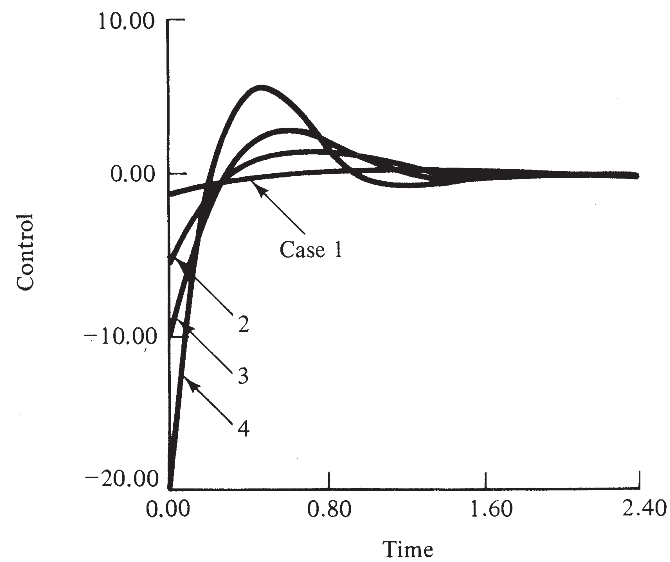
(a) Pitch attitude response for various Q/R ratios

Figure 14.21a



(b) Angle of attack for various Q/R ratios

Figure 14.21b



(c) Elevator deflection history for various Q/R ratios

Figure 14.21c

Applying the result of Problem 14.9 for an equivalent feedback transfer function to Case 3 and assuming that θ is the measured output gives

$$H(s) = \frac{4.4895(s + 3.122 + j2.2715)(s + 3.122 - j2.2715)}{s + 4}$$

This is not physically realizable (is not proper), since the numerator is of higher degree than the denominator. It may be acceptable to add a small time constant term $(\tau s + 1)$ in the denominator to give a realizable feedback transfer function. If τ is sufficiently small (i.e., the added pole is sufficiently far to the left), the contribution of the extra mode may be negligible.

14.11 A linearized aircraft model was introduced in Problem 11.9, where its controllability and observability properties were examined. Its transfer functions were obtained in Problem 12.5 and pole placement feedback controllers for it were found in Problem 13.35. Design a controller using the steady-state optimal regulator approach. For convenience, the system models are repeated here.

$$\mathbf{A} = \begin{bmatrix} -10 & 0 & -10 & 0 \\ 0 & -0.7 & 9 & 0 \\ 0 & -1 & -0.7 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 20 & 2.8 \\ 0 & -3.13 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

The solution involves selecting the appropriate weighting matrices **M, Q,** and **R** and then solving the Riccati Equation 14.39 until steady state is reached. The controller is given by

$$\mathbf{u}^*(t) = -(\tfrac{1}{2})\mathbf{R}^{-1}\mathbf{B}^T[2\mathbf{W}\mathbf{x}(t) + \mathbf{V}]$$

as shown in Figure 14.8. Two issues remain. What method is to be used to solve the Riccati equation? Numerical integration of Eq. (*14.39*) is used until **W** effectively becomes constant. For an asymptotically stable constant coefficient system, a reasonable estimate is that this will occur within four time constants. The time constant is estimated as the reciprocal of the smallest nonzero real part of the eigenvalues of **A.** The eigenvalues of **A** are 0, $-0.7 \pm 3j$, and -10, so the time constant is estimated as $1/0.7 = 1.4$ sec. The integration stepsize ΔT must not be too large, or poor accuracy will result. If it is too small, excessive computer time is required. A method which seems satisfactory is to select ΔT as about $1/(20|\lambda_{\max}|)$, which means that the number of integration steps might be on the order of $T/\Delta T = 80|\lambda_{\max}|/|\lambda_{\min}|$, which can easily exceed 1000. These are just estimates. The magnitudes of the changes in **W** must be monitored to determine when all components have essentially stopped changing.

The second issue is selection of the weighting matrices. For small problems with only a few parameters it may be feasible to parametrically examine the range of possibilities. For most problems a more focused approach is desirable. The expanded quadratic will contain terms of the form $x_i^2 Q_{ii} + u_i^2 R_{ii}$. Similar treatment of the final value terms can be done, but here $M = 0$ is selected because it will have no effect on steady-state answers. If x_i is a position variable with a magnitude of thousands of feet, and if u_i is an angle of say 0.01 radian, it is clear that u_i will have no effect on J unless $R_{ii} >> Q_{ii}$. The point is that scaling units and variable magnitudes are important, as well as the subjective choice of the importance of keeping u_i small compared to keeping x_i small. If all variables in the quadratic cost function are intended to be equally important, then one method is to estimate the maximum possible or allowable values of each variable and use these estimates to select the weights. For the airplane model, suppose that structural limits and prevention of pilot blackout require that the roll rate p be <300 deg/s and the yaw rate r be <18 deg/s. The maximum side-slip angle is estimated as $\beta < 15$ deg and the maximum roll angle as $\phi < 180$ deg. Likewise, there are maximums for the control surface deflection angles, and the hypothetical values assumed here are $\delta_a < 40$ deg and $\delta_r < 10$ deg. Then setting each term in J to unity when the variables are at their limits gives

$$\mathbf{Q} = \text{Diag}[0.0011, 0.308, 0.44, 0.003]$$

$$\mathbf{R} = \text{Diag}[0.0626, 1]$$

The relative magnitudes are all that matter. The above components have been scaled so the largest entry is unity. Considerable rounding off is probably justified because of the gross approximations involved in estimating the maximums. Using the above values for **Q** and **R,** the approximate steady-state feedback gain matrix and the closed-loop eigenvalues it yields are

$$\mathbf{G} = \begin{bmatrix} 0.0358 & 0.0197 & -0.0355 & 0.1942 \\ 0.0001 & -0.3072 & -0.1191 & -0.0001 \end{bmatrix}$$

$$\lambda_i = \{-0.376, -1.181 \pm 2.898j, -10.340\}$$

Cause		x_1	x_2	x_3	x_4
Commands	δ_a	10	0.3	0.5	35
	δ_r	0.03	5.5	1.7	0.01

The resultant system has moved the open-loop pole from the origin to -0.376, which means that the perturbations caused by any step disturbance will decay back to zero. The settling time and damping ratio of the dominant complex poles have been improved considerably and the remaining nondominant pole is not much changed.

If the maximum value of each state variable is multiplied by its two gain values, the resultant commands for δ_a and δ_r can be estimated. These values are valid if only one state variable is at its maximum and all others are zero. This shows that no maximum command is exceeded, unlike the results that can occur if arbitrarily selected **Q** or **R** values are used, or if extreme requests are made of a pole-placement design. In this particular example it is clear that δ_a is primarily controlled by the roll rate x_1 and roll angle x_4, and δ_r is primarily controlled by yaw rate x_2 and side-slip angle x_3 as expected from the physics of the problem.

A series of other sets of Q and R have been analyzed for comparison. With $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = \mathbf{I}$,

$$\mathbf{G} = \begin{bmatrix} 0.6587 & 0.0768 & -0.2610 & 0.9909 \\ 0.0802 & -0.7184 & -0.2743 & 0.0729 \end{bmatrix}$$

$$\lambda_i = \{-0.889, -1.82 \pm j2.6, -22.5\}$$

The dominant complex poles are about the same as the earlier case and the dominant real pole has an improved settling time, but the major difference is the shift in the nondominant pole, which will not be reflected much in system behavior. The required gains are much higher here and commanded deflection angles could easily exceed their limits.

With $\mathbf{Q} = 10\mathbf{I}$ and $\mathbf{R} = 0.1\mathbf{I}$ the resultant eigenvalues are all real $\{-0.995, -1.27, -30.7,$ and $-202.2\}$ and the gains are so high as to be ridiculous. Several are on the order of 9 or 10!

With $\mathbf{Q} = \text{Diag}[0.01, 0.1, 0.1, 0.01]$ and $\mathbf{R} = \text{Diag}[0.1, 1]$ the results are

$$\mathbf{G} = \begin{bmatrix} 0.1160 & 0.0275 & -0.0952 & 0.2956 \\ 0.0012 & -0.1148 & -0.0559 & 0.0004 \end{bmatrix}$$

$$\lambda_i = \{-0.5, -0.8798 \pm 2.965j, -11.823\}$$

The dominant poles are less damped and have a poorer settling time than the first case in spite of the fact that gains are typically higher and command limits could be exceeded.

The Minimum Principle

14.12 A system is described by $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t)$, with $\mathbf{x}(t_0)$ given. Find the necessary conditions which $\mathbf{x}(t)$ and $\mathbf{u}(t)$ must satisfy if they are to minimize

$$J = S(\mathbf{x}(t_f), t_f) + \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t)\, dt$$

The admissible controls must satisfy $\mathbf{u}(t) \in U$. Assume t_f is fixed.

Let $\mathbf{x}^*(t), \mathbf{u}^*(t)$ be the optimal quantities and let $\mathbf{x}(t)$ be arbitrary and let $\mathbf{u}(t)$ be arbitrary but admissible. Let $\mathbf{p}(t)$ be an $n \times 1$ vector of Lagrange multipliers (also called the *costate* or *adjoint* variables). Adjoin the differential constraints to J and call the result J':

$$J' = S(\mathbf{x}(t_f), t_f) + \int_{t_0}^{t_f} \{L(\mathbf{x}, \mathbf{u}, t) + \mathbf{p}^T(t)[\mathbf{f}(\mathbf{x}, \mathbf{u}, t) - \dot{\mathbf{x}}]\}\, dt$$

Since $\mathbf{x}^*, \mathbf{u}^*$ are optimal, $\Delta J = J'(\mathbf{x}, \mathbf{u}) - J'(\mathbf{x}^*, \mathbf{u}^*) \geq 0$ for $\mathbf{x}$ arbitrary and $\mathbf{u} \in U$.

Let $\mathbf{x} = \mathbf{x}^* + \delta\mathbf{x}$. Using Taylor series expansion gives

$$J'(\mathbf{x}^* + \delta\mathbf{x}, \mathbf{u}) = S(\mathbf{x}^*(t_f), t_f) + [\nabla_{\mathbf{x}} S]^T|_{tf}\,\delta\mathbf{x}(t_f) + \int_{t0}^{tf}\Big\{L(\mathbf{x}^*, \mathbf{u}, t)$$
$$+ [\nabla_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{u}, t)^T]\,\delta\mathbf{x} + \mathbf{p}^T\Big[\mathbf{f}(\mathbf{x}^*, \mathbf{u}, t) + \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\delta\mathbf{x} - \dot{\mathbf{x}}^* - \delta\dot{\mathbf{x}}\Big]\Big\}\, dt$$
$$+ \text{higher-order terms}$$

Therefore,

$$\Delta J = (\nabla_{\mathbf{x}} S)^T|_{tf}\,\delta\mathbf{x}(t_f) + \int_{t0}^{tf}\{L(\mathbf{x}^*, \mathbf{u}, t) + \mathbf{p}^T\mathbf{f}(\mathbf{x}^*, \mathbf{u}, t) - L(\mathbf{x}^*, \mathbf{u}^*, t)$$
$$- \mathbf{p}^T\mathbf{f}(\mathbf{x}^*, \mathbf{u}^*, t)\}\, dt + \int_{t0}^{tf}\Big\{[\nabla_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{u}, t)]^T\delta\mathbf{x} + \mathbf{p}^T\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\delta\mathbf{x} - \mathbf{p}^T\delta\dot{\mathbf{x}}\Big\}\, dt$$
$$+ \text{higher-order terms.}$$

When the higher-order terms are dropped, the result is called δJ, the *first variation of J*. The condition for optimality is that $\Delta J \geq 0$ for arbitrary $\delta\mathbf{x}(t)$ and $\mathbf{u}(t) \in U$. If $\mathbf{x}, \mathbf{u}$ are sufficiently close to $\mathbf{x}^*, \mathbf{u}^*$, then the sign of ΔJ is the same as the sign of δJ. By defining the Hamiltonian as $\mathcal{H}(\mathbf{x}^*, \mathbf{u}^*, \mathbf{p}, t) = L(\mathbf{x}^*, \mathbf{u}^*, t) + \mathbf{p}(t)^T\mathbf{f}(\mathbf{x}^*, \mathbf{u}^*, t)$ and using integration by parts, $\int_{t0}^{tf} \mathbf{p}^T\delta\dot{\mathbf{x}}\, dt = \mathbf{p}^T\delta\mathbf{x}\Big|_{t0}^{tf} - \int_{t0}^{tf} \dot{\mathbf{p}}^T\delta\mathbf{x}\, dt$, we obtain

$$\delta J = [\nabla_{\mathbf{x}} S - \mathbf{p}]_{tf}^T\,\delta\mathbf{x}(t_f) + \int_{t0}^{tf}[\mathcal{H}(\mathbf{x}^*, \mathbf{u}, \mathbf{p}, t) - \mathcal{H}(\mathbf{x}^*, \mathbf{u}^*, \mathbf{p}, t)]\, dt$$
$$+ \int_{t0}^{tf}\Big\{\Big[[\nabla_{\mathbf{x}} L(\mathbf{x}^*, \mathbf{u}, t)]^T + \mathbf{p}^T\frac{\partial \mathbf{f}}{\partial \mathbf{x}} + \dot{\mathbf{p}}^T\Big]\,\delta\mathbf{x}\Big\}\, dt$$

In order for this to be nonnegative for arbitrary $\delta\mathbf{x}$, the coefficient of $\delta\mathbf{x}(t)$ inside the integral must vanish on the optimal trajectory.

$$\dot{\mathbf{p}} = -\left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]^T \mathbf{p} - \nabla_{\mathbf{x}} L$$

The term involving $\delta\mathbf{x}(t_f)$ must also vanish. If $\mathbf{x}(t_f)$ is fixed, then $\delta\mathbf{x}(t_f) = \mathbf{0}$. If $\mathbf{x}(t_f)$ is not fixed, then $\delta\mathbf{x}(t_f) \neq \mathbf{0}$, so $\mathbf{p}(t_f) = \nabla_{\mathbf{x}} S|_{tf}$. The remaining integral term must be nonnegative for any admissible $\mathbf{u}$, including a $\mathbf{u}$ which equals $\mathbf{u}^*$ for all except an infinitesimal time interval. Therefore, it is concluded that

$$\mathcal{H}(\mathbf{x}^*, \mathbf{u}, \mathbf{p}, t) \geq \mathcal{H}(\mathbf{x}^*, \mathbf{u}^*, \mathbf{p}, t) \qquad \text{for all } t, \text{ all } \mathbf{u} \in U$$

This constitutes one version of the minimum principle, since $\mathbf{u}^*(t)$ is that $\mathbf{u}$ which minimizes $\mathcal{H}$. If there are no restrictions on $\mathbf{u}$, this reduces to $\partial\mathcal{H}/\partial\mathbf{u} = \partial L/\partial\mathbf{u} + \mathbf{p}^T(\partial\mathbf{f}/\partial\mathbf{u}) = \mathbf{0}$.

14.13 What modifications are necessary in the previous problem if t_f is not fixed in advance, but must be determined as part of the solution?

In forming ΔJ, variations in the final time must now be considered. Let t_f^* be the optimal stopping time and $t_f = t_t^* + \delta t_f$ is the perturbed time. Referring to Figure 14.22, it is apparent that two different variations in $\mathbf{x}$ at the final time must be considered. $\delta\mathbf{x}(t_f)$ is the variation used in Problem 14.12 and $\Delta\mathbf{x}(t_f)$ is the total variation.

$$\Delta J = J'(\mathbf{x}^* + \delta\mathbf{x}, \mathbf{u}, t_f^* + \delta t_f) - J'(\mathbf{x}^*, \mathbf{u}^*, t_f^*)$$
$$= S(\mathbf{x}^*(t_f^*) + \Delta\mathbf{x}_f, t_f^* + \delta t_f) - S(\mathbf{x}^*(t_f^*), t_f^*)$$

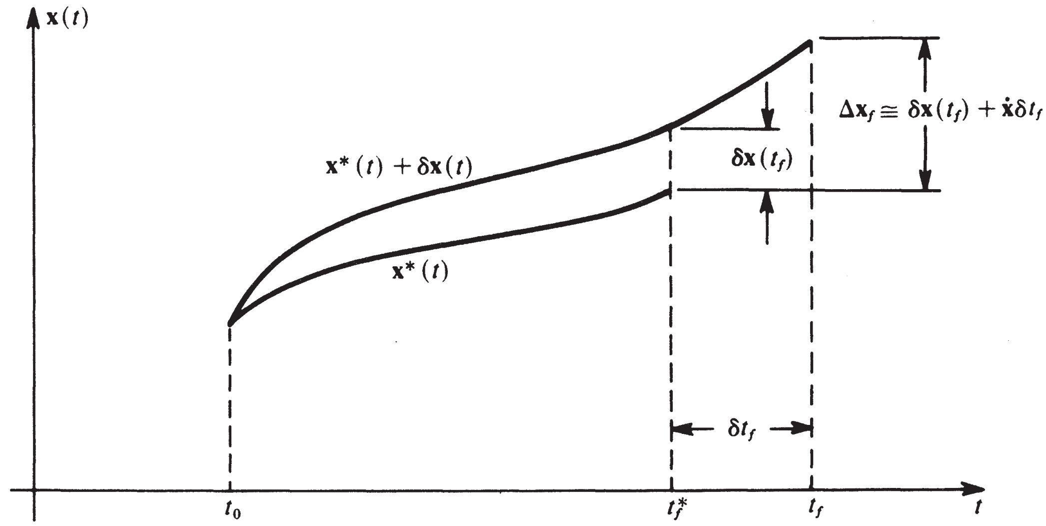


Figure 14.22

$$+\int_{t_0}^{t_f^*+\delta t_f}\{L(\mathbf{x}^*+\delta\mathbf{x},\mathbf{u},t)+\mathbf{p}^T[\mathbf{f}(\mathbf{x}^*+\delta\mathbf{x},\mathbf{u},t)-\dot{\mathbf{x}}^*-\delta\dot{\mathbf{x}}]\}\,dt$$

$$-\int_{t_0}^{t_f^*}\{L(\mathbf{x}^*,\mathbf{u}^*,t)+\mathbf{p}^T[\mathbf{f}(\mathbf{x}^*,\mathbf{u}^*,t)-\dot{\mathbf{x}}^*]\}\,dt$$

In computing δJ, the following relation is used:

$$\int_{t_0}^{t_f^*+\delta t_f}\{\ \}\,dt=\int_{t_0}^{t_f^*}\{\ \}\,dt+\int_{t_f^*}^{t_f^*+\delta t_f}\{\ \}\,dt\cong\int_{t_0}^{t_f^*}\{\ \}\,dt+\delta t_f L$$

Then

$$\delta J=(\nabla_{\mathbf{x}}S)^T|_{t_f}\Delta\mathbf{x}_f-\mathbf{p}^T(t_f)\delta\mathbf{x}(t_f)+\left[\frac{\partial S}{\partial t}+L(\mathbf{x}^*,\mathbf{u},t)\right]\bigg|_{t_f^*}\delta t_f$$

$$+\int_{t_0}^{t_f^*}[\mathcal{H}(\mathbf{x}^*,\mathbf{u},t)-\mathcal{H}(\mathbf{x}^*,\mathbf{u}^*,t)]\,dt$$

$$+\int_{t_0}^{t_f}\left\{\left[[\nabla_{\mathbf{x}}L(\mathbf{x}^*,\mathbf{u},t)]^T+\mathbf{p}^T\frac{\partial\mathbf{f}}{\partial\mathbf{x}}+\dot{\mathbf{p}}^T\right]\delta\mathbf{x}\right\}dt$$

From this, conclusion regarding the two integral terms are the same as in Problem 14.12. The only changes are in the boundary terms. If $\mathbf{x}^*(t_f)$ is free, then using $\delta\mathbf{x}(t_f)=\Delta\mathbf{x}_f-\dot{\mathbf{x}}^*(t_f)\delta t_f$ leads to boundary terms $[(\nabla_{\mathbf{x}}S)^T|_{t_f^*}-\mathbf{p}^T(t_f)]\Delta\mathbf{x}_f+[\partial S/\partial t)+L(\mathbf{x}^*,\mathbf{u},t)+\mathbf{p}^T\dot{\mathbf{x}}^*]|_{t_f}\delta t_f$. The conclusion is that $\mathbf{p}(t_f)=\nabla_{\mathbf{x}}S|_{t_f^*}$ as before, and the additional scalar equation required for determining t_f^* is $\partial S/\partial t+\mathcal{H}|_{t_f}=0$.

If the final state had been restricted in some way, say to lie on a surface $\psi(\mathbf{x}(t_f),t_f)$, then $\Delta\mathbf{x}_f$ and δt_f are interrelated. Their coefficients cannot then be separately set equal to zero, but instead, additional restrictions must be imposed [3, 5].

14.14 Develop an iterative method of solving the two-point boundary value problem which results when the minimum principle is applied to:

$$\dot{\mathbf{x}}=\mathbf{f}(\mathbf{x},\mathbf{u}),\qquad \mathbf{x}(t_0)\text{ known},\qquad t_f\text{ fixed},$$

$$\text{minimize } J=\tfrac{1}{2}[\mathbf{x}(t_f)-\mathbf{x}_d]^T\mathbf{M}[\mathbf{x}(t_f)-\mathbf{x}_d]+\int_{t_0}^{t_f}L\,dt$$

Minimizing $\mathcal{H}$ allows $\mathbf{u}^*(t)$ to be found as a function of $\mathbf{p}(t)$. Then the two-point boundary value problem of Eq. (*14.53*) can be written as $\dot{\mathbf{x}}=\mathbf{f}(\mathbf{x},\mathbf{p}),\ \dot{\mathbf{p}}=\mathbf{h}(\mathbf{x},\mathbf{p});\ \mathbf{x}(t_0)=\mathbf{x}_0,\ \mathbf{p}(t_f)=$

$\mathbf{M}[\mathbf{x}(t_f) - \mathbf{x}_d]$. If $\mathbf{p}(t_0)$ were known, the equations for $\mathbf{x}$ and $\mathbf{p}$ could be solved by numerical integration.

One way to proceed is to assume a $\mathbf{p}(t_0)^{(0)}$, numerically integrate to find $\mathbf{x}(t_f)^{(0)}$ and $\mathbf{p}(t_f)^{(0)}$. Then the terminal conditions can be checked, and the error can be used to estimate a new $\mathbf{p}(t_0)^{(1)}$. Figure 14.23 is used to illustrate the correction procedure [4]. Since $\mathbf{p}(t_f)^{(0)}$ and $\mathbf{x}(t_f)^{(0)}$ are determined by $\mathbf{p}(t_0)^{(0)}$, an unknown functional relation exists among these variables, as implied by the graph. The "slope" of the function multiplied by $\Delta\mathbf{p}(t_0)$ is set equal to the error in the terminal conditions. That is, a Newton-Raphson correction scheme is used.

The new estimate is

$$\mathbf{p}(t_0)^{(1)} = \mathbf{p}(t_0)^{(0)} - \Delta\mathbf{p}(t_0)$$

where

$$\left[\frac{\partial \mathbf{p}(t_f)}{\partial \mathbf{p}(t_0)} - \mathbf{M}\frac{\partial \mathbf{x}(t_f)}{\partial \mathbf{p}(t_0)}\right]^{(0)} \Delta\mathbf{p}(t_0) = \mathbf{p}(t_f)^{(0)} - \mathbf{M}[\mathbf{x}(t_f)^{(0)} - \mathbf{x}_d]$$

The two $n \times n$ sensitivity matrices $\mathbf{S}_\mathbf{p}(t_f) \triangleq \partial\mathbf{p}(t_f)/\partial\mathbf{p}(t_0)$ and $\mathbf{S}_\mathbf{x}(t_f) \triangleq \partial\mathbf{x}(t_f)/\partial\mathbf{p}(t_0)$ can be found by solving two sets of $n \times n$ matrix equations, obtained from the differential equations for $\mathbf{x}$ and $\mathbf{p}$ by interchanging $\partial/\partial\mathbf{p}(t_0)$ and d/dt. They are

$$\dot{\mathbf{S}}_\mathbf{x} = \frac{\partial \mathbf{f}}{\partial \mathbf{x}}\mathbf{S}_\mathbf{x} + \frac{\partial \mathbf{f}}{\partial \mathbf{p}}\mathbf{S}_\mathbf{p}, \qquad \mathbf{S}_\mathbf{x}(0) = [\mathbf{0}]$$

$$\dot{\mathbf{S}}_\mathbf{p} = \frac{\partial \mathbf{h}}{\partial \mathbf{x}}\mathbf{S}_\mathbf{x} + \frac{\partial \mathbf{h}}{\partial \mathbf{p}}\mathbf{S}_\mathbf{p}, \qquad \mathbf{S}_\mathbf{p}(0) = \mathbf{I}_n$$

These can be integrated along with the $\mathbf{x}$ and $\mathbf{p}$ equations. Only the terminal values are needed in the correction scheme, which generalizes to

$$\mathbf{p}(t_0)^{(k+1)} = \mathbf{p}(t_0)^{(k)} - [\mathbf{S}_\mathbf{p}(t_f)^{(k)} - \mathbf{M}\mathbf{S}_\mathbf{x}(t_f)^{(k)}]^{-1}\{\mathbf{p}(t_f)^{(k)} - \mathbf{M}[\mathbf{x}(t_f)^{(k)} - \mathbf{x}_d]\}$$

Success of the method depends on a good initial estimate for $\mathbf{p}(t_0)$ as well as the characteristics of the functions $\mathbf{f}$ and $\mathbf{h}$ and their derivatives.

14.15 Assuming that matrices $\mathbf{A}$ and $\mathbf{B}$ and vectors $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\mathbf{x}_d$ are given, find the solution for the optimal control in Example 14.12.

When $\mathbf{A}$ and $\mathbf{B}$ are constant, the form of the solution for Eq. (*14.54*) is

$$\begin{bmatrix}\mathbf{x}(t)\\ \mathbf{p}(t)\end{bmatrix} = e^{\begin{bmatrix}\mathbf{A} & -1/2\mathbf{B}\mathbf{B}^T\\ \mathbf{0} & -\mathbf{A}^T\end{bmatrix}(t-t_0)}\begin{bmatrix}\mathbf{x}(t_0)\\ \mathbf{p}(t_0)\end{bmatrix}$$

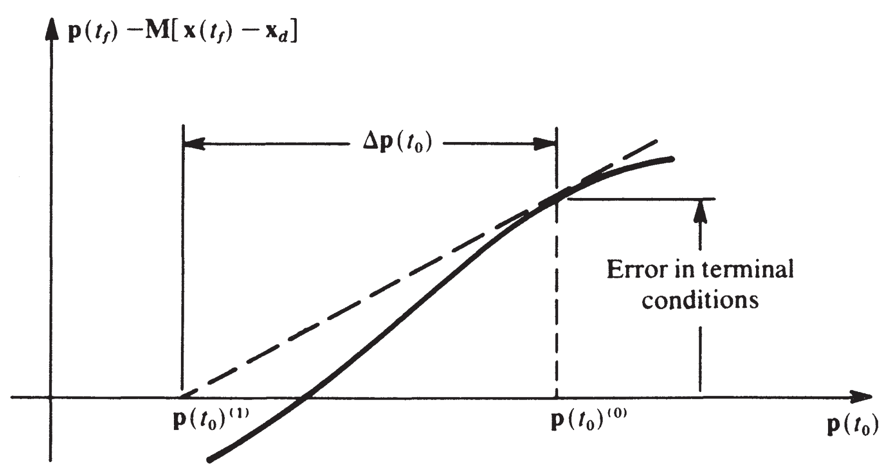


Figure 14.23

For convenience, the $2n \times 2n$ exponential matrix is written in partitioned form as

$$\begin{bmatrix} \boldsymbol{\phi}_{11}(t, t_0) & \boldsymbol{\phi}_{12}(t, t_0) \\ \boldsymbol{\phi}_{21}(t, t_0) & \boldsymbol{\phi}_{22}(t, t_0) \end{bmatrix}$$

This may be computed using the methods of Chapter 8. Actually $\boldsymbol{\phi}_{21}(t, t_0)$ will be the $n \times n$ null matrix for this problem, but the more general form is treated. At the final time,

$$\mathbf{x}(t_f) = \boldsymbol{\phi}_{11}(t_f, t_0)\mathbf{x}_0 + \boldsymbol{\phi}_{12}(t_f, t_0)\mathbf{p}(t_0)$$

$$\mathbf{p}(t_f) = \boldsymbol{\phi}_{21}(t_f, t_0)\mathbf{x}_0 + \boldsymbol{\phi}_{22}(t_f, t_0)\mathbf{p}(t_0)$$

Using the boundary condition $\mathbf{p}(t_f) = 2[\mathbf{x}(t_f) - \mathbf{x}_d]$ gives

$$2[\boldsymbol{\phi}_{11}(t_f, t_0)\mathbf{x}_0 + \boldsymbol{\phi}_{12}(t_f, t_0)\mathbf{p}(t_0) - \mathbf{x}_d] = \boldsymbol{\phi}_{21}(t_f, t_0)\mathbf{x}_0 + \boldsymbol{\phi}_{22}(t_f, t_0)\mathbf{p}(t_0)$$

The unknown $\mathbf{p}(t_0)$ can now be found:

$$\mathbf{p}(t_0) = [\boldsymbol{\phi}_{22}(t_f, t_0) - 2\boldsymbol{\phi}_{12}(t_f, t_0)]^{-1}\{[2\boldsymbol{\phi}_{11}(t_f, t_0) - \boldsymbol{\phi}_{21}(t_f, t_0)]\mathbf{x}_0 - 2\mathbf{x}_d\}$$

With $\mathbf{p}(t_0)$ known, $\mathbf{p}(t)$ and $\mathbf{u}^*(t)$ are given by

$$\mathbf{p}(t) = \boldsymbol{\phi}_{21}(t, t_0)\mathbf{x}_0 + \boldsymbol{\phi}_{22}(t, t_0)\mathbf{p}(t_0)$$

$$\mathbf{u}^*(t) = -\tfrac{1}{2}\mathbf{B}^T\mathbf{p}(t)$$

The optimal control law is open loop since $\mathbf{u}^*(t)$ is expressed as a function of $\mathbf{x}_0$ and $\mathbf{x}_d$.

14.16 Convert the control law of Problem 14.15 to a closed-loop control law.

Instead of writing $\begin{bmatrix} \mathbf{x}(t_f) \\ \mathbf{p}(t_f) \end{bmatrix} = \boldsymbol{\Phi}(t_f, t_0)\begin{bmatrix} \mathbf{x}_0 \\ \mathbf{p}(t_0) \end{bmatrix}$, use $\begin{bmatrix} \mathbf{x}(t_f) \\ \mathbf{p}(t_f) \end{bmatrix} = \boldsymbol{\Phi}(t_f, t)\begin{bmatrix} \mathbf{x}(t) \\ \mathbf{p}(t) \end{bmatrix}$, for a general time t. Repeating much of Problem 14.15 but solving for $\mathbf{p}(t)$ instead of $\mathbf{p}(t_0)$ gives

$$\mathbf{p}(t) = [\boldsymbol{\phi}_{22}(t_f, t) - 2\boldsymbol{\phi}_{12}(t_f, t)]^{-1}\{[2\boldsymbol{\phi}_{11}(t_f, t) - \boldsymbol{\phi}_{21}(t_f, t)]\mathbf{x}(t) - 2\mathbf{x}_d\}$$

Once again $\mathbf{u}^*(t) = -\tfrac{1}{2}\mathbf{B}^T\mathbf{p}(t)$. The feedback control law is illustrated in Figure 14.24.

14.17 Use the minimum principle to find the input voltage $u(t)$ which charges the capacitor of Figure 14.25 from x_0 at $t = 0$ to x_d at a fixed t_f while minimizing the energy dissipated in R. There are no restrictions on $u(t)$.

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

$$J = \int_0^{t_f} i^2 R\, dt = \int_0^{t_f} \frac{[u(t) - x(t)]^2}{R}\, dt$$

The Hamiltonian is $\mathcal{H} = (u - x)^2/R + p(-x/RC + u/RC)$. Minimize $\mathcal{H}$ by setting $\partial\mathcal{H}/\partial u = 0$, or $2(u - x)/R + p/RC = 0$. Therefore, $u^*(t) = -p(t)/2C + x(t)$.

The costate equation is $\dot{p} = -\partial\mathcal{H}/\partial x = 2(u - x)/R + p/RC$.

Eliminating $u^*(t)$, and simplifying, the two-point boundary value problem is $\dot{x} =$

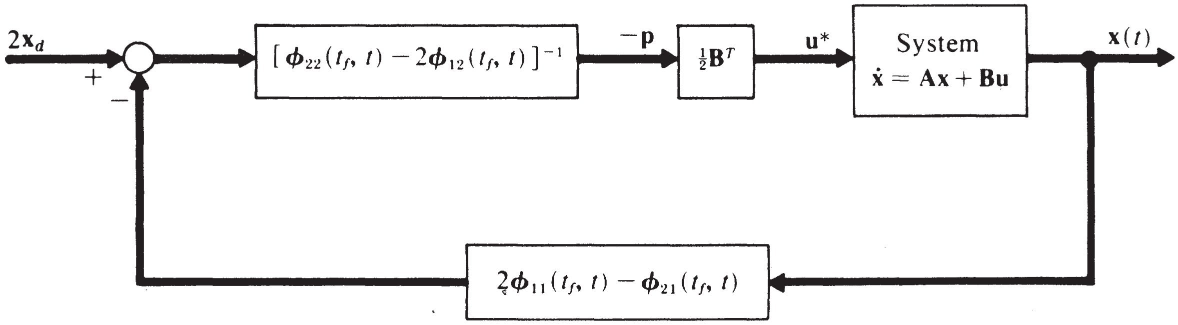


Figure 14.24

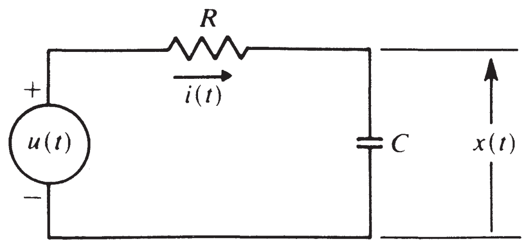


Figure 14.25

$-p/(2RC^2), \dot{p} = 0$, with $x(0) = x_0, x(t_f) = x_d$. This implies that $p(t) = \text{constant}, \alpha$. Therefore, $x(t) = x_0 - \alpha t/(2RC^2)$. From the terminal boundary condition, $\alpha = -2RC^2(x_d - x_0)/t_f$. Therefore, $u^*(t) = RC(x_d - x_0)/t_f + x(t)$. But $x(t) = x_0(1 - t/t_f) + x_d t/t_f$, so $u^*(t) = x_0 + (x_d - x_0)(RC + t)/t_f$.

14.18 A spin-stabilized satellite is wobbling slightly, with components of angular velocity due to the wobble being $x_1(t)$ and $x_2(t)$. The state equations are (see Problems 3.16, and 9.6).

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & -\Omega \\ \Omega & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} u_1(t) \\ u_2(t) \end{bmatrix}$$

Find the control $\mathbf{u}^*(t)$ which drives $\mathbf{x}(t_f)$ to $\mathbf{0}$ at a fixed t_f while minimizing the control energy $J = \int_0^{t_f} \mathbf{u}^T \mathbf{u}\, dt$. There are no restrictions on $\mathbf{u}(t)$.

The Hamiltonian is $\mathcal{H} = \mathbf{u}^T\mathbf{u} + \mathbf{p}^T[\mathbf{A}\mathbf{x} + \mathbf{u}]$. It is minimized by $\mathbf{u}^*(t) = -\frac{1}{2}\mathbf{p}(t)$, where the costate vector satisfies $\dot{\mathbf{p}} = -\partial\mathcal{H}/\partial\mathbf{x} = -\mathbf{A}^T\mathbf{p}$. Note that since $\mathbf{A} = \begin{bmatrix} 0 & -\Omega \\ \Omega & 0 \end{bmatrix}$, we have $-\mathbf{A}^T = \mathbf{A}$.

The two-point boundary value problem is

$$\begin{bmatrix} \dot{\mathbf{x}} \\ \dot{\mathbf{p}} \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\frac{1}{2}\mathbf{I}_2 \\ \mathbf{0} & \mathbf{A} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \mathbf{p} \end{bmatrix} \qquad \text{with } \mathbf{x}(0) = \mathbf{x}_0, \mathbf{x}(t_f) = \mathbf{0}$$

From Problem 8.17, the 4×4 transition matrix for this system is

$$\mathbf{\Phi}(t, 0) = \begin{bmatrix} e^{\mathbf{A}t} & -\dfrac{t}{2}e^{\mathbf{A}t} \\ \mathbf{0} & e^{\mathbf{A}t} \end{bmatrix}, \qquad \text{where } e^{\mathbf{A}t} = \begin{bmatrix} \cos\Omega t & -\sin\Omega t \\ \sin\Omega t & \cos\Omega t \end{bmatrix}$$

Using the terminal boundary condition gives $\mathbf{0} = e^{\mathbf{A}t_f}\mathbf{x}_0 - (t_f/2)e^{\mathbf{A}t_f}\mathbf{p}(0)$, from which $\mathbf{p}(0) = (2/t_f)\mathbf{x}_0$. Using $\mathbf{p}(0)$ gives $\mathbf{p}(t) = e^{\mathbf{A}t}\mathbf{p}(0)$, so $\mathbf{u}^*(t) = -(1/t_f)e^{\mathbf{A}t}\mathbf{x}_0$. When this control is used, the state satisfies $\mathbf{x}^*(t) = (1 - t/t_f)e^{\mathbf{A}t}\mathbf{x}_0$ and the minimum cost is

$$J = \frac{1}{t_f^2}\int_0^{t_f} \mathbf{x}_0^T[e^{\mathbf{A}t}]^T e^{\mathbf{A}t}\mathbf{x}_0\, dt = \mathbf{x}_0^T\mathbf{x}_0/t_f \qquad \text{since } [e^{\mathbf{A}t}]^T = [e^{\mathbf{A}t}]^{-1}$$

14.19 Consider a more general version of Example 14.11. Let the position of the automobile be $x_1(t)$ and the velocity be $x_2(t)$. The admissible controls must satisfy $|u(t)| \le 1$ for all t. Find the $u^*(t)$ which drives the position and velocity to zero simultaneously, in mimimum time.

The system equations are $\dot{x}_1 = x_2, \dot{x}_2 = u$. The Hamiltonian is $\mathcal{H} = 1 + p_1\dot{x}_1 + p_2\dot{x}_2 = 1 + p_1 x_2 + p_2 u$. $\mathcal{H}$ is minimized by selecting $\mathbf{u}^*(t) = -\text{sign}[p_2(t)]$.

The equations for $\mathbf{p}(t)$ are $\dot{p}_1 = -\partial\mathcal{H}/\partial x_1 = 0$ and $\dot{p}_2 = -\partial\mathcal{H}/\partial x_2 = -p_1$. Therefore, $p_1(t) =$ constant, $p_1(0)$, and $p_2(t) = p_2(0) - p_1(0)t$. This indicates that $p_2(t)$ is a linear function of t and changes sign at most once (ignoring the exceptional case where $p_1(0) = p_2(0) = 0$). Therefore, $u^*(t)$ changes sign at most once.

Rather than attempting to determine $p_1(0)$ and $p_2(0)$, the behavior of the system is investigated for $u(t) = +1$ and $u(t) = -1$. With $u = +1$, $x_2(t) = x_2(0) + t$ and $x_1(t) = x_1(0) + x_2(0)t + t^2/2$, or $x_1(t) = \frac{1}{2}x_2^2(t) + x_1(0) - \frac{1}{2}x_2^2(0)$. In the x_1x_2 plane, this represents a family of parabolas open toward the positive x_1 axis. Similarly with $u = -1$, $x_1(t) = -\frac{1}{2}x_2^2(t) + x_1(0) + \frac{1}{2}x_2^2(0)$. This is a family of parabolas open toward the negative x_1 axis. Just one member of each family passes through the specified terminal point $x_1 = x_2 = 0$.

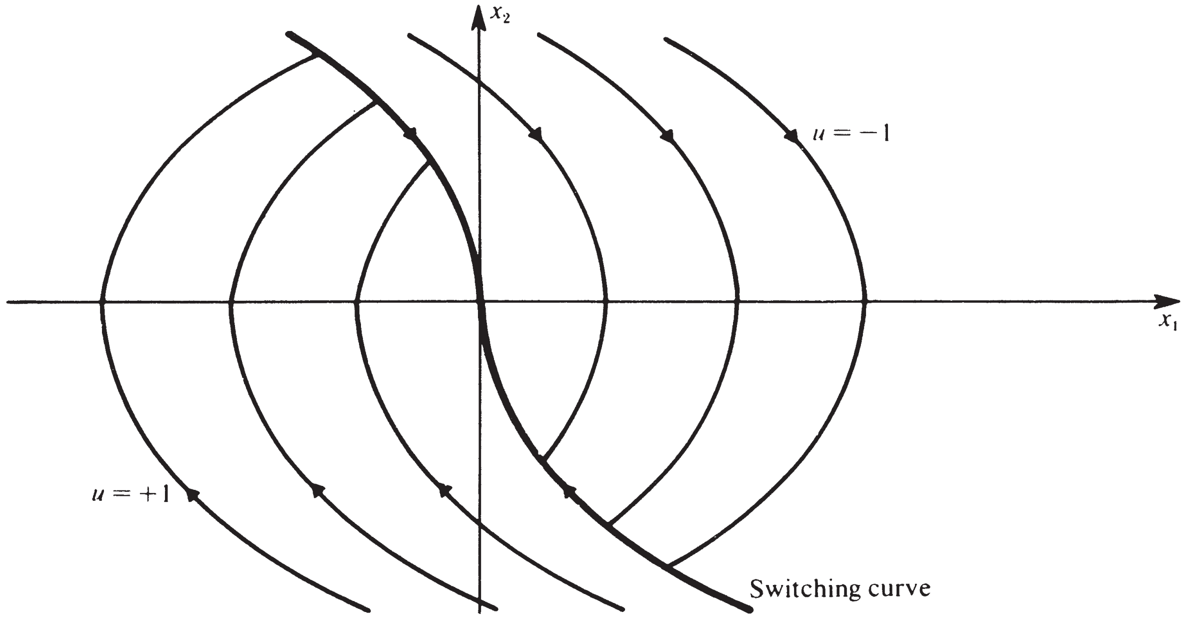


Figure 14.26

Figure 14.26 shows the portion of these two families that are of interest. Segments of the two parabolas through the origin form the switching curve. For any initial state above this curve, $u = -1$ is used until the state intersects the switching curve. Then $u = +1$ is used to reach the origin. For initial states below the switching curve, $u = +1$ is used first, then $u = -1$.

This problem is an example of bang-bang control. It illustrates that the fastest method of coming to a red light and stopping is to use maximum acceleration until the last possible moment, and then use full braking to stop (hopefully) at the intersection. This example also illustrates the remarks in Sec. 14.1. This control is certainly not optimal in terms of tire wear or the number of traffic tickets received.

PROBLEMS

Dynamic Programming

14.20 Use dynamic programming to find the path which moves left to right from point a to point z of Figure 14.27 while minimizing the sum of the costs on each path traveled.

14.21 A student has four hours available to study for four exams. He will earn the scores shown in Table 14.2 for various study times. Use dynamic programming to find the optimal allocation of time in order to maximize the sum of his four scores. Consider only integer numbers of hours.

14.22 Find the control sequence which minimizes $J = [x(2) + 2]^2 + \sum_{k=0}^{1} u(k)^2$ for the scalar system $x(k+1) = \frac{1}{2}x(k) + u(k), x(0) = 10$, no restrictions on $u(k)$. Also find the resulting sequence $x(k)$ and the minimum value of J.

14.23 The same system and initial conditions of Problem 14.22 are considered. Find the optimal control and state sequences which minimize $J = \sum_{k=0}^{4} u(k)^2$ and which give $x(5) = 0$.

14.24 A scalar system is described by $x(k+1) = \frac{1}{2}x(k) + 2u(k)$, with $x(0) = x_0$. Find $x^*(k)$ and $u^*(k)$ which minimize $J = \sum_{k=0}^{2} \{x^2(k) + u^2(k-1)\}$.

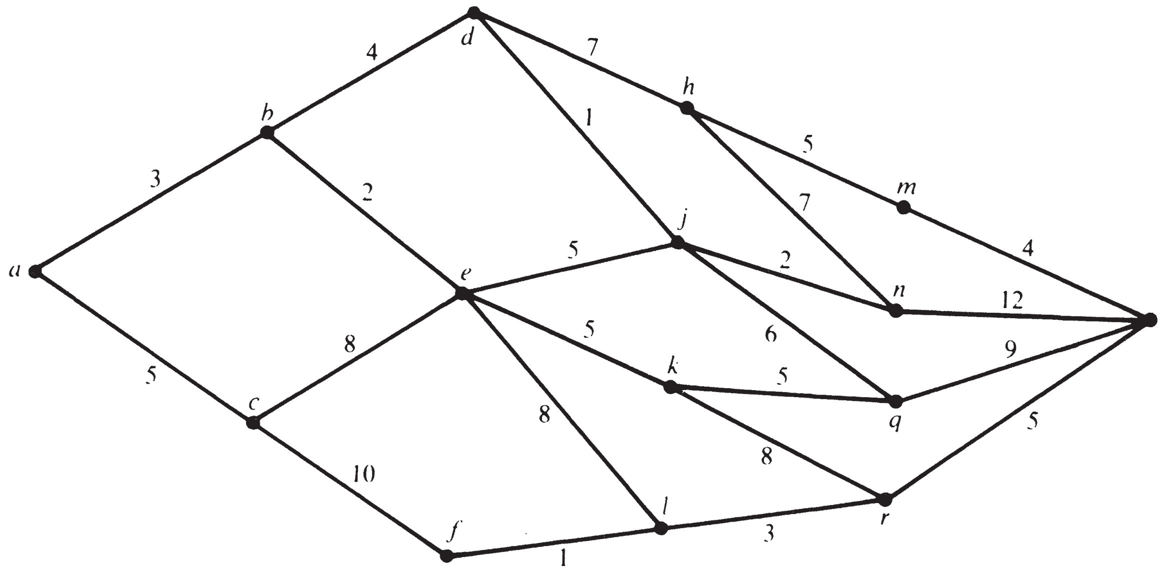


Figure 14.27

14.25 An unstable discrete-time system is described by

$$\mathbf{x}(k+1) = \begin{bmatrix} 1 & 2 \\ -2 & -3 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 1 \\ 0 \end{bmatrix} u(k) \qquad (1)$$

$$y(k) = [0 \quad 1]\mathbf{x}(k) \qquad (2)$$

Find the steady-state feedback gain matrix for the optimal regulator problem, assuming all states are available for use. Also determine the closed-loop eigenvalues that result.

Use the following weights

(a) $\mathbf{Q} = \mathbf{I}, \mathbf{R} = 1$

(b) $\mathbf{Q} = \mathbf{I}, \mathbf{R} = 10$

(c) $\mathbf{Q} = 10\mathbf{I}, \mathbf{R} = 1$

(d) $\mathbf{Q} = 1000\mathbf{I}, \mathbf{R} = 1$

14.26 The system of Problem 14.25 is now assumed to have additive white noises $\mathbf{w}(k)$ on equation (*1*) and $\mathbf{v}(k)$ on equation (*2*). The noise covariance matrices are

$$\mathbf{Q} = \begin{bmatrix} 9 & 0 \\ 0 & 4 \end{bmatrix} \quad \text{and} \quad \mathbf{R} = 16, \text{ respectively}$$

TABLE 14-2 TEST SCORES

Study Hours	Course No.			
	1	2	3	4
0	20	40	40	80
1	45	45	52	91
2	65	57	62	95
3	75	61	71	97
4	83	69	78	98

(a) A Kalman filter (recursive least-squares estimator) is to be used rather than a deterministic observer. Find the steady-state Kalman gain K and the pole locations of the filter's dynamics (i.e., equivalent to the observer poles).

(b) The above filter is used along with the controllers found in Problem 14.25 as indicated by the separation theorem. How do the presence of the filter and the use of estimated states rather than actual states affect the overall system performance in each case? Remember that the separation theorem guarantees that this approach is the best that can be done. But it does not guarantee that the results will be what the controller asked for or that the system will even work well.

14.27 Repeat Problem 14.26 if the measurement noise covariance is reduced to $\mathbf{R} = 0.1$.

14.28 Find the optimal control sequence and the resulting state sequence for the system

$$\mathbf{x}(k+1) = \begin{bmatrix} 1.0 & 0.0 \\ -0.5 & 0.5 \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} 1 \\ -1 \end{bmatrix} u(k)$$

with initial conditions $\mathbf{x}(0) = [4 \quad -2]^T$. The optimal control is to minimize the cost function J defined in Example 14.5, with final time $N = 5$ and

$$\mathbf{M} = \text{Diag}[1000, 1000], \qquad \mathbf{Q} = \text{Diag}[10, 10], \qquad \mathbf{R} = 1, \qquad \mathbf{x}_d = [8 \quad -7]^T,$$

$$\boldsymbol{\eta} = [-1 \quad 1]^T$$

Note the conflict being requested between $\mathbf{x}_d$ and $\boldsymbol{\eta}$. In general, fixed terminal conditions $\mathbf{x}(N) = \mathbf{x}_d$ can be closely approximated by using a sufficiently large $\mathbf{M}$ final weighting matrix.

14.29 Use the discrete-time approximate model of the airplane in Example 13.6 and design a steady-state optimal regulator with

$$\mathbf{Q} = \text{Diag}[0.1, 1.0, 1.0, 0.1] \quad \text{and} \quad \mathbf{R} = \mathbf{I}$$

14.30 A commonly used simplification of the previous system is that $\delta_a = 4\delta_r$. This allows the simpler single-input analysis to be carried out, using as the input matrix $\mathbf{B}$ four times column 1 plus column 2 of the original $\mathbf{B}$ matrix. Using this simplification, find the feedback gains for generating the equivalent control δ_r if

(a) $\mathbf{Q} = \mathbf{I}, \mathbf{R} = 1$

(b) $\mathbf{Q} = \mathbf{I}, \mathbf{R} = 0.1$

(c) $\mathbf{Q} = \mathbf{I}, \mathbf{R} = 25$

(d) $\mathbf{Q} = \text{Diag}[0.1, 1.0, 1.0, 0.1], \mathbf{R} = 1$

14.31 A linear, constant continuous-time system is described by $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}, \mathbf{B} = \mathbf{I}$. Solve the algebraic Riccati equation to find $\mathbf{W}_\infty$ if $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = r\mathbf{I}$, for five cases: $r = 50, 10, 4, 1$, and 0.2. Also determine the constant full feedback gain matrices and the resulting closed-loop eigenvalues.

14.32 Solve the Lyapunov equation $\mathbf{A}^T\mathbf{W} + \mathbf{W}\mathbf{A} = -\mathbf{Q}$, using $\mathbf{Q}$ and $\mathbf{A}$ from the previous problem. Compare the result with the "large r" case $r = 50$.

14.33 A system with $\mathbf{A}$ of Problem 14.31 and $\mathbf{B} = [-1 \quad 1]^T$ is uncontrollable, but since the eigenvalues of $\mathbf{A}$ are $\lambda = -2, -1$, the system is stabilizable. Solve the ARE for $\mathbf{W}_\infty$ and then find $\mathbf{G}_\infty$ and the closed-loop eigenvalues. Use $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = 1$.

14.34 An unstable but controllable continuous-time system has $\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 2 & -3 \end{bmatrix}, \mathbf{B} = \mathbf{I}$. Let $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = r\mathbf{I}$. Find the feedback gain matrix $\mathbf{G}_\infty$ for $r = 50$ and $r = 1000$. Also find the resulting closed-loop eigenvalues. What happens when the Lyapunov equation is used to approximate the ARE for large r in this unstable case?

14.35 For the scalar system $\dot{x} = x + u$, the ARE becomes a simple scalar quadratic. Let $Q = 1$ and

$R = \frac{1}{3}$, and find both roots of this quadratic. Then form the 2×2 Hamiltonian matrix $\mathbf{H}$ and use its eigenvectors to find both $\mathbf{W}_\infty$ and $\mathbf{W}_{ss}$ defined in Sec. 14.4.2. Compare the results.

14.36 Combine the pole placement controller of Example 13.1 and the observer of Example 13.10 into a composite fourth-order system. Find the Kalman observable canonical form to show that the observer modes are not controllable. Also verify that the composite system is observable.

The Minimum Principle

14.37 Use the minimum principle to find the optimal control $u^*(t)$ and the corresponding $x^*(t)$ for the system $\dot{x} = u(t)$ with $x(0) = 0, x(1) = 1$ and no restrictions on $u(t)$. The performance criterion is $J = \int_0^1 (x^2 + u^2)\, dt$.

14.38 Use the minimum principle to solve the optimization problem: $\dot{x}_1 = x_2$, $\dot{x}_2 = u$, $x_1(0) = 1$, $x_2(0) = 1$, $x_1(2) = 0$, $x_2(2) = 0$, minimize $J = \frac{1}{2}\int_0^2 u^2(t)\, dt$ with $u(t)$ unrestricted.

14.39 Solve the following two-point boundary value problem on the interval $0 \le t \le 1$: $\dot{x} = -2x - \frac{1}{2}p$, $\dot{p} = -2x + 2p$, $x(0) = 10$, $p(1) = 4x(1)$.

14.40 The equations of motion for a rocket flying in a vertical plane under the influence of constant gravity g and constant thrust T can be written as $\dot{x}_1 = x_3$, $\dot{x}_2 = x_4$, $\dot{x}_3 = (T/m)\cos u(t)$, $\dot{x}_4 = (T/m)\sin u(t) - g$, where x_1 and x_2 are horizontal and vertical position components, x_3 and x_4 are horizontal and vertical velocity components, and the control $u(t)$ is the thrust angle measured from the horizontal. The rocket is to be flown to a specified terminal altitude with zero vertical velocity at t_f and maximum horizontal velocity. Use the minimum principle to establish that the optimal thrust angle follows the linear-tangent steering law.

$$\tan u^*(t) = \alpha t - \beta \qquad \text{where } \alpha \text{ and } \beta \text{ are constants}$$

(*Hint:* Minimize $J = -x_3(t_f)$.)

14.41 Consider a system which is nonlinear with respect to $\mathbf{x}$, but linear with respect to $\mathbf{u}$, i.e., $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t) + \mathbf{B}\mathbf{u}$. In each of the following cases, use the minimum principle to find the form of $\mathbf{u}^*(t)$ in terms of $\mathbf{p}(t)$.

(a) $J = \int_{t_0}^{t_f} \mathbf{u}^T \mathbf{u}\, dt$, $\mathbf{x}(t_f) = \mathbf{0}$, t_f fixed, $\mathbf{u}^T(t)\mathbf{u}(t) \le 1$ for all t.

(b) Drive $\mathbf{x}(t_f)$ to zero in minimum time, $\mathbf{u}^T(t)\mathbf{u}(t) \le 1$ for all t.

(c) Same as b except each component of $\mathbf{u}$ satisfies $|u_i(t)| \le 1$ for all t.

(d) $J = \int_{t_0}^{t_f} \Sigma |u_i(t)|\, dt$, with $|u_i(t)| \le 1$ for all t.

14.42 A scalar system is described by $\dot{x} = x + u$ with $x(0) = 2$. The admissible controls must satisfy $|u(t)| \le 1$. Use the minimum principle to find $u^*(t)$ which drives $x(t)$ to zero in minimum time.

14.43 Analyze an equivalent single-input model of the aircraft in Problem 14.11 by assuming that $\delta_a = 4\delta_r$. Use $\mathbf{Q} = \text{Diag}[0.1, 1.0, 1.0, 0.1]$ and $\mathbf{R} = 1$.

15

An Introduction to Nonlinear Control Systems

15.1 INTRODUCTION

Most of the text so far has dealt with linear systems. Yet when an engineer is faced with a "real problem," he or she invariably bumps into nonlinearities. Some typical examples are as follows:

1. In positioning a robotic device or in pointing a sensor at a target, the geometry of coordinate transformations comes into play. Sines and cosines are nonlinear functions of their arguments.
2. An actuating motor has inherent current—and hence torque—limitations. Saturation of the control commands is a common nonlinearity. Backlash, hysteresis, and dead zone are other commonly encountered nonlinearities.
3. Many important physical processes are described by nonlinear models. Drag on a moving vehicle is proportional to velocity squared, for example. The voltage-current characteristics of most electronic devices are nonlinear. Coulomb friction is of constant magnitude and always opposes motion, unlike the linear viscous friction model often assumed. Gravitational and electrostatic attraction are inversely proportional to distance squared.
4. Deliberate nonlinearities may be introduced by the control system. On-off relay controllers are common examples.

What are the implications of nonlinearities on the control system analysis and design techniques which have been presented thus far? Many aspects of former chapters are still applicable:

1. The physical modeling techniques using linear graphs apply to nonlinear systems. The individual elemental equations may be nonlinear, but the continuity and compatibility laws apply as before (Chapter 1).

2. The general form of the state equations

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \quad \text{and} \quad \mathbf{y} = \mathbf{h}(\mathbf{x}, \mathbf{u}, t)$$

are still valid (Chapter 3).
3. Knowledge of matrix theory and linear algebra is still essential. A rigorous treatment of nonlinear systems would require much supplemental mathematics, however [1, 2].
4. The Lyapunov stability theory applies to nonlinear systems (Chapter 10).
5. The general formulation of the optimal control problems, using either dynamic programming or the minimum principle, is still valid (Chapter 14).

Many of the linear system results do not apply to nonlinear systems:

1. Superposition does not apply. Knowledge of the system response to initial conditions or to individual inputs does not allow prediction of the total response due to initial conditions plus several simultaneous inputs.
2. Homogeneity does not apply. The response to an input $\alpha\mathbf{u}(t)$ is not just α times the response to $\mathbf{u}(t)$. The response to $\beta\mathbf{x}(t_0)$ is not just β times the response to $\mathbf{x}(t_0)$. The whole concept of designing control systems based on typical test inputs (unit steps, sinusoids, and so on) and then predicting behavior to an actual input by scaling and superposition is generally invalid.
3. The nice correlation between transfer function pole and zero locations and time response behavior is generally invalid.
4. Stability of a system is no longer just a simple function of eigenvalue locations. In fact, it is not proper to speak about stability of a nonlinear system. Rather, the stability of equilibrium points must be investigated, and nonlinear systems may have multiple equilibrium points, some stable and others not.
5. An unforced nonlinear system can possess limit cycles and other behavior not predicted by linear theory.
6. A periodically excited nonlinear system is not restricted to yielding steady-state outputs of the same frequency as the input. Higher harmonics, subharmonics, and even continuous spectra (chaos) can occur in the output of a nonlinear system.

   Jump resonance, beat phenomenon, and other behaviors not predicted by linear theory can occur. Although these are of general interest in nonlinear system dynamics, the design of control systems is usually directed toward the avoidance of such behaviors.
7. Many linear system properties were derived based on full knowledge of the closed-form solution to the linear state equations. In the nonlinear case no known analytical solutions are available except in very rare exceptional cases. In fact the whole question of global existence and uniqueness of solutions to nonlinear differential equations cannot be taken for granted in general [3].
8. Properties such as controllability and observability can no longer be tested for on a global basis by using simple rank tests.

The goal of this chapter is to provide some insight into nonlinear system behavior and to give the engineer some useful tools for attacking certain classes of nonlinear problems. A complete treatment of nonlinear systems and control is not possible. Consideration is restricted to certain classes of nonlinear systems and certain kinds of problems:

1. Many systems with mild, sufficiently smooth nonlinearities can be treated by using a linear approximate model, obtained by linearizing about a known nominal solution or operating point. Most electronic circuit design is based on behavior in the vicinity of an "operating point." Many aircraft, rocket, and spacecraft control systems have been successfully designed using linear behavior in the neighborhood of a nominal trajectory.
2. Some nonlinear systems can be linearized by using a nonlinear state variable feedback controller. The principal advantage of this approach is that known techniques and results for linear systems analysis can then be applied.
3. An introduction to describing functions is presented. This constitutes a useful, albeit an approximate, approach for dealing with many of the unavoidable nonlinearities encountered in real system design.
4. Lyapunov stability theory can be used to analyze systems and even to design controllers of certain types. Application of these very general methods is often limited by the difficulties in finding suitable Lyapunov functions. By restricting attention to systems which are linear except for one nonlinear element, some easy-to-apply results such as the Popov criterion and the circle criterion can be derived. Stability is always a major concern in control system design. It is imperative that the effect of nonlinearities on system stability can be evaluated. The assumptions made in deriving these results are general enough to fit many engineering applications.

Phase-plane representations are used at various times because of the insight they provide. Although general phase-plane analysis techniques exist, they are fully effective only for second-order systems and are not pursued in detail here.

Because of the lack of analytical solutions for nonlinear system equations, simulation takes on much greater significance. Many numerical integration schemes will diverge if the integration step size is too large for the frequency of the signal being integrated. The major danger of applying fixed step-size integration schemes to nonlinear systems whose response frequencies may not be known in advance is that a numerical algorithm instability may be falsely interpreted as a system instability. One commonsense approach suggests that when a simulated response seems to be diverging, the simulation should be retried with a much smaller integration step size.

15.2 LINEARIZATION: ANALYSIS OF SMALL DEVIATIONS FROM NOMINAL

Consider the general nonlinear state variable model of Eqs. (*3.6*) and (*3.7*):

$$\begin{aligned}\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \\ \mathbf{y} &= \mathbf{h}(\mathbf{x}, \mathbf{u}, t)\end{aligned} \tag{15.1}$$

Suppose a nominal solution $\mathbf{x}_n(t)$, $\mathbf{u}_n(t)$, and $\mathbf{y}_n(t)$ is known. The difference between these nominal vector functions and some slightly perturbed functions $\mathbf{x}(t)$, $\mathbf{u}(t)$, and $\mathbf{y}(t)$ can be defined by

$$\delta\mathbf{x} = \mathbf{x}(t) - \mathbf{x}_n(t)$$

$$\delta\mathbf{u} = \mathbf{u}(t) - \mathbf{u}_n(t)$$

$$\delta\mathbf{y} = \mathbf{y}(t) - \mathbf{y}_n(t)$$

Then Eq. (*15.1*) can be written as

$$\begin{aligned}\dot{\mathbf{x}}_n + \delta\dot{\mathbf{x}} &= \mathbf{f}(\mathbf{x}_n + \delta\mathbf{x}, \mathbf{u}_n + \delta\mathbf{u}, t)\\ &= \mathbf{f}(\mathbf{x}_n, \mathbf{u}_n, t) + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_n \delta\mathbf{x} + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right]_n \delta\mathbf{u} + \text{higher-order terms}\end{aligned}$$

$$\begin{aligned}\mathbf{y}_n + \delta\mathbf{y} &= \mathbf{h}(x_n + \delta\mathbf{x}, \mathbf{u}_n + \delta\mathbf{u}, t)\\ &= \mathbf{h}(\mathbf{x}_n, \mathbf{u}_n, t) + \left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]_n \delta\mathbf{x} + \left[\frac{\partial \mathbf{h}}{\partial \mathbf{u}}\right]_n \delta\mathbf{u} + \text{higher-order terms}\end{aligned}$$

where $[\ ]_n$ means the derivatives are evaluated on the nominal solutions. Since the nominal solutions satisfy Eq. (*15.1*), the first terms in the preceding Taylor series expansions cancel. For sufficiently small $\delta\mathbf{x}$, $\delta\mathbf{u}$, and $\delta\mathbf{y}$ perturbations, the higher-order terms can be neglected, leaving the linear equations

$$\begin{aligned}\delta\dot{\mathbf{x}} &= \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_n \delta\mathbf{x} + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{u}}\right]_n \delta\mathbf{u}\\ \delta\mathbf{y} &= \left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]_n \delta\mathbf{x} + \left[\frac{\partial \mathbf{h}}{\partial \mathbf{u}}\right]_n \delta\mathbf{u}\end{aligned} \tag{15.2}$$

If $\mathbf{x}_n(t) = \mathbf{x}_e = \text{constant}$ and if $\mathbf{u}_n(t) = 0 = \delta\mathbf{u}(t)$, then the stability of the equilibrium point $\mathbf{x}_e$ is governed by

$$\delta\dot{\mathbf{x}} = \left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_n \delta\mathbf{x} \tag{15.3}$$

For this case, the Jacobian matrix $[\partial\mathbf{f}/\partial\mathbf{x}]$ is constant and its eigenvalues determine system stability in the neighborhood of $\mathbf{x}_e$, in the following sense. If all λ_i have negative real parts, the equilibrium point is asymptotically stable for sufficiently small perturbations. If one or more eigenvalues have positive real parts, the equilibrium point is unstable. If one or more of the eigenvalues are on the $j\omega$ axis and all others are in the left-half plane, no conclusion about stability can be drawn from this linear model. Whether the actual behavior of the system is divergent or convergent will depend upon the neglected higher-order terms in the Taylor series expansion. Thus, except for the borderline $j\omega$ axis case, stability of the nonlinear Eq. (*15.1*) is the same as the linearized model Eq. (*15.2*), at least in a small neighborhood of the equilibrium point. Problem 15.1 proves these results and gives precise conditions under which they apply.

EXAMPLE 15.1 Find the equilibrium points for the system described by

$$\ddot{y} + (1 + y)\dot{y} - 2y + 0.5y^3 = 0$$

Then evaluate the linearized Jacobian matrix at each equilibrium point and determine the stability characteristics from the eigenvalues.

Letting $x_1 = y$ and $x_2 = \dot{y}$ gives the state variable model

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ 2x_1 - 0.5x_1^3 - (1 + x_1)x_2 \end{bmatrix} = \mathbf{f}(\mathbf{x})$$

Equilibrium points are solutions of $\mathbf{f}(\mathbf{x}) = \mathbf{0}$, so each must have $x_2 = 0$ and $2x_1 - 0.5x_1^3 = 0$. The three solutions are $x_{e1} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $x_{e2} = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$ and $x_{e3} = \begin{bmatrix} -2 \\ 0 \end{bmatrix}$. The Jacobian matrix is

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 2 - \dfrac{3x_1^2}{2} - x_2 & -(1 + x_1) \end{bmatrix}$$

so that

$$\left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_1 = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$$

Its eigenvalues are at $+1$ and -2, so this is a saddle point (see Figure 10.3).

$$\left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_2 = \begin{bmatrix} 0 & 1 \\ -4 & -3 \end{bmatrix}$$

Its eigenvalues are at $-\frac{3}{2} \pm j\sqrt{7}/2$, so this point is a stable focus (see Figure 10.4).

$$\left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_3 = \begin{bmatrix} 0 & 1 \\ -4 & 1 \end{bmatrix}$$

Its eigenvalues are at $\frac{1}{2} \pm j\sqrt{15}/2$, so this point is an unstable focus (see Figure 10.4). If this system has initial conditions exactly at any one of the three equilibrium points, the state will remain there indefinitely in the absence of disturbances. For any other initial condition the state will eventually settle to $\mathbf{x} = [2 \quad 0]^T$. Figure 15.1 shows one representative phase plot which begins near the unstable focus, with $\mathbf{x}(0) = [-2, 0.1]^T$ and settles at the stable focus. ■

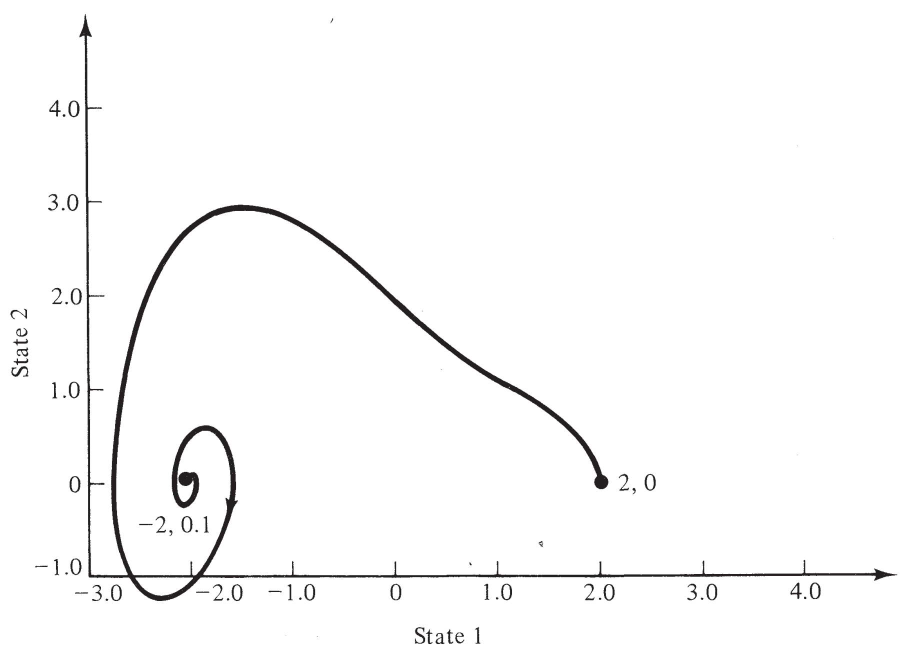


Figure 15.1

If a time-varying nominal solution $\{\mathbf{x}_n(t), \mathbf{u}_n(t), \mathbf{y}_n(t)\}$ is used (perhaps as obtained from numerical solution of Eq. (*15.1*)), then the Jacobian matrices of Eq. (*15.2*) will also be time-varying in general. As was pointed out in Sec. 10.7 the stability of linear time-varying systems is not as straightforward as the linear, constant case.

With $\delta\mathbf{u}(t)$ restricted to zero, Eq. (*15.2*) can be used to investigate the passive behavior of perturbed trajectories. It is of interest to know whether a trajectory $\mathbf{x}(t)$ will passively return to $\mathbf{x}_n(t)$ (i.e., asymptotic stability) or will remain within some bounded neighborhood of it (i.e., stability i.s.L.) or will diverge from it (i.e., unstable). These types of analysis must always be used with caution because of the assumptions made regarding $\delta\mathbf{x}(t)$ remaining small.

The input perturbation $\delta\mathbf{u}(t)$ can be used to actively control the behavior of $\delta\mathbf{x}(t)$, thus forcing it to return to and remain at or near zero. Thus the linearizing assumption that $\delta\mathbf{x}(t)$ is small can be made somewhat self-fulfilling. A linear feedback control law, $\delta\mathbf{u}(t) = -\mathbf{K}\delta\mathbf{x}(t)$, could be used, and a typical implementation is shown in Figure 15.2. The overall goal is to maintain the trajectory near the known, precomputed nominal in spite of initial condition perturbations or input disturbances. The gain matrix $\mathbf{K}$ in the control law could be computed using pole placement techniques of Sec. 13.4. If the closed-loop poles are forced to be sufficiently stable, then $\delta\mathbf{x}(t)$ will rapidly return to $\mathbf{0}$ after any upset. Alternatively, the gain $\mathbf{K}$ could be found as the result of an optimal regulator design problem, as discussed in Sec. 14.4.

EXAMPLE 15.2 The equations for the orbit-plane motion of a satellite in orbit about a planet with an ideal inverse-square gravity field are

$$\ddot{r} - \dot{\theta}^2 r = -\frac{\mu}{r^2} + a_r$$

$$r\ddot{\theta} + 2\dot{r}\dot{\theta} = a_i$$

where a_r and a_i are the radial and in-track components of any acceleration terms due to thrust, drag, gravitational anomalies, and the like. These will be treated as components of the input vector $\mathbf{u}$. Determine the equations that describe small perturbations about a circular orbit of radius R.

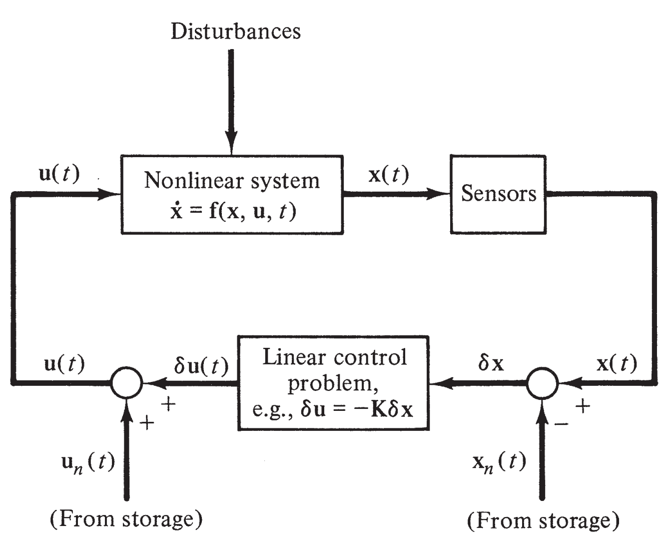


Figure 15.2

Let the state vector be $\mathbf{x} = [r \quad \theta \quad \dot{r} \quad \dot{\theta}]^T$. Then the nonlinear state equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} x_3 \\ x_4 \\ x_4^2 x_1 - \mu/x_1^2 + u_1 \\ -\dfrac{2x_3 x_4}{x_1} + \dfrac{u_2}{x_1} \end{bmatrix} = \mathbf{f}(\mathbf{x}, \mathbf{u})$$

The Jacobian matrices are

$$\frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ x_4^2 + \dfrac{2\mu}{x_1^3} & 0 & 0 & 2x_4 x_1 \\ \dfrac{2x_3 x_4}{x_1^2} - \dfrac{u_2}{x_1^2} & 0 & \dfrac{-2x_4}{x_1} & \dfrac{-2x_3}{x_1} \end{bmatrix}$$

and

$$\frac{\partial \mathbf{f}}{\partial \mathbf{u}} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & \dfrac{1}{x_1} \end{bmatrix}$$

For a circular nominal orbit, $x_{1n} = R$, $x_{3n} = \dot{R} = 0$, and $x_{4n} = \dot{\theta}_n = \omega$. To maintain R constant, $\dot{R}$ and $\ddot{R}$ must be zero, which leads to the relation $\omega = \mu/R^3$. x_{2n} does not appear explicitly in the linearized equations but clearly will increase linearly with time, since its derivative $x_{4n} = \omega$ is constant. The nominal values of both components of $\mathbf{u}$ are zero. Using these results gives the linear perturbation model

$$\delta\dot{\mathbf{x}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 3\omega^2 & 0 & 0 & 2R\omega \\ 0 & 0 & \dfrac{-2\omega}{R} & 0 \end{bmatrix} \delta\mathbf{x} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & \dfrac{1}{R} \end{bmatrix} \delta\mathbf{u}$$

The eigenvalues of $[\partial\mathbf{f}/\partial\mathbf{x}]_n$ can be found to be $\lambda_i = \{0, 0, j\omega, -j\omega\}$. Since they are all on the $j\omega$ axis, the linear model gives inconclusive results regarding stability of the nonlinear system. The linear model is very useful for studying the effect of perturbations away from the nominal circular orbit and remains accurate for sizable changes in altitude, as long as δr is small compared to the (large) nominal R value. Rocket thrusters can be used to actively drive observed perturbations back to zero. In order to predict the future effect of state perturbations, it is useful to know that the transition matrix is

$$\Phi(t, 0) = \begin{bmatrix} 4 - 3\cos\omega t & 0 & \dfrac{\sin\omega t}{\omega} & \dfrac{-2R(\cos\omega t - 1)}{\omega} \\ \dfrac{6(\sin\omega t - \omega t)}{R} & 1 & \dfrac{-2(\cos\omega t - 1)}{R\omega} & \dfrac{4\sin\omega t}{\omega} - 3t \\ 3\omega\sin\omega t & 0 & \cos\omega t & 2R\sin\omega t \\ \dfrac{6\omega(\cos\omega t - 1)}{R} & 0 & \dfrac{-2\sin\omega t}{R} & 4\cos\omega t - 3 \end{bmatrix}$$

The effect of the accelerations $\delta\mathbf{u}$ over a period T is given by $\int_0^T \mathbf{\Phi}(T, \tau)[\partial\mathbf{f}/\partial\mathbf{u}]\delta\mathbf{u}(\tau)\,d\tau$, and for $\delta\mathbf{u}$ constant over the interval $[0, T]$, this can be integrated to obtain

$$\int_0^T \mathbf{\Phi}(T, \tau)\left[\frac{\partial\mathbf{f}}{\partial\mathbf{u}}\right]\delta\mathbf{u}(\tau)\,d\tau = \begin{bmatrix} \dfrac{1-\cos\omega T}{\omega^2} & \dfrac{-2(\sin\omega T - \omega T)}{\omega^2} \\ \dfrac{-2(\sin\omega T - \omega T)}{\omega^2 R} & \dfrac{4(1-\cos\omega T)}{R\omega^2} - \dfrac{3T^2}{2R} \\ \dfrac{\sin\omega T}{\omega} & \dfrac{2(1-\cos\omega T)}{\omega} \\ \dfrac{-2(1-\cos\omega T)}{R\omega} & \dfrac{4\sin\omega T}{R\omega} - \dfrac{3T}{R} \end{bmatrix} \delta\mathbf{u}$$

■

15.3 DYNAMIC LINEARIZATION USING STATE FEEDBACK

In the previous section local linearization of a nonlinear system was investigated. We now consider the problem of synthesizing a control input $\mathbf{u}(t)$, which will cause the system

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u}, t) \tag{15.4}$$

to have a response which matches some specified template system. That is, let $\mathbf{y}(t) = \mathbf{H}\mathbf{x}(t)$. It is desired that $\mathbf{y}(t)$ match as closely as possible the response of the specified template system

$$\dot{\mathbf{y}}_d = \mathbf{g}(\mathbf{y}_d, \mathbf{y}, t) \tag{15.5}$$

In a typical example, the $\mathbf{g}$ function might specify a linear system,

$$\dot{\mathbf{y}}_d = \mathbf{F}\mathbf{y}_d + \mathbf{G}\mathbf{v} \tag{15.6}$$

with $\mathbf{v}(t)$ being perhaps a step function input and with the response possessing certain desirable transient characteristics. If a control input $\mathbf{u}(t)$ can be found to achieve the goal $\dot{y} \equiv \dot{\mathbf{y}}_d$, then the original system will behave as a linear system. This is what we term *dynamic linearization.*

Define the error $\mathbf{e}(t) = \mathbf{H}\mathbf{x}(t) - \mathbf{y}_d(t)$. Then

$$\begin{aligned} \dot{\mathbf{e}}(t) &= \mathbf{H}\dot{\mathbf{x}}(t) - \dot{\mathbf{y}}_d(t) \\ &= \mathbf{H}\mathbf{f}(\mathbf{x}, \mathbf{u}, t) - \mathbf{g}(\mathbf{y}_d, \mathbf{v}, t) \end{aligned} \tag{15.7}$$

Suppose for the moment that $\mathbf{H} = \mathbf{I}$. When $\dot{\mathbf{e}}$ is set to zero, it may be possible to solve the resulting equation for the unknown input $\mathbf{u}$ in terms of known or measurable quantities $\mathbf{x}$, $\mathbf{y}_d$, and $\mathbf{v}$. If this is accomplished, the feedback-modified system will have the same derivative as the template system. If the template system is linear, then the original system will have been linearized. In essence the nonlinearities of the original system are canceled and replaced by the desired linear terms. This form of dynamic linearization has been known for many years [4, p. 560].

EXAMPLE 15.3 [5–7] It is desired that the first-order nonlinear system

$$\dot{x} = x + u + xu$$

behave like the linear system

$$\dot{y}_d = -\sigma y_d \qquad \text{with initial condition } y_d(0) = 10$$

For this scalar system let $y(t) = x(t)$. Setting $\dot{x} = \dot{y}_d$ leads to $u(t) = [-x(t) - \sigma y_d(t)]/[1 + x(t)]$, provided $x(t) \neq -1$. At least two potential problems exist with this scheme. (1) Even if the derivaties can be made to match exactly, the initial condition $x(0)$ may not match $y_d(0)$ for a variety of reasons. The exact initial conditions may not be known due to measurement error, or the desire may be to have the system respond like the template system regardless of its initial $x(0)$ value. Of course, matching derivatives does not mean matching response curves. This will be addressed in the sequel. (2) The resulting control law for $u(t)$ has a singularity at $x(t) = -1$. An infinite amount of control would be required at this point. Truxal [4] pointed out that forcing a nonlinear system to respond like a linear system generally means that components must be overdesigned to allow the avoidance of nonlinear behavior. The singularity of this example is an extreme case of this. ■

The problem of initial condition mismatch, either deliberate or unintentional, can be addressed by adding a convergence factor matrix **S**, as follows. Instead of setting $\dot{\mathbf{e}} = \mathbf{0}$, we require that

$$\dot{\mathbf{e}} = \mathbf{S}\mathbf{e} \tag{15.8}$$

In a similar manner to the development of Chapter 13 for state variable observers, the matrix **S** is specified with asymptotically stable eigenvalues. Then $\mathbf{e}(t) \to \mathbf{0}$, and thus $\mathbf{Hx}(t) \to \mathbf{y}_d(t)$ at a rate controlled by choice of **S**. Note that the previous development is a special case with $\mathbf{S} \equiv [0]$. The equation for finding the control $\mathbf{u}(t)$ is thus

$$\mathbf{Hf}(\mathbf{x}, \mathbf{u}, t) = \mathbf{g}(\mathbf{y}_d, \mathbf{v}, t) + \mathbf{S}[\mathbf{Hx}(t) - \mathbf{y}_d(t)] \tag{15.9}$$

The existence of a solution of Eq. (*15.9*) for $\mathbf{u}(t)$ can be established in certain cases by using the implicit function theorem [8], which establishes sufficient conditions on the function **f**. The solvability is also influenced by the number of independent equations that need to be satisfied relative to the number of unknown control components. This explains the existence of the matrix **H**. It will not generally be possible to match all n components of **x** to an n-dimensional $\mathbf{y}_d$ vector when there are only $r < n$ components in the control vector **u**. General conditions for solvability are addressed in Problems 15.5, 15.6, and 15.7. For any specific problem, a direct attempt to solve for **u** will often be the most expedient method of determining whether or not such a solution can be found, and this is the approach presented in the example problems. Assuming the existence of a solution, the control law will be of the feedback form

$$\mathbf{u}(t) = \mathbf{u}(\mathbf{x}(t), \mathbf{y}_d(t), \mathbf{v}(t), \mathbf{H}, \mathbf{S}) \tag{15.10}$$

and the procedure is obviously a model-matching or model-tracking scheme, as shown in Figure 15.3.

Note that when a linear system is used as the template, then Eq. (*15.9*) becomes

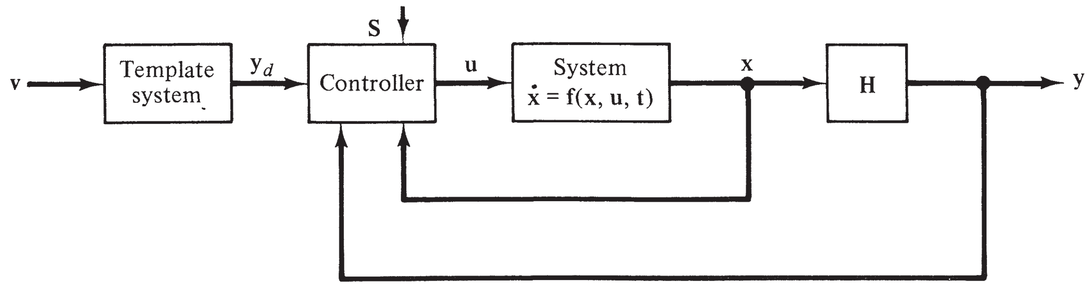


Figure 15.3

$$\begin{aligned}\mathbf{Hf}(\mathbf{x}, \mathbf{u}, t) &= \mathbf{F}\mathbf{y}_d + \mathbf{G}\mathbf{v} + \mathbf{S}[\mathbf{H}\mathbf{x} - \mathbf{y}_d] \\ &= [\mathbf{F} - \mathbf{S}]\mathbf{y}_d + \mathbf{S}\mathbf{H}\mathbf{x} + \mathbf{G}\mathbf{v}\end{aligned} \tag{15.11}$$

If the convergence matrix **S** is selected equal to **F**, then $\mathbf{y}_d$ is not directly required, and the need to synthesize the template system is removed. If $\mathbf{S} = [\mathbf{0}]$ is selected, a major feedback path is eliminated. Figure 15.4a shows the general form, with **S** presumably being selected to be somewhat faster than **F**—i.e., eigenvalues more negative. Figure 15.4b shows the result when $\mathbf{S} = \mathbf{F}$, and Figure 15.4*c* shows the configuration with $\mathbf{S} = [\mathbf{0}]$. In all cases the "solve" box refers to finding the input **u** which satisfies Eq. *(15.11)*.

EXAMPLE 15.4 The scalar system of Example 15.3 is reconsidered, but now the convergence factor S is included so that for all initial conditions, $x(0)$, $x(t)$ will ultimately approach the desired response $y_d(t)$. Set $\dot{e} = \dot{x} - \dot{y}_d = Se$, where S is a negative real number. Then

$$x + u + xu + \sigma y_d = S(x - y_d)$$

from which, if $x(t) \neq -1$,

$$u(t) = \frac{S[x(t) - y_d(t)] - x(t) - \sigma y_d}{1 + x(t)}$$

Substituting this back into the system equations gives the coupled pair

$$\begin{bmatrix}\dot{x} \\ \dot{y}_d\end{bmatrix} = \begin{bmatrix}S & -(\sigma + S) \\ 0 & -\sigma\end{bmatrix}\begin{bmatrix}x \\ y_d\end{bmatrix}$$

The 2×2 transition matrix is easily found using methods of Chapter 8:

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

Then the system response is

$$x(t) = e^{-\sigma t} y_d(0) + [x(0) - y_d(0)]e^{St}$$

This is the desired template response, $e^{-\sigma t} y_d(0)$, plus an initial condition mismatch term, which dies out at a rate determined by the convergence factor S. If $S << -\sigma$, this term will quickly die out, leaving only the desired response. ■

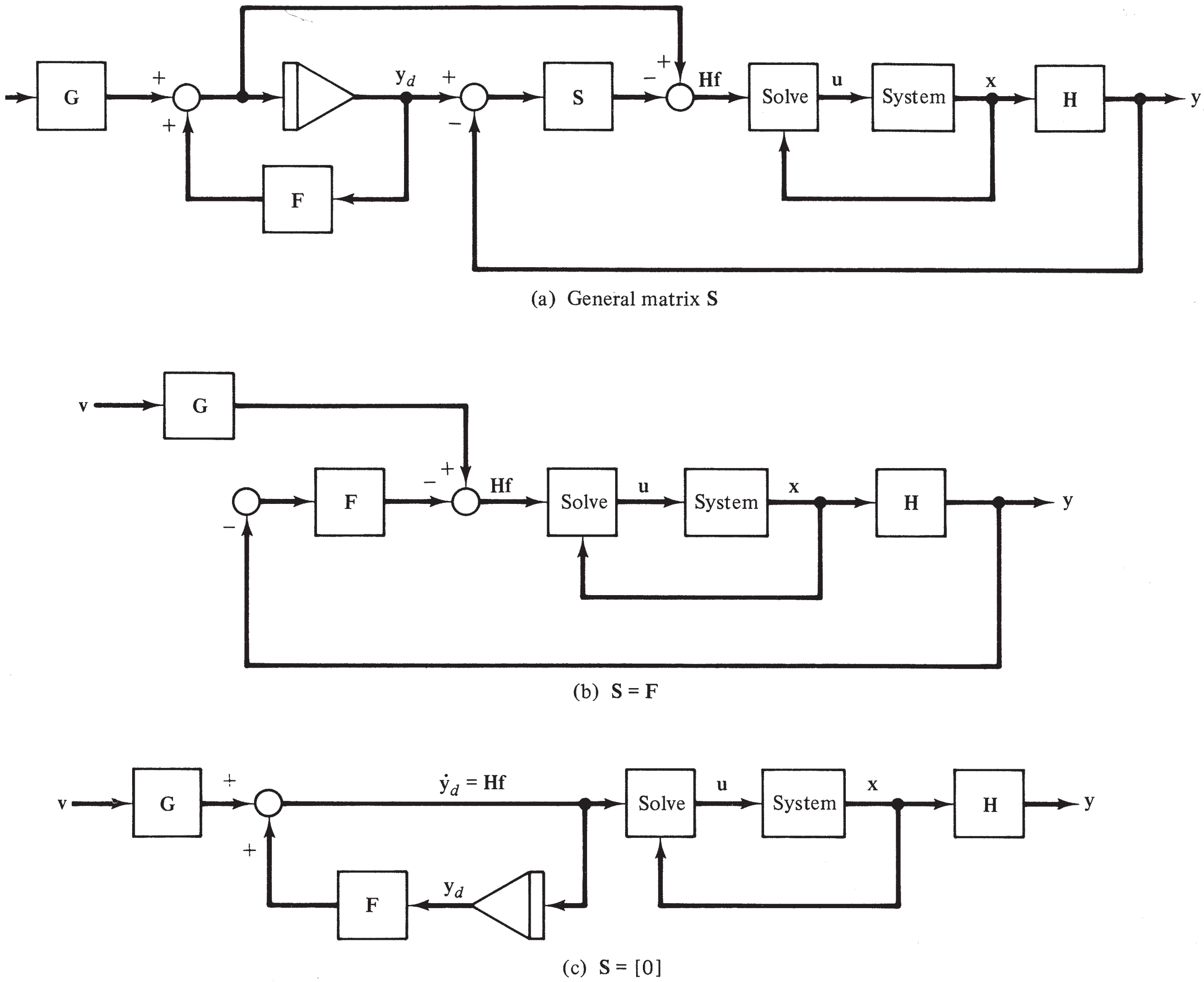


Figure 15.4

15.4 HARMONIC LINEARIZATION: DESCRIBING FUNCTIONS [4, 9–11]

Harmonic linearization is an approximate method of analyzing certain kinds of nonlinear systems using well-understood linear methods. It differs from the small perturbation linearization method of Sec. 15.2 in that signal amplitudes are not restricted to be small perturbations from known nominal values or equilibrium points. Rather, large amplitude signals can be treated, provided they are nearly sinusoidal and certain additional conditions are satisfied. Initially the systems under discussion will be limited to nth-order unforced linear systems with one scalar-valued nonlinearity, $n(x_i)$, which depends on a single state x_i as its input.

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} - \mathbf{B}n(x_i) \tag{15.12}$$

The minus sign is arbitrary, since it could have been absorbed into the definition of the input column matrix **B.** The chosen form allows representing Eq. (*15.12*) in the traditional single loop, negative-feedback block diagram form of Figure 15.5, with $G(s) = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B}$. $G(s)$ is the linear transfer function from the output u of the nonlinearity to x_i, which plays the role of the system output y, and thus $y = \mathbf{C}\mathbf{x}$ determines the appropriate matrix **C.** If the input e to the nonlinearity is *assumed* sinusoidal,

$$e(t) = E \sin \omega t$$

the output $u(t)$ will generally also be periodic but not just a pure sinusoid. The first few terms of its Fourier series expansion are given as

$$u(t) = b_0 + a_1 \sin \omega t + b_1 \cos \omega t + a_2 \sin 2\omega t + b_2 \cos 2\omega t + a_3 \sin 3\omega t + b_3 \cos 3\omega t + \cdots \qquad (15.13)$$

The coefficients are given by

$$a_k = \frac{1}{T}\int_{-T}^{T} \sin(k\omega t) n(t)\,dt \qquad \text{where } T = \frac{\pi}{\omega}$$

$$b_k = \frac{1}{T}\int_{-T}^{T} \cos(k\omega t) n(t)\,dt \qquad \text{for } k > 0$$

$$b_0 = \frac{1}{2T}\int_{-T}^{T} n(t)\,dt$$

The output of the linear portion of the system—namely, $G(s)$—will likewise be the sum of terms, one from each component in the expansion for u. The *essential* assumption upon which the validity of harmonic linearization is based is that $G(s)$ is of a sufficiently low-pass nature, so that all harmonics are attenuated to a negligible level. Only the fundamental frequency terms survive the trip around the loop and have an effect on the input to the nonlinearity. Furthermore, many interesting nonlinearities have a zero average value, so that b_0 is zero. (This assumption can be removed.) Then the important part of the time-domain nonlinearity output caused by the input $E \sin \omega t$ is just $a_1 \sin \omega t + b_1 \cos \omega t$. The ratio of the Laplace transforms of the nonlinearity output and input is thus $N(s) = [a_1 \omega + b_1 s]/E\omega$. This ratio has no meaning except for sinusoidal signals, and when $s = j\omega$ it reduces to

$$N(j\omega) = \frac{a_1}{E} + \frac{b_1 j}{E} = \frac{\sqrt{a_1^2 + b_1^2}\, e^{j\varphi}}{E}$$

where

$$\varphi = \tan^{-1}\left(\frac{b_1}{a_1}\right) \qquad (15.14)$$

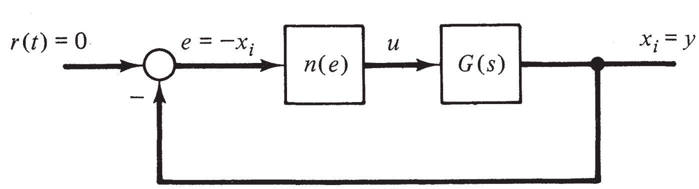


Figure 15.5

This is the *describing function* for the nonlinearity. It is the ratio of the magnitude of the fundamental output component to the mangitude of the input to the nonlinearity and as such plays the role of an equivalent gain. To emphasize that it depends on the input magnitude E as well as the frequency, we will write it as $N(E, \omega)$ from now on. Notice that the phase shift φ will be zero and the describing function N will be purely real for any single-valued odd symmetric nonlinearity, since then $b_k = 0$. In this case all even harmonic terms a_k are also zero, by symmetry. Thus the first neglected term is the a_3 term, which (1) is usually much smaller than a_1 to begin with and (2) is more strongly attenuated by $G(s)$ because of its higher frequency. Many odd symmetric nonlinearities are very accurately accounted for using the describing function approach.

For many nonlinearities, N is not really a function of frequency but only of input amplitude E. In those cases the nonlinearity is replaced by a real but signal-dependent gain, N. Viewing the describing function as an equivalent gain is most helpful in the applications of describing functions to stability analysis that follow.

EXAMPLE 15.5 Derivation of a Describing Function A large number of odd-symmetric nonlinearities of engineering significance can be represented by three linear segments, as shown in Figure 15.6*a* and as determined by three slopes K_1, K_2, and K_3 and two breakpoints α and β. For example, the dead zone nonlinearity in Figure 15.6*b* has $\alpha = 0$, $\beta = 1$, $K_2 = 0$, $K_3 = 1$, and K_1 arbitrary. The saturation characteristic of Figure 15.6*c* has $\beta = \infty$, $K_2 = 0$, and K_3 arbitrary. The relay of Figure 15.6*d* (which has the same shape as coulomb friction and preload) has $\alpha = 0$, $\beta = 0$ (or some small ϵ to prevent a discontinuity), K_1 arbitrary, $K_2 = 1/\epsilon$, and $K_3 = 0$. Many other combinations can be formed, such as a relay with dead zone, by proper choice of these five parameters. In order to derive the describing function for this nonlinearity, ten separate time increments must be considered for each full (normalized) cycle $t' = \omega t \in [0, 2\pi]$ in the most general case. (Sketch the input sine wave and identify when its magnitude falls in each segment of the nonlinearity.) In the following discussion, the prime is dropped from the normalized t' for convenience.

1. $0 < t < t_\alpha$, input sinusoid value $e(t) < \alpha$, output $n(e) = n_1(t) = K_1 E \sin t$
2. $t_\alpha < t < t_\beta, \alpha < e(t) < \beta, n(e) = n_2(t) = K_2 E[\sin t - \sin t_\alpha] + K_1 E \sin t_\alpha$
3. $t_\beta < t < t_{\beta'}, e(t) > \beta', n(e) = n_3(t) = K_3 E[\sin t - \sin t_\beta] + K_2 E[\sin t_\beta - \sin t_\alpha] + K_1 E \sin t_\alpha$
4. $t_{\beta'} < t < t_{\alpha'}, \alpha < e(t) < \beta$ (same as region 2)
5. $t_{\alpha'} < t < \pi, 0 < e(t) < \alpha$ (same as region 1)

The last five regions divide the negative half-cycle $[\pi, 2\pi]$ into the same regions as (1) through (5). The values for t_α and t_β are functions of the input amplitude E and are given by

$$t_\alpha = \sin^{-1}\left(\frac{\alpha}{E}\right) \text{ if } E > \alpha \quad \text{and} \quad t_\alpha = \frac{\pi}{2} \text{ otherwise}$$

$$t_\beta = \sin^{-1}\left(\frac{\beta}{E}\right) \text{ if } E > \beta \quad \text{and} \quad t_\beta = \frac{\pi}{2} \text{ otherwise}$$

Symmetry gives the other times, such as

$$t_{\beta'} = \pi - t_\beta \quad \text{and} \quad t_{\alpha'} = \pi - t_\alpha$$

The kth harmonic coefficient can be evaluated from

$$a_k = \left(\frac{4}{\pi}\right)\left\{\int_0^{t_\alpha} n_1(t) \sin(kt)\,dt + \int_{t_\alpha}^{t_\beta} n_2(t) \sin(kt)\,dt + \int_{t_\beta}^{\pi/2} n_3(t) \sin(kt)\,dt\right\}$$

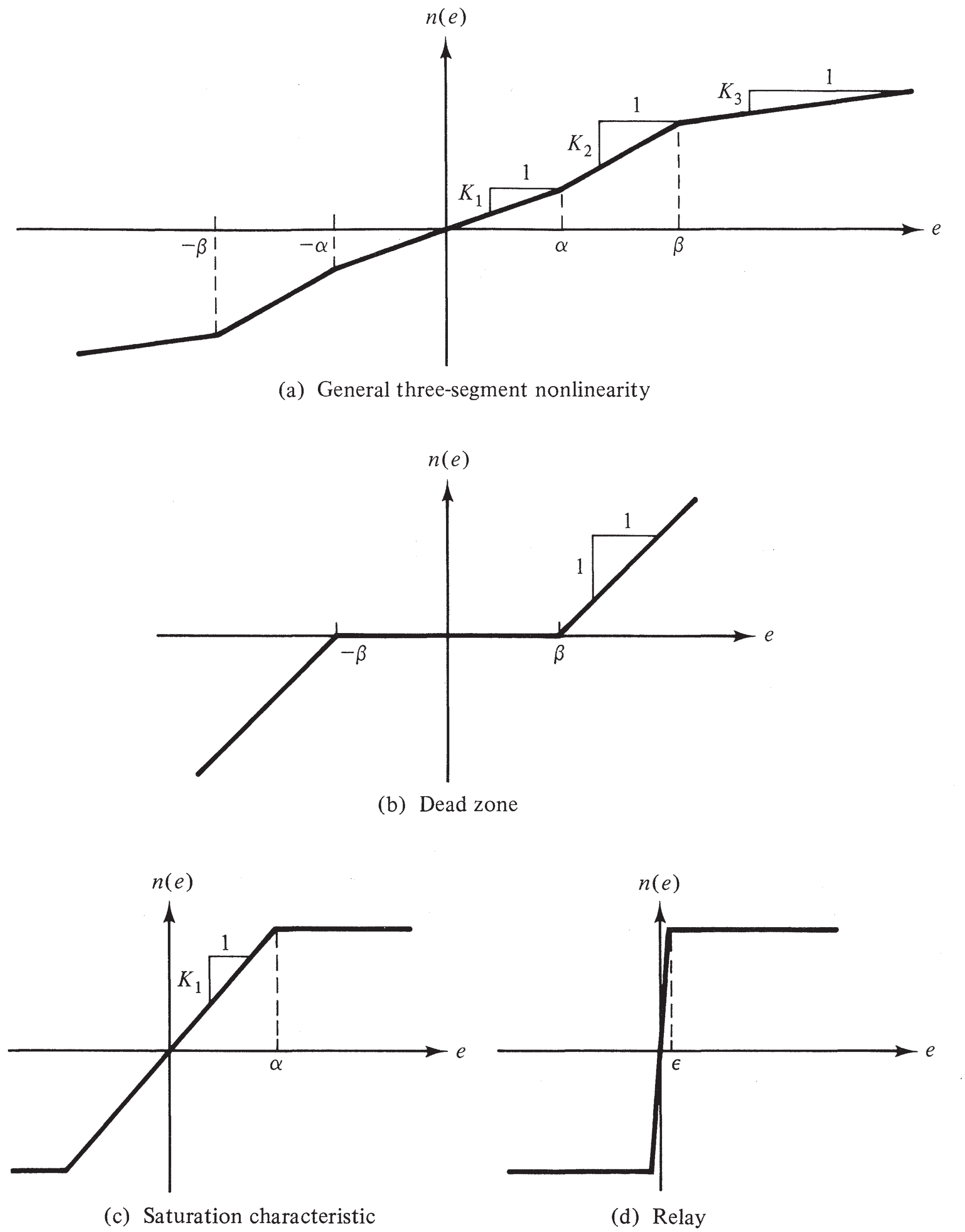


Figure 15.6 Some common single-valued odd symmetric nonlinearities.

Evaluation of these integrals for $k = 1$ gives the describing function, which is independent of frequency ω:

$$N(E) = \frac{a_1}{E} = \frac{2(K_1 - K_2)[t_\alpha + \sin(2t_\alpha)/2]}{\pi}$$

$$+ \frac{2(K_2 - K_3)[t_\beta - \sin(2t_\beta)/2]}{\pi} + K_3 + \frac{2K_2 \sin(2t_\beta)}{\pi}$$

Note that if E is less than α, then both t_α and t_β are equal to $\pi/2$, and $N(E)$ reduces to just K_1 as it should. If $t_\alpha = 0$ and $t_\beta = \pi/2$, then $N(E)$ reduces to K_2. A computer is helpful in evaluating all except these special limiting cases. ■

The describing functions for the preceding class of nonlinearities are real and independent of the excitation frequency. This is not always true, as demonstrated by the following multiple-valued nonlinearity and by the example in Problem 15.10.

EXAMPLE 15.6 Relay with Hysteresis Figure 15.7 shows a common model for a relay with hysteresis. Assuming $E > \alpha$, there are three critical time periods during a typical cycle of the input sine wave:

$$0 < t < t_1: \quad 0 < e(t) < \alpha, \quad e(t) \text{ rising to } \alpha, \text{ for which } N(e) = -M$$

$$t_1 < t < t_2: \quad -\alpha < e(t) < \alpha, \quad e(t) \text{ falling toward } -\alpha, \text{ for which } N(e) = M$$

$$t_2 < t < 2\pi: \quad e(t) < \alpha, \quad e(t) \text{ again rising toward } \alpha, N(e) = -M$$

From Figure 15.7 it is seen that $t_1 = \sin^{-1}(\alpha/E)$ and $t_2 = \sin^{-1}(-\alpha/E) = \pi + t_1$. The two fundamental Fourier coefficients are thus

$$a_1 = \left(\frac{M}{\pi}\right)\left\{-\int_0^{t_1} \sin t\,dt + \int_{t_1}^{t_2} \sin t\,dt - \int_{t_2}^{2\pi} \sin t\,dt\right\} = \frac{4M\cos(t_1)}{\pi}$$

$$= \left(\frac{4M}{\pi}\right)\sqrt{1-\left(\frac{\alpha}{E}\right)^2}$$

$$b_1 = \left(\frac{M}{\pi}\right)\left\{-\int_0^{t_1} \cos t\,dt + \int_{t_1}^{t_2} \cos t\,dt - \int_{t_2}^{2\pi} \cos t\,dt\right\} = -\frac{4M\sin(t_1)}{\pi}$$

$$= -\frac{4M\alpha}{\pi E}$$

so that the describing function is

$$N(\omega, E) = \left(\frac{4M}{\pi E}\right)\left\{\sqrt{1-\left(\frac{\alpha}{E}\right)^2} - \frac{j\alpha}{E}\right\} \qquad \text{for } E > \alpha$$

When $E < \alpha$, a_1, b_1, and N are all zero. ■

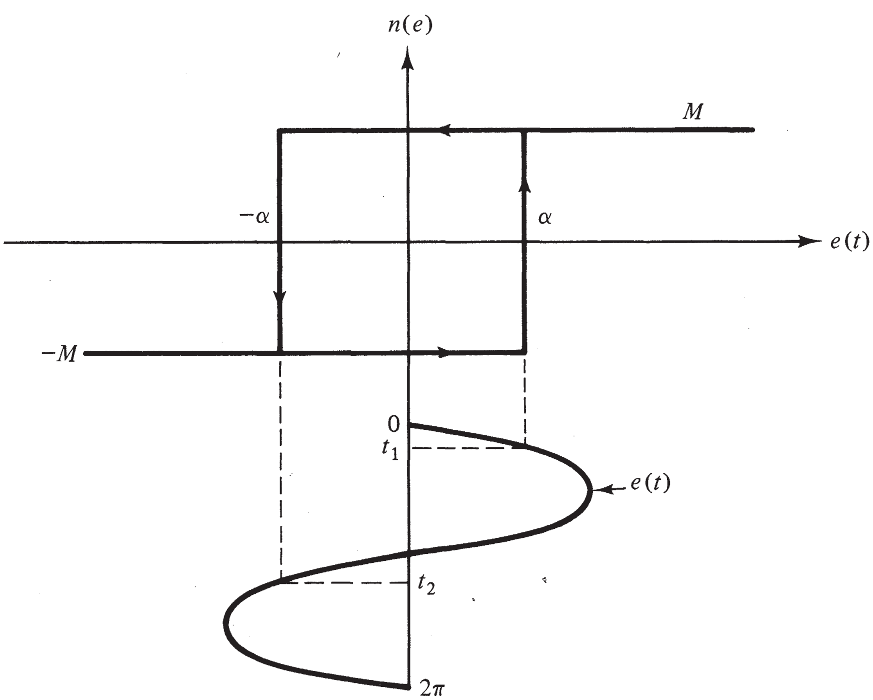


Figure 15.7

Other examples of describing functions are given in the problems. Extensive tabulations for every conceivable type of nonlinearity can be found in the references [9, 13]. For actual system components, describing functions can be determined experimentally.

15.5 APPLICATIONS OF DESCRIBING FUNCTIONS

A distinctive characteristic of nonlinear systems is the possibility of exhibiting limit cycles. A limit cycle is a self-excited, self-sustaining periodic oscillation. A stable limit cycle is one for which trajectories slightly perturbed from it are attracted back to it. An unstable limit cycle is one for which perturbed trajectories do not return to it. Rather, they approach other limit cycles, equilibrium points, grow without bound, or approach some other bounded but nonperiodic trajectories. The later case is now referred to as *chaos* [14–17]. The major uses of describing functions are (1) to investigate the stability of systems like the one in Figure 15.5, (2) to investigate the possible existence of limit cycles and to predict their amplitude and frequency when they exist, and (3) to modify or compensate the linear system to prevent the occurrence of undesirable limit cycles (or chaos?). A limit cycle is by definition periodic, although the waveform is not necessarily sinusoidal. If the limit cycle has a strong fundamental Fourier component, then describing functions can provide acceptably accurate results in their analysis. Consider the system described by Eq. (*15.12*), with the nonlinearity replaced by its describing function approximation, N. The unforced closed-loop system is described by

$$\dot{\mathbf{x}} = (\mathbf{A} - \mathbf{B}[\mathbf{N}]\mathbf{C})\mathbf{x} \tag{15.15}$$

The output matrix $\mathbf{C}$ is used to select the components of $\mathbf{x}$ which are input to the nonlinearities. At this point the factor $[\mathbf{N}]$ could be a matrix of describing functions representing multiple nonlinearities, $\mathbf{n}(\mathbf{x}) \approx [\mathbf{N}]\mathbf{C}\mathbf{x}$. The closed-loop eigenvalues satisfy

$$|\mathbf{I}\lambda - (\mathbf{A} - \mathbf{B}[\mathbf{N}]\mathbf{C})| = 0 \tag{15.16}$$

This can be rearranged as

$$\begin{aligned} |\mathbf{I}\lambda - \mathbf{A}|\cdot|\mathbf{I} + (\mathbf{I}\lambda - \mathbf{A})^{-1}\mathbf{B}\mathbf{C}[\mathbf{N}]| &= |\mathbf{I}\lambda - \mathbf{A}|\cdot|\mathbf{I} + [\mathbf{N}]\mathbf{C}(\mathbf{I}\lambda - \mathbf{A})^{-1}\mathbf{B}| \\ &= |\mathbf{I}\lambda - \mathbf{A}|\cdot|\mathbf{I} + [\mathbf{N}]G(\lambda)| = 0 \end{aligned}$$

The whole concept of describing functions is based on periodic solutions, so this result has meaning only for complex conjugate sets of λ values. Thus the condition for existence of periodic solutions is that

$$|\mathbf{I} + [N]G(j\omega)| = 0 \tag{15.17}$$

In the simplest case of one nonlinearity, this determinant is just a scalar equation $1 + NG(j\omega) = 0$. Points at which $NG(j\omega) = -1$ identify potential limit cycles. These points can be determined either analytically or graphically (by using Nyquist, Bode, or log magnitude versus angle plots of $G(j\omega)$ and comparing them with the critical point $-1/N$ from the describing function, rather than the usual point -1). That is, $G(j\omega) = -1/N$ can be analyzed, or alternatively inverse Nyquist plots can be used with $N = -1/G(j\omega)$.

Real, Frequency-Independent Describing Functions

In cases such as Example 15.5, where N is purely real and independent of frequency, all the usual conclusions apply. In particular, the number of unstable closed-loop poles Z_r (called Z because they are zeros of the characteristic equation) can be determined from Nyquist's criterion as

$$Z_r = P_r - N_c \tag{15.18}$$

where P_r is the number of right-half plane open-loop poles of G and N_c is the number of counterclockwise encirclement of the critical point $-1/N$ by the polar plot of $G(j\omega)$ as ω ranges over $-\infty$ to ∞. In the case of real, frequency-independent $N(E, \omega)$, root locus techniques can also be used (with caution) on $NG(s) = -1$. The describing function N plays the role of a varying gain. Since this is valid only for $s = j\omega$, the root locus results are helpful only in predicting behavior near the $j\omega$ axis crossover points. Attempts to correlate frequency response results with transient response behavior using known root locus methods are generally not very satisfactory.

The General Case

When $N(E, \omega)$ is complex and frequency-dependent, the Nyquist or Bode method can be applied without modification to find the limit cycle locations. Stability conclusions drawn from Eq. (*15.18*) can be wrong, however. The number of encirclements of $NG(j\omega)$ is no longer determined solely by the behavior of $G(j\omega)$. The safest approach is always to rely on the basic equation $NG(j\omega) = -1$, but this means the advantage of seeing separately the effects of the linear G and the nonlinearity is lost. In some cases it may be possible to separate $N(E, \omega)$ into a product $N(E, \omega) = N_1(E)N_2(\omega)$, with N_1 real. Then a modified $G'(j\omega) = G(j\omega)N_2(\omega)$ can be used in $G'(j\omega) = -1/N_1(E)$. Equation (*15.18*) can now be used again with $G'(j\omega)$, which will generally not have the same number of encirclements as $G(j\omega)$. See Problems 15.11 and 15.12 for an example. The root locus procedure must be similarly modified. The polar form of $N(E, j\omega) = Ke^{j\varphi}$ shows that if a root locus approach were to be attempted using $G(j\omega)$, the 180° locus would not suffice. It would need to be modified by the phase shift φ, which is generally frequency-dependent. This approach is usually not worth the required effort. In those cases where it is possible to form $G'(s)$, the root locus approach can be useful.

EXAMPLE 15.7 A single-loop system has $G(s) = 2/[s(s+p)]$ cascaded with the hysteresis-type relay of Example 15.6, with $\alpha = M = 1$. If this system can exhibit a limit cycle, find its approximate amplitude and frequency by using describing functions.

For this scalar case Eq. (*15.17*) gives the limit cycle requirement as $4/(\pi E)\{\sqrt{1 - 1/E^2} - j/E\} = -j\omega(j\omega + p)/2 = (\omega^2 - j\omega p)/2$. From the imaginary part we find that $\pi\omega p/8 = 1/E^2$. From the real part we find $\omega^2/2 = 4/(\pi E)\sqrt{1 - 1/E^2}$. Combining gives

$$\omega^3 + p^2\omega - \frac{8p}{\pi} = 0$$

Table 15.1 gives the results for three different pole locations p, as well as numerical results obtained by simulating the actual nonlinear system.

TABLE 15.1

p	4	1	0.5
Describing function predictions:			
ω (rad/s)	0.62161	1.1245	1.0071
Period $T = 2\pi/\omega$ (s)	10.1079	5.5875	6.239
E	1.012	1.5	2.2487
Simulation results:			
Period T	9.3	5.6	6.4
E	1.044	1.6	2.32
Percent error in:			
T	8.7	0.22	2.5
E	3.06	6.25	3.1

Figure 15.8 shows the time response of the input to the nonlinearity for the cases with $p = 4$ and $p = 0.5$. The later case is much more nearly sinusoidal, and this explains the higher accuracy of the describing function in predicting the period in this case. Figure 15.9 shows the phase-plane plot of the limit cycle behavior, with $p = 0.5$, for initial conditions inside and outside the limit cycle. Note that a simple root locus plot shows that this system is always asymptotically stable for any *real* gain N. A real equivalent gain cannot explain the sustained oscillations exhibited here. ■

Although some interesting and useful aspects of describing function analysis of nonlinear systems have been presented here and in the chapter-end problems, space

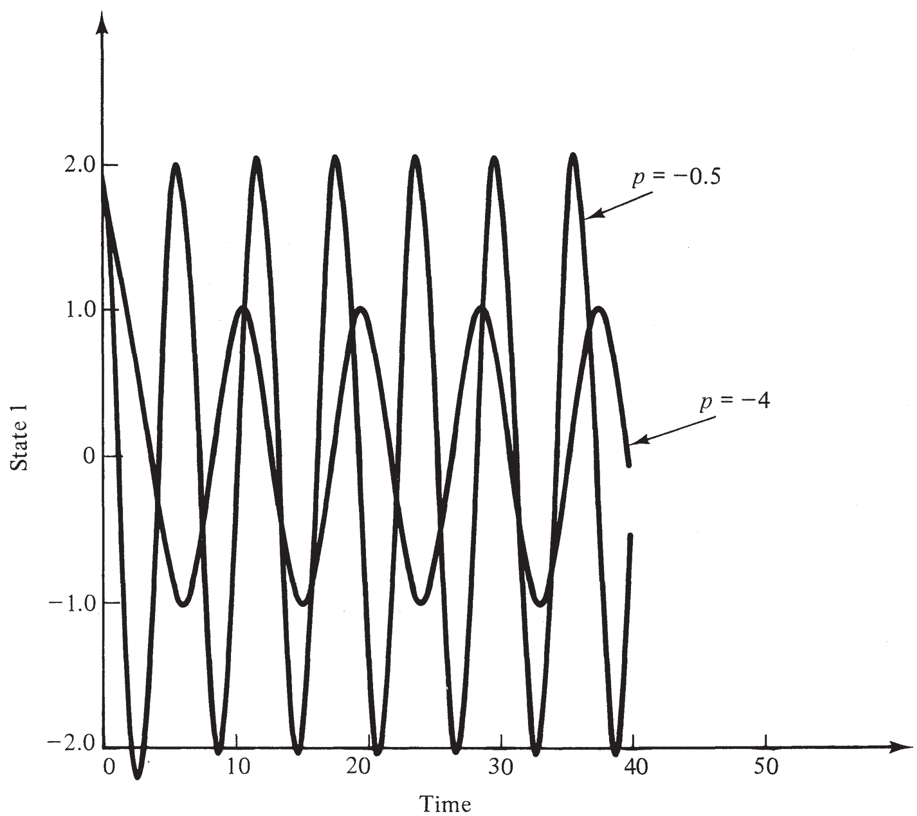


Figure 15.8

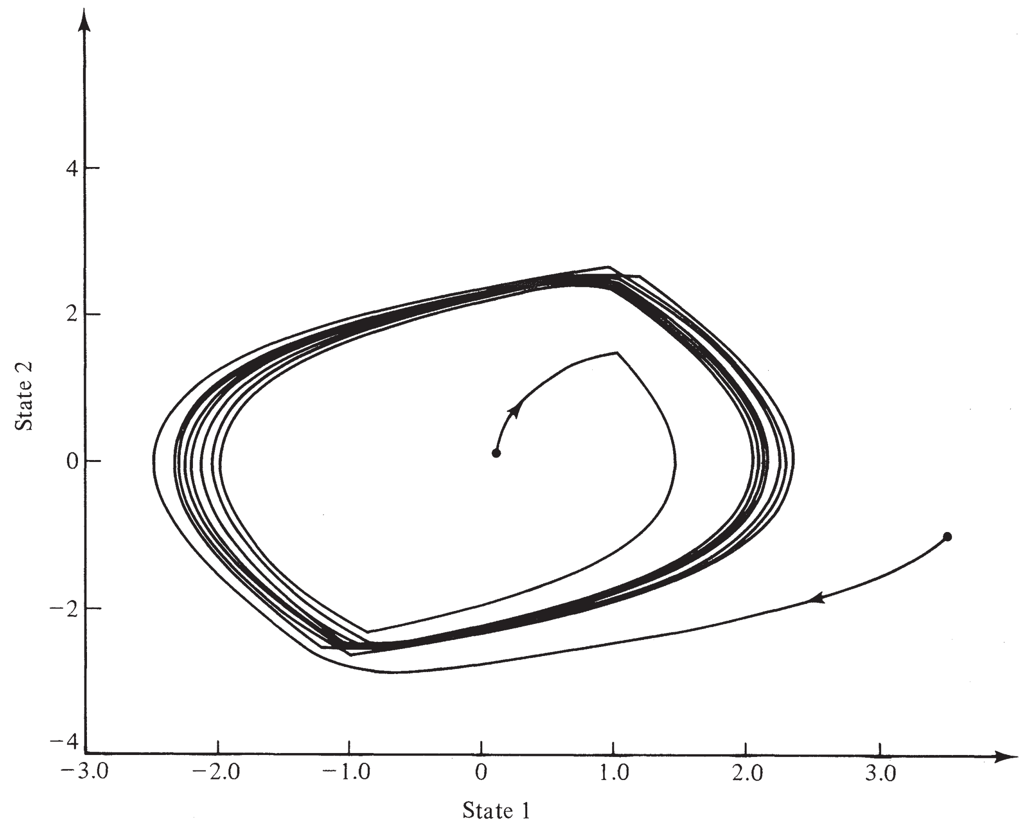


Figure 15.9

limitations have forced omission of several additional considerations, which can be found in references such as 9 and 11. Nonsymmetrical nonlinearities which yield a nonzero average value is one such topic. Another has to do with nonzero external inputs. Harmonic analysis techiques can be applied to such forced, closed-loop systems, and additional interesting phenomena such as jump resonances and subharmonic responses then arise. Significant space has been devoted in the problems to the application of describing functions to systems with (potentially) chaotic behavior. The field of chaos in dynamic systems is still evolving, and only a small segment of it has been discussed here. In control problems the usual goal would be to *avoid* such chaotic behavior (although it can inadvertently creep into various adaptive and self-learning control systems, which are invariably nonlinear [18, 19]). Describing functions are presented here as a tool for understanding and predicting the possibility of chaos in some nonlinear systems. Then, compensation or redesign techniques can be used to avoid such potentially troublesome behavior. Finally, the very useful topic of dual input describing functions has not been developed here. The concept involves the deliberate insertion of a second sinusoidal input, *dither,* to a nonlinearity. Dither is usually of a much higher frequency than the fundamental frequencies of interest. If the assumed low-pass nature of the linear system is present, the high-frequency dither component will have negligible direct effect on most of the system. It can have a pronounced effect on the behavior of the nonlinearity. For example a nonlinearity with deadband, as in Figure 15.6*b,* will have no output until the amplitude of the input $e(t)$ exceeds β. If a high-frequency dither signal $d(t)$ is superimposed onto $e(t)$, then some output will appear whenever $e(t) + d(t)$ exceeds β, so the effective gain of the nonlinearity has been changed. The low-pass filtered envelope of this $e(t) + d(t)$ output can provide positive benefits to system behavior [9, 11].

15.6 LYAPUNOV STABILITY THEORY AND RELATED FREQUENCY DOMAIN RESULTS

The direct method of Lyapunov [20–23] was presented in Sec. 10.6 with sufficient generality to allow application to nonlinear systems being considered here. Although all the stability and instability theorems apply without change, the determination of suitable Lyapunov functions is less straightforward. The solution of Lyapunov's equation, Eq. (*10.22*), to find quadratic Lyapunov functions is generally not adequate. A variety of methods for obtaining Lyapunov functions have been proposed [12, 24–26], and a few of these are illustrated in Problems 15.17 through 15.20. The variable gradient method [27, 28] introduced in Problem 10.14 and illustrated in Problem 10.15 is one fairly general approach. Problems 15.21 through 15.23 demonstrate the variable gradient method on several nonlinear systems.

Lyapunov analysis of nonlinear systems differs from linear systems in that linear system stability conclusions are global. In nonlinear problems, a given equilibrium point may have a finite or local domain of attraction. That is, initial states within that domain will converge to the equilibrium point. Initial points outside the domain of attraction will diverge from it, settle into some sort of limit cycle around it, or perhaps undergo some other more complex behavior (chaos). Lyapunov stability theorems can be used to estimate the extent of the domains of stable attraction. Lyapunov instability theorems can be used to estimate the extent of the unstable (repulsive) behavior around an equilibrium point. We now use the variable gradient method to illustrate this.

EXAMPLE 15.8 Consider the Van der Pol equation $\ddot{y} + \mu(y^2 - 1)\dot{y} + \beta y = 0$. In state variable form the system equations are

$$\dot{x}_1 = x_2 \quad \text{and} \quad \dot{x}_2 = -\mu(x_1^2 - 1)x_2 - \beta x_1$$

This system has a single equilibrium point at the origin. Assume that the gradient of the still unknown Lyapunov function V is

$$\nabla V = \begin{bmatrix} \alpha_{11}x_1 + \alpha_{12}x_2 \\ \alpha_{21}x_1 + \alpha_{22}x_2 \end{bmatrix}$$

The final coefficient α_{22} can be set to 2 without loss of generality, and this will be seen to ensure that V is quadratic in x_2. This will be done here. Assuming that V is not an explicit function of time, the time rate of change of V is

$$\dot{V} = [\nabla V]^T\dot{\mathbf{x}} = [\alpha_{11} - 2\beta - \mu\alpha_{21}(x_1^2 - 1)]x_1x_2 - \alpha_{21}\beta x_1^2$$
$$- [\alpha_{12} - 2\mu(x_1^2 - 1)]x_2^2$$

The troublesome cross terms can be eliminated by selecting

$$\alpha_{11} = 2\beta + \mu\alpha_{21}(x_1^2 - 1)$$

leaving

$$\dot{V} = -\alpha_{21}\beta x_1^2 + [\alpha_{12} - 2\mu(x_1^2 - 1)]x_2^2$$

One easy way to proceed would be to select $\alpha_{12} = \alpha_{21} = 0$. Then, if $\mu > 0$, there is a region defined by $x_1^2 < 1$ in which $\dot{V}$ will be positive. With these choices, $\nabla V^T = [2\beta x_1 \quad 2x_2]$. The so-called curl equations $\partial(\nabla V)_i/\partial x_j = \partial(\nabla V)_j/\partial x_i$ are automatically satisfied, and the line integral easily gives $V = \int(\nabla V)_1\,dx_1 + \int(\nabla V)_2\,dx_2 = \beta x_1^2 + x_2^2$. This is a positive definite function, and a given value of V defines an ellipse. The largest such ellipse which satisfies $x_1^2 < 1$ defines the region Ω, where V and $\dot{V}$ are both positive. The equation for the family of ellipses has no x_1x_2 term so x_1 and x_2 are principal axes. Setting $x_2 = 0$ shows that the maximum value of V is β. Then setting $x_1 = 0$ shows that the maximum $x_2 = \sqrt{\beta}$. By the instability theorem, Theorem 10.6, the system is unstable and Ω is an estimate of the region of repulsion. If μ is negative, the origin is a point of stable equilibrium. The same region Ω would then be an estimate of the region of attraction, since at all points within it, we have $V > 0$ and $\dot{V} < 0$. ■

EXAMPLE 15.9 A different estimate of the regions of attraction or repulsion for the Van der Pol equation can be found by selecting different coefficients for the gradient of V. The same selection for α_{11} is again made to eliminate the cross terms. Now, a tentative choice for $\alpha_{12} = \alpha_{21} = -2\mu$ will give $\dot{V} = -2\mu[x_2 - \beta]x_1^2$. The selected gradient is

$$\nabla V = \begin{bmatrix} [-2\mu^2(x_1^2 - 1) + 2\beta]x_1 - 2\mu x_2 \\ -2\mu x_1 + 2x_2 \end{bmatrix}$$

The curl equations are again satisfied automatically, and a line integral gives

$$V = \left[\beta + \mu^2 - \frac{\mu x_1^2}{2}\right]x_1^2 - 2\mu x_1 x_2 + x_2^2$$

If we define $w = [\beta + \mu^2 - \mu x_1^2/2]$, then

$$V = \mathbf{x}^T\begin{bmatrix} w & -\mu \\ -\mu & 1 \end{bmatrix}\mathbf{x}$$

Using principle minors, this is seen to be positive definite provided $w > 0$ and $w - \mu^2 > 0$—that is, as long as $x_1^2 < 2\beta/\mu$. Likewise, (when $\mu > 0$) $\dot{V} \geq 0$ provided that $x_2^2 < \beta$. It is seen that the limit on x_1 is different here than in the previous example, and the limit on x_2 is the same. Here the figure determined by V = constant is no longer a simple ellipse. The estimate of Ω is given by the largest closed contour V = constant that satisfies the stated limits on x_1 and x_2. The actual boundary of Ω is the interior of a limit cycle. It is of irregular shape, somewhat similar to the limit cycle in Figure 15.9. See Reference 11 or 12 for the precise shape. ■

As demonstrated here and in the problems, finding a suitable Lyapunov function may require a bit of ingenuity. If attention is restricted to a certain subclass of nonlinear systems, then fairly general sufficient conditions for stability can be, and have been, derived by using Lyapunov methods. One such approach, valid for systems with a single nonlinearity, uses a Lyapunov function which is a quadratic (as in the linear system case) plus an integral of the nonlinearity. This approach is usually associated with the name Lur'e [20, 23, 28]. One simple demonstration is given.

Consider the system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} - \mathbf{B}n(y)$, where $y = \mathbf{C}\mathbf{x}$ is a scalar input to the nonlinearity. Assume that $\mathbf{A}$ is asymptotically stable. It is also assumed that $n(0) = 0$, so that the origin is the only equilibrium point. Then by setting $\mathbf{Q} = \mathbf{I}$ and solving Lyapunov's equation (10.22), a positive definite matrix $\mathbf{P}$ can be found. In order to study the stability of the nonlinear system, a Lyapunov function is selected as $V = \mathbf{x}^T\mathbf{P}\mathbf{x} + \int_0^y n(\xi)\,d\xi$. Clearly, if the integral term is nonnegative for all $\mathbf{x}$, then V will be positive definite. The time-derivative is given by

$$\begin{aligned}\dot{V} &= \mathbf{x}^T[\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A}]\mathbf{x} - 2n(y)\mathbf{B}^T\mathbf{P}\mathbf{x} + n(y)\dot{y} \\ &= -\mathbf{x}^T\mathbf{x} - n(y)[2\mathbf{B}^T\mathbf{P} - \mathbf{C}\mathbf{A}]\mathbf{x} - \mathbf{C}\mathbf{B}n(y)^2\end{aligned}$$

The question of stability for this class of problems reduces to a determination of whether the nonlinear term $n(y)$ can cause $\dot{V}$ to lose the negative-definiteness which the quadratic term in $\mathbf{x}$ would give. Other strategies for selecting $\mathbf{P}$ may also be used.

EXAMPLE 15.10 Consider the system

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ -8 & -6 \end{bmatrix}\mathbf{x} - \begin{bmatrix} 0 \\ 1 \end{bmatrix} n(x_1)$$

The solution of $\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} = -\mathbf{I}$ gives

$$\mathbf{P} = \begin{bmatrix} \frac{9}{8} & \frac{1}{16} \\ \frac{1}{16} & \frac{3}{32} \end{bmatrix}$$

Since the input to $n(\ )$ is x_1, $\mathbf{C} = [1 \quad 0]$, so $\mathbf{CB} = 0$; then

$$\dot{V} = -x_1^2 - x_2^2 - \frac{n(x_1)x_1}{8} - x_2\left[x_2 + \left(\frac{13}{16}\right)n(x_1)\right]$$

This can be rearranged into

$$\dot{V} = -\frac{n(x_1)x_1}{8} - \mathbf{x}^T\begin{bmatrix} 1 & \dfrac{13n(x_1)}{32x_1} \\ \dfrac{13n(x_1)}{32x_1} & 1 \end{bmatrix}\mathbf{x}$$

The first term is negative for all x_1 except 0 for any nonlinearity which lies in the first and third quadrant of $n(x_1)$ versus x_1 space and which satisfies $n(0) = 0$. The quadratic-form term is negative definite, and hence the system is asymptotically stable, provided $n(x_1)^2 < x_1^2(\frac{32}{13})^2$. ■

The preceding result is extremely conservative. Recall that Lyapunov theorems give sufficient conditions, not necessary conditions. Problem 15.20 uses the variable gradient method and includes this system as a special case, with $f(x_1) = 6$ and $g(x_1) = 8x_1 + n(x_1)$. There it is shown that a sufficient condition for global asymptotic stability is that $n(x_1)x_1 > 0$—that is, any nonlinearity in the first and third quadrants. An even more general result is given in Reference 28, where the condition is that $\int n(\xi)\,d\xi > 0$. This would include, for example, a nonlinearity such as $n(x_1) = e^{-x_1}\sin(x_1)$ which does not stay within the first and third quadrants. The results derived from the Lur'e approach of using a quadratic plus integral Lyapunov function obviously depend upon the particular quadratic selected. It has been shown that the strongest results that can be obtained by using a Lur'e-type Lyapunov function are those contained in the

Popov criterion [9, 12, 23]. The Popov criterion and the closely similar *circle criterion* are examples of frequency domain stability conditions. The interested reader should consult Reference 23 and the references therein for proofs and to learn about the difficult step of converting standard Lyapunov function results into frequency domain conditions for stability. There are several advantages of the frequency domain results to be given next: (1) They circumvent the difficult step of finding a Lyapunov function; (2) they apply without special regard to the order of the system; and (3) they rely on Nyquist-type stability analysis methods, which are well known from linear systems analysis. The principle disadvantage is that they do not always apply. These conditions are applicable to so-called sector nonlinearities—that is, linearities that are contained between two straight lines through the origin. Hysteresis-type nonlinearities and nonlinearities involving products of several state variables are notable examples that cannot be treated by these methods. To be specific, attention is restricted to systems like the one in Figure 15.5. The nonlinearity is assumed to be a single-valued, piecewise continuous function which satisfies $n(0) = 0$ and $K_l < n(e)/e < K_u$ for $e \neq 0$, where K_l and K_u are slopes of the straight lines which provide the lower and upper bounds of the nonlinearity. This is sometimes stated as $n(e)$ belongs to the sector $[K_l, K_u]$.

Popov's Stability Criterion

If the linear portion of the system $G(s)$ is a proper, asymptotically stable transfer function and if $K_l = 0$, $K_u < \infty$ and $n(e)$ is not an explicit function of time, then Popov's criterion can be applied. It ensures that the system is globally asymptotically stable if a real number q and an arbitrarily small positive δ can be found such that

$$\text{Re}\{(1 + j\omega q)G(j\omega)\} + 1/K_u \geq \delta > 0 \qquad (15.19)$$

If $G(s)$ has simple poles *on* the $j\omega$ axis, Popov's criterion remains valid provided the lower bounding slope K_l is larger than some arbitrarily small positive ϵ. Popov's criterion can be given a useful graphical interpretation. To do so we define a Popov locus for the system, which is a modification of the familiar Nyquist locus. The real part is unchanged, and the imaginary part is multiplied by ω. That is, define $G^*(j\omega) = \text{Re}\{G(j\omega)\} + j\omega\,\text{Im}\{G(j\omega)\}$. Then, Eq. (*15.19*) becomes $\text{Re}\{G^*(j\omega)\} - q\,\text{Im}\{G^*(j\omega)\} + 1/K_u \geq \delta > 0$. This indicates that a plot of the Popov locus must lie entirely to the right of a straight line through the point $-1/K_u$ and having a finite slope $1/q$.

EXAMPLE 15.11 Find the maximum upper sector bound K_u for the nonlinearity $n(e)$ of Figure 15.5 for which Popov's criterion can assure stability if $G(s) = [s^3 + 13s^2 + 72s + 160]/[s(s^4 + 6s^3 + 26s^2 + 56s + 80)]$.

A portion of the Popov locus is shown in Figure 15.10. The real-axis crossover occurs at $\omega = 2.866$, with $|G^*(j\omega)| = 0.965$. Thus any value of $-1/K_u$ more negative than -0.965 admits a finite positive sloping line that is totally to the left of the locus. That is, $K_u = 1/0.965 = 1.036$ is the maximum slope of the upper sector bound for which Popov can guarantee stability. Note that since $G(s)$ has one pole on the $j\omega$ axis, the lower sector slope must be larger than some positive ϵ. ■

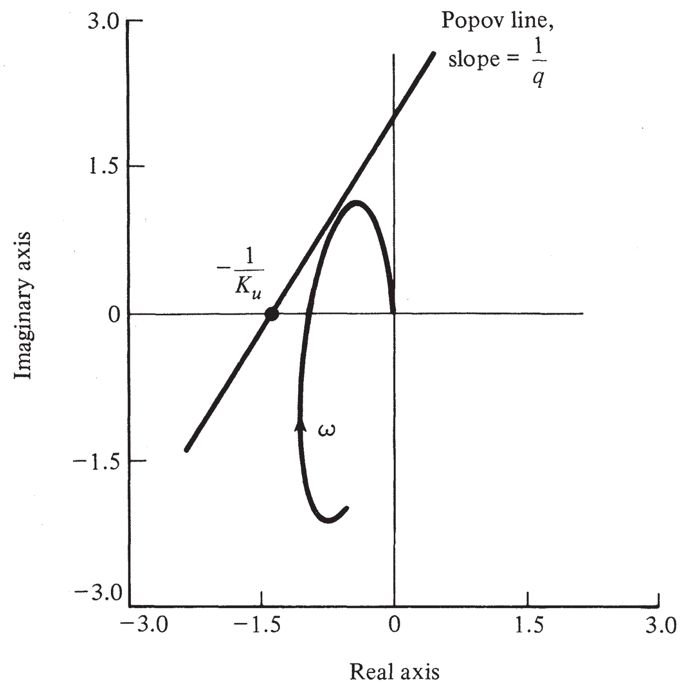


Figure 15.10 Popov locus.

It is pointed out [9] that the output of the system in Figure 15.5 will still asymptotically approach zero for nonzero reference inputs $r(t)$ into the system, provided that $r(t)$ is bounded, uniformly continuous, and square integrable. This rules out any type of constant or bias input component, but in such a case a zero asymptotic output would not be a reasonable expectation on purely intuitive grounds.

The Circle Criterion

The circle stability criterion applies to the same general type of system as discussed before, but with somewhat different specific conditions. The nonlinearity may now be a function of time, $n(e, t)$, but for all time it remains inside the bounding sector $[K_l, K_u]$. The linear system need not be stable, but it is assumed that $G(s)$ is strictly proper (i.e., the state model matrix $\mathbf{D}$ is zero) and there are no common numerator and denominator factors (i.e., the full-order state model is both controllable and observable). A critical circle or disk will be utilized in place of the critical point $-1 + j0$ of the Nyquist criterion for linear system stability. The critical circle cuts the real axis at points $-1/K_u$ and $-1/K_l$ and has a diameter given by $1/K_l - 1/K_u$. The usual Nyquist polar plot, and not the modified Popov locus, is used with the circle criterion. The circle criterion states that the polar plot of $G(j\omega)$ must not cut the critical circle, and the encirclements or lack of encirclements of the critical circle by the locus follow the usual rules associated with the point -1 for linear systems. Several subcases are possible:

(a) If both K_l and K_u are positive, the critical circle is entirely to the left of the origin.

(b) If both K_l and K_u are negative, the critical circle will be entirely to the right of the origin.

(c) When K_l and K_u have opposite signs, the critical circle will have the origin as an internal point.

If the open-loop system $G(s)$ has no right-half plane poles ($P_r = 0$ in Eq. (*15.18*)), then stability requires that the critical circle not be encircled. This means that in cases (a) and (b), the polar plot must stay outside the critical circle. In case (c) the polar plot must stay entirely within the critical circle. If $G(s)$ has unstable poles, then the polar plot still must not intersect the critical circle but must have the correct number of encirclements of it, $N_c \equiv P_r$ as given by Eq. (*15.18*), in order to give $Z_r = 0$ and thus assure asymptotic stability.

EXAMPLE 15.12 Examine the system of Example 15.11 using the circle criterion. The positive-frequency portion of the Nyquist polar plot is shown in Figure 15.11. Since the open-loop system has no unstable poles, the critical circle must not be encircled by the polar plot; i.e., $N_c = 0$ is required. The real-axis crossover is the same as in the previous example because $G^*(j\omega)$ and $G(j\omega)$ have equal real parts. Thus the limit on the upper-sector slope must be less than 1.036. The exact limit depends on the lower-sector bound K_l. If $K_l = 0$, then an infinite-radius circle (i.e., a vertical line) results. This vertical line would need to be positioned further to the left than -1.036 to avoid intersecting the locus. Also note that if both sector bounds have negative slopes, the locus will not intersect the critical circle but will encircle it once in the clockwise direction. The sufficient conditions for stability cannot be satisfied by this system with any negative slope. ■

Sector-type nonlinearities contain all linear gains between K_l and K_u as special cases. Stability for any nonlinearity in this class then necessarily requires that the system be stable for this range of linear gains. The Nyquist stability criterion must therefore hold for all critical points $[-1/K_l, -1/K_u]$. These are all the points along the real-axis diameter of the critical circle. *Aizerman's conjecture* [9, 12, 23] was that a

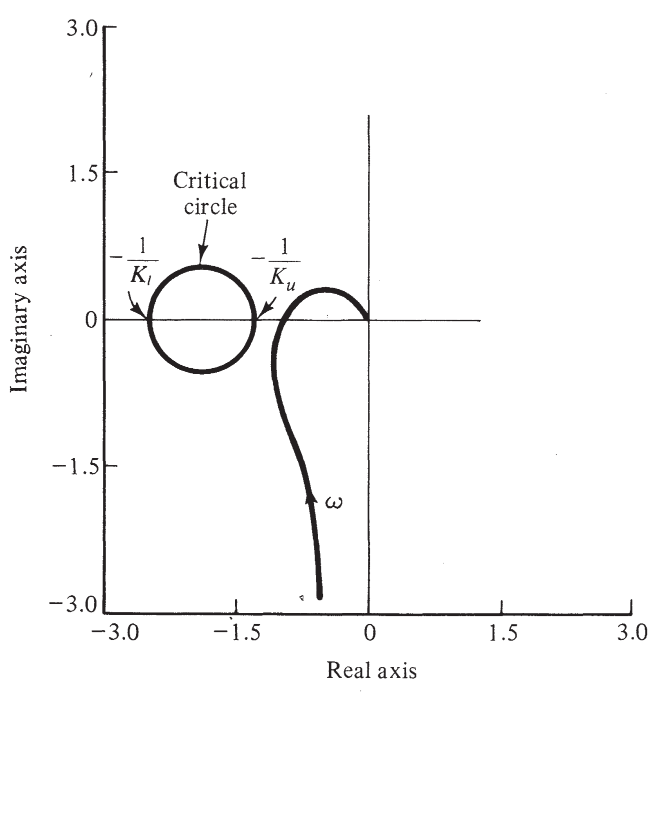


Figure 15.11 Nyquist plot.

nonlinear system with a sector nonlinearity would be stable if the linear system was stable for the entire range of gains in the sector. Aizerman's conjecture is generally false. The circle criterion indicates that not just the real-axis diameter but rather the entire critical circle must be checked for intersections with, or encirclements by, the polar plot.

Some extensions of these single nonlinearity results to systems with multiple nonlinearities are available. A brief survey may be found in Appendix C of Reference 9.

REFERENCES

1. Brockett, R. W.: "Nonlinear Systems and Differential Geometry," *Proceedings of the IEEE,* Vol. 64, No. 1, January 1976, pp. 61–72.
2. Isidori, A.: *Nonlinear Control Systems: An Introduction,* Springer-Verlag, Berlin, 1985.
3. Coddington, E. A. and N. Levinson: *Theory of Ordinary Differential Equations,* McGraw-Hill, New York, 1955.
4. Truxal, J. G.: *Automatic Control Systems Synthesis,* McGraw-Hill, New York, 1955, Chapters 10 and 11.
5. Boye, A. J. and W. L. Brogan: "A Method of Controlling Deterministic Nonlinear Systems," *Proc. American Control Conference,* Boston, Mass. June 1985.
6. Boye, A. J. and W. L. Brogan: "Application of a Nonlinear System Controller to Flight Control," *29th Midwestern Symposium on Circuits and Systems,* Lincoln, Nebraska, August 1986.
7. Boye, A. J. and W. L. Brogan: "A Nonlinear System Controller," *International Journal of Control,* Vol. 44, No. 5, December 1986, pp. 1209–1218.
8. Taylor, A. E. and R. W. Mann: *Advanced Calculus,* Xerox, Boston, Mass. 1972.
9. Atherton, D. P.: *Nonlinear Control Engineering,* Van Nostrand Reinhold Co., New York, 1975.
10. D'Azzo, J. J. and C. H. Houpis: *Feedback Control System Analysis and Synthesis,* 1st ed., McGraw-Hill, New York, 1960, Chapter 18.
11. Elgerd, O. I.: *Control Systems Theory,* McGraw-Hill, New York, 1967, Chapter 8.
12. Power, H. M. and R. J. Simpson: *Introduction to Dynamics and Control,* McGraw-Hill, New York, 1978, Chapter 9.
13. Gibson, J.: *Nonlinear Automatic Control,* McGraw-Hill, New York, 1963.
14. Guckenheimer, J. and P. Holmes: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields,* Springer-Verlag, New York, 1983.
15. Schuster, H. G.: *Deterministic Chaos,* Physik-Verlag, Weinheim, F. R. of Germany, 1984.
16. Chua, L. O. and R. N. Madan: "Sights and Sounds of Chaos," *IEEE Circuits and Devices Magazine,* January 1988, pp. 3–13.
17. Bartissol, P. and L. O. Chua: "The Double Hook," *IEEE Transactions on Circuits and Systems,* Vol. 35, No. 12, December 1988, pp. 1512–1522.
18. Mareels, I. M. Y. and R. R. Bitmead: "Bifurcation Effects in Robust Adaptive Control," *IEEE Transactions on Circuits and Systems,* Vol. 35, No. 7, July 1988, pp. 835–841.
19. Salam, F. M. A. and Shi Bai: "Complicated Dynamics of a Prototype Continuous-Time Adaptive Control System," *IEEE Transactions on Circuits and Systems,* Vol. 35, No. 7, July 1988, pp. 842–849.
20. LaSalle, J. and S. Lefschetz: *Stability by Liapunov's Direct Method,* Academic Press, New York, 1961.
21. 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.
22. 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.
23. Willems, J. L.: *Stability Theory of Dynamical Systems,* John Wiley, New York, 1970.
24. Reiss, R. and G. Geiss: "The Construction of Liapunov Functions," *IEEE Transactions on Automatic Control,* Vol. AC-8, No. 4, October 1963, pp. 382–383.
25. 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. 365–84.
26. 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.
27. Schultz, D. G. and J. E. Gibson: "The Variable Gradient Method for Generating Liapunov Functions," *AIEE Trans.,* Part II, Vol. 81, September 1962, pp. 203–210.
28. Ogata, K.: *State Space Analysis of Control Systems,* Prentice Hall, Englewood Cliffs, N.J., 1967, Chapter 8.
29. Baumann, W. T. and W. J. Rugh: "Feedback Control of Nonlinear Systems by Extended Linearization," *IEEE Transactions on Automatic Control,* Vol. AC-31, No. 1, January 1986, pp. 40–46.
30. Chang, L-C: *Probabilistic Error Model of Robot End-Effector's Dynamic State,* Unpublished Ph.D Dissertation, University of Nebraska, Lincoln, August 1987.
31. Matsumoto, T., L. O. Chua, and M. Komuro: "A Double Scroll," *IEEE Transactions on Circuits and Systems,* Vol. 32, No. 8, August 1985, pp. 797–818.

ILLUSTRATIVE PROBLEMS

Nonlinear System Stability

15.1 Let the origin be an equilibrium point of a slightly nonlinear system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$. If this system is linearized about the origin, perhaps using Taylor's series, then $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{h}(\mathbf{x})$, where $\mathbf{A}$ is the Jacobian matrix $[\partial\mathbf{f}/\partial\mathbf{x}]$ evaluated at $\mathbf{x} = 0$, and $\mathbf{h}(\mathbf{x})$ represents higher-order terms. Show that if $\|\mathbf{h}(\mathbf{x})\| \le \alpha\|\mathbf{x}\|$ for some positive constant α, then asymptotic stability of the linear equation $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ implies asymptotic stability of the nonlinear equation as well [21, 23].

Let $\mathbf{\Phi}(t, \tau) = e^{\mathbf{A}(t-\tau)}$ be the transition matrix of the linear equation. Then treating $\mathbf{h}(\mathbf{x})$ as a forcing term, the solution for the nonlinear system can be written as

$$\mathbf{x}(t) = \mathbf{\Phi}(t, t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t, \tau)\mathbf{h}(\mathbf{x}(\tau))\, d\tau$$

Therefore,

$$\|\mathbf{x}(t)\| \le \|\mathbf{\Phi}(t, t_0)\|\cdot\|\mathbf{x}(t_0)\| + \int_{t_0}^{t} \|\mathbf{\Phi}(t, \tau)\|\, \|\mathbf{h}(\mathbf{x}(\tau))\|\, d\tau$$

Asymptotic stability of the linear part ensures that $\|\mathbf{\Phi}(t, \tau)\|$ is bounded by a decaying exponential, $\|\mathbf{\Phi}(t, \tau)\| \le Me^{-k(t-\tau)}$ for all $t \ge \tau$, all $\tau \ge t_0$. Using this and the bound on $\|\mathbf{h}(\mathbf{x})\|$ leads to

$$\|\mathbf{x}(t)\| \le Me^{-k(t-t_0)}\|\mathbf{x}(t_0)\| + \int_{t_0}^{t} \alpha Me^{-k(t-\tau)}\|\mathbf{x}(\tau)\|\, d\tau$$

Multiplying by the positive function e^{kt} leaves the sense of the inequality unchanged, so

$$e^{kt}\|\mathbf{x}(t)\| \le Me^{kt_0}\|\mathbf{x}(t_0)\| + \int_{t_0}^{t} \alpha Me^{k\tau}\|\mathbf{x}(\tau)\|\, d\tau \tag{1}$$

Call the right-hand side of equation (*1*) $U(t)$ for convenience. Note that $\dot{U}(t) = \alpha M e^{kt}\|\mathbf{x}(t)\| = \alpha M$ times left-hand side of equation (*1*). Hence

$$\dot{U}/(\alpha M) \le U \quad \text{or} \quad dU/U \le \alpha M\, dt$$

Integrating both sides gives $\ln(U(t)/C) \le \alpha M(t - t_0)$, where the integration constant C is the value of U at $t = t_0$, namely, $C = Me^{kt_0}\|\mathbf{x}(t_0)\|$. Then

$$U(t) \le Ce^{\alpha M(t - t_0)} \le Me^{kt_0}\|\mathbf{x}(t_0)\|e^{\alpha M(t - t_0)}$$

is an explicit upper bound for the right-hand side of equation (*1*). Therefore,

$$e^{kt}\|\mathbf{x}(t)\| \le Me^{kt_0}\|\mathbf{x}(t_0)\|e^{\alpha M(t - t_0)}$$

or

$$\|\mathbf{x}(t)\| \le Me^{-(k - \alpha M)(t - t_0)}\|\mathbf{x}(t_0)\|$$

Thus $\|\mathbf{x}(t)\| \to 0$ provided the bound on $\|\mathbf{h}(\mathbf{x})\|$ is sufficiently small, i.e., if $\alpha < k/M$. The constant α must satisfy $\alpha \ge \|\mathbf{h}(\mathbf{x})\|/\|\mathbf{x}\|$. Since $\mathbf{h}(\mathbf{x})$ is composed of second- or higher-order terms in components of $\mathbf{x}$, this restriction can be made as small as we please by restricting $\|\mathbf{x}\|$ to be sufficiently small.

Thus asymptotic stability of the linear part of the system implies asymptotic stability of the nonlinear system in a sufficiently small neighborhood of the origin. It can also be shown that if the Jacobian matrix has one or more right-half plane eigenvalues, then the equilibrium point is unstable. In summary, the stability of the nonlinear system in the neighborhood of an equilibrium point is the same as the linearized portion, with one exception. If one or more eigenvalues are on the imaginary axis and all others are in the left-half plane, the linearized system gives no definite information about stability. Second-order (and higher) terms need to be evaluated in this case.

15.2 Derive the nonlinear differential equation which relates the angle θ of the inverted pendulum [29] mounted on a cart to the input force f_x on the cart. See Figure 15.12*a* and the free-body diagrams of the two members in Figure 15.12*b* and *c*. The pendulum is of length L and mass m, and its moment of inertia about the center of gravity is J. The cart has mass M. The horizontal and vertical displacements of the center of gravity of the pendulum are $x = X + L/2 \sin\theta$ and $y = L/2 \cos\theta$. Letting the reaction forces at the pendulum support point be F_x and F_y, summing forces on the pendulum gives

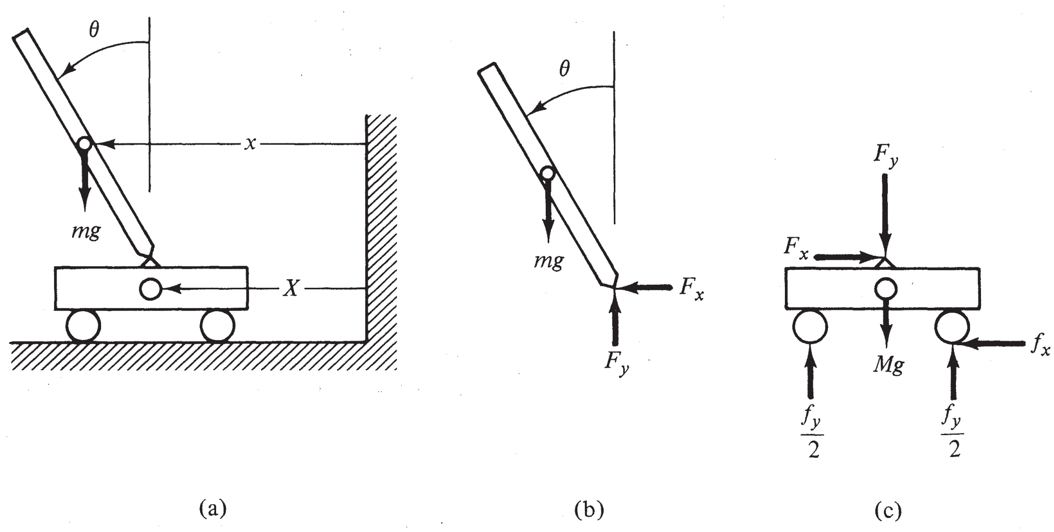


Figure 15.12

$$F_x = m\ddot{X} + \frac{mL}{2}\ddot{\theta}\cos\theta - \frac{mL}{2}\dot{\theta}^2\sin\theta \tag{1}$$

$$F_y - mg = -\frac{mL}{2}\ddot{\theta}\sin\theta - \frac{mL}{2}\dot{\theta}^2\cos\theta \tag{2}$$

$$\frac{F_y L}{2}\sin\theta - \frac{F_x L}{2}\cos\theta = J\ddot{\theta} \tag{3}$$

The external force imparted to the cart by its drive wheels is f_x. By summing horizontal forces on the cart, one obtains

$$\ddot{X} = \frac{f_x - F_x}{M} \tag{4}$$

Substituting Eq. (*4*) into (*1*) and then combining that result with Eq. (*2*) and (*3*) gives a nonlinear second-order differential equation relating the input force f_x to the pendulum angle θ,

$$\left\{J + \frac{mL^2}{4}\sin^2\theta + \frac{mML^2}{4(M+m)}\cos^2\theta\right\}\ddot{\theta} = \frac{mgL}{2}\sin\theta - \frac{m^2L^2}{2(m+M)}\dot{\theta}^2\sin(2\theta) - \frac{mL\cos\theta f_x}{2(M+m)}$$

If the pendulum has a uniform mass distribution, $J = mL^2/12$, and then Eq. (*5*) reduces to

$$\ddot{\theta} = \frac{\{2g\sin\theta - mL/[2(m+M)]\dot{\theta}^2\sin(2\theta) - 2\cos\theta f_x/(M+m)\}}{4L/3 - mL\cos^2\theta/(M+m)} \tag{6}$$

If the length L is written as twice the half-length, Eq. (*6*) becomes identical to the result in [29].

15.3 **(a)** Let $x_1 = \theta$, $x_2 = \dot{\theta}$, and $u = f_x$. Derive the state equations for linear perturbations of the pendulum in Problem 15.2. Use $\theta = \dot{\theta} = u = 0$ as the nominal status.

(b) Use the linearized constant-coefficient perturbations to design state feedback control gains that yield closed-loop poles at $\lambda = -4$ and -5. Use the parameter values [29] $m = 2$ kg, $M = 8$ kg, $L = 1$ m, and $g = 9.8$ m/s.

(c) Use the linear state feedback gains of part (b) with the actual nonlinear system given in part (a). Use computer simulation to investigate the transient response of θ, $\dot{\theta}$, and u for initial values of θ of 0.5, 1, 1.2, and 1.25 rad. Let $\dot{\theta}(0) = 0$.

(a) The nonlinear state equations are

$$\dot{\mathbf{x}} = \begin{bmatrix} x_2 \\ \dfrac{2g\sin x_1 - mL/[2(m+M)]x_2^2\sin(2x_1) - 2\cos x_1 f_x/(M+m)}{D} \end{bmatrix} = \mathbf{f}(\mathbf{x}, u) \tag{1}$$

where the denominator is $D = 4L/3 - mL\cos^2\theta/(M+m)$. Then

$$\left[\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\right]_n = \begin{bmatrix} 0 & 1 \\ \dfrac{2g}{4L/3 - mL/(M+m)} & 0 \end{bmatrix}, \quad \left[\frac{\partial\mathbf{f}}{\partial u}\right]_n = \begin{bmatrix} 0 \\ \dfrac{-2}{4L(M+m)/3 - mL} \end{bmatrix}$$

(b) Using the pole placement algorithm of Sec. 13.4, the feedback gain matrix is found to be $\mathbf{K} = [-211.333 \quad -51]$. If, instead, an infinite horizon **LQ** optimum problem (Sec. 14.4.2) had been specified with weights $\mathbf{Q} = \mathbf{I}$ and $\mathbf{R} = 1$, the gain would then be $\mathbf{K} = [-196 \quad -47.142]$, and the resulting closed-loop poles of the linear system would be $\lambda = [-4.07$ and $-4.24]$.

(c) When the first set of gains is used to form $u(t) = -\mathbf{Kx}$ in Eq. (*1*) with values $x_1(0) = 0.5$, 1, and 1.2 and with $x_2(0) = 0$, the three response curves of Figure 15.13*a* are obtained for $x_1 = \theta(t)$. Figure 15.13*b* gives the corresponding $x_2(t) = \dot{\theta}(t)$. Figure 15.13*c* shows the input

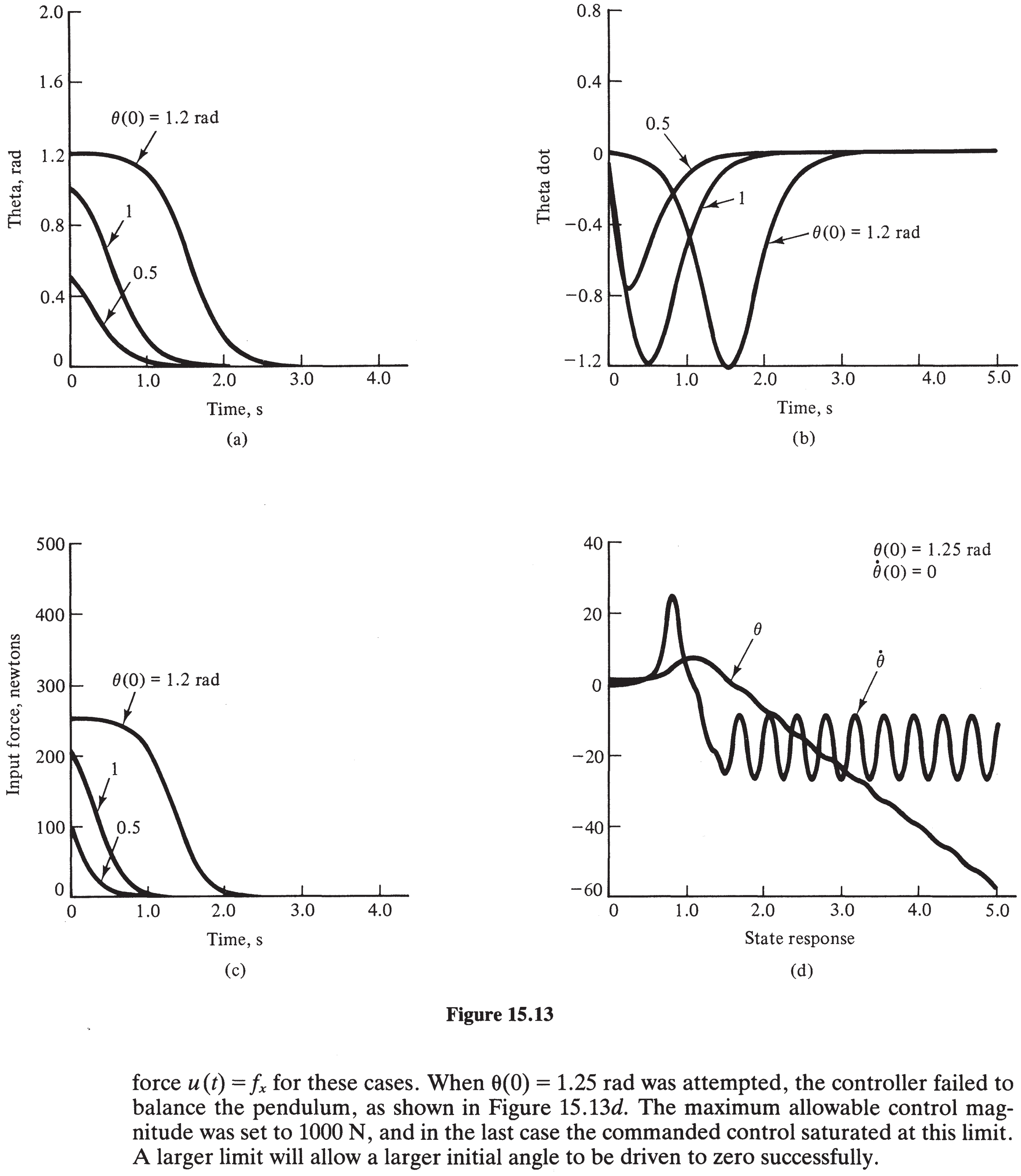


Figure 15.13

force $u(t) = f_x$ for these cases. When $\theta(0) = 1.25$ rad was attempted, the controller failed to balance the pendulum, as shown in Figure 15.13*d*. The maximum allowable control magnitude was set to 1000 N, and in the last case the commanded control saturated at this limit. A larger limit will allow a larger initial angle to be driven to zero successfully.

15.4 The two-link mechanism in Figure 15.14 demonstrates nonlinearities typical of many robot devices. The equations governing the motion as a function of the two input joint torques can be derived by any of several methods [30]. By summing forces and torques on each link it can be shown that

$$T_1 = \left[\frac{m_1}{2} + m_2\right] gL_1 \cos\theta + \frac{m_2 g L_2}{2} \cos\psi$$

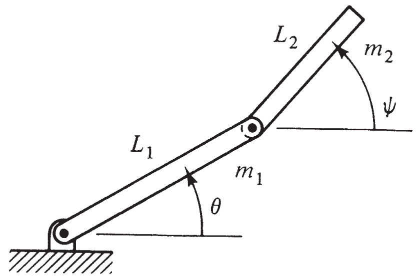


Figure 15.14

$$+\ddot{\theta}\left\{\frac{m_1 L_1^2}{3} + m_2\left[L_1^2 + \frac{L_1 L_2 \cos(\psi - \theta)}{2}\right]\right\} + \ddot{\psi}\left\{m_2\left[\frac{L_2^2}{3} + \frac{L_1 L_2 \cos(\psi - \theta)}{2}\right]\right\}$$

$$-\frac{\dot{\psi}^2 m_2 L_1 L_2 \sin(\psi - \theta)}{2} + \frac{\dot{\theta}^2 L_1 L_2 \sin(\psi - \theta)}{2} \quad (1)$$

and

$$T_2 = \frac{m_2 g L_2 \cos(\psi)}{2} + \frac{\ddot{\theta} m_2 L_1 L_2 \cos(\psi - \theta)}{2} + \frac{m_2 L_2^2 \ddot{\psi}}{3} + \frac{m_2 L_1 L_2 \dot{\theta}^2 \sin(\psi - \theta)}{2} \quad (2)$$

Let $x_1 = \theta$, $x_2 = \psi$, $x_3 = \dot{\theta}$, and $x_4 = \dot{\psi}$ and write equations (*1*) and (*2*) in state variable form.

Since $\dot{x}_3$ and $\dot{x}_4$ appear in both equations, we must first separate out these variables. Rewrite Eqs. (*1*) and (*2*) as

$$D_{11}(\mathbf{x})\ddot{\theta} + D_{12}(\mathbf{x})\ddot{\psi} = F_1(\mathbf{x}) + T_1$$

$$D_{21}(\mathbf{x})\ddot{\theta} + D_{22}(\mathbf{x})\ddot{\psi} = F_2(\mathbf{x}) + T_2$$

where the coefficients of the second derivatives have been redefined as D_{ij} and where all other terms except the torques have been combined into the definitions of the two F_i terms. The state equations can then be written as

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \\ \dot{x}_4 \end{bmatrix} = \begin{bmatrix} x_3 \\ x_4 \\ \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}^{-1} \begin{bmatrix} F_1(\mathbf{x}) + T_1 \\ F_2(\mathbf{x}) + T_2 \end{bmatrix} \end{bmatrix} = \mathbf{f}(\mathbf{x}, \mathbf{T})$$

These nonlinear equations can be linearized for analyzing small motions about some nominal conditions by using the methods of Sec. 15.2, leading to

$$\delta\dot{\mathbf{x}} = [\partial\mathbf{f}/\partial\mathbf{x}]_n\,\delta\mathbf{x} + [\partial\mathbf{f}/\partial\mathbf{T}]_n\,\delta\mathbf{T}$$

Details of evaluating the partial derivatives $\partial\mathbf{f}/\partial\mathbf{x}$ and $\partial\mathbf{f}/\partial\mathbf{T}$ are left as an exercise; they may also be found in [30].

15.5 Consider an nth-order nonlinear system which is linear in the r control components, $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}) + \mathbf{Bu}$. Investigate conditions under which it is possible to find a control function satisfying $\mathbf{f}(\mathbf{x}) + \mathbf{Bu} - \mathbf{Fy}_d - \mathbf{Gv} = \mathbf{S}(\mathbf{x} - \mathbf{y}_d)$ as required in dynamic linearization.

Rewriting this as $\mathbf{Bu} = \mathbf{w}$, where $\mathbf{w} = -\mathbf{f}(\mathbf{x}) + \mathbf{Fy}_d + \mathbf{Gv} + \mathbf{S}(\mathbf{x} - \mathbf{y}_d)$, the results of Chapter 6 indicate that a unique solution exists for any *arbitrary* $\mathbf{w}$ if and only if $\mathbf{B}$ is square and nonsingular. This is a very restrictive condition which is rarely met with control systems of order higher than one. Again from Chapter 6 results, for a *specific* $\mathbf{w}$, solutions will exist if rank$[\mathbf{B}]$ = rank$[\mathbf{B} \mid \mathbf{w}]$. A unique $\mathbf{u}$ will exist for $\mathbf{w}$ if, in addition, rank$[\mathbf{B}] = r$, the number of control components. Some influence can be exerted on $\mathbf{w}$ through choice of $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{S}$. What is desired is an ability to find $\mathbf{u}$ for arbitrary $\mathbf{x}$ and $\mathbf{y}_d$, and this can be achieved under certain conditions if compatible choices are made for the forms of $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{S}$. For example, consider an nth-order system in phase variable form, with one input, $\overset{(n)}{y} = f(y, \dot{y}, \ldots, \overset{(n-1)}{y}) + u$, which becomes in state variable form

$$\dot{\mathbf{x}} = \begin{bmatrix} x_2 \\ x_3 \\ x_4 \\ \vdots \\ x_n \\ f(\mathbf{x}) \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} u$$

Clearly rank$[\mathbf{B}] = 1$ and $[\mathbf{B} \mid (\mathbf{F} - \mathbf{S})\mathbf{y}_d + \mathbf{Sx} - \mathbf{f}(\mathbf{x}) + \mathbf{Gv}]$ will also have rank $= 1$ for all $\mathbf{x}$ and $\mathbf{y}_d$ if both $\mathbf{F}$ and $\mathbf{S}$ are in companion form and if the first $n - 1$ rows of $\mathbf{Gv}$ are zero. This ensures that the first $n - 1$ rows of $\mathbf{Bu} = \mathbf{w}$ are identically zero. The last row gives a scalar equation that can be solved for $u(t)$.

The preceding results easily generalize to any number r of coupled pth-order scalar differential equations. We illustrate with just two equations in terms of two input variables,

$$\dddot{y}_1 = f_1(\ddot{y}_1, \dot{y}_1, y_1, \dot{y}_2, y_2) + b_{11} u_1 + b_{12} u_2$$

$$\ddot{y}_2 = f_2(\ddot{y}_1, \dot{y}_1, y_1, \dot{y}_2, y_2) + b_{21} u_1 + b_{22} u_2$$

By picking state variables $x_1 = y_1$, $x_2 = \dot{y}_1$, $x_3 = \ddot{y}_1$, $x_4 = y_2$, and $x_5 = \dot{y}_2$, the state equations take the form

$$\dot{\mathbf{x}} = \begin{bmatrix} x_2 \\ x_3 \\ f_1(\mathbf{x}) + b_{11} u_1 + b_{12} u_2 \\ x_4 \\ f_2(\mathbf{x}) + b_{21} u_1 + b_{22} u_2 \end{bmatrix}$$

Since rows 1, 2, and 4 of the matrix $\mathbf{B}$ are zero, rank$[\mathbf{B}] \leq 2$. Consideration of the solvability condition rank$[\mathbf{B}] =$ rank$[\mathbf{B} \mid \mathbf{w}]$ suggests a compatible choice for $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{S}$ should make rows 1, 2, and 4 of $\mathbf{w}$ zero also. This can be accomplished by selecting

$$\mathbf{F} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ F_{31} & F_{32} & F_{33} & F_{34} & F_{35} \\ 0 & 0 & 0 & 0 & 1 \\ F_{51} & F_{52} & F_{53} & F_{54} & F_{55} \end{bmatrix}, \quad \mathbf{S} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ S_{31} & S_{32} & S_{33} & S_{34} & S_{35} \\ 0 & 0 & 0 & 0 & 1 \\ S_{51} & S_{52} & S_{53} & S_{54} & S_{55} \end{bmatrix}, \quad \mathbf{G} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ G_{31} & G_{32} \\ 0 & 0 \\ G_{51} & G_{52} \end{bmatrix}$$

Thus there are only two nontrivial rows which contribute to the solution for $\mathbf{u}$, and a unique solution exists for all $\mathbf{x}$, $\mathbf{y}_d$, and $\mathbf{v}$ provided the 2×2 matrix $[b_{ij}]$ is nonsingular.

15.6 The system in Problem 15.5 is now generalized to allow it to be nonlinear in the control variables: $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, \mathbf{u})$. What can be said about the existence of a dynamic linearizing control law?

If the function $\mathbf{f}(\mathbf{x}, \mathbf{u})$ is continuous in some region $\mathcal{R}$ containing the point $\mathbf{x}_0$, $\mathbf{u}_0$, and if it is continuously differentiable in $\mathcal{R}$, the results of Problem 15.5 remain valid in the region $\mathcal{R}$, with $[\partial\mathbf{f}/\partial\mathbf{u}]_0$ replacing the previous matrix $\mathbf{B}$. This follows from the implicit function theorem [8]. If there are r input components and n states with $r < n$, then the $n \times r$ matrix $[\partial\mathbf{f}/\partial\mathbf{u}]_0$ must have full rank r. The augmented matrix $\mathbf{W}$ must also have rank r. This can sometimes be ensured by selecting $\mathbf{F}$, $\mathbf{G}$, and $\mathbf{S}$ in a form compatible with the given nonlinear state equations, just as in the case of linear control terms in Problem 15.5. This restricts the number and location of the nonlinear terms which can be dealt with in the state equations. Problem 15.7 shows a system where this condition cannot be met. When rank$[\mathbf{B}] \neq$ rank$[\mathbf{W}]$, the linearizing equations are inconsistent and no exact solution exists. Two choices may be considered in this situation. Certain equations can be ignored or combined so that a smaller set of consistent equations remains. This can be accomplished by premultiplying $\mathbf{f}(\mathbf{x}, \mathbf{u})$ by an $r \times n$ matrix $\mathbf{H}$. This means that only r states, or combinations of states, are being matched to an rth-order template system. Of course the resulting nth-order closed-loop system will generally not be linear in this case. The second possibility is to use a least-squares approximate solution to the full nth-order template-matching problem. This can be done provided **B** is of full rank, but the result will not be exactly linear, and its performance may not be satisfactory. Both possibilities are demonstrated in Problem 15.7.

15.7 Consider the problem of applying dynamic linearization to the system

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & -1 & -1 \\ 1 & 0 & 0 \\ 0 & 0 & -4 \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ ux_2 \\ 2 + x_1x_3 \end{bmatrix}$$

For this system, $\partial\mathbf{f}/\partial u = [0 \quad x_2 \quad 0]^T$ has rank 1, except at $x_2 = 0$. Clearly any choice for $u(t)$ will have no effect on the third component of $\dot{\mathbf{x}}$ in this case, so exact linearization by matching to a third-order linear template is not possible. By selecting

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{or} \quad \mathbf{H} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

we can match a second-order template system to either the first two components of **x**, or with the second **H**, we would match x_1 to y_{d_1} and $x_2 + x_3$ to y_{d_2}. The second **H** is selected, along with a template system described by

$$\dot{\mathbf{y}}_d = \begin{bmatrix} 0 & -1 \\ 2 & -3 \end{bmatrix}\mathbf{y}_d + \begin{bmatrix} 0 \\ 1 \end{bmatrix}v$$

This is a stable system with eigenvalues at $\lambda = -1$ and -2. The steady-state solution for a step input of magnitude V is $\mathbf{y}_d = [-V/2 \quad 0]^T$. The convergence matrix is initially selected as

$$\mathbf{S} = \begin{bmatrix} 0 & -1 \\ 25 & -10 \end{bmatrix}$$

but $\mathbf{S} = \mathbf{F}$ will also be tested. Except when $x_2 = 0$, the control is given by

$$u(t) = \frac{2y_{d_1} - 3y_{d_2} + v + S_{21}(x_1 - y_{d_1}) + S_{22}(x_2 + x_3 - y_{d_2}) - x_1 + 4x_3 - 2 - x_1x_3}{x_2}$$

The nonlinear system equations were numerically integrated using fourth-order Runge-Kutta with a step size of 0.02. The magnitude of $u(t)$ was limited to 100. Figure 15.15*a* compares y_{d_1} and the state x_1 obtained with the two matrices **S** mentioned. Figure 15.15*b* shows y_{d_2} along with two sets of x_2, x_3 responses. The control command is shown in Figure 15.15*c*. It saturated immediately with the fast **S** and remained at either 100 or -100 for the first second and thereafter remained well within the bounds. When $\mathbf{S} = \mathbf{F}$ was selected, the system response was much slower and smoother. The commanded control signal gradually increased until saturation occurred at about 3 s.

Attempting to least-squares fit to a third-order template system in this problem is not productive, since $u = [\mathbf{B}\mathbf{B}^T]^{-1}\mathbf{B}^T\{\dots\}$ amounts to ignoring the first and third components of the matching equations.

15.8 Use the method of dynamic linearization to control the inverted pendulum of Problems 15.2 and 15.3. Compare the results with the linear perturbation controller of Problem 15.3.

The controller in Problem 15.3 was designed to yield closed-loop eigenvalues at $\lambda = -4$ and -5. For comparison purposes, a linear template system which has these same eigenvalues is specified, namely,

$$\dot{\mathbf{y}}_d = \begin{bmatrix} 0 & 1 \\ -20 & -9 \end{bmatrix}\mathbf{y}_d + \begin{bmatrix} 0 \\ 1 \end{bmatrix}v.$$

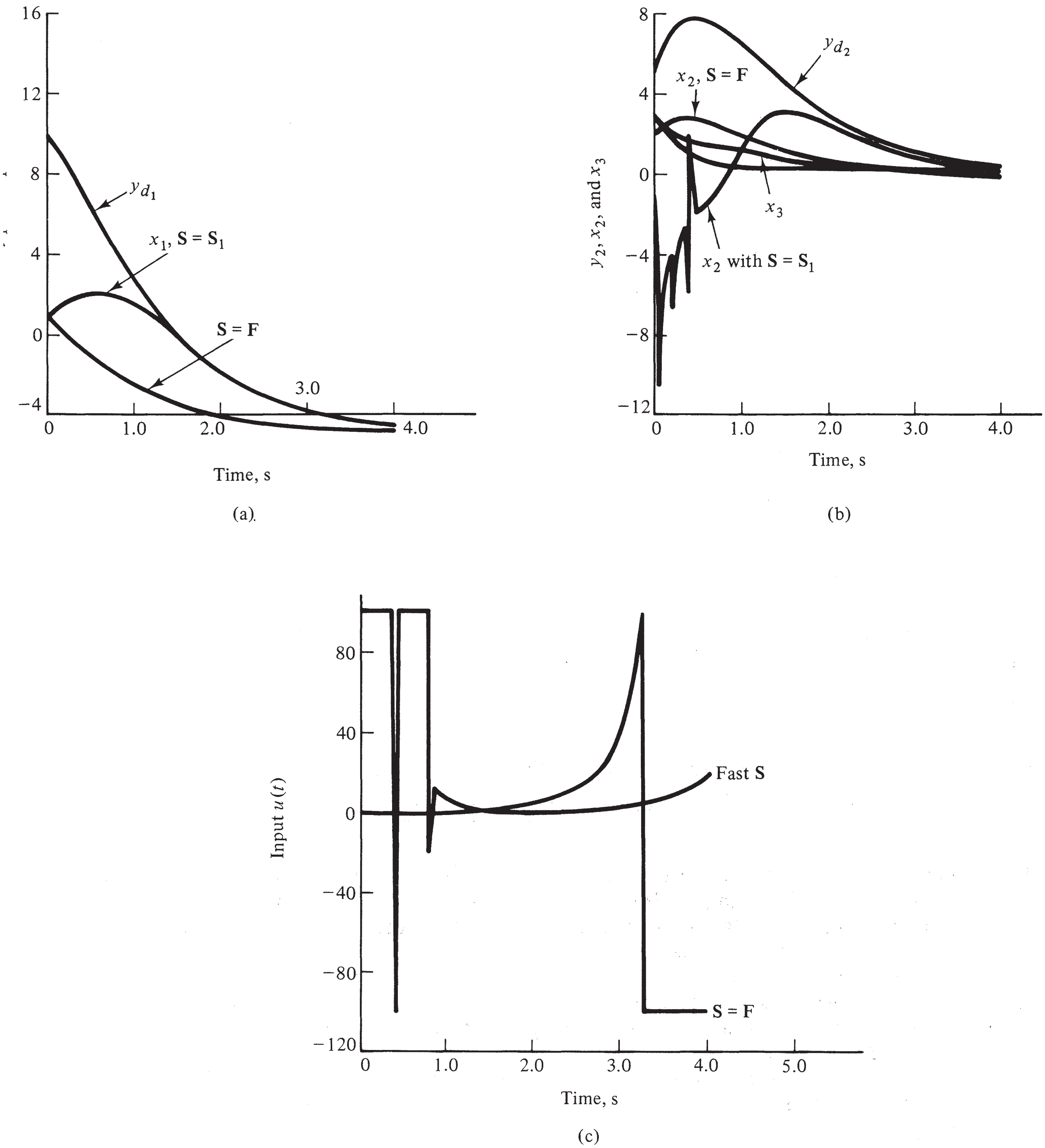


Figure 15.15

The nonzero convergence matrix **S** will also be specified to be in the companion form

$$\mathbf{S} = \begin{bmatrix} 0 & 1 \\ S_{21} & S_{22} \end{bmatrix}$$

This is not a totally arbitrary choice. The first component of Eq. (*15.11*) here requires that the first rows of **F** and **S** agree so that $x_2 = y_{d_2} + x_2 - y_{d_2}$. The second component of Eq. (*15.11*) is the only one containing the unknown control $u(t)$. If we define $w = (F_{21} - S_{21})y_{d_1} + (F_{22} - S_{22})y_{d_2} + S_{21}x_1 + S_{22}x_2 + v$ and the denominator D of Problem 15.3, this second component can be written as

$$2g \sin x_1 - \frac{mLx_2^2 \sin(2x_1)}{2(m+M)} - \frac{2 \cos x_1 u}{M+m} = Dw$$

from which, if $\cos x_1 \neq 0$, i.e., if $x_1 = \theta \neq \pi/2$, the control law is

$$u = \frac{(m+M)\{2g \sin x_1 - mL/[2(m+M)]x_2^2 \sin(2x_1) - Dw\}}{\cos x_1}$$

Unless stated otherwise, $v = 0$ in the following simulation tests of this control law. Three values of the matrix **S** are tested.

1. With $S_{21} = -100$, $S_{22} = -20$ ($\lambda = -10$ is a double root). The typical result of Figure 15.16*a* is obtained, using $\mathbf{x}(0) = [-0.5 \quad 0]^T$ and $\mathbf{y}_d(0) = [0.5 \quad 0]^T$. It is seen that $\mathbf{x}(t)$ and $\mathbf{y}_d(t)$ come together in a fraction of a second, as determined by **S**, and that they both settle to zero in less than 2 s, as determined by **F**.
2. With $\mathbf{S} \equiv \mathbf{F}$ but all other parameters as before, the response of Figure 15.16*b* is obtained. The settling times are noticeably longer.
3. With $\mathbf{S} = [0]$ and all other parameters as before, $x_2(t) \rightarrow y_{d_2}(t)$, but the first component of **x**—i.e., θ—does not go to zero, as shown in Figure 15.16*c*. The derivatives of **x** and $\mathbf{y}_d$ do match, and since the second components have equal initial conditions, these terms remain in perfect agreement. Having θ converge to -1 rad is not a satisfactory solution to the balancing problem. This points out the utility of the convergence matrix **S**.

The preceding three tests were repeated with the template input $v(t)$ being a periodic square wave. Results are shown in Figures 15.17*a, b, c*. Again, the first two cases successfully track the desired response, but (2) is considerably slower and allows a much bigger transient error to build up before settling. With $\mathbf{S} = [\mathbf{0}]$, the pendulum fails to achieve vertical balance. When the original angle error is allowed to increase to 1.48 rad, the controller is no longer able to balance the pendulum (the first **S** was used), but at $\theta(0) = 1.47$ rad, it performed correctly. Note that this is a much larger error that could be nulled by the linear controller of Problem 15.3. The actual response is shown in Figure 15.18, with θ oscillating around $\pi/2$; the input force history commanded by the controller is a square wave oscillating between the ± 1000-N limits.

15.9 Find the describing function for the general hysteresis-type nonlinearity in Figure 15.19*a*, which can be described by the five parameters α, β, K_1, K_2, and K_3. Note that the symmetry of the parallelogram gives the expression $\gamma = [K_1(\beta + \alpha) + K_2(\beta - \alpha)]/2$. Assume the amplitude E of the input sinusoid is larger than β.

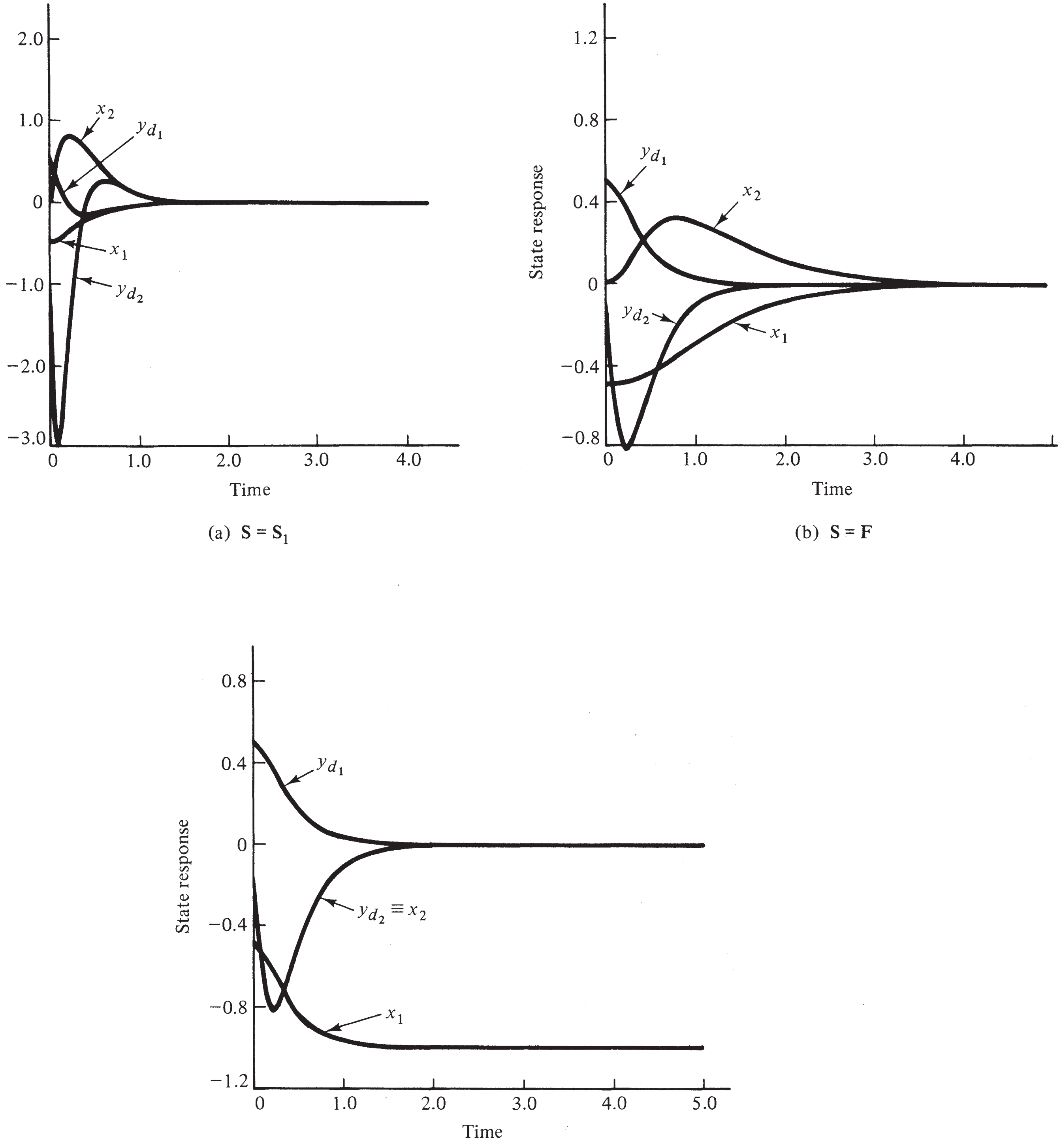
There are seven time segments of interest in each cycle of the sine wave in Figure 15.19*b*.

1. $0 < t < t_1$; $0 < e(t) < \alpha$, $n(e) = -\gamma + K_1[\beta + e(t)]$
2. $t_1 < t < t_2$; $\alpha < e(t) < \beta$, $n(e) = -\gamma + K_1[\alpha + \beta] + K_2[e(t) - \alpha]$
3. $t_2 < t < t_3$; $\beta < e(t)$, $n(e) = \gamma + K_3[e(t) - \beta]$
4. $t_3 < t < t_4$; $-\alpha < e(t) < \beta$, $n(e) = \gamma + K_1[e(t) - \beta]$
5. $t_4 < t < t_5$; $-\beta < e(t) < -\alpha$, $n(e) = -\gamma + K_2[e(t) + \beta]$
6. $t_5 < t < t_6$; $e(t) < -\beta$, $n(e) = -\gamma + K_3[e(t) + \beta]$
7. $t_6 < t < 2\pi$; $-\beta < e(t) < 0$, $n(e) = -\gamma + K_1[e(t) + \beta]$

where $t_1 = \sin^{-1}(\alpha/E)$, $t_2 = \sin^{-1}(\beta/E)$, $t_3 = \pi - t_1$, $t_4 = \pi + t_1$, $t_5 = \pi + t_2$, $t_6 = 2\pi - t_2$. These times are all normalized times.

The actual times are obtained by dividing by the frequency ω. Using $e(t) = E \sin(t)$, the Fourier coefficient a_1 is composed of the sum of seven terms, each of the form

$$\left(\frac{1}{\pi}\right)\int \{[C_0 + C_1 \sin t] \sin t\}\, dt = \left(\frac{1}{\pi}\right)\left\{C_0(\cos t_a - \cos t_b) + C_1(t_b - t_a) + \frac{C_1}{4}[\sin(2t_a) - \sin(2t_b)]\right\}$$



(a) $\mathbf{S} = \mathbf{S}_1$

(b) $\mathbf{S} = \mathbf{F}$

(c) $\mathbf{S} = [0]$

Figure 15.16

This will be defined as a function $S_f(t_a, t_b, C_0, C_1)$ to avoid repetition. Then a_1 is the sum of seven $S_f(\ )$ functions, each with appropriate arguments. For example, on the first interval, $t_a = t_1$, $t_b = t_2$, $C_0 = -\gamma + K_1\beta$, and $C_1 = K_1 E$. On interval 2, $t_a = t_2$, $t_b = t_3$, $C_0 = -\gamma + K_1(\alpha + \beta) - K_2\alpha$, and $C_1 = K_2 E$. The others are similar. The coefficient b_1 is given similarly by the sum of seven terms, each of the form

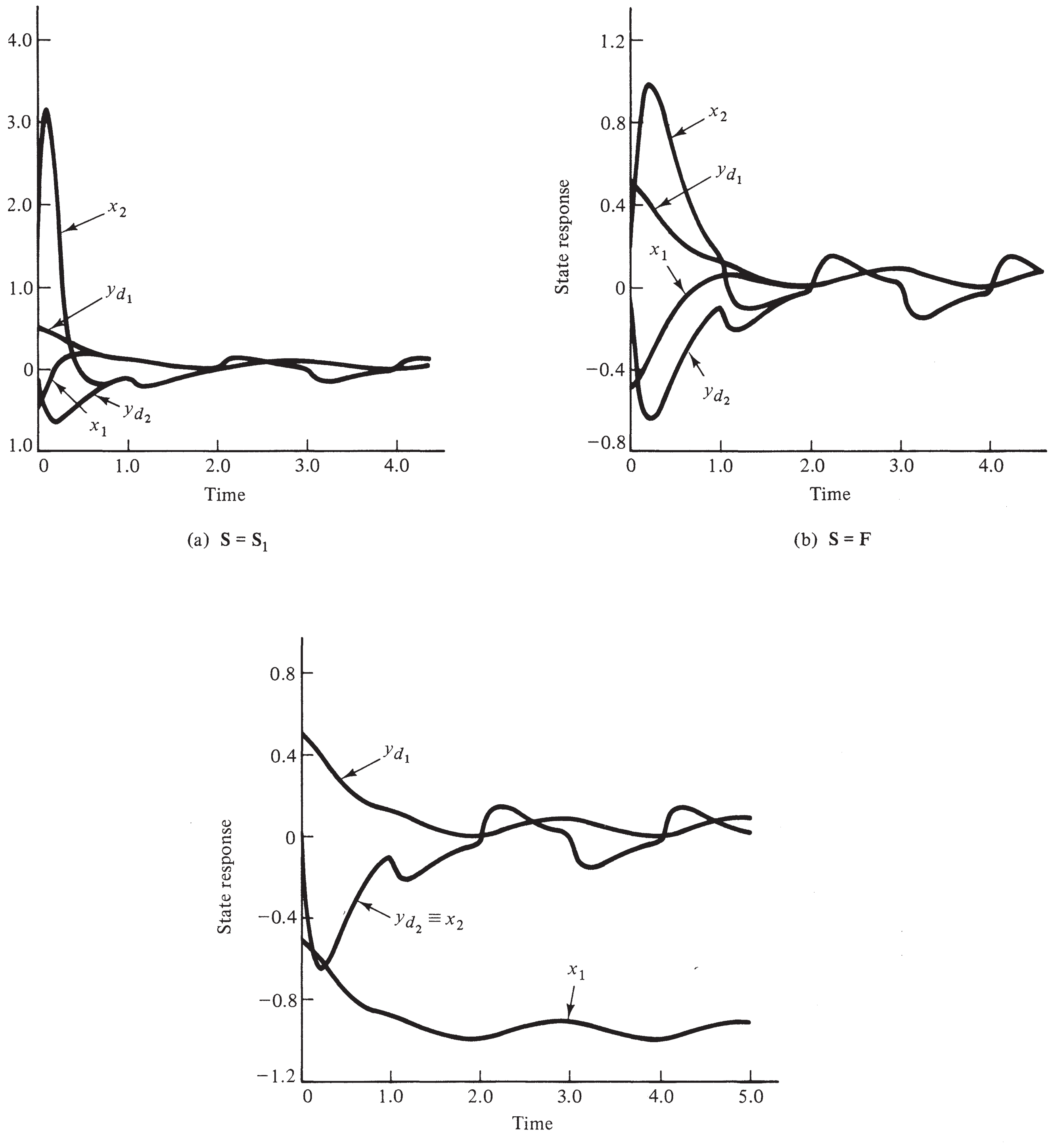


Figure 15.17

$$\left(\frac{1}{\pi}\right)\int\{[C_0 + C_1 \sin t] \cos t\}\, dt = \left(\frac{C_0}{\pi}\right)(\sin t_b - \sin t_a) + \frac{C_1}{(4\pi)}[\cos(2t_a) - \cos(2t_b)]$$

$$\equiv C_f(t_a, t_b, C_0, C_1)$$

The same arguments are used to evaluate the seven cosine coefficients $C_f(\ )$ and the sine coefficients $S_f(\ )$ on each interval.

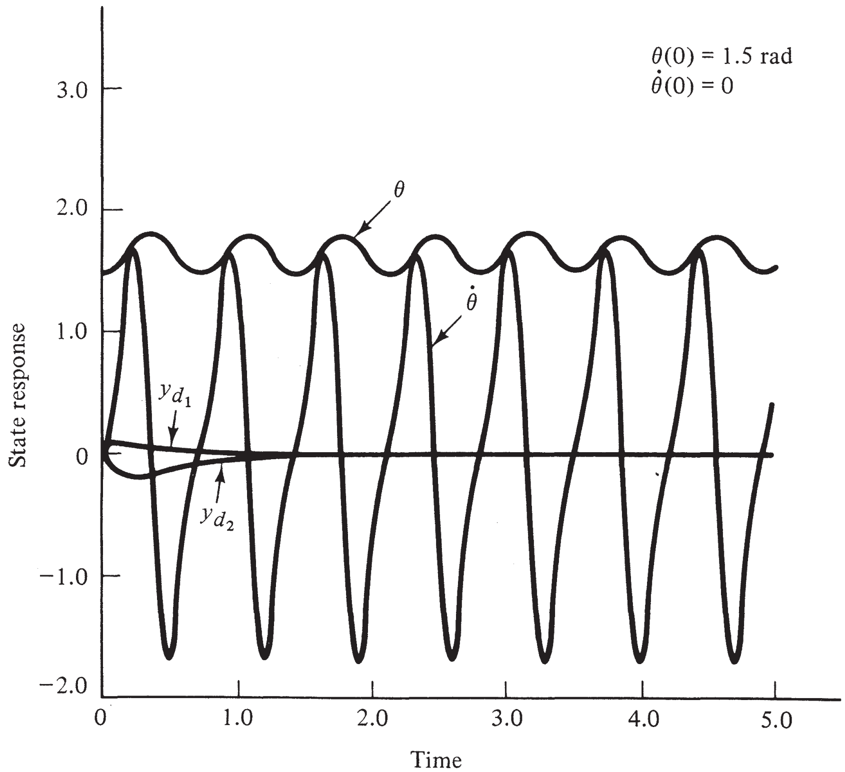


Figure 15.18

15.10 Certain nonlinearities with multiple but related inputs can be treated as if there were a single input, and the previous methods of finding describing functions can be applied. Demonstrate this by finding the describing function for $n(e, \dot{e}) = e^2 \dot{e}$.

If $e(t) = E \sin(\omega t)$, then $\dot{e}(t) = E\omega \cos(\omega t)$, so that $n(e, \dot{e})$ can be treated as having a single input $n(e) = E^3 \omega \sin^2(\omega t) \cos(\omega t)$. In this case no integration is required to find the describing function. Trigonometric identities give

$$n(e) = \left(\frac{E^3 \omega}{2}\right) \sin(\omega t) \sin(2\omega t) = \left(\frac{E^3 \omega}{4}\right)[\cos(\omega t) - \cos(3\omega t)]$$

The fundamental component is now obvious. It has a 90° phase shift relative to the input. Thus the describing function is

$$N(E, \omega) = \left(\frac{E^2 \omega}{4}\right) e^{j\pi/2} = \frac{E^2 \omega j}{4}$$

This example is frequency-dependent and purely imaginary.

Since the product of two sinusoids of frequencies ω_1 and ω_2 produces terms with frequencies $\omega_1 - \omega_2$ and $\omega_1 + \omega_2$, the foregoing procedure will fail to produce a valid describing function in many cases. For example $e(t)^2 = E^2 \sin^2(\omega t)$ gives a dc term and a double-frequency term but no fundamental component.

15.11 Use the describing function of Problem 15.10 to analyze the Van der Pol equation for possible limit cycles.

$$\ddot{y} + \mu(y^2 - 1)\dot{y} + \beta y = 0; \text{ both } \mu \text{ and } \beta \text{ are positive.}$$

Separating the linear and nonlinear parts gives $\ddot{y} - \mu\dot{y} + \beta y = -\mu y^2 \dot{y}$, which has the representation of Figure 15.5, with $n(e) = \mu y^2 \dot{y}$ and $G(s) = 1/[s^2 - \mu s + \beta]$. Thus the condition for existence of a limit cycle is that $-1/G(j\omega) = N(E, \omega)$. Equating imaginary parts gives $\mu\omega = \mu\omega E^2/4$, from which the approximate amplitude of the limit cycle is $E = 2$. By equating

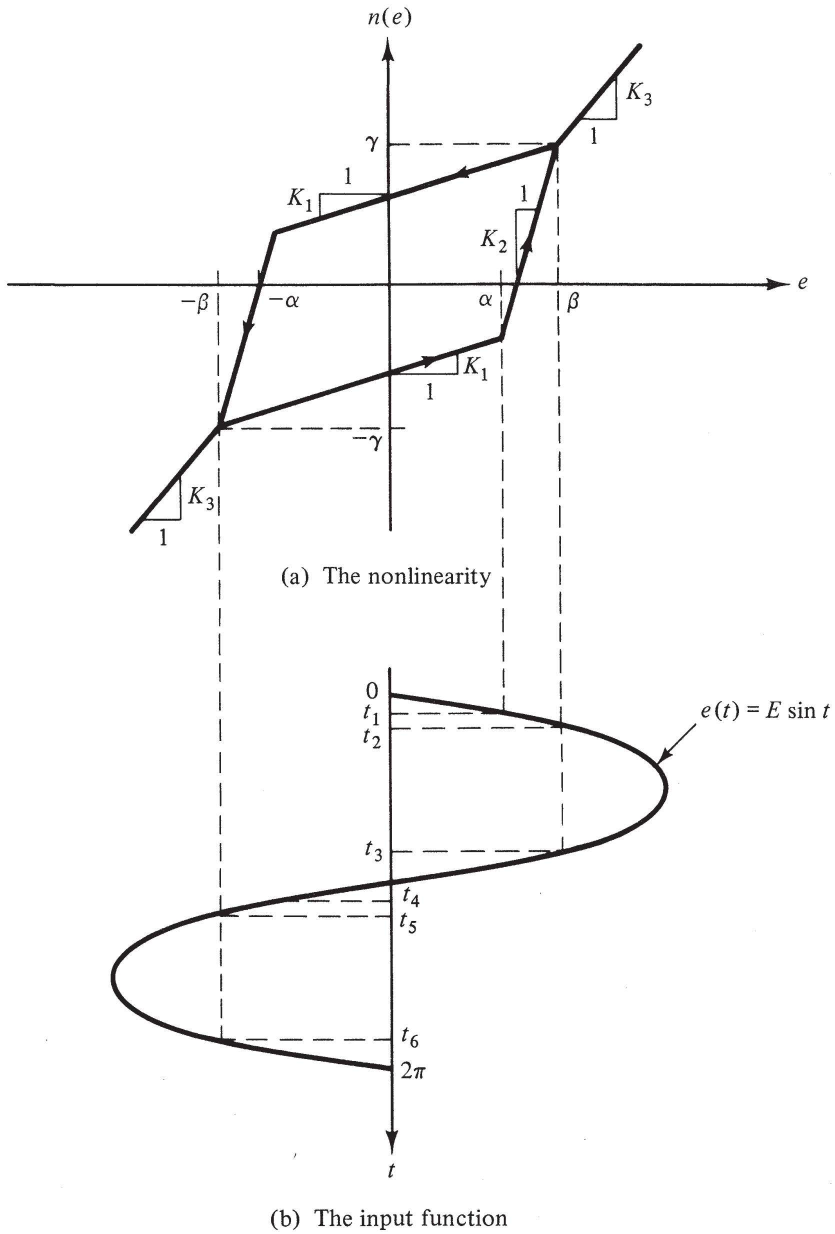


Figure 15.19

real parts, it is found that the limit cycle frequency (at least the fundamental) is $\omega = \sqrt{\beta}$. Sketches of the Nyquist plot and the $-1/N$ locus are given in Figure 15.20. The solution just found identifies the intersection of these two curves.

15.12 Is the limit cycle of Problem 15.11 stable or unstable?

Any critical point inside the closed contour formed by the plot of $G(j\omega)$ in Figure 15.20, for $-\infty < \omega < \infty$, is encircled once in the counterclockwise direction—i.e., $N_c = 1$. Any critical point outside the closed contour has $N_c = 0$. Since $G(s)$ has two open-loop right-half plane poles, $P_r = 2$. Thus application of Nyquist's criterion (Eq. (*15.18*)) to $G(j\omega)$, using $-1/N$ as the critical point, indicates the following:

1. The closed-loop system is unstable, with $Z_r = 2$ if $-1/N$ is outside the closed contour. This is *correct.*
2. The closed-loop system is unstable with $Z_r = 1$ if $-1/N$ is inside the closed contour. This is *incorrect.*

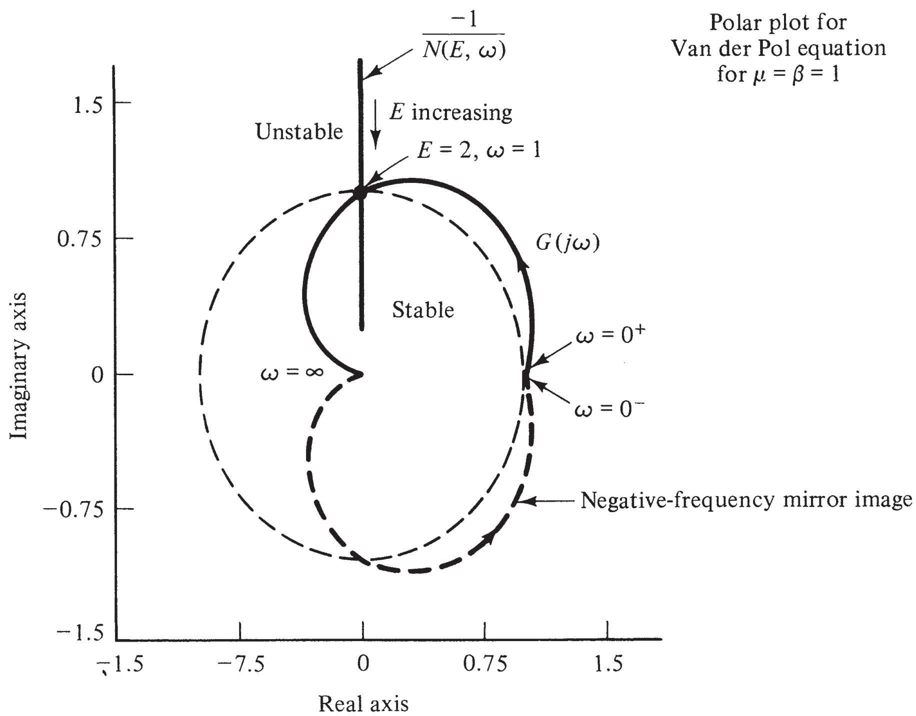


Figure 15.20

Note that $N(E, \omega)$ factors into $N_1(E)N_2(\omega)$. To analyze this system correctly, a modified

$$G'(j\omega) = \frac{j\omega}{-\omega^2 - \mu j\omega + \beta}$$

which incorporates the frequency-dependent part of $N(E, \omega)$, is defined. The remaining $N_1(E) = E^2/4$ will be used to define a real critical point $-1/N_1$. The polar plot of $G'(j\omega)$ is a double loop, one as ω varies from 0 to ∞ and the other, the mirror image for negative ω. Figure 15.21 shows the modified plot. Now critical points inside the contour have $N_c = 2$, and hence $Z_r = 0$, indicating stability. Critical points outside the contour are unstable. Thus stability of the equivalent gain closed-loop system requires that the critical point $-1/N$ must be *inside* the contour. (This is opposite the usual situation.) Now the stability of the limit cycle can be determined. Assume that E and ω are at the intersection point, i.e., we have a limit cycle. Then if the amplitude E increases slightly, $-1/N$ moves inside the contour, causing the system to become asymptotically stable and causing E to decrease back toward the limit cycle value. If E decreases a little, $1/N$ increases, and the critical point $-1/N$ moves outside the Nyquist contour. This gives an unstable system, causing E to grow back toward the intersection point. Thus the limit cycle is stable.

15.13 The circuit of Figure 15.22 has been widely used as an example of chaotic behavior [16, 17]. It is linear except for one nonlinear resistor. Its current will be modeled as a nonlinear function of the voltage across it, $i = n(v_{C1})$. The nonlinearity is piecewise linear, of the type treated in Example 15.5. Selecting states shown in the diagram, the state equations for this system are

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \left(\frac{1}{RC_1}\right)(x_2 - x_1) - \frac{n(x_1)}{C_1} \\ \left(\frac{1}{RC_2}\right)(x_1 - x_2) + \frac{x_3}{C_2} \\ -\left(\frac{1}{L}\right)x_2 \end{bmatrix}$$

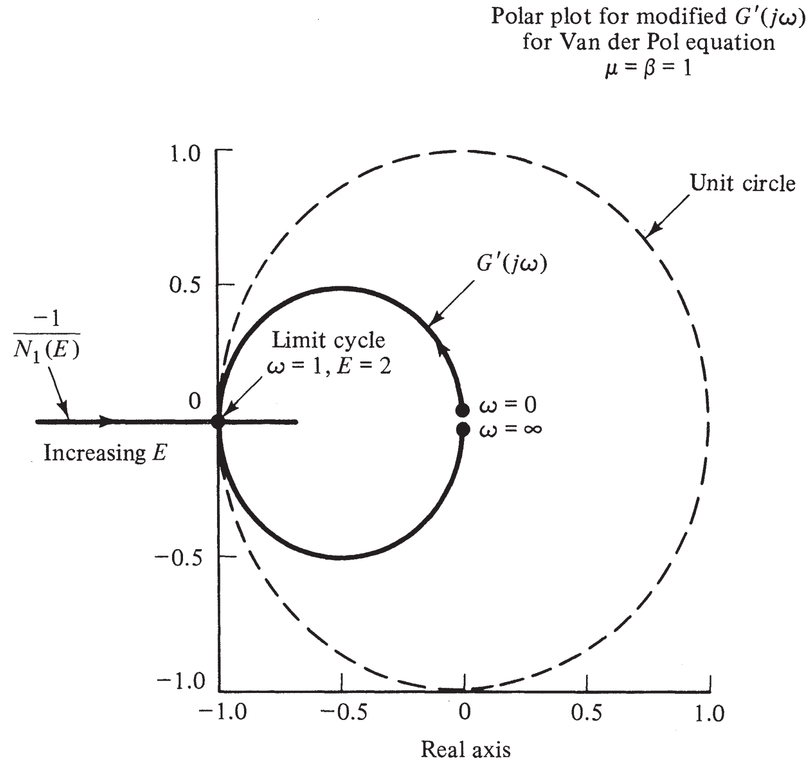


Figure 15.21

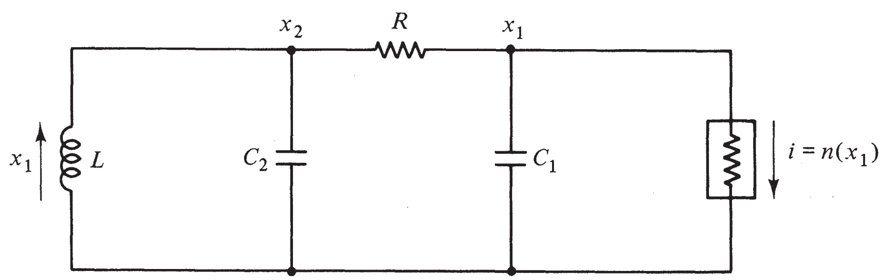


Figure 15.22 Chua's circuit example

Use Chua's values, $C_1 = \frac{1}{9}$, $C_2 = 1$, $L = \frac{1}{7}$, $\alpha = 1$, $K_1 = -0.8$, $R = 1/0.7$, $K_2 = -0.5$, $\beta = \infty$, and K_3 arbitrary. Find the equilibrium points of this nonlinear system and determine their stability (for small perturbations).

Using the given values, the state equations reduce to

$$\dot{\mathbf{x}} = \begin{bmatrix} -6.3 & 6.3 & 0 \\ 0.7 & -0.7 & 1 \\ 0 & -7 & 0 \end{bmatrix}\mathbf{x} - \begin{bmatrix} 9 \\ 0 \\ 0 \end{bmatrix} n(x_1) = \mathbf{A}\mathbf{x} - \mathbf{B}n(x_1)$$

$\dot{\mathbf{x}} = \mathbf{0}$ implies that $x_2 = 0$, $-6.3x_1 - 9n(x_1) = 0$ and that $0.7x_1 + x_3 = 0$. One equilibrium point is $\mathbf{x}_{e_1} = \mathbf{0}$. Other solutions are found from $x_1 = -9n(x_1)/6.3$. On the first linear segment, $n(x_1) = -0.8x_1$, and the only solution is $x_1 = 0$. On the second segment, $n(x_1) = -0.8 - 0.5(x_1 - 1)$ from which we find $x_1 = \pm 1.5$. This gives $\mathbf{x}_{e_2} = [1.5 \quad 0 \quad -1.05]^T$ and $\mathbf{x}_{e_3} = [-1.5 \quad 0 \quad 1.05]^T$. The Jacobian matrix is just $\mathbf{A}$ with A_{11} changed to $A_{11} - 9\partial n/\partial x_1$. Since $\partial n/\partial x_1$ is just K_1, the Jacobian at $\mathbf{x}_{e_1}$ is

$$[\partial \mathbf{f}/\partial \mathbf{x}]_1 = \begin{bmatrix} 0.9 & 6.3 & 0 \\ 0.7 & -0.7 & 1 \\ 0 & -7 & 0 \end{bmatrix}$$

and the eigenvalues are $\lambda = 1.552, -0.676 \pm j1.897$. At both $\mathbf{x}_{e_2}$ and $\mathbf{x}_{e_3}$, $\partial n(x_1)/\partial x_1 = K_2 = -0.5$, so the Jacobian at both these points is

$$\left[\frac{\partial \mathbf{f}}{\partial \mathbf{x}}\right]_2 = \begin{bmatrix} -1.8 & 6.3 & 0 \\ 0.7 & -0.7 & 1 \\ 0 & -7 & 0 \end{bmatrix},$$

and the eigenvalues are $\lambda = -2.759, 0.1297 \pm j2.1329$. All three equilibrium points are unstable, but their behavior would be expected to be different. Perturbations from the first would be expected to be dominated by a growing exponential $e^{1.552t}$, whereas the other two would be expected to exhibit growing oscillations of the form $e^{0.1297t}\sin(2.1329t)$.

15.14 Use the describing function approach to analyze the Chua circuit of Problem 15.13.

The results of Example 15.5 are used to calculate the describing function $N(E)$ for the nonlinearity $n(x_1)$. It is convenient here to examine $G(j\omega) = -1/N(E)$, and for that purpose $-1/N(E)$ is plotted in Figure 15.23. The transfer function from the output of the nonlinearity to x_1 is found by using $\mathbf{C} = [1 \quad 0 \quad 0]$, along with $\mathbf{A}$ and $\mathbf{B}$ of Problem 15.13. This gives

$$G(s) = \mathbf{C}[s\mathbf{I} - \mathbf{A}]^{-1}\mathbf{B} = \frac{9[s^2 + 0.7s + 7]}{[s^3 + 7s^2 + 7s + 44.1]}$$

The poles are at $s = -6.9105$ and $s = -0.0447 \pm j2.526$. The zeros are at $s = -0.35 \pm j2.6225$. Because of the lightly damped complex poles and zeros at frequencies rather close together, rapid changes in the magnitude and phase of the transfer function can be expected at nearby frequencies. The real-axis crossover points are crucial and are found analytically.

$$G(j\omega) = \frac{9\{(-\omega^2 + 7) + 0.7j\omega\}}{(44.1 - 7\omega^2) + j(7\omega - \omega^3)} \tag{1}$$

Multiplication by the complex conjugate of the denominator and setting the imaginary component to zero gives the equation for real axis crossover points, $0.7\omega(44.1 - 7\omega^2) - (-\omega^2 + 7)(7\omega - \omega^3) = 0$. One root of this equation is $\omega_1 = 0$. Factoring this term out and rearranging gives $\omega^4 - 9.1\omega^2 + 18.13 = 0$. This quadratic in ω^2 has two real roots, and their square roots give the positive frequencies at which $G(j\omega)$ crosses the real axis, $\omega_2 = 1.71642$ and $\omega_3 = 2.4807$. Using these in the real component of (*1*) gives the magnitudes at the crossover points as $G(0) = 1.42$, $G(j\omega_2) = 1.554$, and $G(j\omega_3) = 7.4455$. A sketch of the Nyquist polar plot

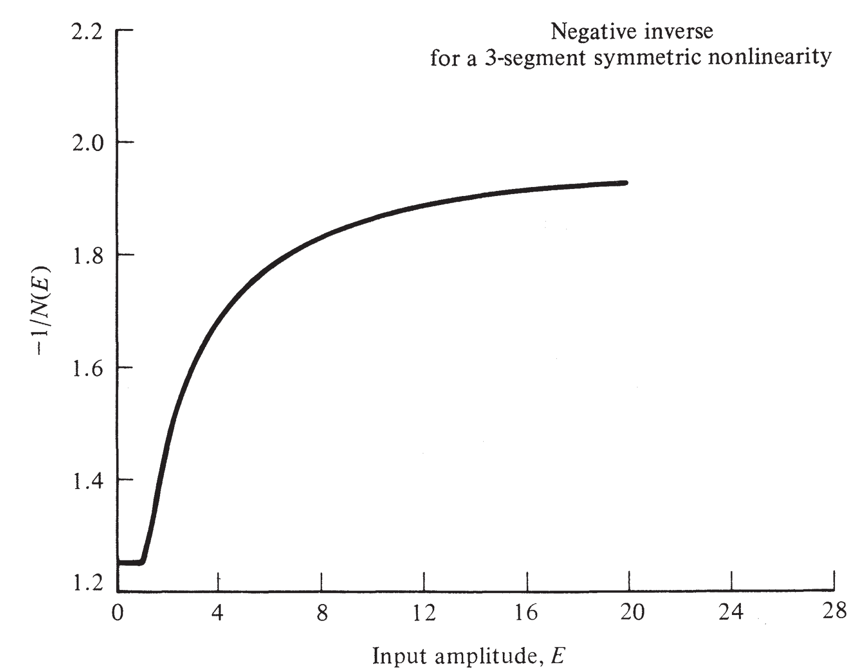


Figure 15.23

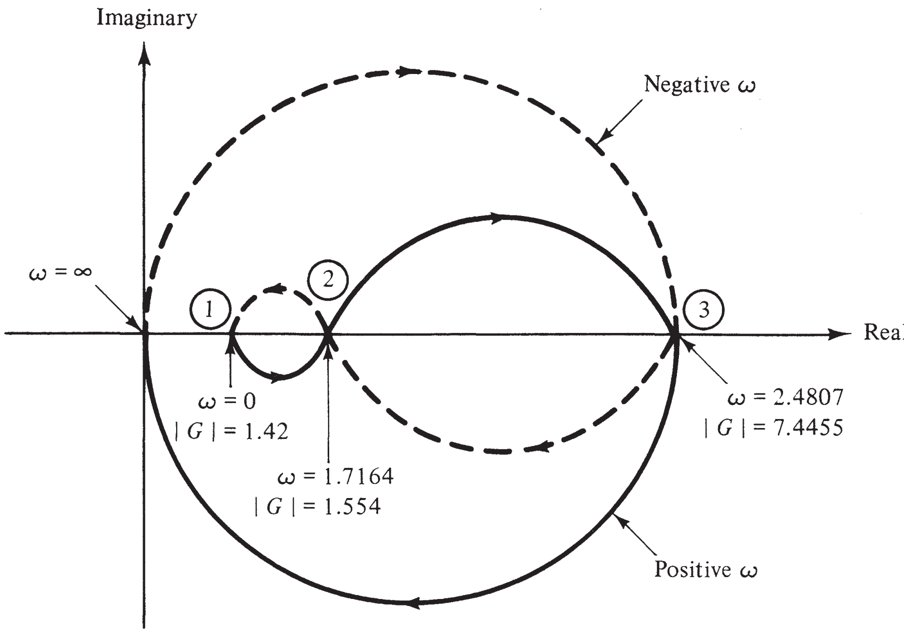


Figure 15.24 Polar plot of $G(j\omega)$

(not to scale) is given in Figure 15.24. The Nyquist criterion, Eq. (*15.18*), indicates that stability requires no encirclements of the critical point. For this system there are zero encirclements if the critical point is outside the entire plot or inside the small loop between crossover points 1 and 2. There are two clockwise encirclements ($N_c = -2$) for all critical points between crossover points 2 and 3, and $N_c = -1$ for all critical points between the origin and the first crossover point. For the Chua nonlinearity, the plot of $-1/N$ in Figure 15.23 shows a continuum of critical points between 1.25 and 2. In this range there are two intersections of $G(j\omega)$ and $-1/N$. The first one occurs for $\omega = 0$, and the second occurs for $\omega = 1.71642$. These intersections indicate potential limit cycles according to the describing function approximations. To check the stability of these two limit cycle points, assume we are at point 1 and a small increase occurs in the amplitude of x_1. This moves the critical point into a stable region, and the perturbed amplitude should decay back toward point 1. A small decrease in amplitude moves $-1/N$ into an unstable region, and the amplitude would be expected to increase back toward point 1. Thus point 1 appears to indicate a stable dc "limit cycle." Similar considerations at intersection point 2 show this to be an unstable limit cycle condition. In particular if the amplitude is perturbed away from point 2 to the right, it will not return to point 2, but neither can it proceed to another limit cycle, since point 3 is not an intersection of G and $-1/N$. Some more complicated behavior is indicated for a range of frequencies above $\omega = 1.7164$ but below $\omega = 2.4807$. The actual response to this system with the selected values is chaotic. The response contains a continuous spectrum of frequencies, not just discrete limit cycle frequencies. A small segment of the time response, starting with $\mathbf{x}(0) = [0.01 \quad 0 \quad 0]^T$, is shown in Figure 15.25. Although clearly not sinusoidal, counting peaks in various regions shows about 4 cycles in 10 s, or a period of 2.5 s. This crudely indicates some frequency content at about the frequency of crossover point 3. Superimposed are a range of lower modulating frequencies. Two-dimensional phase-plane plots of various pairs of states are given in Figures 15.26*a*, *b*, and *c*. Tendencies to oscillate about the equilibrium points $\mathbf{x}_{e_2}$ and $\mathbf{x}_{e_3}$ found in Problem 15.13 are clearly evident.

15.15 Add a modifying gain K to Chua's system of Problem 15.14 so that we now have $Kn(x_1)$ being fed into the same linear system. Use the describing function technique to predict behavior for various values of K.

The effect of K could be included in a modified $G(j\omega)$ plot or a modified $N' = KN$. We select the latter case for discussion. The real axis crossover frequencies are unchanged, and by selecting K we can cause $-1/N'$ to intersect with the crossover points in various ways:

1. If $K < 1.25/7.445 = 0.1679$, the entire interval $-1/N'$ is to the right of point 3, and stable behavior is predicted (no intersections, no limit cycles predicted).

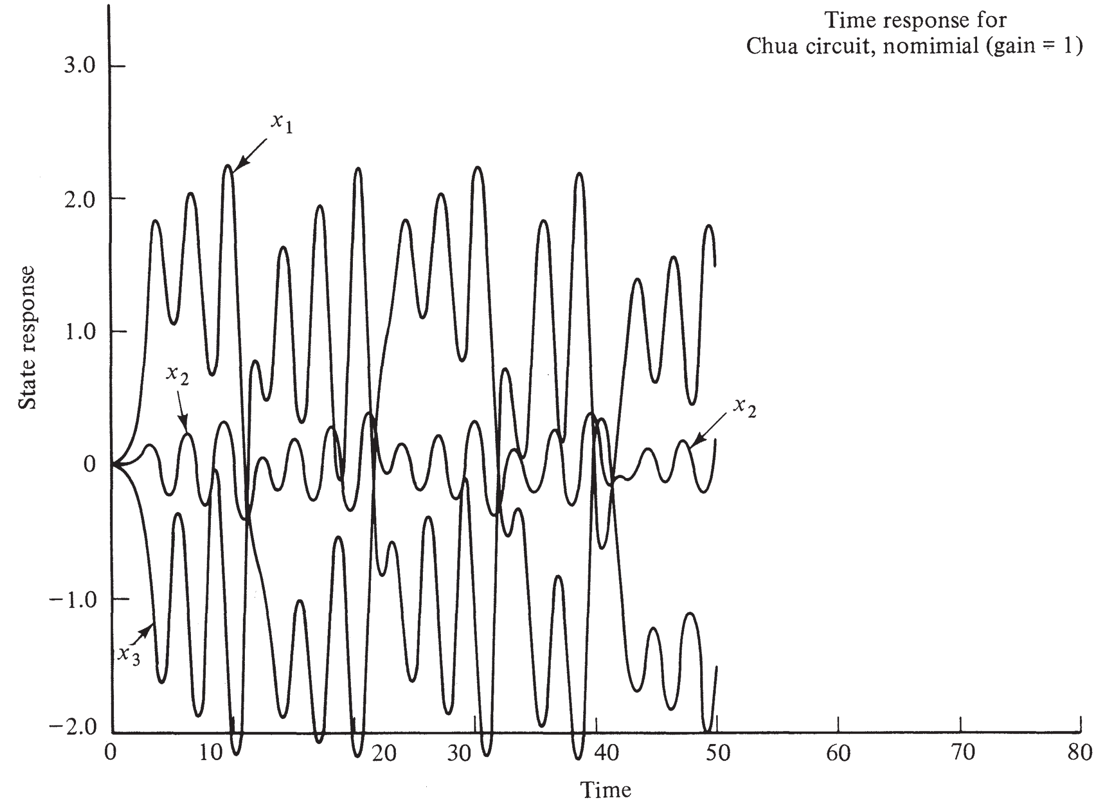


Figure 15.25

2. If $K > 2/1.42 = 1.408$, the entire interval $-1/N'$ falls between the origin and point 1, giving an unambiguous prediction of an unstable system (no intersections, no limit cycles predicted).
3. If $1.287 < K < 1.408$, the interval $-1/N'$ has just one intersection with $G(j\omega)$ at point 1, and some sort of stable dc ($\omega = 0$) behavior is predicted in steady state. The amplitude E required to make this happen can be predicted by using Figure 15.23 to find E for which $-1/N' = 1.42$. With $K = 1.3$, for example, this means $-1/N = 1.3(1.42) = 1.846$. This corresponds to $E \approx 8$ or 9, but it is not clear what this might mean for $\omega = 0$.
4. If $0.16789 < K < 0.2686$, $-1/N'$ has a single intersection with $G(j\omega)$ at point 3, and this predicts that a stable limit cycle with a frequency of $\omega = 2.4807$ rad/s. The amplitude E is also predicted by using Figure 15.23 to estimate the value of E at which $-1/N'(E)$ takes on the value 7.4455, that is, where $-1/N = 7.4455(K)$. With $K = 0.188$, the estimate is $E \approx 1.7$.

One simulation case in each of these categories is provided in Figures 15.27*a, b, c,* and *d.* In each case the general character of the describing function predictions are borne out reasonably well. The precision of the demarcation points is not clear, since a very slow growth or decay in amplitude takes a long time to notice. Also recall that the amplitude predictions are only for state x_1.

15.16 Use describing functions to investigate the possibility of limit cycles for the control loop of Figure 15.5 if $G(s) = (s + 20)/[s(s + 2)(s + 4)]$ and the nonlinearity is of the type given in Problem 15.9, with $\alpha = 0$, $\beta = 1$, $K_1 = 1$, $K_2 = 2$, and $K_3 = 1$.

The result in Problem 15.9 assumed that the amplitude E exceeded β. To analyze this system, the nonlinearity behavior must also be specified for amplitudes smaller than β. Here it is assumed that a set of smaller parallelograms nested inside the one shown in Figure 15.19*a* applies. The slope changes occur whenever $\dot{e}$ changes sign (at the maximum and minimum amplitudes) and whenever e changes sign. Under these assumptions the polar plot of $-1/N$ of

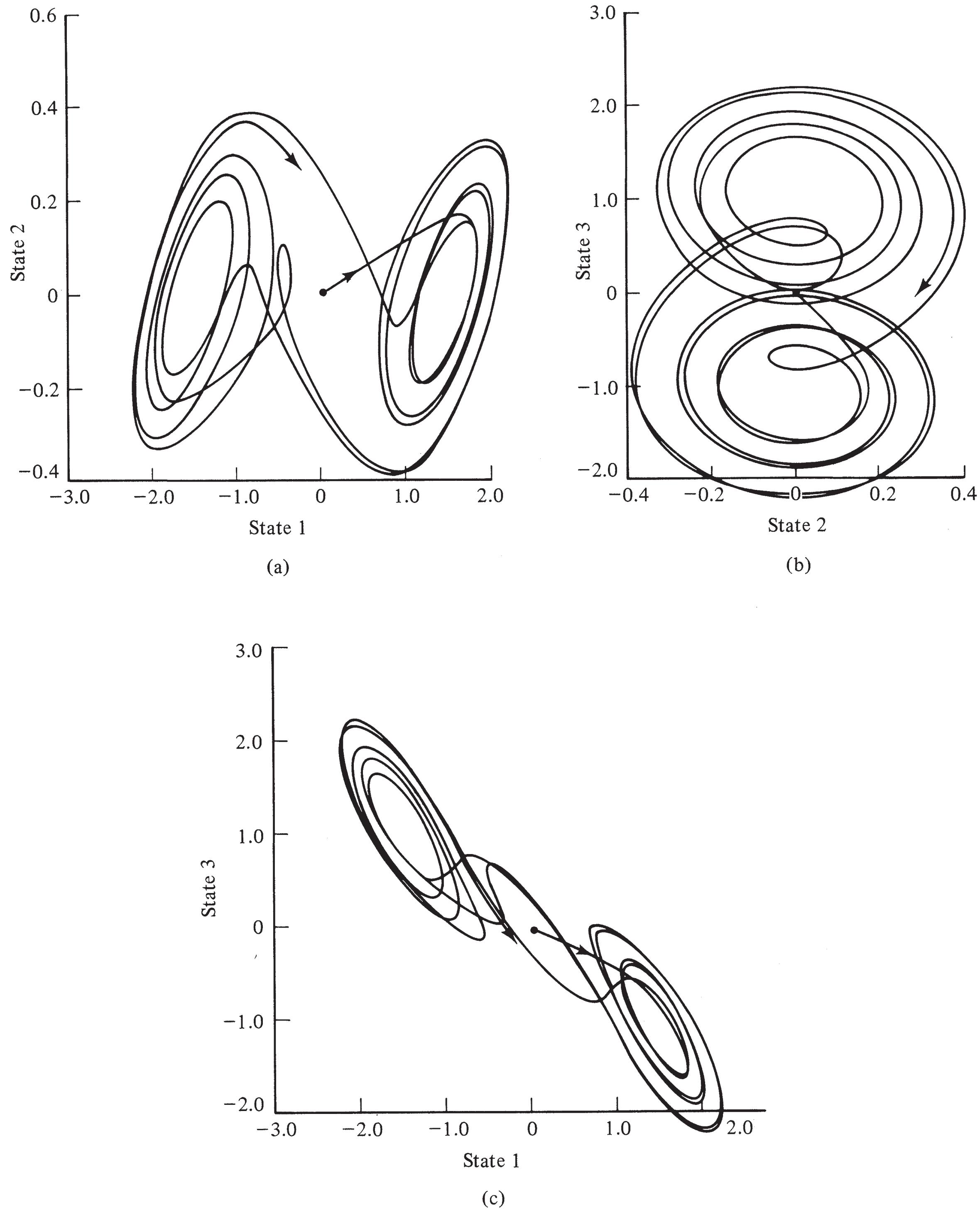
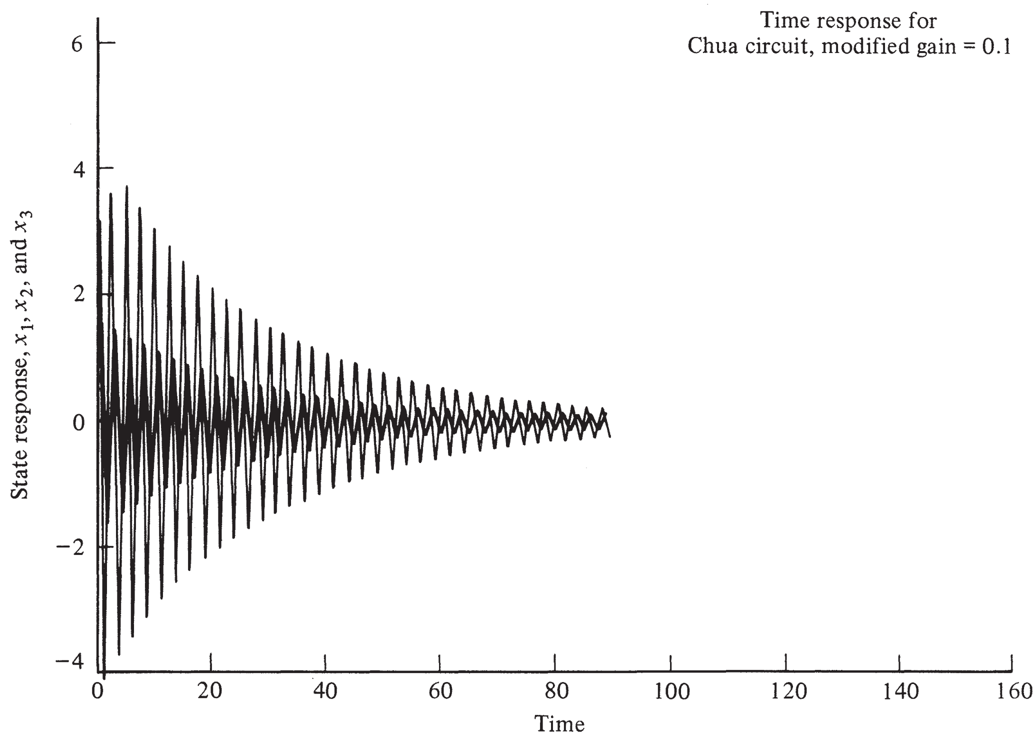
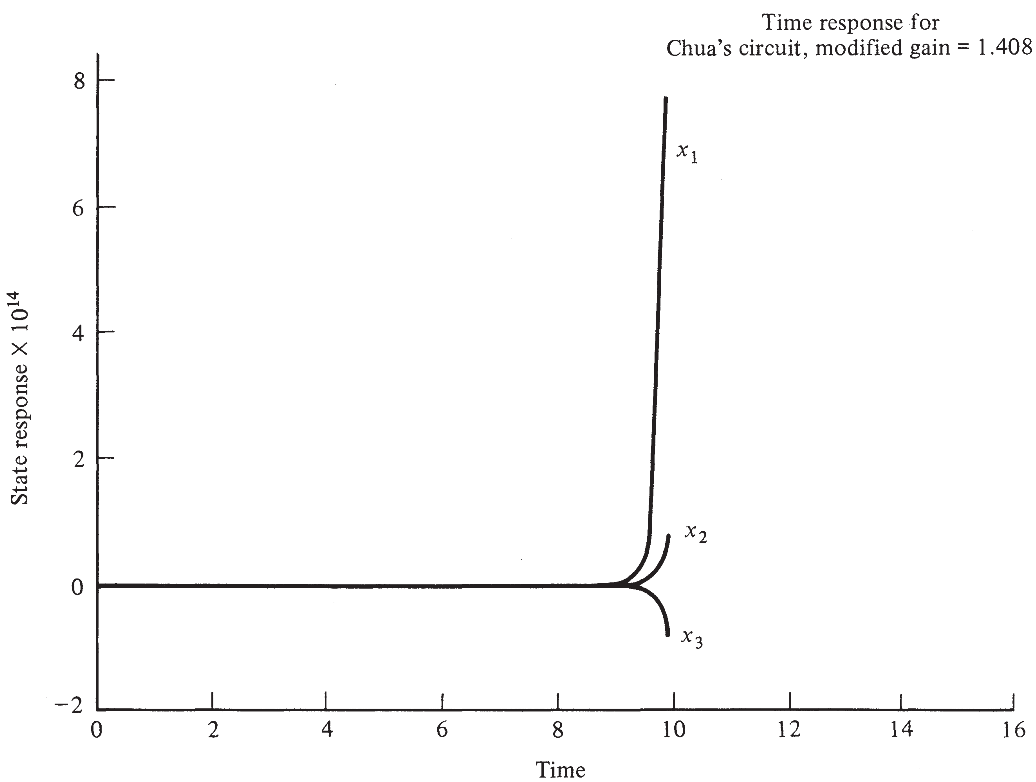


Figure 15.26

Figure 15.28 is obtained. Superimposed on it is a portion of the polar plot for $G(j\omega)$. There is one intersection, and it indicates existence of a stable limit cycle with an amplitude of $E \approx 0.8$ and a frequency of $\omega \approx 1.85$ rad/s. A simulation of this system generated the time response shown in Figure 15.29. This shows an x_1 amplitude of the expected magnitude, about 0.8, but the period of oscillation is 2.8 s, indicating $\omega \approx 2.24$, about 20% higher than predicted by the polar plot intersection. The probable reason is that the behavior of $n(e)$ is not really determined by e

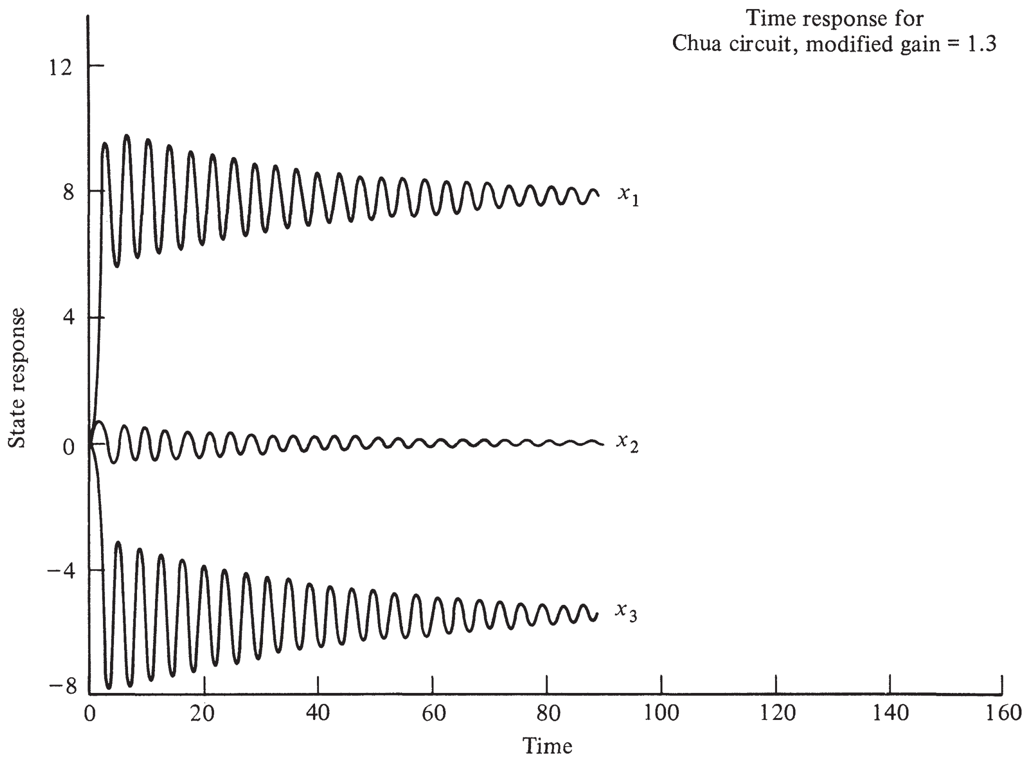


(a)

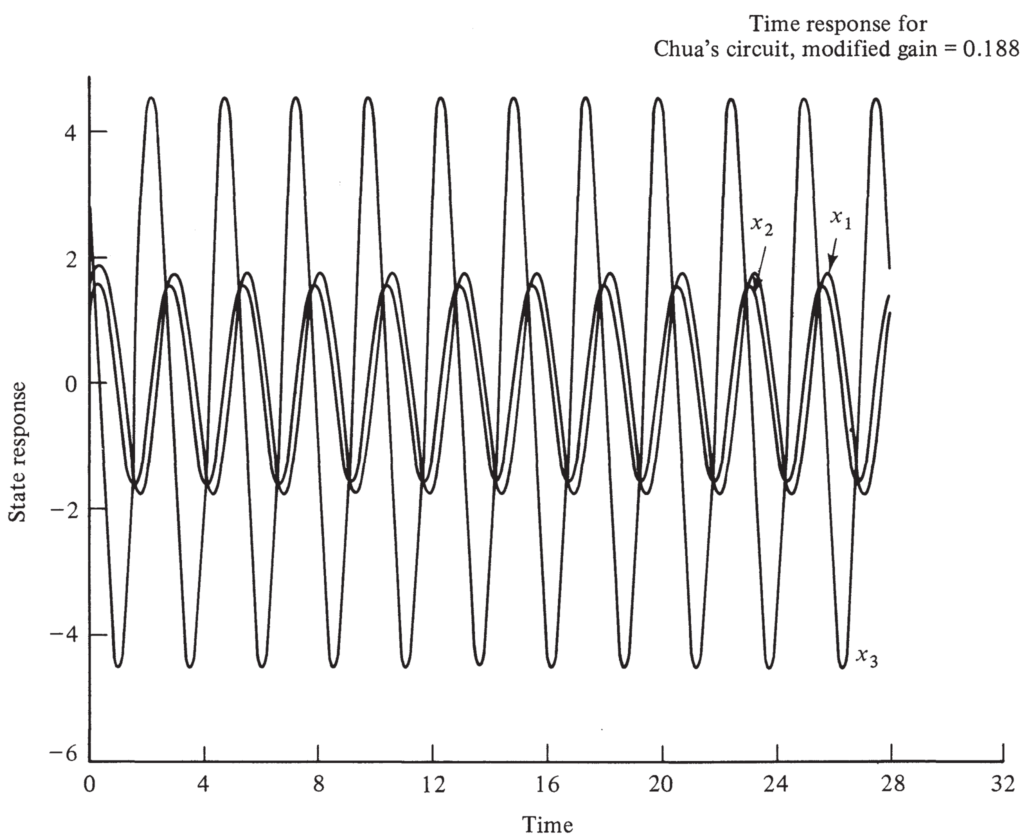


(b)

Figure 15.27



(c)



(d)

Figure 15.27 (Continued)

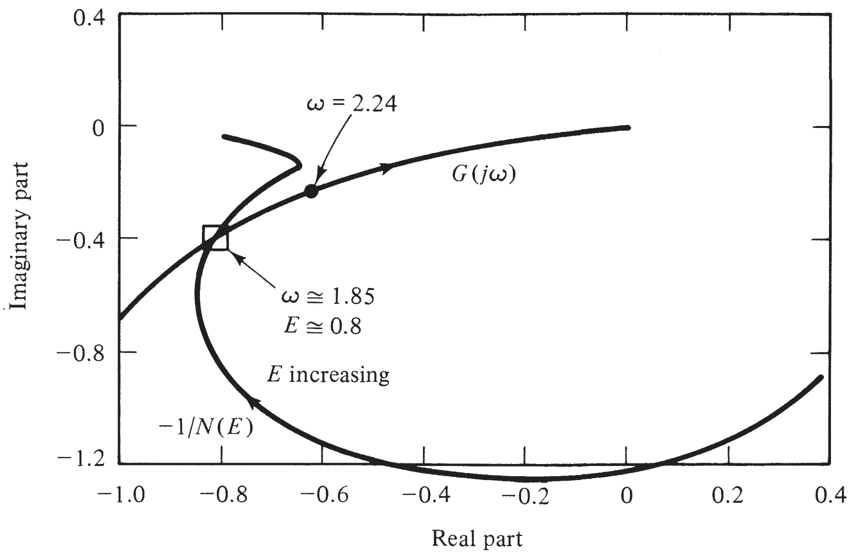


Figure 15.28

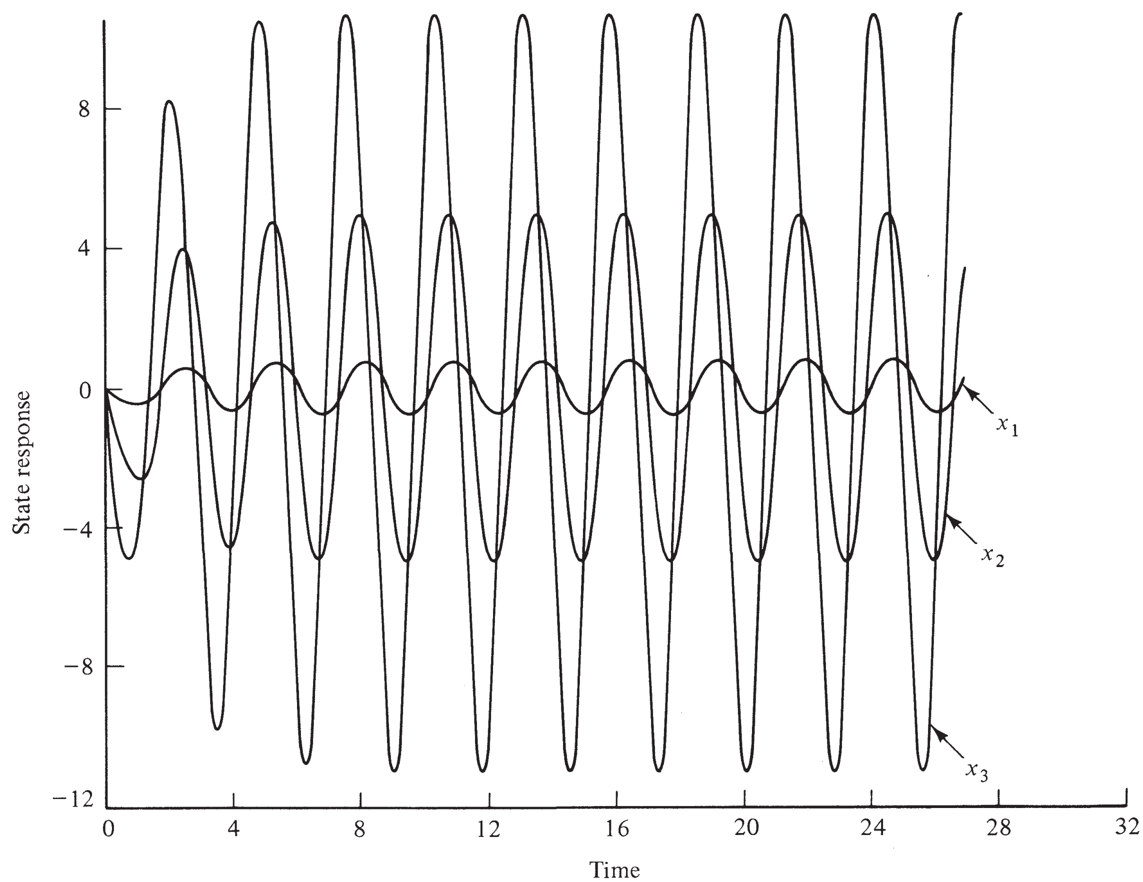


Figure 15.29

alone, nor even by both e and $\dot{e}$. Figure 15.30 shows four different signals, all having the same instantaneous value of $e(t)$. Two have positive $\dot{e}$ and two have negative $\dot{e}$. In order to determine correctly which parallelogram is being traversed, more history about $e(t)$ is required. In the simulation the most recent sign change in $\dot{e}(t)$ was used to determine the maximum amplitude of the current parallelogram being traversed. Other similar errors of approximation in hysteresis-type nonlinearities are discussed more fully in Reference 9.

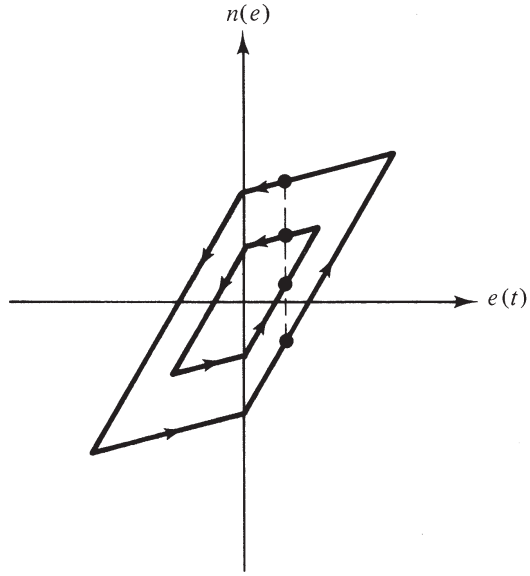


Figure 15.30

15.17 Investigate the following nonlinear system for stability:

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -g(x_2) - f(x_1) \tag{1}$$

The stability of an *equilibrium point* must be investigated. Since there could be several, it is not really proper to speak of *system stability*. Equilibrium points satisfy $\dot{\mathbf{x}} = \mathbf{0}$, so $x_{2_e} = 0$ is required, and then $f(x_{1_e}) = -g(0)$. It is assumed that $g(0) = 0$ and that $f(x_1) = 0$ only at $x_1 = 0$. Thus the origin is the only equilibrium point, by assumption.

If x_1 is thought of as a position, then x_2 is a velocity, and the above equations might represent a unit mass connected to a nonlinear spring and damper. The spring force is $f(x_1)$ and the damper force is $g(x_2)$. This analogy suggests trying a Lyapunov function composed of a kinetic energy-like term with x_2^2 and a potential spring energy term (equal to work done by $f(x_1)$):

$$V(\mathbf{x}) = c_1 x_2^2 + c_2 \int_0^{x_1} f(\xi)\, d\xi$$

This term is positive definite if $c_1 > 0, c_2 > 0$ and if $f(x_1)$ always has the same sign as x_1, for example, any odd function of x_1. Then

$$\dot{V} = 2c_1 x_2 \dot{x}_2 + c_2 f(x_1)\dot{x}_1$$

Using equation (*1*) gives

$$\dot{V} = 2c_1 x_2[-g(x_2) - f(x_1)] + c_2 f(x_1)x_2$$

Selecting $c_2 = 2c_1$ gives $\dot{V}(\mathbf{x}) = -c_2 x_2 g(x_2)$. $\dot{V}$ is negative semidefinite if $g(x_2)$ always has the same sign as x_2. If this is true, stability i.s.L. is ensured by Theorem 10.1. Actually a slight generalization of Theorem 10.2 is possible. If, instead of $\dot{V}(\mathbf{x})$ being negative definite, $\dot{V}(\mathbf{x})$ can be shown to be always negative *along any trajectory of the system,* asymptotic stability can still be concluded [20].

In this problem, $\dot{V} = 0$ only if $x_2 = 0$, and is negative otherwise. But, if $x_2 \equiv 0$, then $\dot{x}_2 = 0$ also and this requires that $f(x_1) = 0$. By assumption, this means $x_1 = 0$, so $\dot{V} < 0$ for all possible $\mathbf{x}(t)$ trajectories, except at the equilibrium point $\mathbf{x} = \mathbf{0}$. It is concluded that this system is asymptotically stable if $f(x_1)x_1 > 0$ and $g(x_2)x_2 > 0$ for all $\mathbf{x} \neq \mathbf{0}$. Further, since $V(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$, the stability is global.

15.18 Use Lyapunov's direct method to study the stability of the origin $\mathbf{x}=\mathbf{0}$ for the system [21] described by

$$\dot{x}_1 = x_2 - ax_1(x_1^2+x_2^2)$$
$$\dot{x}_2 = -x_1 - ax_2(x_1^2+x_2^2) \qquad (1)$$

A trial Lyapunov function is assumed as $V(\mathbf{x}) = c_1x_1^2 + c_2x_2^2$, with c_1 and c_2 unspecified but positive constants. Then $V(\mathbf{x})$ is positive definite and $V(\mathbf{x})\to\infty$ as $\|\mathbf{x}\|\to\infty$. The time derivative is

$$\dot{V}(\mathbf{x}) = 2c_1x_1\dot{x}_1 + 2c_2x_2\dot{x}_2$$

Using Eq. (*1*), this becomes

$$\dot{V}(\mathbf{x}) = 2c_1x_1[x_2 - ax_1(x_1^2+x_2^2)] + 2c_2x_2[-x_1 - ax_2(x_1^2+x_2^2)]$$

If the selection $c_1 = c_2$ is made, then the troublesome x_1x_2 product terms cancel, leaving $\dot{V}(\mathbf{x}) = -2ac_1(x_1^2+x_2^2)^2$. If the constant a is positive, $\dot{V}(\mathbf{x})$ is negative definite and the origin is globally asymptotically stable by Theorem 10.3.

15.19 Derive conditions which ensure asymptotic stability of the origin for the system in Eq. (*1*) [24]. Use the integration technique of Problem 10.11 to determine a suitable Lyapunov function:

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = x_3, \qquad \dot{x}_3 = -(x_1+cx_2)^n - bx_3 \qquad (1)$$

Try $\dot{V}(\mathbf{x}) = -x_3^2$. Then $V(\mathbf{x}) = \int_{t_1}^{t} \dot{V}(\mathbf{x})\,dt = -\int_{t_1}^{t} x_3\dot{x}_2\,dt$. Integration by parts gives

$$V(\mathbf{x}) = -x_3x_2 + \int x_2\dot{x}_3\,dt = -x_3x_2 - \int x_2(x_1+cx_2)^n\,dt - b\int x_2x_3\,dt$$

But $x_3 = \dot{x}_2$, so $\int_{t_1}^{t} x_2x_3\,dt = \int_0^{x_2} x_2\,dx_2 = x_2^2/2$. Adding and subtracting $\int cx_3(x_1+cx_2)^n\,dt$ gives

$$V(\mathbf{x}) = -x_3x_2 - \int (x_2+cx_3)(x_1+cx_2)^n\,dt + \int cx_3(x_1+cx_2)^n\,dt - \frac{bx_2^2}{2}$$

From Eq. (*1*), $x_2 + cx_3 = \dot{x}_1 + c\dot{x}_2$ and $(x_1+cx_2)^n = -\dot{x}_3 - bx_3$. Therefore,

$$V(\mathbf{x}) = -x_3x_2 - \frac{(x_1+cx_2)^{n+1}}{n+1} - \frac{bx_2^2}{2} - c\int x_3\dot{x}_3\,dt - bc\int x_3^2\,dt$$
$$= -x_3x_2 - \frac{(x_1+cx_2)^{n+1}}{n+1} - \frac{bx_2^2}{2} - \frac{cx_3^2}{2} - bc\int x_3^2\,dt$$

This $V(\mathbf{x})$ is not positive definite; in fact, it can be made negative definite. Therefore, a modified function is selected as

$$V'(\mathbf{x}) = -V(\mathbf{x}) - bc\int_{t_1}^{t} x_3^2\,dt = \frac{(x_1+cx_2)^{n+1}}{n+1} + \frac{bx_2^2}{2} + \frac{cx_3^2}{2} + x_2x_3$$
$$= \frac{(x_1+cx_2)^{n+1}}{n+1} + \frac{b}{2}\left(x_2 + \frac{x_3}{b}\right)^2 + \frac{(bc-1)x_3^2}{2b}$$

This is positive definite if $b>0, bc-1>0$ and if $n+1$ is any even positive integer. Also, $\dot{V}'(\mathbf{x}) = -\dot{V}(\mathbf{x}) - bcx_3^2 = -(bc-1)x_3^2$. The same conditions ensure that $\dot{V}'(\mathbf{x}) \le 0$ for all $\mathbf{x}$. Since $x_3 \equiv 0$ requires $\dot{x}_3 = 0$ and $\dot{x}_2 = 0$, $x_1 =$ constant, it is seen that $x_3 = 0$ holds along a solution only at the point $\mathbf{x} = \mathbf{0}$. The above conditions thus ensure asymptotic stability.

For another direct-integration method of generating Lyapunov functions, see the discussion of Parks' method [12].

15.20 Consider the nonlinear system $\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x})$, and assume that $\mathbf{f}(\mathbf{x})$ equals zero only at $\mathbf{x}=\mathbf{0}$. Then

$\mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{x})$ is a positive definite function of $\mathbf{x}$ and can serve as a potential Lyapunov function. Find sufficient conditions for asymptotic stability.

Since $V = \mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{x})$, $\dot{V} = \dot{\mathbf{f}}(\mathbf{x})^T\mathbf{f}(\mathbf{x}) + \mathbf{f}(\mathbf{x})^T\dot{\mathbf{f}}(\mathbf{x})$. But $\dot{\mathbf{f}}(\mathbf{x}) = [\partial\mathbf{f}/\partial\mathbf{x}]\dot{\mathbf{x}} = [\partial\mathbf{f}/\partial\mathbf{x}]\mathbf{f}$, so that

$$\dot{V} = \mathbf{f}^T\left\{\left[\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\right]^T + \left[\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\right]\right\}\mathbf{f}$$

Thus, by Theorem 10.3 the origin is asymptotically stable if $[\partial\mathbf{f}/\partial\mathbf{x}]^T + [\partial\mathbf{f}/\partial\mathbf{x}]$ is negative definite—i.e., has all its eigenvalues strictly in the left-half plane for all $\mathbf{x}$. Furthermore, if $\mathbf{f}(\mathbf{x})^T\mathbf{f}(\mathbf{x}) \to \infty$ as $\|\mathbf{x}\| \to \infty$, then the origin is *globally* asymptotically stable. This result is attributed to Krasovskii [23, 28].

15.21 Use the variable gradient technique to investigate the stability of the nonlinear system (see pages 59 and 67 of Reference 20) described by

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -f(x_1)x_2 - g(x_1) \tag{1}$$

It is known that $g(0) = 0$ and that $\mathbf{x} = \mathbf{0}$ is the only equilibrium point. The gradient is assumed to be of the form

$$\nabla_\mathbf{x} V = \begin{bmatrix} \alpha_{11}x_1 + \alpha_{12}x_2 \\ \alpha_{21}x_1 + x_2 \end{bmatrix}$$

The curl equations require $\partial\nabla V_1/\partial x_2 = \partial\nabla V_2/\partial x_1$, or

$$x_1\frac{\partial\alpha_{11}}{\partial x_2} + \alpha_{12} + x_2\frac{\partial\alpha_{12}}{\partial x_2} = x_1\frac{\partial\alpha_{21}}{\partial x_1} + \alpha_{21}$$

Using the assumed gradient and Eq. (*1*) gives

$$\dot{V} = (\nabla_\mathbf{x} V)^T\begin{bmatrix} x_2 \\ -f(x_1)x_2 - g(x_1) \end{bmatrix}$$
$$= \alpha_{11}x_1x_2 + \alpha_{12}x_2^2 - \alpha_{21}f(x_1)x_1x_2 - f(x_1)x_2^2 - g(x_1)\alpha_{21}x_1 - g(x_1)x_2$$

This expression should be made at least negative semidefinite.

One possible solution begins by setting $\alpha_{21} = 0$. Then the curl equations are satisfied if $\alpha_{12} = 0$ and if α_{11} is not a function of x_2. By setting $\alpha_{11} = g(x_1)/x_1$, we obtain $\dot{V} = -f(x_1)x_2^2$, which is negative semidefinite if $f(x_1) \ge 0$ for all x_1. Then $\nabla_\mathbf{x} V = [g(x_1) \quad x_2]^T$ and

$$V = \int_0^{x_1} g(\xi)\,d\xi + \int_0^{x_2} \xi\,d\xi = \int_0^{x_1} g(\xi)\,d\xi + \frac{x_2^2}{2}$$

Thus V is positive definite, and hence a Lyapunov function, if $g(x_1)x_1 > 0$ for all $x_1 \ne 0$.

The following conclusions regarding stability can be drawn:

1. Stable i.s.L. if $g(x_1)x_1 > 0, f(x_1) \ge 0$, for all $\mathbf{x} \ne \mathbf{0}$ by Theorem 10.1.
2. Asymptotically stable if, in addition, $f(x_1)$ and $g(x_1) = 0$ only at $x_1 = 0$. This ensures that $\dot{V} \ne 0$ on any solution of equation (*1*) except at $\mathbf{x} = \mathbf{0}$.
3. Globally asymptotically stable if, in addition, $\int_0^{x_1} g(\xi)\,d\xi \to \infty$ as $|x_1| \to \infty$.

15.22 Use Lyapunov's direct method to investigate the stability of the nonlinear time-varying system [27] $\ddot{x} + a\dot{x} + g(x, t)x = 0$.

The state equations are $\dot{x}_1 = x_2, \dot{x}_2 = -ax_2 - g(x_1, t)x_1$.

Assume that $\nabla_\mathbf{x} V = \begin{bmatrix} \alpha_{11}x_1 + \alpha_{12}x_2 \\ \alpha_{21}x_1 + x_2 \end{bmatrix}$. Then since $V(\mathbf{x})$ may be an explicit function of time,

$$\dot{V}(\mathbf{x}) = (\nabla_\mathbf{x} V)^T\dot{\mathbf{x}} + \partial V/\partial t$$
$$= \alpha_{11}x_1x_2 + \alpha_{12}x_2^2 - a\alpha_{21}x_1x_2 - g(x_1, t)x_1^2\alpha_{21} - ax_2^2 - g(x_1, t)x_1x_2 + \partial V/\partial t$$

In order to remove the x_1x_2 product terms, set $\alpha_{11} = a\alpha_{21} + g(x_1, t)$. The curl equations can then be satisfied if $\alpha_{21} = \alpha_{12} =$ constant. Then

$$V(\mathbf{x}) = \int_0^{x_1} [a\alpha_{21} + g(x_1, t)]x_1 dx_1 + \int_0^{x_2} [\alpha_{21}x_1 + x_2]\, dx_2 \Big|_{x_1 = \text{constant}}$$

$$= a\alpha_{21}\frac{x_1^2}{2} + \alpha_{21}x_1x_2 + \frac{x_2^2}{2} + \int_0^{x_1} g(x_1, t)x_1\, dx_1$$

If $\int_0^{x_1} g(x_1, t)x_1\, dx_1 > 0$ for all x_1 and t, then $V(\mathbf{x}) > \frac{1}{2}(x_2 + \alpha_{21}x_1)^2 + \frac{1}{2}(a\alpha_{21} - \alpha_{21}^2)x_1^2$. Thus $V(\mathbf{x})$ is positive definite if $a > 0$ and $a\alpha_{21} > \alpha_{21}^2$. This is ensured by selecting $\alpha_{21} = a - \epsilon$, where ϵ is a small positive number. Checking the time derivative,

$$\frac{dV}{dt} = a\alpha_{21}x_1\dot{x}_1 + \alpha_{21}\dot{x}_1x_2 + \alpha_{21}x_1\dot{x}_2 + x_2\dot{x}_2 + g(x_1, t)x_1\dot{x}_1 + \int_0^{x_1} \frac{\partial g(x_1, t)}{\partial t} x_1\, dx_1$$

Using $\alpha_{21} = a - \epsilon$ and the differential equations for $\dot{x}_1$ and $\dot{x}_2$ gives

$$\dot{V} = -\epsilon x_2^2 - (a - \epsilon)g(x_1, t)x_1^2 + \int_0^{x_1} \frac{\partial g(x_1, t)}{\partial t} x_1\, dx_1$$

This expression is negative definite if $g(x_1, t) > 0$ for all x_1, t and if the integral term is sufficiently small. That is, if

$$\int_0^{x_1} \frac{\partial g(x_1, t)}{\partial t} x_1\, dx_1 < ag(x_1, t)x_1^2$$

This will be true, for example, if $\max_{x_1, t} [\partial g(x_1, t)/\partial t] < 2ag(x_1, t)$ for all x_1 and t. Additionally, if $g(x_1, t)$ is bounded for all x_1 and t, then $V(\mathbf{x}, t)$ can be bounded as required by condition (4) of Theorem 10.5. If these conditions are satisfied, the system is uniformly globally asymptotically stable.

PROBLEMS

15.23 Use dynamic linearization to design a controller for the first order nonlinear system $\dot{x} = x^2u$ so that the resulting response mimics the linear system $\dot{y}_d = -2y_d + v$, where v is a unit step function. Investigate various initial conditions for the actual system and for the template system. What impact do these have on potential singularities in the linearizing control law?

15.24 The desired response for the second-order nonlinear system described by

$$\dot{x}_1 = x_2^2 + u \quad \text{and} \quad \dot{x}_2 = x_2 u$$

is intended to mimic the uncoupled linear system

$$\dot{y}_{d_1} = -y_{d_1} + v, \qquad \dot{y}_{d_2} = -4y_{d_2} + v$$

Investigate the dynamic linearization procedure, noting that a single control variable is being asked to satisfy two conflicting equations. Use a least-squares approximation to find a control law, and test its response in a simulation. Let $v(t)$ be a unit step function and set $\mathbf{y}_d(0) = \mathbf{0}$.

15.25 Use dynamic linearization to design a controller for the nonlinear orbit equations of Example 15.2. Select a linear template system satisfying $\ddot{r} + \alpha\dot{r} + \beta r = \beta R$ and $\ddot{\theta} + \kappa\dot{\theta} + \sigma\theta = \sigma\mu t/R^3$. Select values for α, β, κ, and σ so that r and θ smoothly approach their steady-state values in about one-fourth of an orbit revolution.

15.26 Verify that the limit cycle found in Example 15.7 is stable.

15.27 Repeat the analysis of Problem 15.13 of the system equilibrium points, but reverse the order of the two slopes in the nonlinearity, so that $K_1 = -0.5$ and $K_2 = -0.8$.

15.28 Use describing functions to investigate the possibility of limit cycles (and chaotic behavior), as was done in Problem 15.14, but with the modified nonlinearity of Problem 15.27. Results derived in Example 15.5 yield the plot of $-1/N$ given in Figure 15.31.

15.29 Repeat the analysis of Problem 15.15 but with the nonlinearity of Problems 15.27 and 15.28.

15.30 Another version of Chua's chaotic circuit [31] has a three-segment nonlinearity of the type treated in Example 15.5. The breakpoints and slopes are now $\alpha = 1$, $\beta = 5$, $K_1 = -0.8$, $K_2 = -0.5$, and $K_3 = +2$. The linear part of the circuit is the same as in Problem 15.13, with all the same values. Find the equilibrium points and investigate their stability. Use describing functions to investigate the possibility of limit cycles or other strange oscillations. A plot of the describing function is given in Figure 15.32.

15.31 Repeat Problem 15.30, but with new parameters values $R = 0.71$, $L = 1$, $C_1 = C_2 = 0.1$, $\alpha = 1$, $\beta = 2$, $K_1 = -6$, $K_2 = -5$, and $K_3 = +10$. The describing function is plotted in Figure 15.33.

15.32 A system with the configuration of Figure 15.5 has a nonlinearity with dead zone and saturation, described by Figure 15.6*a* with $\alpha = 0.25$, $\beta = 2.5$, $K_1 = K_3 = 0$, and $K_2 = 1$. The linear portion is described by $G(s) = K(s+8)/[s(s+2)(s+4)(s+6)]$. Find the range of gain values K for which the linear system, without the nonlinearity, is unstable. Show that the presence of the nonlinearity can stabilize the system. Also find a value of K where limit cycles can exist. Determine the amplitude and frequency of the stable limit cycle.

15.33 Consider the nonlinearity of Problem 15.32 along with $G(s) = 10(s^2 + 4s + 68)/[s(s+2)(s+4)]$. Use root locus analysis to find the amplitude and frequency of a stable limit cycle by treating the describing function $N(E)$ as a part of the root locus gain $KN(E)$ and finding the $j\omega$ axis crossover points.

15.34 **(a)** Using precisely the same steps as in Problem 10.11c, find a Lyapunov function and its time derivative for the nonlinear system [20].

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = x_3, \qquad \dot{x}_3 = -F(x_2)x_3 - ax_2 - bx_1$$

(b) Determine sufficient conditions for asymptotic stability.

15.35 A special case of the Lorenz equation, widely used as an example of chaotic behavior, is given by [14, 15]

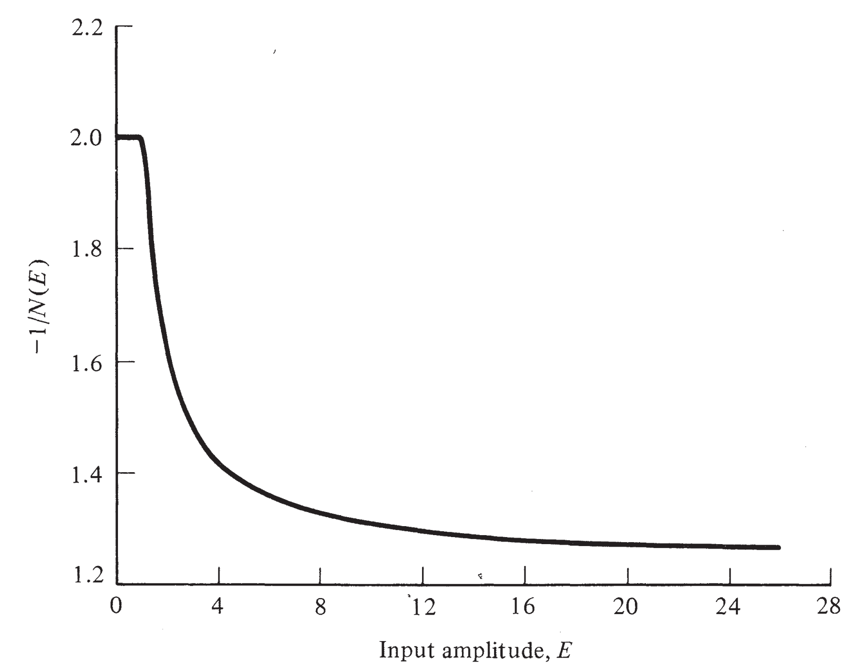


Figure 15.31

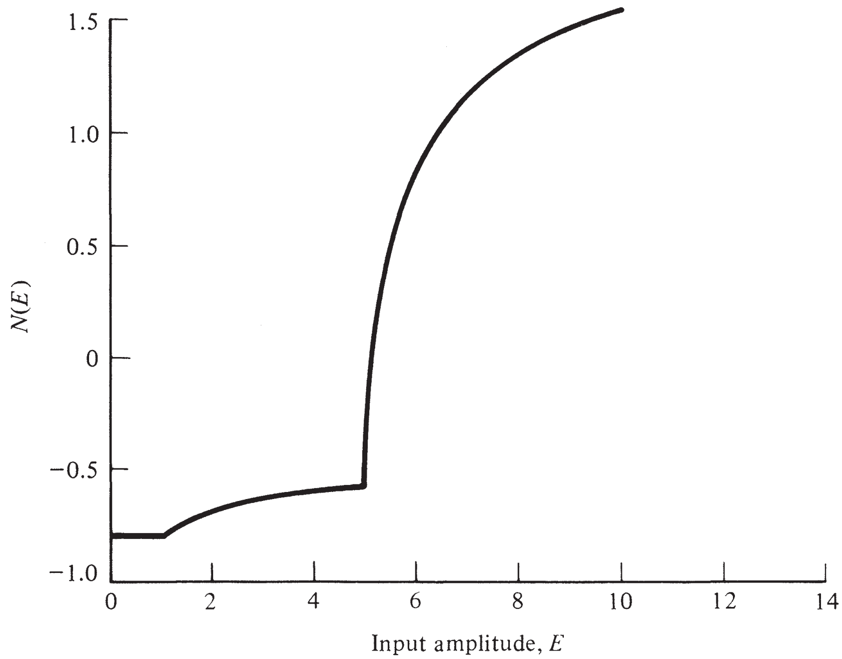


Figure 15.32

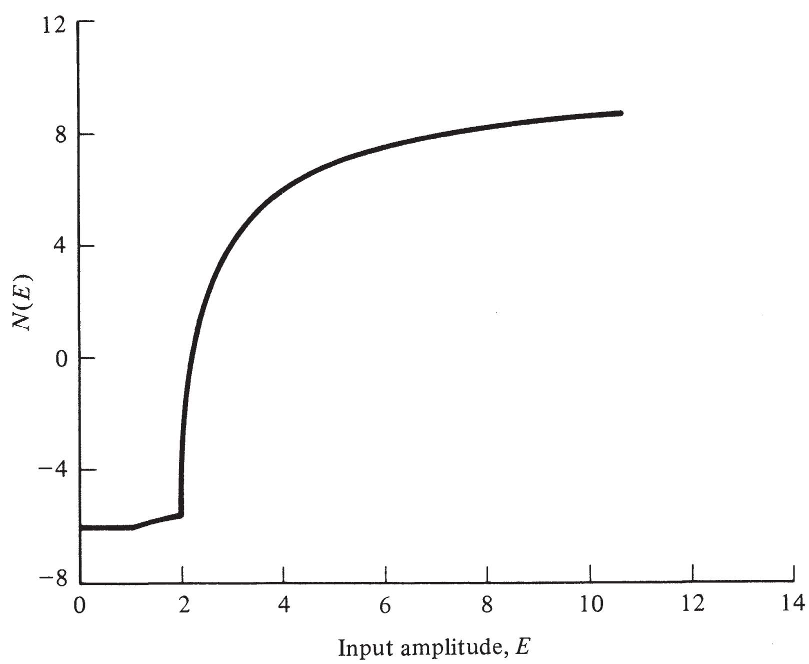


Figure 15.33

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} -10 & 10 & 0 \\ r & -1 & 0 \\ 0 & 0 & -\frac{8}{3} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ -x_1 x_3 \\ x_1 x_2 \end{bmatrix}$$

(a) Show that if $r < 1$, there is only one equilibrium point located at the origin, and it is locally stable.
(b) Use a suitable Lyapunov function to estimate the domain of stable attraction to this equilibrium point. Set $r = 0.5$ for simplicity.
(c) The more interesting, chaotic behavior occurs for values of r near 28. Show that there are then three equilibrium points, all of which are locally unstable.
(d) Estimate the domain of instability which surrounds the origin by using Lyapunov techniques.

15.36 The divergence theorem [8] states that for a sufficiently smooth vector field $\mathbf{f}(\mathbf{x})$ defined over a sufficiently smooth closed volume Υ with surface area S composed of incremental area elements dA with outward pointing normal vectors $\mathbf{n}(\mathbf{x})$,

$$\iiint_{\Upsilon} \operatorname{div}(\mathbf{f})\, dV = \iint_{S} \mathbf{f}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})\, dA$$

where $\operatorname{div}(\mathbf{f}) = \partial f_1/\partial x_1 + \partial f_2/\partial x_2 + \cdots + \partial f_n/\partial x_n$ is the divergence of $\mathbf{f}$. Let $\mathbf{f} = \dot{\mathbf{x}}$ and note that $\iint \mathbf{f}(\mathbf{x}) \cdot \mathbf{n}(\mathbf{x})\, dA = \iint \dot{\mathbf{x}}\Delta t \cdot \mathbf{n}(\mathbf{x})\, dA/\Delta t \to d\Upsilon/dt$. Apply this to the vector $\mathbf{f}$ of the Lorenz equations and show that the volume Υ containing an arbitrary set of initial states satisfies $d\Upsilon/dt = -\frac{41}{3}\Upsilon$. This means that the volume that contains the solution trajectories through these initial states is always decreasing according to $\Upsilon(t) = \Upsilon(0)e^{-(41/3)t}$ [15].

15.37 Apply the divergence theorem of the previous problem to a linear system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ and show that the trajectories which originally occupy a volume $\Upsilon(0)$ satisfy $\Upsilon(t) = \Upsilon(0)e^{\alpha t}$, where $\alpha = \operatorname{trace}\{\mathbf{A}\}$. From this show that "bounded volume" result for solution trajectories does not in any way imply bounded norms for $\mathbf{x}(t)$ and hence does not imply any kind of stability. *Hint:* Consider a specific unstable case such as $\mathbf{A} = \operatorname{diag}[2 + 2j,\ 2 - 2j,\ -10]$. Trajectories initially in a three-dimensional volume diverge to infinity, but in a planar subspace, with zero volume.

Answers to Problems

CHAPTER 1

1.21 $Q = \frac{A}{\rho g}\dot{P}$

1.22 $v(t) = v(t_0) + \int_{t_0}^{t} Q(\tau)\, d\tau$

1.23 $v_0 = L(df_3/dt) + Rf_3$

1.24 $y(s)/u(s) = [Cs + 1/R_1]/[Cs + (1/R_1 + 1/R_2)]$

1.26 $Y(z) = \sum_{n=0}^{\infty} e^{-0.1nT} e^{-nsT} = \frac{1}{1 - z^{-1}e^{-0.2}} = \frac{z}{z - e^{-0.2}}$

1.27 $H(z) = \frac{10z^2(z + 0.5)}{(z - 0.2)(z - 0.4)(z - 0.8)}$
Poles at $z = 0.2, 0.4, 0.8$
Zeros at 0., 0., -0.5
Stable

1.28 $y(nT) = 156.25 - 2.916(0.2)^n + 30(0.4)^n - 173.33(0.8)^n$

CHAPTER 2

2.25 $3 < K < 9$

2.26 $K = 20$, $\omega = 2$ rad/s

2.27 With $K = \sqrt{327{,}680} \cong 572$, the s row of Routh's array is zero. The auxiliary equation then is $44.2s^2 + 2288 = 0$, indicating the cross-over frequency is $\omega = 7.19$ rad/s.

2.28 **(a)** Type 0, $K_p = 20$ db $= 10$, $K_v = 0$, $K_a = 0$ **(b)** Type 1, $K_p = \infty$, $K_v = -20$ db $= 0.1$, $K_a = 0$ **(c)** Type 2, $K_p = \infty$, $K_v = \infty$, $K_a = -60$ db $= 0.001$

2.30 Stable for all K, but very low stability margins for small values of K.

2.31 For $K > 7.5$ there is one clockwise and one counterclockwise encirclement. Therefore, $N = 0$ and system is stable. There are two unstable roots for $0 < K < 7.5$.

2.32 Gain margin $\cong 1.76$, i.e., $K > 8800$ causes system instability. Phase margin $\cong 13°$. $\omega \cong 15$ rad/s for a $-180°$ phase angle.

2.33 $K = 9.65$, $\omega = 11.8$ rad/s (*Hint:* Dominant roots at $-\alpha \pm j\omega$; $e^{-\alpha} = 0.0015$.)

2.34 Nonminimum phase

2.35 Direct realization: $y(t_k) = -0.5y(t_{k-1}) + E(t_{k-1}) - 0.5E(t_{k-2})$

Parallel realization:
$$y_1(t_k) = E(t_{k-1})$$
$$y_2(t_k) = -0.5y_2(t_{k-1}) + E(t_{k-1})$$
$$y(t_k) = -y_1(t_k) + 2y_2(t_k)$$

Cascade realization:
$$E_1(t_k) = E(t_{k-1})$$
$$y(t_k) = E_1(t_k) - 0.5[E_1(t_{k-1}) + y(t_{k-1})]$$

2.36 $C(nT) \cong 13.3333 - 20(0.5)^n + 6.6667(-0.5)^n$

2.37 (a) $C(z) = 1 + (\alpha + \beta - a - b)z^{-1} + (ab + \alpha^2 + \alpha\beta + \beta^2 - a\alpha - b\alpha - a\beta - b\beta)z^{-2} + \cdots$. Therefore, $C(0) = 1$, $C(T) = (\alpha + \beta - a - b)$ and $C(2T) = ab + \alpha^2 + \alpha\beta + \beta^2 - a\alpha - b\alpha - a\beta - b\beta$.

(b) $$C(nT) = \frac{ab}{\alpha\beta}\delta_{0n} + \left[\frac{\alpha^2 - a\alpha - b\alpha + ab}{\alpha(\alpha - \beta)}\right]\alpha^n + \left[\frac{\beta^2 - a\beta - b\beta + ab}{\beta(\beta - \alpha)}\right]\beta^n$$

where $\delta_{0n} = \begin{cases} 0 & \text{if } n \neq 0 \\ 1 & \text{if } n = 0 \end{cases}$

(c) $C(nT) \to 0$ as $n \to \infty$

2.38 (a) $$\frac{C(z)}{R(z)} = \frac{0.004845K(z + 0.9672)}{z^2 - (1.9239 - 0.00484K)z + (0.90484 + 0.004686K)}$$

(b)

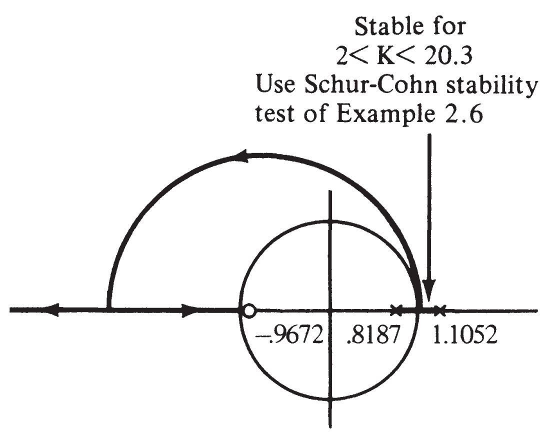


Figure A2.38

2.39 Let [ ]* indicate the Z-transform of whatever is inside [ ].

$$E(z) = \frac{[RG_1]^*}{1 + [G_1G_2H_1]^* + [G_1G_2G_3H_2]^*}$$

$$C(s) = \frac{[RG_1]^* G_2G_3G_5/[1 + G_5H_3]}{1 + [G_1G_2H_1]^* + [G_1G_2G_3H_2]^*} + \frac{RG_4G_5}{1 + G_5H_3}$$

$C(z)$ is the same except put an * on $[G_2G_3G_5/[1 + G_5H_3]]$ and $[RG_4G_5/[1 + G_5H_3]]$

2.40 $$G_c(z) = \frac{Az^2 + Bz + D}{(1 + \tau/T)z^2 - (2\tau/T + 1)z + \tau/T}$$

where
$$A = K_D/T + K_P(1 + \tau/T) + K_I(T + \tau)$$
$$B = -2(K_D/T + K_P\tau/T) - K_P - K_I\tau$$
$$D = K_D/T + K_P\tau/T$$

For simplicity, let $\tau \to 0$. Then one obtains an algorithm of the form

$$E_1(t_k) = E_1(t_{k-1}) + \alpha E(t_k) - \beta E(t_{k-1}) + \gamma E(t_{k-2})$$

CHAPTER 3

3.19 $\mathbf{A}_1 = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix}$, $\mathbf{B}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\mathbf{C}_1 = [1 \quad 0]$, $D_1 = 0$

$\mathbf{A}_2 = \begin{bmatrix} -a & 1 \\ -b & 0 \end{bmatrix}$, $\mathbf{B}_2 = \mathbf{B}_1$, $\mathbf{C}_2 = \mathbf{C}_1$, $D_2 = D_1$

3.20 Let $x_1 = y_1$, $x_2 = y_2$, $x_3 = \dot{y}_2$, $\mathbf{x} = [x_1 \quad x_2 \quad x_3]^T$, $\mathbf{y} = [y_1 \quad y_2]^T$, $\mathbf{u} = [u_1 \quad u_2]^T$,

$$\mathbf{A} = \begin{bmatrix} -3 & -3 & 0 \\ 0 & 0 & 1 \\ 0 & -3 & -4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

3.21 Of the many possible answers, one is obtained by setting $x_1 = y_1$, $x_2 = \dot{y}_1 - u_2$, $x_3 = y_2 - 2u_1$.

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

3.22 $\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \left[\begin{array}{cc|c} -1 & 1 & 0 \\ 0 & -1 & 0 \\ \hline 0 & 0 & -6 \end{array}\right] \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \left[\begin{array}{c} 0 \\ 1 \\ \hline 1 \end{array}\right] u$ and $y = [\frac{1}{5} \quad -\frac{1}{25} \quad \frac{1}{25}]\mathbf{x}$

3.23 With x_1 = voltage across C_1, x_2 = current through L_1, x_3 = current through L_2,

$$u = i_s, \quad \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1/(C_1 + C_2) & -1/(C_1 + C_2) \\ -1/L_1 & -R/L_1 & 0 \\ 1/L_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ R/L_1 \\ 0 \end{bmatrix} u$$

3.24 x_1, x_2 = voltage across C_1, C_2. x_3 = current through L. Then

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} \dfrac{-1}{C_1(R_1 + R_3)} & 0 & \dfrac{R_1}{C_1(R_1 + R_3)} \\ \dfrac{NK}{R_4 C_2} & -\dfrac{1}{R_4 C_2} & 0 \\ \dfrac{-R_1}{L(R_1 + R_3)} & 0 & \dfrac{-(R_1 R_2 + R_1 R_3 + R_2 R_3)}{L(R_1 + R_3)} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 0 \\ -\dfrac{NK}{R_4 C_2} \\ \dfrac{1}{L} \end{bmatrix} v_s$$

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

3.25 $\dot{x}_1 = \dfrac{1}{C_1}\left[-\dfrac{x_1}{R_1 + R_2} - \dfrac{R_1}{R_1 + R_2} x_3 + \dfrac{u_1(t)}{R_1 + R_2} \right]$

$$\dot{x}_2 = \frac{-1}{C_2} u_2(t)$$

$$\dot{x}_3 = \frac{1}{L}\left[\frac{R_1 x_1}{R_1 + R_2} - \frac{(R_1 R_2 + R_1 R_3 + R_2 R_3) x_3}{R_1 + R_2} + \frac{R_2}{R_1 + R_2} u_1(t) - R_3 u_2(t) \right]$$

3.26 $x_1(k) = y_1(k)$, $x_2(k) = y_1(k+1) + 10 y_1(k) - 2u_1(k) - y_2(k)$, $x_3(k) = y_2(k)$. Then

$$\mathbf{x}(k+1) = \begin{bmatrix} -10 & 1 & 1 \\ -3 & 0 & -2 \\ 4 & 0 & -4 \end{bmatrix} \mathbf{x}(k) + \begin{bmatrix} 2 & 0 \\ 1 & 0 \\ -1 & 2 \end{bmatrix} \mathbf{u}(k)$$

$$\mathbf{y}(k) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}(k)$$

3.27 Let $x_1(k) = y(k), x_2(k) = y(k+1)$. Then

$$\mathbf{x}(k+1) = \begin{bmatrix} 0 & 1 \\ -2 & -3 \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k), \quad y(k) = [1 \quad 0]\mathbf{x}(k)$$

Let $x_1(k) = y(k),\ x_2(k) = y(k+1) + 3y(k)$. Then

$$\mathbf{x}(k+1) = \begin{bmatrix} -3 & 1 \\ -2 & 0 \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} 0 \\ 1 \end{bmatrix} u(k), \quad y(k) = [1 \quad 0]\mathbf{x}(k)$$

CHAPTER 4

4.35 $i_1 = \dfrac{(h_{22}R_L + 1)}{\Delta}v_1,\ i_2 = \dfrac{h_{21}}{\Delta}v_1,\ v_2 = \dfrac{-h_{21}R_L}{\Delta}v_1$, where $\Delta = (h_{11}h_{22} - h_{12}h_{21})R_L + h_{11}$

4.36 $v_1 = \dfrac{1}{\Delta}[h_{11}(1 + h_{22}R_L) - h_{12}h_{21}R_L]i_1,\ i_2 = \dfrac{h_{21}i_1}{\Delta},\ v_2 = -\dfrac{1}{\Delta}h_{21}R_L i_1$, where $\Delta = 1 + h_{22}R_L$

4.37 $|\mathbf{A}| = -468$

4.38 $$\mathbf{A}^{-1} = \frac{1}{17}\begin{bmatrix} 3 & 4 & -3 \\ 7 & -19 & 10 \\ -2 & 3 & 2 \end{bmatrix},\ \mathbf{B}^{-1} = \frac{1}{58}\begin{bmatrix} 59 & 2 & 3 & 4 & 5 \\ 1 & 60 & 3 & 4 & 5 \\ -4 & -8 & 46 & -16 & -20 \\ -2 & -4 & -6 & 50 & -10 \\ -8 & -16 & -24 & -32 & 18 \end{bmatrix}$$

$$\mathbf{C}^{-1} = \left[\begin{array}{cc|cc|cc} \frac{1}{5} & \frac{-2}{5} & 0 & 0 & 0 & 0 \\ \frac{2}{5} & \frac{1}{5} & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{7} & \frac{-5}{7} & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & \frac{-3}{55} & \frac{8}{55} \\ 0 & 0 & 0 & 0 & \frac{8}{55} & \frac{-3}{55} \end{array}\right]$$

4.39 $$\mathbf{A}(s) = \begin{bmatrix} \frac{1}{s} & \frac{1}{s^2} \\ \frac{1}{s+a} & \frac{b\beta}{s^2+\beta^2} \\ \frac{2}{s^3} & \frac{s+1}{(s+1)^2+\beta^2} \end{bmatrix}, \quad \mathbf{B}(s) = \begin{bmatrix} \frac{s}{s^2-\beta^2} & \frac{\beta}{s^2-\beta^2} \\ \frac{1}{(s+a)^2} & \frac{s}{s^2+\beta^2} \end{bmatrix}$$

4.40 $\mathbf{A}(t) = [t^4 \quad t\cos\beta t],\ \mathbf{B}(t) = \begin{bmatrix} e^{-t} - e^{-2t} & -e^{-t} + 2e^{-2t} \\ 0 & \delta(t) \end{bmatrix}$

4.41 **(a)** $\mathbf{S} = \begin{bmatrix} 2 & -0.5 & 1 \\ 0 & 2.78388 & 1.61645 \\ 0 & 0 & 2.32101 \end{bmatrix}$ **(b)** Not possible, **A** not positive definite.

(c) $$\mathbf{S} = \begin{bmatrix} 4 & 1 & 0.25 & -0.25 & 0.75 \\ 0 & 3 & 1.25 & 0.75 & -0.91666 \\ 0 & 0 & 4.83477 & 0.64636 & 0.40505 \\ 0 & 0 & 0 & 3.15551 & 2.41267 \\ 0 & 0 & 0 & 0 & 3.10035 \end{bmatrix}$$

4.42 $\mathbf{H}(s) = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} (s+1)^2 & 0 \\ 0 & s+1 \end{bmatrix}^{-1} = \begin{bmatrix} (s+1)^2 & 0 \\ 0 & s+1 \end{bmatrix}^{-1}\begin{bmatrix} 1 & s+1 \\ 0 & 1 \end{bmatrix}$

4.43 $\mathbf{H}(s) = \begin{bmatrix} 1 & 1 \\ s+1 & -1 \end{bmatrix}\begin{bmatrix} (s+1)(s+2) & -2(s+3) \\ 0 & s+3 \end{bmatrix}^{-1}$. Other valid answers may be found.

4.44 Using the observable canonical form for each subsystem gives

$$\mathbf{A} = \begin{bmatrix} -3 & 1 & 0 & 0 \\ -6 & 0 & 0 & 0 \\ 1 & 0 & -4 & 1 \\ 2 & 0 & -3 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 1 \\ 5 \\ 0 \\ 0 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \mathbf{D} = [0].$$

Using the controllable canonical form for each subsystem gives

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -6 & -3 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 5 & 1 & -3 & -4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 5 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}, \mathbf{D} = [0].$$

4.45

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -6 & -3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 5 & 1 & -3 & -4 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 & -6 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & -2 & -1 & 10 & -1 & -1 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}, \mathbf{D} = [0], \text{ and}$$

$$\mathbf{C} = \begin{bmatrix} 5 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -2 & 1 \end{bmatrix}$$

4.46 $\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -6 & -3 & -2 & -1 \\ 0 & 0 & 0 & 1 \\ 5 & 1 & -3 & -4 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \mathbf{C} = \begin{bmatrix} 5 & 1 & 0 & 0 \\ 0 & 0 & 2 & 1 \end{bmatrix}$

4.47 $Q \approx Q_n + (c/2)\sqrt{\rho g/h_n}\,\delta h$, where $h = h_n + \delta h$ and $Q_n = c\sqrt{\rho g h_n}$

4.48 **(a)** $\theta = \cos^{-1}\left[\dfrac{\langle(\mathbf{r}_1 - \mathbf{x}), (\mathbf{r}_2 - \mathbf{x})\rangle}{\|\mathbf{r}_1 - \mathbf{x}\|\,\|\mathbf{r}_2 - \mathbf{x}\|}\right]$

(b) $\delta\theta = [\nabla_{\mathbf{x}}\theta]_n^T\,\delta\mathbf{x}$. Letting $\mathbf{r}_1 - \mathbf{x} = \mathbf{u}$ and $\mathbf{r}_2 - \mathbf{x} = \mathbf{v}$, the three components of $\nabla_{\mathbf{x}}\theta$ are given by

$$\frac{\partial\theta}{\partial x_i} = \frac{(u_i + v_i)\|\mathbf{u}\|\,\|\mathbf{v}\| - (v_i\|\mathbf{u}\|^2 + u_i\|\mathbf{v}\|^2)\cos\theta}{\|\mathbf{u}\|^2\|\mathbf{v}\|^2\sin\theta}$$

CHAPTER 5

5.39 **(a)** $|\mathbf{G}| = 16$

(b) $\hat{\mathbf{v}}_1 = \dfrac{1}{\sqrt{14}}[1 \quad 2 \quad 3]^T$, $\hat{\mathbf{v}}_2 = \dfrac{1}{\sqrt{35}}[1 \quad -5 \quad 3]^T$, $\hat{\mathbf{v}}_3 = \dfrac{1}{\sqrt{10}}[-3 \quad 0 \quad 1]^T$

5.40 $\mathbf{r}_1 = \begin{bmatrix} \frac{5}{4} & \frac{1}{4} & -\frac{1}{4} \end{bmatrix}^T$, $\mathbf{r}_2 = \begin{bmatrix} -\frac{1}{4} & -\frac{1}{4} & \frac{1}{4} \end{bmatrix}^T$, $\mathbf{r}_3 = [-3 \quad 0 \quad 1]^T$

5.41 $\mathbf{z} = \dfrac{5}{\sqrt{14}}\hat{\mathbf{v}}_1 - \dfrac{23}{\sqrt{35}}\hat{\mathbf{v}}_2 - \dfrac{21}{\sqrt{10}}\hat{\mathbf{v}}_3$

5.42 $\mathbf{z} = \frac{37}{4}\mathbf{x}_1 - \frac{13}{4}\mathbf{x}_2 - 21\mathbf{x}_3$

5.43 $\mathbf{r} = \frac{1}{10}[3 \quad -1 \quad 0]^T$, $\mathbf{r}_2 = \frac{1}{70}[-7 \quad 19 \quad -10]^T$, $\mathbf{r}_3 = \frac{1}{7}[0 \quad -1 \quad 2]^T$

5.44 $\mathbf{r}_1 = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}^T$, $\mathbf{r}_2 = \begin{bmatrix} 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix}^T$, $\mathbf{r}_3 = [1 \quad 0 \quad -1]^T$, $\mathbf{z} = 2\mathbf{x}_1 - \mathbf{x}_2 + 5\mathbf{x}_3$

5.46

$$\mathbf{G} = \begin{bmatrix} 2.5784502E+03 & 6.3518805E+02 & -1.1518515E+02 & 3.3449918E+02 \\ 6.3518805E+02 & 6.2262292E+02 & 6.9457092E+01 & 8.6028557E+01 \\ -1.1518515E+02 & 6.9457092E+01 & 2.9569002E+01 & -1.9252985E+01 \\ 3.3449918E+02 & 8.6028557E+01 & -1.9252985E+01 & 1.1678545E+02 \end{bmatrix}$$

$|\mathbf{G}| = 3.12183 \times 10^8$

The orthonormal basis vectors $\mathbf{v}_i$ are

$$\begin{bmatrix} 8.7438685E-01 \\ 2.5207549E-01 \\ 2.9540095E-02 \\ -4.1356131E-01 \end{bmatrix}, \begin{bmatrix} -1.4671828E-01 \\ 8.4976387E-01 \\ 4.4605288E-01 \\ 2.3960790E-01 \end{bmatrix}, \begin{bmatrix} -3.4713072E-01 \\ 4.2903343E-01 \\ -6.5270102E-01 \\ -5.1904905E-01 \end{bmatrix}, \begin{bmatrix} 3.0564949E-01 \\ 1.7403936E-01 \\ -6.1167425E-01 \\ 7.0862055E-01 \end{bmatrix}$$

Then $\mathbf{x} = -0.6777\mathbf{v}_1 + 12.712\mathbf{v}_2 - 14.736\mathbf{v}_3 + 31.548\mathbf{v}_4$. Note that the sum of the squares of the components of $\mathbf{x}$ in the original and the new coordinates is 1374.5.

5.47 The Grammian is

$$\begin{bmatrix} 4.0000000E+00 & 4.0008001E+00 & -7.9900002E+00 \\ 4.0008001E+00 & 4.0016012E+00 & -7.9916024E+00 \\ -7.9900002E+00 & -7.9916024E+00 & 1.5960100E+01 \end{bmatrix}$$

Its determinant is not zero, but very small, because the three vectors are very nearly colinear. To proceed with these poorly conditioned vectors as a basis set would produce very unreliable results.

5.48 An orthonormal basis for the subspace is, from Gram-Schmidt,

$$\begin{bmatrix} 7.3127246E-01 \\ 3.6563623E-01 \\ -3.6563623E-01 \\ 7.3127240E-02 \\ 4.3876347E-01 \end{bmatrix}, \begin{bmatrix} -1.1037266E-01 \\ -2.8796402E-01 \\ -7.9833186E-01 \\ -4.9211112E-01 \\ -1.5933463E-01 \end{bmatrix}, \begin{bmatrix} 5.2741766E-01 \\ 1.2584069E-01 \\ 2.0743863E-01 \\ -2.8101894E-01 \\ -7.6419383E-01 \end{bmatrix}$$

The components of $\mathbf{x}_1$ and $\mathbf{x}_2$, expressed in these coordinates, are

$$\begin{bmatrix} 1.2431632E+00 \\ -1.8481143E+00 \\ -1.8451571E-01 \end{bmatrix}, \begin{bmatrix} 4.8995252E+00 \\ 3.3418725E+00 \\ -1.1238300E+01 \end{bmatrix}$$

5.49 Dimension is 2.

5.50 $\|\mathbf{x}_n\| = \frac{5}{\sqrt{14}}$, $x_1 = \frac{-5}{7}$, $x_2 = \frac{-15}{14}$, $x_3 = \frac{5}{14}$

5.51 $\mathbf{y}_p = \begin{bmatrix} \frac{1}{2} & 2 & \frac{9}{2} \end{bmatrix}^T$

5.52 The sine and/or cosine functions form an orthogonal basis set, $\{\mathbf{v}_i, i = 1, \infty\}$. The reciprocal basis $\{\mathbf{r}_i\}$ differs only by normalizing constants. The inner product is that defined in Problem 5.22 and the Fourier coefficients can be thought of as components of an infinite dimensional vector.

5.53 $\mathbf{A}$ must be an $n \times n$ real matrix which satisfies $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$. This last condition is the definition of a positive definite matrix.

5.55 $\langle \mathbf{f}, \mathbf{g} \rangle = \int_a^b \bar{\mathbf{f}}^T(\tau)\mathbf{g}(\tau)\, d\tau$, $\|\mathbf{f}\| = \langle \mathbf{f}, \mathbf{f} \rangle^{1/2}$

5.58 $\mathbf{A} = \begin{bmatrix} 0 & -1 & 2 \\ -3 & 2 & 2 \\ 0 & 2 & 3 \end{bmatrix}$

5.59 No, it depends upon the fact that $\mathbf{B}^{-1} = \mathbf{B}^T$.

5.60 $\mathbf{A}^{-1} = \dfrac{1}{\sin^2 \alpha} \begin{bmatrix} \hat{\mathbf{u}}^T - \hat{\mathbf{v}}^T \cos \alpha \\ \hat{\mathbf{v}}^T - \hat{\mathbf{u}}^T \cos \alpha \\ \mathbf{w}^T \end{bmatrix}$

5.61 $\dot{\mathbf{T}}_{BE} = \begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} \mathbf{T}_{BE}$, where $\boldsymbol{\omega} = [\omega_x \quad \omega_y \quad \omega_z]^T$ is the angular velocity of the body coordinates with respect to the inertial coordinates, which could be measured with body-mounted rate gyros. (*Hint:* Let $\mathbf{x}_I = [1 \quad 0 \quad 0]^T$ be a fixed inertial vector. The same vector expressed in body coordinates is $\mathbf{x}_B$, the first column of $\mathbf{T}_{BE}$. Then use Euler's formula for differentiating a vector in rotating coordinates,

$$\left.\frac{d\mathbf{x}}{dt}\right|_I = \left.\frac{d\mathbf{x}}{dt}\right|_B + \boldsymbol{\omega} \times \mathbf{x}_B$$

Repeat for two other vectors $\mathbf{y}_I$ and $\mathbf{z}_I$ and use the result of Problem 5.6 for the cross product.)

5.62 $\mathbf{y}_r = \frac{1}{7}[6 \quad 19 \quad 32]^T$, $\mathbf{y}_p = \frac{1}{14}[34 \quad 5 \quad 53]^T$

5.65 **(a)** *Hint:* Result of Problem 7.37 will be helpful. **(b)** *Hint:* Results of Problem 7.26 will be helpful.

5.66 Ignoring the boundary terms, the formal adjoint equation is $\mathbf{y}(k-1) = \mathbf{A}_k^T\, \mathbf{y}(k)$. Note that the time index k on the adjoint equation runs backward.

CHAPTER 6

6.30 $\mathbf{x} = \frac{1}{27}[40 \quad -49 \quad 48]^T$

6.31 $\mathbf{x}_2(0) = \phi_{22}^{-1}(T)[\mathbf{x}_2(T) - \phi_{21}(T)\mathbf{x}_1(0)]$, where $\boldsymbol{\Phi}(T)$ is partitioned into $\left[\begin{array}{c|c} \phi_{11}(T) & \phi_{12}(T) \\ \hline \phi_{21}(T) & \phi_{22}(T) \end{array}\right]$.

6.33 $\mathbf{x} = [-4\alpha \quad 0 \quad \alpha]^T$, α an arbitrary scalar

6.34 $\mathbf{x}_1 = [-1 \quad 0 \quad 1]^T$, $\mathbf{x}_2 = [-1 \quad 1 \quad 0]^T$ and any linear combination of these.

6.35 Yes, since $r_A = 1$, $\mathbf{x} = [-2\alpha \quad \alpha]^T$, α arbitrary.

6.36 The degeneracy of $\mathbf{A}$ is 3.; the null space basis is

$$\begin{bmatrix} 0.0000000E+00 \\ -1.0000000E+00 \\ 1.1764708E-01 \\ 4.1176471E-01 \\ 0.0000000E+00 \end{bmatrix}, \begin{bmatrix} -1.0000000E+00 \\ 0.0000000E+00 \\ -7.6470608E-01 \\ 8.2352948E-01 \\ 0.0000000E+00 \end{bmatrix}, \begin{bmatrix} 0.0000000E+00 \\ 0.0000000E+00 \\ 6.4705873E-01 \\ 7.6470590E-01 \\ -1.0000000E+00 \end{bmatrix}$$

6.37 **(a)** $x_1 = \frac{13}{5}$, $x_2 = \frac{1}{5}$ **(b)** $x_1 = \frac{84}{35}$, $x_2 = \frac{14}{35}$

6.38 $x = 2$. Column space is one dimensional with basis $[2 \quad 1]^T$. $\mathbf{y} - \mathbf{A}x$ is perpendicular to this line, i.e., $\mathbf{A}x$ is the orthogonal projection of $\mathbf{y}$ onto this line.

6.40 $a = 1$, $b = 3$

6.41 $a = \frac{32}{37}$, $b = \frac{107}{37}$, $\|\mathbf{e}\|^2 = \frac{8}{37}$

6.42 $a = \frac{104}{109}$, $b = \frac{323}{109}$

6.43 $i(0) \cong 17.3$

6.44 $\mathbf{a} = [2.6286 \quad 0.082145 \quad -0.003572]^T$, $\|\mathbf{y}_e\| = 0.4326$, and the estimated final GPA is 3.057, which is probably more realistic even though the residual error is larger here.

6.45 $C = 5.3704$, $\alpha = 1.077$, estimated mile time is 5 min, 22 s

6.46 After processing first four equations $\mathbf{x}_4 = [-.566 \quad -.033 \quad .533]^T$, and after five $\mathbf{x}_5 = [-.472 \quad -.127 \quad .627]^T$. Roughly these same values were obtained using several initial estimates for **x**, as long as the first **P** is very large. Disagreement with Example 6.7 after $k = 4$ would not be surprising since unreliable results can be given by the recursive equations whenever **A** is not full rank.

6.47 Introduction of a suitable weighting matrix can be handled by defining a weighted inner product, $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle_Q = \mathbf{x}_1^T \mathbf{Q} \mathbf{x}_2$. This new inner product modifies the meaning of $\mathscr{A}^*$, since $\langle \mathbf{y}, \mathscr{A}(\mathbf{x}) \rangle_Q = \langle \mathscr{A}^*(\mathbf{y}), \mathbf{x} \rangle_Q$. With this definition for $\mathscr{A}^*$, the previous results still apply.

6.48 $\mathbf{X} = \begin{bmatrix} -0.5 & 0.25 \\ 0.25 & -0.25 \end{bmatrix}$.

6.49 No solution exists. **B** has $+1$ as an eigenvalue and **A** has -1 as an eigenvalue. The 4×4 **Q** matrix is singular, of rank 3 and the augmented **W** matrix has rank 4. The equations are inconsistent and no solution exists.

CHAPTER 7

7.39 $\lambda_i = 1, -2, 3$. $\mathbf{x}_i = [-1 \quad 1 \quad 1]^T$, $[11 \quad 1 \quad -14]^T$, and $[1 \quad 1 \quad 1]^T$. $\mathbf{J} = \text{diag}[1, -2, 3]$

7.40 $\lambda_1 = \lambda_2 = 2$, $m = 2$, $q = 2$, $\mathbf{x}_1 = [1 \quad 0]^T$, $\mathbf{x}_2 = [0 \quad 1]^T$

7.41 $|\mathbf{A} - \mathbf{I}\lambda| = 0 \Rightarrow \lambda^3(2 - \lambda) = 0$. Therefore, $\lambda = 0$ is an eigenvalue with algebraic multiplicity 3 and index $k = 2$. $\text{Rank}(\mathbf{A} - \mathbf{I}\lambda)|_{\lambda=0} = 2$. Therefore, $q = 2$, so there are two eigenvectors $\mathbf{x}_1, \mathbf{x}_3$ and one generalized eigenvector associated with $\lambda = 0$, and they are solutions of $\mathbf{A}^2\mathbf{x} = \mathbf{0}$. $\lambda_4 = 2$ has the eigenvector $\mathbf{x}_4$. These are shown as columns of

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

7.42 $\lambda_1 = 3 + j$, $\lambda_2 = 3 - j$, $\lambda_3 = 6$, $\mathbf{x}_1 = [6 - 2j \quad 2 - 4j \quad 0]^T$, $\mathbf{x}_2 = \bar{\mathbf{x}}_1$, $\mathbf{x}_3 = [0 \quad 0 \quad 1]^T$, $\mathbf{J} = \text{diag}[3 + j,\ 3 - j,\ 6]$

7.43 No. Although both matrices have $\lambda_1 = \lambda_2 = 2$, the Jordan form for **A** is $\mathbf{J} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$. Since **B** is already in Jordan form and since $\mathbf{J} \neq \mathbf{B}$, they are not similar.

7.44 $\lambda_1 = 15$, $\mathbf{x}_1 = [1 \quad 1 \quad -1]^T$, $\lambda_2 = 9$, $\mathbf{x}_2 = [1 \quad -2 \quad -1]^T$, $\lambda_3 = 3$, $\mathbf{x}_3 = [1 \quad 0 \quad 1]^T$

7.45 $\lambda_1 = 5.049$, $\lambda_2 = 0.643$, $\lambda_3 = 0.308$, $\mathbf{x}_1 = [1 \quad 0.802 \quad 0.445]^T$, $\mathbf{x}_2 = [1 \quad -0.555 \quad -1.247]^T$, $\mathbf{x}_3 = [1 \quad -2.247 \quad 1.802]^T$

7.46 **(a)** *Hint:* Set $\lambda = 0$ in the definition of $\Delta(\lambda)$ and in $\Delta'(\lambda)$. **(b)** Use results of Problem 7.15 and note: (i) the only factor in $\Delta(\lambda)$ which involves all the diagonal elements a_{ii}, and which gives rise to all λ^{n-1} terms, is of the form $(a_{11} - \lambda)(a_{22} - \lambda) \cdots (a_{nn} - \lambda)$, and (ii) the form of the coefficient of λ^{n-1}, and (iii) $c_i' = (-1)^n c_i$.

7.47 $\lambda_i = \{-4.33945,\ 2.169727 \pm 2.4745j\}$, $\mathbf{x}_1 = [0.69505 \quad -0.355598 \quad -1]^T$, $\mathbf{x}_2 = [0.863702 - 0.181845j \quad -j \quad 0.1867736 - 0.116646j]^T$ and $\mathbf{x}_3 = \bar{\mathbf{x}}_2$. $\mathbf{J} = \text{Diag}[\lambda_1, \lambda_2, \lambda_3]$

7.48 $\lambda_i = \{1.48431, 6.257845 \pm 1.123484j\}$, $\mathbf{x}_1 = [0.97929 \quad -1 \quad -0.4636]^T$, $\mathbf{x}_2 = [-0.5107 - 0.41175j \quad -0.30786 - 0.17794 \quad -1]^T$, and $\mathbf{x}_3 = \bar{\mathbf{x}}_2$. $\mathbf{J} = \text{Diag}[\lambda_1, \lambda_2, \lambda_3]$

7.49 **(a)** Negative definite, **(b)** positive semidefinite, **(c)** negative semidefinite, **(d)** positive definite, **(e)** indefinite

7.50 $\mathbf{x} = (1/\sqrt{2})[-1 \quad 1]^T$ (an eigenvector), $\mathbf{Q}_{\max} = 4 = \lambda_{\max}$ = eigenvalue associated with **x**.

7.51 From equation (7.5), $\mathbf{E}_i = \mathbf{x}_i \rangle \langle \mathbf{r}_i$. This is a projection since $\mathbf{E}_i \mathbf{E}_i = \mathbf{E}_i$ and $\mathbf{E}_i \mathbf{E}_j = 0$ with $i \neq j$.

7.52 Similar to the preceding problem except the eigenspaces are now mutually orthogonal.

CHAPTER 8

8.25 $\mathbf{A}^{-1}=\frac{1}{6}[-\mathbf{A}^2+2\mathbf{A}+3\mathbf{I}]=\frac{1}{6}\begin{bmatrix}-2 & 4 & 0\\ 2 & 2 & 0\\ 3 & -3 & 3\end{bmatrix}$

8.26 $\begin{bmatrix}e^{-t} & \frac{1}{4}(e^{t}-e^{-t})\\ 0 & e^{t}\end{bmatrix}$

8.27 $\begin{bmatrix}e^{-3t} & 2te^{-3t}\\ 0 & e^{-3t}\end{bmatrix}$

8.29 $\begin{bmatrix}\frac{1}{2}(e^{7t}+e^{-t}) & \frac{5}{8}(e^{7t}-e^{-t})\\ \frac{2}{5}(e^{7t}-e^{-t}) & \frac{1}{2}(e^{7t}+e^{-t})\end{bmatrix}$

8.30 $\frac{1}{3}\begin{bmatrix}e^{-t}+2e^{-4t} & e^{-t}-e^{-4t}\\ 2(e^{-t}-e^{-4t}) & 2e^{-t}+e^{-4t}\end{bmatrix}$

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

8.32 $\begin{bmatrix}-3e^{-t}+4e^{-2t} & -6(e^{-t}-e^{-2t}) & 0\\ 2(e^{-t}-e^{-2t}) & 4e^{-t}-3e^{-2t} & 0\\ 0 & 0 & e^{-3t}\end{bmatrix}$

8.33 $\begin{bmatrix}\alpha_0 & \alpha_1 & \alpha_2\\ \alpha_2 & \alpha_0-3\alpha_2 & \alpha_1+3\alpha_2\\ \alpha_1+3\alpha_2 & -3\alpha_1-8\alpha_2 & \alpha_0+3\alpha_1+6\alpha_2\end{bmatrix}$ with $\begin{cases}\alpha_0=e^{t}-te^{t}+\frac{1}{2}t^2e^{t}\\ \alpha_1=te^{t}-t^2e^{t}\\ \alpha_2=\frac{1}{2}t^2e^{t}\end{cases}$

8.34

$$e^{At}=\begin{bmatrix}0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 0.1 & 0.1054 & -0.0738 & 0.1\end{bmatrix}+e^{-t}\begin{bmatrix}1 & 0.1047 & 0.9739 & 0\\ 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ -0.1 & -0.0105 & -0.0974 & 0\end{bmatrix}$$

$$+e^{-0.7t}\sin(3t)\begin{bmatrix}0 & 0.3246 & -0.3142 & 0\\ 0 & 0 & 3 & 0\\ 0 & -0.333 & 0 & 0\\ 0 & -0.0571 & -0.2847 & 0\end{bmatrix}$$

$$+e^{-0.7t}\cos(3t)\begin{bmatrix}0 & -0.1047 & -0.9739 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & -0.0949 & 0.1712 & 0\end{bmatrix}$$

8.35 $\mathbf{A}^k=(0.5)^k\begin{bmatrix}0 & 4 & -4\\ 0 & 2 & -2\\ 0 & 1 & -1\end{bmatrix}+\begin{bmatrix}0 & -1 & 2\\ 0 & -1 & 2\\ 0 & -1 & 2\end{bmatrix}$

8.38

$$e^{\mathbf{A}t}=\begin{bmatrix}e^{3t}[\cos(t)+\sin(t)] & -2e^{3t}\sin(t) & 0\\ e^{3t}\sin(t) & e^{3t}[\cos(t)-\sin(t)] & 0\\ 0 & 0 & e^{6t}\end{bmatrix}$$

With $t = 0.2$ this gives $\mathbf{A}_1 \approx \begin{bmatrix} 2.147797 & -0.72399 & 0 \\ 0.361999 & 1.423799 & 0 \\ 0 & 0 & 3.320117 \end{bmatrix}$ with

$\lambda = \{3.320117, 1.785798 \pm 0.36199j\} = \{e^{1.2}, e^{0.6}[\cos(0.2) \pm \sin(0.2)j]\}$

8.41 An easy approximation is given by using the dominant eigenvalue, λ_d, that is, the eigenvalue with the smallest nonzero real part. This mode is called dominant since it will take longest to decay. Let the real part of λ be $-r$. Then, solving $e^{-rt} = 0.02$ gives $T_s \approx -\ln(0.02)/r \approx 4/r$. The term $1/r$ is an approximation of the dominant time constant τ_d. A 1% definition of settling is often used, and this changes the 4 to 4.6. This rough approximation may not be sufficiently accurate if the one mode is not truly dominant [4].

8.42 Nyquist's sampling theorem requires that $\omega_{\max} T < \pi$, that is, there must be at least two samples per period of the highest frequency to avoid aliasing [4] and loss of information. This theoretical limit is almost never enough in controls problems. For adequate representation of states a rough rule of thumb is more like 6 to 10 samples per period, or $T < 2\pi/6\omega_{\max}$. The same 6 to 10 rule is often applied to the number of samples per dominant time constant as well. The value used for $\omega_{\max}$ is usually interpreted as the highest *significant* modal frequency to be retained in the model. The distinction is necessary because some physical problems involve many or even an infinite number of frequencies due to structural vibration or other causes [5]. Other modes may have such a large negative real part of λ that they rapidly decay to insignificance.

CHAPTER 9

9.29 $y_1(t) = +\frac{9}{2}te^{-3t} + \frac{25}{4}e^{-3t} - \frac{21}{4}e^{-t}$; $y_2(t) = \frac{7}{2}e^{-t} - \frac{3}{2}e^{-3t}$

9.30 $y_1(t) = 11e^{-2t} - 3e^{-3t} - 7e^{-t}$; $y_2(t) = \frac{7}{2}e^{-t} - \frac{3}{2}e^{-3t}$

9.31 $\mathbf{x}(t) = \begin{bmatrix} (10 - 9t)e^{-t} \\ (1 - 9t)e^{-t} \end{bmatrix}$

9.32 $\mathbf{x}(t) = \begin{bmatrix} \frac{83}{9}e^{-t} - \frac{25}{3}te^{-t} + \frac{7}{9}e^{2t} \\ \frac{8}{9}e^{-t} - \frac{25}{3}te^{-t} + \frac{1}{9}e^{2t} \end{bmatrix}$

9.33 $\mathbf{x}(t_f) = \mathbf{0}$. Note that $\mathbf{u}(t) = -\frac{1}{t_f}\mathbf{\Phi}(t, 0)\mathbf{x}(0)$.

9.35 $\mathbf{\Phi}(t, 0) = e^{-t/2}\begin{bmatrix} \cos\frac{3}{2}t - \frac{1}{3}\sin\frac{3}{2}t & -\frac{5}{3}\sin\frac{3}{2}t \\ \frac{2}{3}\sin\frac{3}{2}t & \cos\frac{3}{2}t + \frac{1}{3}\sin\frac{3}{2}t \end{bmatrix}$

9.37 $x_1(k) = 2 - 5k + 5k(3 - k)$, $x_2(k) = 5 + 10k$, $x_3(k) = 10$

9.38 $\mathbf{x}(k) = \begin{bmatrix} 0 \\ 10 \end{bmatrix} + \sum_{j=1}^{k} (0.632)(0.368)^{k-j}\begin{bmatrix} 1 \\ -1 \end{bmatrix}$

9.39 $\mathbf{x}(k) = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}\mathbf{q}(k)$ gives $\mathbf{q}(k+1) = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}\mathbf{q}(k) + \begin{bmatrix} 1 \\ 0 \end{bmatrix}u(k)$

$\mathbf{y}(k) = [4 \quad 1]\mathbf{q}(k)$

9.40 No. The system matrix $\mathbf{A}$ is singular and consequently $\mathbf{\Phi}^{-1}$ does not exist.

9.41 $\dot{\mathbf{x}} = \begin{bmatrix} 0 & 1 \\ 0 & -1/\tau \end{bmatrix}\mathbf{x} + \begin{bmatrix} 0 \\ K/\tau \end{bmatrix}u(t)$

$$\mathbf{x}(k+1) = \begin{bmatrix} 1 & \tau(1 - e^{-\Delta t/\tau}) \\ 0 & e^{-\Delta t/\tau} \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} \Delta t - \tau(1 - e^{-\Delta t/\tau}) \\ 1 - e^{-\Delta t/\tau} \end{bmatrix}Ku(k)$$

9.42 $\mathbf{x}(k+1) = \begin{bmatrix} 1 - K\Delta t + K\tau(1-e^{-\Delta t/\tau}) & \tau(1-e^{-\Delta t/\tau}) \\ -K(1-e^{-\Delta t/\tau}) & e^{-\Delta t/\tau} \end{bmatrix}\mathbf{x}(k)$

$$+ \begin{bmatrix} \Delta t - \tau(1-e^{-\Delta t/\tau}) \\ 1 - e^{-\Delta t/\tau} \end{bmatrix} Kr(k)$$

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

9.43 $\mathbf{A} = \begin{bmatrix} 0.3666 & 0.1010 & 0.0110 \\ 0 & 0.6703 & 0.1341 \\ 0 & 0 & 0.6703 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0.0136 \\ 0.1802 \\ 0.1648 \end{bmatrix}$

9.44 One of many possible answers, obtained by using a parallel realization, is

$$\mathbf{x}(k+1) = \begin{bmatrix} 0.04979 & 0 & 0 \\ 0 & 0.22313 & 0 \\ 0 & 0 & 0.60653 \end{bmatrix}\mathbf{x}(k) + \begin{bmatrix} 0.2121 \\ 0.1934 \\ 0.0649 \end{bmatrix} u(k)$$

$$y(k) = [0.04979 \quad -0.22313 \quad 0.60653]\mathbf{x}(k)$$

CHAPTER 10

10.16 **(a)** Eigenvalues are $\lambda_1 = -4, \lambda_2 = 2$. Every term in $\mathbf{\Phi} \to \infty$, as e^{2t} when $t \to \infty$. Therefore, $\|\mathbf{\Phi}(t, 0)\| \to \infty$ and the system is unstable. **(b)** Eigenvalues are $\lambda = -1, -2$, and -3. $\|\mathbf{\Phi}(t, 0)\| \to 0$ as $t \to \infty$ and the system is globally asymptotically stable. **(c)** Eigenvalues are $\lambda = -\frac{1}{4}$ and $-\frac{3}{4}$. The system is asymptotically stable.

10.17 **(a)** $\lambda_1 = 1, \lambda_2 = \frac{1}{2}$. Stable i.s.L. but not asymptotically stable; observable but not controllable. **(b)** $\lambda_1 = -1, \lambda_2 = -3, \lambda_3 = -5$. Asymptotically stable, uncontrollable and unobservable. **(c)** $\lambda = 1 \pm \sqrt{21}$. System is unstable, completely controllable and observable. These results illustrate that stability, controllability and observability are independent system properties. One property does not imply or require any of the others.

10.18 $u_1(t) = -M_1 \operatorname{sign}[7x_1(t) - x_2(t)]$, $u_2(t) = -M_2 \operatorname{sign}[-8x_1(t) + 19x_2(t)]$

CHAPTER 11

11.26 System is completely controllable and completely observable.

11.27 **(a)** Completely controllable, **(b)** completely controllable, **(c)** not completely controllable.

11.28 **(a)** Not controllable, not observable. This system will reappear in Example 12.2. **(b)** Not controllable; it is observable. This system is a special case of Problem 12.1b with $\alpha = 1$. **(c)** Both controllable and observable. See also Problems 12.8 and 12.9.

11.29 Yes. When put into state variable form, the system is completely controllable.

11.30 Completely controllable and completely observable for all finite values of k_1 and b_2 except when $k_1 b_2 = 1$.

11.31 Controllable, unless $R_3 C_2 = R_2 C_1$. Observable unless $R_1 = 0$. In these cases pole-zero cancellation occurs.

11.32 **(a)** Uncontrollable and unobservable. Two identical *RLC* systems in parallel. **(b)** Not completely controllable, due to pole-zero cancellation. Completely observable.

11.33 *Hint:* Use the semigroup property of transition matrices.

11.34 No. Rank of **P** is only 2.

11.35 The system *is* observable. The Kalman observable canonical form matrices are

$$\mathbf{A}=\begin{bmatrix}0.5 & 0.866 & 2.4495\\ -0.28867 & -0.5 & 2.8284\\ -0.8165 & -1.4142 & 2.0\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}3.5355 & 2.1213\\ 2.8577 & 1.2247\\ 1.1547 & 0\end{bmatrix},\ \mathbf{C}=\begin{bmatrix}1.4142 & 0 & 0\\ 0.7071 & 1.2247 & 0\end{bmatrix}$$

11.36 **(a)** Controllable. The Kalman controllable canonical matrices are

$$\mathbf{A}=\begin{bmatrix}-4. & -1.9466 & -1.4868\\ -0.64888 & -6.1053 & -3.1355\\ 0.84958 & 4.0657 & 1.1053\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}1.5275 & 0.43643\\ 0 & 1.3452\\ 0 & 0\end{bmatrix},$$

$$\mathbf{C}=\begin{bmatrix}0.6546 & 2.01778 & 3.0822\\ 0.65465 & 4.99134 & 5.3533\end{bmatrix}$$

(b) Not observable. Kalman observable canonical matrices are

$$\mathbf{A}=\begin{bmatrix}-1.0714 & 0.37115 & 0\\ -4.925 & -3.9286 & 0\\ -5.5549 & -1.06905 & -4\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}0.26726 & 0.80178\\ -0.66865 & 0.87438\\ 1.34715 & 0.7698\end{bmatrix},\ \mathbf{C}=\begin{bmatrix}3.74165 & 0 & 0\\ 7.216 & 1.3887 & 0\end{bmatrix}$$

CHAPTER 12

12.19 $\dot{\mathbf{x}}=\begin{bmatrix}-1/R_1C-1/R_2C & -1/C\\ 1/L & 0\end{bmatrix}\mathbf{x}+\begin{bmatrix}-1/R_2C\\ 1/L\end{bmatrix}u,\ y=[1\quad 0]\mathbf{x}+u,$

$$H(s)=\frac{s(s+1/R_1C)}{s^2+(1/R_1C+1/R_2C)s+1/LC}$$

12.20 $\dot{\mathbf{x}}=\begin{bmatrix}-1/CR_1 & 0 & 1/C & 0\\ 0 & -1/CR_1 & 0 & -1/C\\ -1/L & 0 & -R_2/L & -R_2/L\\ 0 & 1/L & -R_2/L & -R_2/L\end{bmatrix}\mathbf{x}+\begin{bmatrix}0 & 0\\ 0 & 0\\ 1/L & 0\\ 0 & 1/L\end{bmatrix}\begin{bmatrix}u_1\\ u_2\end{bmatrix},\ \mathbf{y}=\begin{bmatrix}1 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\end{bmatrix}\mathbf{x}$

12.21 $\mathbf{H}(s)=\begin{bmatrix}0 & 0\\ 1/(s+3) & 0\end{bmatrix}$

12.22 First-order state equation $\dot{x}_1=[-x_1/R_1+u_1/R_1+u_2]/C,\ y_1=x_1,\ y_2=x_1+u_2R_2,$

$$\mathbf{H}(s)=\frac{\begin{bmatrix}1/R_1 & 1\\ 1/R_1 & 1+R_2C(s+1/R_1C)\end{bmatrix}}{(s+1/R_1C)C}$$

12.23 $\mathbf{D}=\begin{bmatrix}0 & 0 & 1\\ 0 & 0 & 1\end{bmatrix}$. One possible choice for $\{\mathbf{A},\mathbf{B},\mathbf{C}\}$ is

$$\mathbf{A}=\begin{bmatrix}-1 & 0 & 0 & 0\\ 0 & -1 & 0 & 0\\ 0 & 0 & -2 & 0\\ 0 & 0 & 0 & -3\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}1 & 1 & -1\\ 0 & 1 & 0\\ 0 & 1 & 0\\ 0 & 1 & 4\end{bmatrix},\ \mathbf{C}=\begin{bmatrix}1 & 0 & -1 & 0\\ 0 & \frac{1}{2} & 0 & -\frac{1}{2}\end{bmatrix}$$

12.24 **(a)** is reducible since it is completely observable but not controllable. **(b)** is reducible since it is completely controllable but not observable. **(c)** is irreducible. It is completely controllable and observable.

12.25 $\mathbf{D}=\begin{bmatrix}1&0&0\\0&0&0\end{bmatrix}$. One possible solution for $\{\mathbf{A},\mathbf{B},\mathbf{C}\}$ is

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

12.26 $\mathbf{D}=\begin{bmatrix}0&0\\0&0\end{bmatrix}$. One possible solution for $\{\mathbf{A},\mathbf{B},\mathbf{C}\}$ is

$$\mathbf{A}=\left[\begin{array}{cc|c|c}-1&1&0&0\\0&-1&0&0\\\hline 0&0&-1&0\\\hline 0&0&0&0\end{array}\right],\ \mathbf{B}=\left[\begin{array}{cc}0&0\\1&0\\\hline 1&1\\\hline 1&0\end{array}\right],\ \mathbf{C}=\left[\begin{array}{cc|c|c}1&-2&1&0\\0&0&1&1\end{array}\right]$$

Another solution has the same $\mathbf{A}$, but $\mathbf{B}=\left[\begin{array}{cc}-1&1\\1&0\\\hline 1&1\\\hline 1&0\end{array}\right]$ and $\mathbf{C}=\left[\begin{array}{cc|c|c}1&0&0&0\\0&0&1&1\end{array}\right]$. A third solution is given by Chen (see page 251 of Reference 4).

12.27 One eighth-order irreducible realization is

$$\dot{\mathbf{x}}=\left[\begin{array}{cccc|ccc|c}0&1&0&0&&&&\\0&0&1&0&&\mathbf{0}&&\\0&0&0&1&&&&\\0&0&0&0&&&&\\\hline &&&&0&1&0&\\&\mathbf{0}&&&0&0&1&\\&&&&0&0&0&\\\hline &&&&&&&0\end{array}\right]\mathbf{x}+\left[\begin{array}{ccc}0&0&0\\-1&0&1\\0&0&0\\1&1&-2\\\hline 0&0&0\\2&0&-2\\1&0&1\\\hline 1&0&-1\end{array}\right]\mathbf{u}$$

$$\mathbf{y}=\left[\begin{array}{cccc|ccc|c}1&0&0&0&0&0&0&1\\1&1&0&0&\frac{1}{2}&0&0&1\\1&0&-1&0&-1&-5&1&10\end{array}\right]\mathbf{x}$$

See Reference 4 for a detailed solution.

12.28 $$\mathbf{T}(z)=\begin{bmatrix}\dfrac{0.5z^2-1.3z+1.245}{(z-1)}&-(z+0.3)\\\dfrac{1.5z^2-1.2z+0.145}{(z-1)}&2(z-0.5)\end{bmatrix}\frac{1}{(z^2-z+0.89)}$$

12.29 One possible answer is

$$\mathbf{A}=\begin{bmatrix}1&0&0&0\\0&0.3679&0&0\\0&0&1&1\\0&0&-0.5&0\end{bmatrix}\ \mathbf{B}=\begin{bmatrix}1\\1\\-0.9878\\-1.3761\end{bmatrix}\ \mathbf{C}=\begin{bmatrix}3.1641&-2.1762&1&0\\1.8984&-0.8984&0&0\end{bmatrix}\ \mathbf{D}=[0]$$

12.30 $$\mathbf{H}(z)=\begin{bmatrix}(z-1)(z-0.5)&0\\0&(z-0.5)^2(z-0.7)\end{bmatrix}^{-1}\begin{bmatrix}0.2z&3z(z+0.3)\\z+0.8&z(z-0.5)(z-0.7)\end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1.5 & 1 & 0 & 0 & 0 \\ -0.5 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1.7 & 1 & 0 \\ 0 & 0 & -0.95 & 0 & 1 \\ 0 & 0 & 0.175 & 0 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0.2 & 5.4 \\ 0 & -1.5 \\ 0 & 0.5 \\ 1 & -0.6 \\ 0.8 & 0.175 \end{bmatrix},$$

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 & 3 \\ 0 & 1 \end{bmatrix}$$

12.31 Original eighth-order system: Same as in Problem 12.10 except for the 2,2 block representation of h_{12}. Those blocks change to $\mathbf{A}_3 = \begin{bmatrix} 0 & 1 \\ -0.5 & 1.5 \end{bmatrix}$ and $\mathbf{C}_3 = \begin{bmatrix} -1.5 & 5.4 \\ 0 & 0 \end{bmatrix}$. **B** and **D** are unchanged. Original Rank(P) = 6, Rank(Q) = 5. Use **Q** to select fifth-order reduced system with

$$\mathbf{A} = \begin{bmatrix} 1.263911 & 0 & 0.150202 & 0 \\ 0 & 0.489418 & 0 & 8.13180 \\ -1.34220 & 0 & 0.236089 & 0 \\ 0 & -0.78092 & 0 & 1.220540 \\ 0 & 0.2418970 & 0 & 0.164580 \end{bmatrix},$$

$$\mathbf{B} = \begin{bmatrix} 0.035663 & 0.962905 \\ 0 & 0.363697 \\ 0.56055 & -0.267254 \\ 0.894503 & 0.004733 \\ 0.438318 & 0.174045 \end{bmatrix}, \mathbf{C}^T = \begin{bmatrix} 5.608 & 0 \\ 0 & 1.3748 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix},$$

and $\mathbf{D} = \begin{bmatrix} 0 & 3 \\ 0 & 1 \end{bmatrix}$. This realization is also controllable and hence irreducible.

12.32 $\mathbf{H}' = \begin{bmatrix} \dfrac{3}{s+1} & \dfrac{2s(s^2+2s+10)(s+5)}{(s+1)(s^2+2s+10)} \end{bmatrix} \begin{bmatrix} (s^2+2s+10)(s+5) & 0 \\ 0 & (s^2+2s+10)(s+5) \end{bmatrix}^{-1}$

can be reduced to $\mathbf{H}' = \begin{bmatrix} \dfrac{3}{s+1} & \dfrac{2s}{s+1} \end{bmatrix} \begin{bmatrix} (s^2+2s+10)(s+5) & 0 \\ 0 & s+5 \end{bmatrix}^{-1}$ from which a minimal realization is

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -50 & -20 & -7 & 0 \\ 0 & 0 & 0 & -5 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{C}^T = \begin{bmatrix} 15 & 1 \\ 3 & 1 \\ 0 & 0 \\ -10 & -4 \end{bmatrix}, \mathbf{D} = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix}$$

12.33 The original eighth-order realization is

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ -50 & -20 & -7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & -50 & -20 & -7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -5 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}, \mathbf{C}^T = \begin{bmatrix} 3 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 0 \\ -10 & 0 \\ 0 & -4 \end{bmatrix}$$

and $\mathbf{D} = \begin{bmatrix} 0 & 2 \\ 0 & 1 \end{bmatrix}$. For this system, Rank(P) = 4 and Rank(Q) = 6. The fourth-order controllable realization is

$$\mathbf{A}=\begin{bmatrix}-7 & 0 & -2 & -5\\ 0 & -5 & 0 & 0\\ 1 & 0 & 0 & 0\\ 0 & 0 & 1 & 0\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}1.4142 & 0\\ 0 & 1.4142\\ 0 & 0\\ 0 & 0\end{bmatrix},\ \mathbf{C}^T=\begin{bmatrix}0 & 0\\ -7.071 & -2.8284\\ 0 & 0.7071\\ 2.1213 & 0.7071\end{bmatrix}$$

and $\mathbf{D}=\begin{bmatrix}0 & 2\\ 0 & 1\end{bmatrix}$. This realization is observable and hence irreducible.

12.34 After reduction to the minimal determinant degree, one answer is

$$\mathbf{H}(s)=\begin{bmatrix}(s+1)(s+3) & -.25(s+2)(3s+5)\\ 0 & (s+1)(s+2)(s+5)\end{bmatrix}^{-1}\begin{bmatrix}-1.5(s+2) & .25(s-1)\\ 2(s+2)(s+5) & (s+1)^2\end{bmatrix}$$

From this, a fifth-order realization is

$$\mathbf{A}=\begin{bmatrix}-4 & 1 & -0.25 & 0 & 0\\ -3 & 0 & 0.25 & 0 & 0\\ 0 & 0 & -8 & 1 & 0\\ 0 & 0 & -17 & 0 & 1\\ 0 & 0 & -10 & 0 & 0\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}-1.5 & 0.25\\ -3 & -0.25\\ 2 & 1\\ 14 & 2\\ 20 & 1\end{bmatrix},\ \mathbf{C}^T=\begin{bmatrix}1 & 0\\ 0 & 0\\ 0.75 & 1\\ 0 & 0\\ 0 & 0\end{bmatrix},\ \mathbf{D}=[0]$$

12.35 $\mathbf{H}(s)=\begin{bmatrix}(s+2)^2(s+5) & -1\\ 0 & (s+2)\end{bmatrix}^{-1}\begin{bmatrix}0 & (s+2)^2\\ 1 & 0\end{bmatrix}$

$$\mathbf{A}=\begin{bmatrix}-9 & 1 & 0 & 0\\ -24 & 0 & 1 & 0\\ -20 & 0 & 0 & 1\\ 0 & 0 & 0 & -2\end{bmatrix},\ \mathbf{B}=\begin{bmatrix}0 & 1\\ 0 & 4\\ 0 & 4\\ 1 & 0\end{bmatrix},\ \mathbf{C}=\begin{bmatrix}1 & 0 & 0 & 0\\ 0 & 0 & 0 & 1\end{bmatrix}$$

CHAPTER 13

13.22 $\mathbf{K}=\begin{bmatrix}1 & -5 & -1\\ 0 & 9 & 3\end{bmatrix}$. (*Hint:* Define $\alpha=\dfrac{1}{\lambda+3}$ and let $\lambda\to -3$ after finding $\mathbf{G}^{-1}$.)

13.24 One solution is $\mathbf{K}'=[17\ \ 0]^T$, another is $\mathbf{K}'=[0\ \ 17]^T$.

13.25 One solution is $\mathbf{K}'=\begin{bmatrix}0 & 6\\ 9 & 0\end{bmatrix}$. For another solution using a slightly different method, see [6].

13.26 $\mathbf{K}'=[-1/\epsilon^2\ \ 1/\epsilon^2]$

13.27 No, because $\mathbf{N}=\begin{bmatrix}0 & 1\\ 0 & 1\end{bmatrix}$ is singular [11].

13.28 No, $\mathbf{N}=\begin{bmatrix}1 & 0\\ 1 & 0\end{bmatrix}$ is singular [11].

13.29 $\mathbf{F}_d=\begin{bmatrix}1 & -1\\ 0 & 1\end{bmatrix}$, $\mathbf{K}_d=\begin{bmatrix}-2 & -2 & 0\\ 0 & -3 & -4\end{bmatrix}$, $\mathbf{H}(s)=\begin{bmatrix}1/s & 0\\ 0 & 1/s^2\end{bmatrix}$

13.30 $\mathbf{K}=[1\ \ 0.5]$

13.31 $\mathbf{K}=[1.34\ \ 0.4]$

13.32 Two choices are

$$\mathbf{K}=\begin{bmatrix}0.4 & -0.3\\ 0.25 & 0.15\end{bmatrix}\ \text{and}\ \begin{bmatrix}0.3 & -0.4\\ 0.15 & 0.25\end{bmatrix}$$

13.33 Using the nested integrator method (Example 3.6) gives

$$\mathbf{A} = \begin{bmatrix} -3 & 1 & 0 & 0 & 0 \\ -2 & -2 & 2 & 1 & 0 \\ 2 & 0 & -6 & 0 & 1 \\ 4 & -4 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 5 & 1 \\ 0 & 1 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad \mathbf{D} = [0]$$

One choice for **K** is

$$\begin{bmatrix} -2.1251 & -67.1349 & -26.9825 & -12.045 & -15.107 \\ 4.8752 & 171.368 & 70.0198 & 34.4558 & 40.8934 \end{bmatrix}$$

A choice for the full state observer is

$$\mathbf{L} = \begin{bmatrix} 6.3127 & 1.0443 & -0.0136 \\ -35.2083 & 13.0842 & 1.5313 \\ -43.6735 & -2.2322 & 9.6030 \\ -288.0065 & 49.2646 & -2.4314 \\ -367.7839 & -17.7946 & 59.789 \end{bmatrix}$$

13.36 $p(z) = z^2 + 0.3z - 0.009$; $q(z) = 0.3990z - 0.1515$; $e(z) = 0.3z - 0.001$
13.37 $p(z) = z + 0.09$; $q(z) = 0.31z - 0.035$; $e(z) = p(z) - d(z) = 0.01$

CHAPTER 14

14.20 $g_a = 21$ is minimum cost. Optimal path is a, b, e, l, r, z.
14.21 2 hr on course 1, score = 65; 0 hr on course 2, score = 40; 1 hr on course 3, score = 52; 1 hr on course 4, score = 91. Total max. score = 248.
14.22 $u^*(0) = -1$, $u^*(1) = -2$, $x^*(0) = 10$, $x^*(1) = 4$, $x^*(2) = 0$, $J^* = 9$. For comparison, if no control is used, $u(k) = 0$, $x(0) = 10$, $x(1) = 5$, $x(2) = 2.5$, and $J = 20.25$.
14.23

$k =$	0	1	2	3	4	5
$u^*(k)$	−0.0147	−0.0293	−0.0587	−0.1173	−0.2346	—
$x^*(k)$	10.0	4.985	2.463	1.173	0.469	0.0

14.24 $u^*(0) = -21x_0/104$, $u^*(1) = -x_0/52$, $x^*(0) = x_0$, $x^*(1) = 5x_0/52$, $x^*(2) = x_0/104$
14.25 The gain matrices and eigenvalues are: **(a)** $[-1.620 \quad -1.959]$, $0.190 \pm j0.148$ **(b)** $[-1.199 \quad -1.413]$, $-0.4 \pm j0.26$ **(c)** $[-1.755 \quad -2.136]$, -0.033, -0.0212 (both real) **(d)** $[-1.775 \quad -2.162]$, 0.0044, -0.229 (both real)
14.26 **(a)** $\mathbf{K}_{ss}^T = [-0.5202 \quad 0.8588]$. Filter eigenvalues are $0.2319 \pm j0.2956$. **(b)** Based on the filter pole locations, the filter will be much slower than controllers c and d and a little slower than a. Therefore, the filter will have a major impact on system response, maybe even be the dominant effect. In case b the filter is somewhat faster than the controller, and the total system will act more nearly like the full state controller.
14.27 **(a)** $\mathbf{K}_{ss}^T = [-0.4561 \quad 0.9977]$. Eigenvalues are $0.0404 \pm j0.0257$. **(b)** Because of the lower noise level, the filter response is much faster. The effect on the control loop will be essentially negligible and the controller will work about as well as if it had the actual state to work with rather than estimate. Stated differently, with this low noise level the estimated states will agree very closely with the true values.

14.28 On backward pass find

k	k'	$\mathbf{G}^T$	$v = -\mathbf{FV}$
5	0	—	—
4	1	[0.74963 −0.24988]	7.4963
3	2	[0.72159 −0.23296]	−0.5684
2	3	[0.72140 −0.23305]	−0.9369
1	4	[0.72140 −0.23305]	−0.9537
0	5	[0.72140 −0.23305]	−0.9544

Then on the forward pass, starting with the given $\mathbf{x}(0)$, find

k	$\mathbf{x}(k)^T$	$u(k) = -\mathbf{G}(k)\mathbf{x}(k) + v(k)$
0	[4 −2]	−4.3061
1	[−0.3061 1.3061]	−0.4284
2	[−0.7345 1.2345]	−0.1192
3	[−0.8538 1.1038]	0.3048
4	[−0.5490 0.6740]	8.0762
5	[7.5273 −7.4648]	—

14.29 $\mathbf{G} = \begin{bmatrix} 0.0422 & 0.0484 & -0.1059 & 0.2630 \\ 0.0013 & -0.4759 & -0.7321 & -0.0005 \end{bmatrix}$

$\lambda_i = \{0.104, 0.898, 0.6068 \pm j0.3519\}$

The feedback matrix **G** and the resulting eigenvalues are:
(a) [0.0271 −0.0032 −0.1099 0.1255], $\{0.0025, 0.8196, 0.721 \pm j0.479\}$
(b) [0.0276 −0.0039 −0.1119 0.1292], $\{2.76 \times 10^{-4}, 0.818, 0.721 \pm j0.4788\}$
(c) [0.0208 0.0004 −0.0833 0.1052], $\{0.0447, 0.8437, 0.7197 \pm j0.4819\}$
(d) [0.0227 −0.1097 −0.1507 0.1120], $\{0.0222, 0.8306, 0.699 \pm j0.455\}$. For reference the open-loop eigenvalues are $\{1.0, 0.135, 0.7175 \pm j0.4909]$

14.31

r	$\mathbf{W}_\infty$	$\mathbf{G}_\infty$	λ_{cl}
50	$\begin{bmatrix} 1.2276 & 0.2421 \\ 0.2421 & 0.24699 \end{bmatrix}$	$\mathbf{W}_\infty/50$	−1.9896, −1.0398 note: open-loop $\lambda = -2, -1$
10	$\begin{bmatrix} 1.1508 & 0.21573 \\ 0.21573 & 0.23686 \end{bmatrix}$	$\mathbf{W}_\infty/10$	−1.9396, −1.1992
4	$\begin{bmatrix} 1.04327 & 0.17995 \\ 0.17995 & 0.22322 \end{bmatrix}$	$\mathbf{W}_\infty/4$	−1.6583, −1.6583
0.2	$\begin{bmatrix} 0.42801 & 0.20482 \\ 0.20482 & 0.76754 \end{bmatrix}$	$5\mathbf{W}_\infty$	$-2.9535 \pm j1.10677$

14.32 $\mathbf{W} = \begin{bmatrix} 1.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}$

14.33 $\mathbf{W}_\infty = \begin{bmatrix} 0.80127 & 0.14102 \\ 0.14102 & 0.212813 \end{bmatrix}$, $\mathbf{G}_\infty = [-0.66025 \quad 0.071797]$ and $\lambda = -2, -1.73205$.

14.34 With $r = 50$, $\mathbf{G}_\infty = \begin{bmatrix} 1.0582 & 0.29701 \\ 0.29701 & 0.08639 \end{bmatrix}$, $\lambda_{cl} = -0.5805, -3.564$. With $r = 1000$, $\mathbf{G}_\infty =$ $\begin{bmatrix} 1.0419 & 0.2925 \\ 0.2925 & 0.08229 \end{bmatrix}$, $\lambda_{cl} = -0.5625, -3.560$. Note: Open-loop eigenvalues are at $\lambda = +0.56155$, -3.5616

14.35 The roots of the quadratic are $\mathbf{W} = 1$ and $-\frac{1}{3}$. Only one solution is positive definite. The eigenvector method gives $\mathbf{W}_\infty = 1$ and $\mathbf{W}_{ss} = -\frac{1}{3}$.

14.36 The canonical **A** matrix is $\mathbf{A}' = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ 0 & \mathbf{A}_{22} \end{bmatrix}$. Also $\mathbf{B}' = \begin{bmatrix} \mathbf{B}_1 \\ 0 \end{bmatrix}$

and $\mathbf{C} = [0 \quad 0.7071 \quad -0.56904 \quad 0.41975]$

where $\mathbf{A}_{11} = \begin{bmatrix} -7 & -6 \\ 2 & 0 \end{bmatrix}$, $\mathbf{A}_{12} = \begin{bmatrix} -59.4519 & 31.42803 \\ -15.2902 & 11.2787 \end{bmatrix}$.

$\mathbf{A}_{22} = \begin{bmatrix} -39.19376 & 33.12394 \\ -29.37606 & 23.19376 \end{bmatrix}$ is uncontrollable and has $\lambda = -8, -8$. $\mathbf{A}_{11}$ has $\lambda = -3, -4$.

14.37 $u^*(t) = \dfrac{\cosh t}{\sinh(1)}$, $x^*(t) = \dfrac{\sinh t}{\sinh(1)}$

14.38 $u^*(t) = -\frac{7}{2} + 3t$, $x_1^*(t) = 1 + t - 7t^2/4 + t^3/2$, $x_2^*(t) = 1 - 7t/2 + 3t^2/2$

14.39 $x(t) \cong 10 \cosh \sqrt{5}t - 10.065 \sinh \sqrt{5}t$, $p(t) \cong 5.011 \cosh \sqrt{5}t - 4.462 \sinh \sqrt{5}t$

14.41 **(a)** $u^*(t) = -\frac{1}{2}\mathbf{B}^T\mathbf{p}(t)$ if $\|\frac{1}{2}\mathbf{B}^T\mathbf{p}(t)\| \le 1$; $\mathbf{u}^*(t) = -\mathbf{B}^T\mathbf{p}(t)/\|\mathbf{B}^T\mathbf{p}(t)\|$ otherwise. Linear control with saturation. **(b)** $u^*(t) = -\mathbf{B}^T\mathbf{p}(t)/\|\mathbf{B}^T\mathbf{p}(t)\|$ for all t unless $\mathbf{B}^T\mathbf{p}(t) \equiv \mathbf{0}$ on a finite time interval (singular control case). Optimal control is always on the boundary of U except in the case of singular control. **(c)** Each component satisfies $u_i^*(t) = -\text{sign}[\mathbf{B}^T\mathbf{p}(t)]_i$, where $[\mathbf{B}^T\mathbf{p}(t)]_i$ is the ith component of $\mathbf{B}^T\mathbf{p}(t)$. This is the so-called bang-bang control and is valid provided $[\mathbf{B}^T\mathbf{p}(t)]_i \neq 0$. If $[\mathbf{B}^T\mathbf{p}(t)]_i \equiv 0$ on a finite interval, we have singular control. **(d)** $u_i^*(t) = 0$ when $|[\mathbf{B}^T\mathbf{p}(t)]_i| \le 1$: $u_i^*(t) = -\text{sign}[\mathbf{B}^T\mathbf{p}(t)]_i$ otherwise.

14.42 No solution exists. It is not possible to reach the origin with this unstable system and the bounded admissible controls unless $|x(0)| < 1$. The minimum principle provides only necessary conditions. It does not guarantee existence of a solution.

14.43 $\mathbf{G} = [0.2215 \quad -0.3735 \quad 0.2227 \quad 0.3162]$, $\lambda_i = \{-0.926, -28.15, -0.9165 \pm j2.88\}$

CHAPTER 15

15.23 With $e = x - y_d$, and using $\dot{e} = -\alpha e$, the control is $u(t) = \{-2y_d(t) + v(t) - \alpha[x(t) - y_d(t)]\}/x(t)^2$ provided $x \neq 0$. For $x(0) > 0$ the singularity at $x(t) = 0$ causes no problem in reaching the desired response.

15.24 The error equation $\dot{\mathbf{e}} = \mathbf{S}\mathbf{e}$, with diagonal S, gives $\begin{bmatrix} 1 \\ x_2 \end{bmatrix}\mathbf{u} = \begin{bmatrix} v - x_2^2 - y_{d_1} + S_{11}(x_1 - y_{d_1}) \\ v - 4y_{d_2} + S_{22}(x_2 - y_{d_2}) \end{bmatrix}$.

Least squares solution gives $u(t) = \{v - x_2^2 - y_{d_1} + S_{11}(x_1 - y_{d_1}) + x_2[v - 4y_{d_2} + S_{22}(x_2 - y_{d_2})]\}/(1 + x_2^2)$

15.25 Pick all $S_{ij} = 0$ except $S_{13} = S_{24} = 1$ and S_{31}, S_{33}, S_{42} and S_{44}. The first two error equations are automatically satisfied and the last two give

$$u_1(t) = S_{31}(x_1 - y_1) + S_{33}(x_3 - y_3) - x_4^2 x_1 + \mu/x_1^2 - \alpha y_3 - \beta y_1 + \beta R$$

$$u_2(t) = x_1\{S_{42}(x_2 - y_2) + S_{44}(x_4 - y_4) + 2x_3x_4/x_1 - \kappa y_4 - \sigma y_2 - \sigma\mu t/R^3\}$$

Parameter values must be selected. The orbit period is $T = 2\pi R^3/\mu$. The time constants τ for the template system are selected to give $4\tau = T/4$. For simplicity set $\alpha = \kappa$ and $\beta = \sigma$. Damping of $\zeta = 0.707$ gives $\beta = \{8\sqrt{2}\mu/(\pi R^3)\}^2$ and $\alpha = 16\mu/(\pi R^3)$. The convergence factors are selected so that the error settles twice as fast as the template system. This gives $S_{31} = S_{42} = 4\beta$ and $S_{33} = S_{44} = 2\alpha$. There are many other acceptable answers.

15.27 $\mathbf{x}_{e1} = \mathbf{0}$, $\mathbf{x}_{e2}$ and $\mathbf{x}_{e3} = [\pm 3 \quad 0 \quad 2.1]^T$. The Jacobian matrix evaluated at $\mathbf{x}_{e1}$ here is the same as in Problem 15.13 evaluated at $\mathbf{x}_{e2}$, and likewise the Jacobian here for $\mathbf{x}_{e2}$ is the same as in Problem 15.13 at $\mathbf{x}_{e1}$.

15.28 The plot of $-1/N(E)$ still ranges from 1.25 to 2, but in the *opposite* direction as a function of E. Intersections with $G(j\omega)$ indicate a stable limit cycle at $\omega = 1.71642$ with $E \approx 2.7$ and an unstable limit cycle at $\omega = 0, E \approx 4$. Simulation indicates that for small initial conditions a response very similar to the predicted limit cycle (but with a small dc offset) develops. For $x_1(0)$ larger than about 1.7 an unstable growth ensues.

15.29 Refer to Figure 15.24.

Range of K	Intersections of $G(j\omega)$ and $-1/N$	Predicted Behavior
$K < 0.16788$	None. $-1/N$ to right of point 3	Stable for all I.C.
0.16788, 0.2686	Intersects point 3	Unstable limit cycle‡
0.2686, 0.804	None. $-1/N$ between pts. 2,3	Unstable for all I.C.
0.804, 0.880	Intersects point 2	Stable limit cycle $\omega = 1.7164$
0.880, 1.287	Intersects both 1 and 2	Chaotic?
1.287, 1.408	Intersects 1.	Unstable limit cycle‡
$K > 1.408$	None. $-1/N$ between 0 and 1	Unstable for all I.C.

‡ Whether actual behavior is stable or unstable is dependent upon initial conditions (I.C.)

15.30 The plot of $-1/N$ now continues along the positive real axis, approaching ∞ for an amplitude E somewhat larger than 5 (the point where the effective gain $N(E)$ crosses through 0). The $-1/N$ plot then resumes at $-\infty$ and proceeds to -0.5 as E continues to increase. One new intersection with $G(j\omega)$ at $\omega = 2.4807$ indicates a stable limit cycle. Simulation shows behavior is the same as that of Problem 15.14 for smaller initial conditions, because the second nonlinearity breakpoint occurs at a large amplitude. For initial conditions sufficiently large the predicted stable limit cycle is observed here, whereas the system of Problem 15.14 goes unstable.

15.31 The circuit still appears to be chaotic for some initial conditions.

15.32 Routh's criterion shows the system is unstable for $K > 30.69$. The auxiliary equation shows the frequency at which the $j\omega$ axis is crossed is $\omega = 2.56$. Example 15.5 results yield a describing function which increases from 0 at an amplitude $E = 0.25$ to a maximum of 0.873 at $E = 2.51$ and then decreases again. For $30.69 < K < 35.15$ the effective gain $N(E)K$ is less than 30.69 and the system is stabilized by the nonlinearity. For each $K > 35.15$ there are two amplitudes E which give the critical $N(E)K = 30.69$, and thus two potential limit cycles. The larger E, where $N(E)$ is decreasing, gives a stable limit cycle. One example has $K = 40$, $\omega = 2.56$ and $E = 3.3$ as a stable limit cycle, excited if $y(0) > 1.35$.

15.33 The root locus has two $j\omega$ axis crossings and at each there are two amplitudes E which give the critical gain values.

ω	$10N(E)$	$N(E)$	approximate E (see Example 15.5)
6.1127	7.372	0.7372	1.2 (stable) or 3.5 (unstable)
3.8096	1.628	0.1628	0.345 (unstable) or 17.0 (stable)

Two *stable* limit cycles are possible, as noted. Initial conditions will determine which one, if any, is excited.

15.34 **(a)** $V'(\mathbf{x}) = \frac{1}{2b}(ax_2 + bx_1)^2 + \frac{1}{2ab}(ax_3 + bx_2)^2 + \int_0^{x_2} \left[F(x_2) - \frac{b}{a}\right] x_2\, dx_2$

$$\dot{V}'(\mathbf{x}) = -\frac{a}{b}\left[F(x_2) - \frac{b}{a}\right] x_3^2$$

(b) Asymptotically stable if $a > 0$, $b > 0$, and $F(x_2) > b/a$ for all x_2.

15.35 **(a)** $\mathbf{f}(\mathbf{x}) = \mathbf{0} \Rightarrow (r - 1 - x_3)x_1 = 0$ and $8x_3/3 = x_1^2$. One solution is $\mathbf{x}_{e_1} = \mathbf{0}$. If $x_1 \neq 0$, then $x_3 = r - 1$ and $x_1 = \pm[8(r-1)/3]^{0.5}$. For $r < 1$ the only real solution is $x_{e_1} = \mathbf{0}$.

$$\partial\mathbf{f}/\partial\mathbf{x} = \begin{bmatrix} -10 & 10 & 0 \\ r - x_3 & -1 & -x_1 \\ x_2 & x_1 & -\frac{8}{3} \end{bmatrix}$$

At $\mathbf{x} = \mathbf{0}$, $|I\lambda - \partial\mathbf{f}/\partial\mathbf{x}| = (\lambda + \frac{8}{3})[\lambda^2 + 11\lambda + 10(1 - r)]$. All roots stable if $r < 1$.
(b) One answer, using variable gradient is $\nabla\mathbf{V} = [x_1 + 0.1x_2 \quad 0.1x_1 + 2x_2 \quad 2x_3]^T$, $\dot{\mathbf{V}}$ is negative definite if $x_1^2 < 20(\frac{16}{3})$. $\mathbf{V} = x_1^2/2 + 0.1x_1x_2 + x_2^2 + x_3^2$ is positive definite for all $\mathbf{x}$.
(c) With $r = 28$, part (a) gives two more equilibrium points. $\mathbf{x}_e = [\pm 8.485 \quad \pm 8.485 \quad 27]^T$. Eigenvalues of $\partial\mathbf{f}/\partial\mathbf{x}$ are: at $\mathbf{x}_{e_1}$; $\lambda = -\frac{8}{3}, 11.828, -22.828$ (saddle point) at $\mathbf{x}_{e_2}$ and $\mathbf{x}_{e3}$; $\lambda = -13.845, 0.0892 \pm j10.165$
(unstable focus).

15.37 The trace of $\mathbf{A}$ can be negative while still having one or more unstable eigenvalues. The suggested example has $\text{Tr}(\mathbf{A}) = -6$, yet the solution diverges as an unstable focus in the x_1, x_2 plane as e^{2t}. Even though $\|\mathbf{x}(t)\| \to \infty$, the infinite oscillatory growth is confined to a zero volume (plane).

Index

A

B

C

D

E

F

G

K

L

M

N

O

P

Q

R

S

T

U

V

W

Z