gedness, one often uses a relay which is only "on" or "off":

$$u = -\operatorname{sgn} y \triangleq \begin{Bmatrix} -1, & y > 0 \\ +1, & y < 0 \end{Bmatrix} = -\frac{y}{|y|}$$

(where sgn = *signum* = "sign of").

a. Apply the force $u = -\operatorname{sgn} y$, and sketch the phase-plane paths. Is this control satisfactory?

b. Apply the force $u = -\operatorname{sgn}(y + \dot{y})$, and repeat part a. Include the *switching curve* $\dot{y} = -y$ on your sketch. Show especially that for this *control law* there is a particular solution curve that has *exactly one* reversal of sign of control, $u$. Show also that for other curves there comes a time when *no solution* is possible. Can you account for this? What do you think happens here? What would a *physical* relay do?

c. Show finally a switching curve that *never* has more than one reversal of sign of control.

*Remarks:* This nonlinear feedback control problem is a famous one in the theory of automatic control. The fact that the solution curve in part c is *time-optimal* in a certain sense was first proved in the mid-1950s by D. Bushaw and was a spur to the study of optimization techniques. For example, now the result can be obtained by an easy application of the Pontryagin maximum principle (see, e.g., [15, pp. 182–184]).

## 2.3 INITIAL CONDITIONS FOR ANALOG-COMPUTER SIMULATION; OBSERVABILITY AND CONTROLLABILITY FOR CONTINUOUS- AND DISCRETE-TIME REALIZATIONS

In the first two parts of this section we shall study some simple problems that will introduce two basic concepts—controllability and observability—associated with any realization $\{A, b, c\}$. This discussion will call on some of the matrix theory material in Secs. 2, 4, and 8 of the Appendix.

In Sec. 2.3.3, we further explore and reinforce the ideas and results of all the previous sections by reexamining them for discrete-time systems. We can notice many parallels and also some simplifications in the discrete-time case. In particular, we show how discrete-time simulations can be used to better understand several earlier results and also to naturally indicate some useful new ones.

An important thing to notice in all the discussions is the value of the special realizations of Sec. 2.1 in understanding and illuminating some apparently purely mathematical problems.

As a partial review, Sec. 2.3.4 contains several worked examples to illustrate how various earlier results can be combined to solve some interesting problems.

### 2.3.1 Determining the Initial Conditions; State Observability

In Sec. 2.1 we found various forms in which a system described by a $n$th-order input-output differential equation could be realized. However, we had not completed the simulations because of our assumption that the initial values were zero. The problem now facing us is to explain how given nonzero initial conditions $\{y(0-), \dot{y}(0-), \ldots, y^{(n-1)}(0-)\}$ can be translated into appropriate initial values $\{x_i(0-)\}$ for the state variables of any particular realization (or, more concretely, for the integrators in the corresponding analog-computer simulation).

Clearly the first step is to find a convenient expression for $\{y(t), \ldots, y^{(n-1)}(t)\}$. It is easy to see that with

$$\dot{x}(t) = Ax(t) + bu(t), \qquad y(t) = cx(t) \tag{1}$$

we can write

$$\begin{aligned} y(t) &= cx(t) \\ \dot{y}(t) &= c\dot{x}(t) = cAx(t) + cbu(t) \\ \ddot{y}(t) &= cA\dot{x}(t) + cb\dot{u}(t) \\ &= cA^2x(t) + cAbu(t) + cb\dot{u}(t) \end{aligned}$$

and so on, which can be conveniently arranged in matrix form as, say,

$$\mathcal{Y}(t) = \mathcal{O}x(t) + \mathbf{T}\mathcal{U}(t) \tag{2}$$

where

$$\mathcal{Y}(t) \triangleq [y(t) \quad \dot{y}(t) \quad \cdots \quad y^{(n-1)}(t)]' \tag{2a}$$

$$\mathcal{U}(t) \triangleq [u(t) \quad \dot{u}(t) \quad \cdots \quad u^{(n-1)}(t)]' \tag{2b}$$

$$\mathcal{O} \triangleq \mathcal{O}(c, A) \triangleq [c' \quad A'c' \quad \cdots \quad (A')^{n-1}c']' \tag{2c}$$

and

$$\mathbf{T} = \text{a lower triangular Toeplitz matrix with first column } [0 \quad cb \quad \cdots \quad cA^{n-2}b]' \tag{2d}$$

Assuming that $\mathcal{U}(0-) = 0$, we then have

$$\mathcal{Y}(0-) = \mathcal{O}x(0-) \tag{3}$$

and the question is whether we can find an *initial-state* vector $x(0-)$ for a given vector $\mathcal{Y}(0-)$. This is a problem in the theory of linear equations (cf. Sec. 4 in the Appendix), and the question is whether $\mathcal{Y}(0-)$ can be formed by a linear combination of the $n$ columns of the matrix $\mathcal{O}$.

If these $n$ columns are all linearly independent, i.e., if the $n \times n$ matrix $\mathcal{O}$ is nonsingular, then we can always find $x(0-)$, for *any* $n$-vector $\mathcal{Y}(0-)$, as $x(0-) = \mathcal{O}^{-1}\mathcal{Y}(0-)$.

On the other hand, if $\mathcal{O}$ is singular (not of full rank), some columns of $\mathcal{O}$ will be linearly dependent on others, and therefore we can only find a solution $x(0-)$ for special choices of $\mathcal{Y}(0-)$, namely, those that lie in the (column) range space of $\mathcal{O}$.

Now in a differential equation it is generally important to have full freedom of choice in the initial conditions, and this means that of the many possible state-space realizations of the differential equation we should be most interested only in those for which the matrix $\mathcal{O}$ is such that

$$\mathcal{O}(c, A) \text{ has full rank} \tag{4}$$†

Such realizations are said to be *observable*, and the matrix $\mathcal{O}(c, A)$ is called the *observability matrix* of the pair $\{c, A\}$. (These concepts will appear in a more general framework soon.) We first note that, in the present case where $\mathcal{O}$ is a square matrix, this is equivalent to the condition that

$$\mathcal{O}(c, A) \text{ is nonsingular} \tag{4a}$$

**The State-Observability Problem.** The significance of condition (4) stands out even more clearly in a different but closely related problem. Suppose we have a physical system described by the state-space equations

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), \qquad t \geq 0 \\ y(t) &= cx(t), \qquad x(0) = x_0 \end{aligned}$$

We assume that we know the matrices $\{A, b, c\}$ and also the input and output functions $\{u(t), t \geq 0\}$, $\{y(t), t \geq 0\}$. The problem is to determine the states $\{x(t), t \geq 0\}$. Such problems can arise in many contexts, e.g., for diagnostic reasons or for purposes of control; several examples will be given in this book.

Now this problem is very close to the initial-condition problem discussed above, because really the only unknown in our problem is the *initial* state $x_0$: Knowing $x_0$, $\{A, b, c\}$, and $\{u(t), t > 0\}$, we could set up the equation

$$\dot{x}(t) = Ax(t) + bu(t), \qquad x(0) = x_0$$

and thus obtain $x(t)$ as a function of $t$.

In fact, if we can determine its value at any time, say $t_1$, then we can again obtain $x(t)$, $t \geq 0$, by solving the differential equation

$$\dot{x}(t) = Ax(t) + bu(t), \qquad x(t_1) \text{ given}$$

†We write $\mathcal{O}(c, A)$ from time to time to emphasize the dependence on the given realization.

Now determining $x(t_1)$ brings us back to Eq. (2), which we shall write as

$$\mathcal{O}(c, A)x(t_1) = \mathcal{Y}(t_1) - \mathbf{T}\mathcal{U}(t_1)$$

However, when $\mathcal{O}$ is singular (not of full rank), there is a slight difference from the analog-computer problem. In that problem, we had to assume that the right-hand side [$\mathcal{Y}(0-)$ in that case] was in the (column) range space of the matrix $\mathcal{O}$; otherwise the equations had no solution. In the present problem, we start with a realization $\{A, b, c\}$ and an input $u(\cdot)$ and *obtain* $y(\cdot)$ *from it*: Therefore $\mathcal{Y}(t_1) - \mathbf{T}\mathcal{U}(t_1)$ has to lie in the range of $\mathcal{O}$.† Thus, whether or not $\mathcal{O}$ is singular, our equation (2) will always have a solution in the state-determination problem. This would seem to dilute the significance of condition (4), except for a fact that we have not yet taken into account: When $\mathcal{O}$ is singular, there will be more than one solution of the consistent equations (2). When $\mathcal{O}$ is singular, it has a nontrivial *null-space*; i.e., there will exist non-zero vectors $\theta$ such that

$$\mathcal{O}(c, A)\theta = 0$$

Any such $\theta$ can be added to a given solution to obtain yet another solution. This degeneracy is of less consequence in the analog-computer problem, where any one of these solutions will suffice to match the initial conditions $\mathcal{Y}(0-)$. However, in the state-determination problem, also called the problem of "observing" the states, this loss of uniqueness is usually fatal. For one thing, we are generally interested in the actual values of the states and not in a whole family of possible values; second, this loss of uniqueness would make meaningless the determination of the state behavior over any time interval.

In other words, the nonsingularity of the matrix $\mathcal{O}(c, A)$ is crucial to the problem of observing the states. For these reasons $\mathcal{O}(c, A)$ is called the *observability matrix*, and a realization $\{A, b, c\}$ with a full-rank (nonsingular) observability matrix is said to be *observable*.

The question arises of whether matters can be helped when $\mathcal{O}(c, A)$ is not of full rank by using further derivatives $\{y^{(n)}(t), y^{(n+1)}(t), \text{etc.}\}$. The answer is no. Taking more derivatives effectively just gives us more equations for $x(t)$, with coefficients $cA^n$, $cA^{n+1}$, and so on. But here we recall the important Cayley-Hamilton theorem, which says that for any $n \times n$ matrix $A$, $A^n$ is a linear combination of the lower powers of $A$, $\{A^0(= I), A^1, \ldots, A^{n-1}\}$—see Sec. 8 in the Appendix. Therefore all terms $\{cA^{n+i}, i \geq 0\}$ will be linearly dependent on the rows of $\mathcal{O}(c, A)$, and no more information about $x(t)$ can be obtained by considering higher-order derivatives of $y(t)$.

†Of course we are assuming no errors in the measurement of $y(\cdot)$ and the calculation of $\dot{y}(\cdot)$, $\ddot{y}(\cdot)$, etc. With measurement error we have a different (statistical) problem, which we shall not treat here. See, however, some brief remarks in Secs. 6.2, 9.2 and the final paragraph of this section.

In fact we can say much more: Since $y(\cdot)$ satisfies a differential equation, it is a fairly well-behaved function and can be proved to have a Taylor series expansion around any point. But this means that knowledge of $y(\cdot)$ and all its derivatives at any point $t$ will actually determine the function $y(\cdot)$ at all other points in its region of definition. Therefore if we cannot determine $x(t)$ from $\{y(t), \dot{y}(t), \ldots\}$, i.e., if the realization is not observable, then we cannot determine $x(t)$ in any way even from knowledge of $y(\cdot)$ over a whole (future or past) interval.†

It is important to stress that not all realizations need be observable—it depends entirely on the particular pair $\{c, A\}$. However, the two realizations in Sec. 2.1.1 that we called observability [cf. Fig. 2.1-7 and Eq. (2.2-9)] and observer [cf. Fig. 2.1-9 and Eq. (2.2-5)] forms have a particular significance here: They are always guaranteed observable.

By direct calculation we find that

$$\mathcal{O}_{ob} = \mathcal{O}(c_{ob}, A_{ob}) = I \tag{5}$$

so that, for this realization, the determination of $x(0-)$ is trivial, $x(0-) = \mathcal{Y}(0-)$. Examination of the block diagram (Fig. 2.1-7) will show graphically why this is so [recall our assumption that $\mathcal{U}(0-) = 0$].‡ The simplicity that the *observability form* brings to the state-observation (or state-determination) problem is the reason for its name.

The observer form is not quite so convenient,§ though it is simple enough: Direct calculation will yield the nice formula

$$\mathcal{O}_o^{-1} = \mathcal{O}^{-1}(c_o, A_o) = \mathcal{A}_- \tag{6}$$

where

$$\mathcal{A}_- = \text{a lower triangular Toeplitz matrix with first column } [1 \quad a_1 \quad \cdots \quad a_{n-1}]'$$

Again this fact stands out fairly vividly if the block diagram of the observer form (Fig. 2.1-9) is now examined (starting with the last integrator).

However, scrutiny of the block diagrams for the controller (Fig. 2.1-4) and controllability (Fig. 2.1-10) forms will show that the state-determination problem is no longer so easy, and we really have to solve some simultaneous

†Note that we have shown this only for constant-parameter continuous-time realizations. In Sec. 2.3.3, we shall see that a more refined analysis can be carried out for discrete-time systems.

‡It may be worth noting here explicitly that in the problem of state determination at any time, not just $t = 0-$, one can assume $u(\cdot) \equiv 0$, without loss of generality, since the effect of a nonzero (but known) $u(\cdot)$ is merely to change the right-hand side in the key equation $\mathcal{O}x(t) = \mathcal{Y}(t) - \mathcal{T}\mathcal{U}(t)$ to some new known vector. This is often a very convenient assumption.

§"The" problem for which the observer form is most natural will be described in Sec. 4.1.

equations to find $x(t)$. In fact we have to invert $\mathcal{O}(c, A)$, and for these forms it may or may not happen that $\mathcal{O}(c, A)$ has full rank (i.e., is invertible). But if this is the case, why then should we bother with these forms? One reason is that the duality we noted earlier in Sec. 2.2.1 indicates that there must be some problems for which these forms are most "natural"; more generally, several factors enter into the choice of realizations in practical problems.

In this connection, we should note that except perhaps at the point $t = 0-$, where they may be specified, it is clearly difficult to *measure* the values $\{y(t), \dot{y}(t), \ldots, y^{(n-1)}(t), u(t), \ldots, u^{(n-2)}(t)\}$, since differentiation amplifies "noise." Therefore more realistic observation schemes will have to be developed, with inevitable loss in the accuracy of the state "estimates"; a way of doing this will be developed later in Sec. 4.1 (see also Sec. 9.2). Moreover, we should note that the discrete-time analog of the procedure of this section is quite realistic, as will soon be shown in Sec. 2.3.3.

**Reprise.** We have defined a realization $\{A, b, c\}$, or just a pair $\{c, A\}$, as being *observable* if its observability matrix $\mathcal{O}(c, A)$ has full rank. The presence or absence of this property, together with the fact that the observability and observer realizations are clearly observable, will be helpful in many problems of system theory. In this section we discussed two simple examples of such problems. One was the determination of the initial state of a realization given an arbitrary set of initial conditions $\{y(0-), \ldots, y^{(n-1)}(0-)\}$. The other was the determination of the state at any time given full knowledge of the input and output functions for $t > 0-$. We only gave an idealized (physically unrealistic) solution of the second problem. While more realistic solutions will be developed later (Secs. 4.1 and 9.2), it is worth emphasizing that both problems provide alternative equivalent characterizations of the property of observability. As we shall see on several later occasions, the whole point of having alternative characterizations is that they further illuminate a property, and moreover in new problems one characterization might be more convenient to use than another.

### 2.3.2 Setting Up Initial Conditions; State Controllability

We have seen that for an observable realization the proper initial conditions (and, in fact, the state at any time) for the realization can be calculated from the input and output functions and their derivatives. It is then natural to ask [prompted also by the duality that we have noticed between the observer (observability) and controller (controllability) forms] how, for a given simulation, we can actually set up any desired initial conditions, i.e., how we can find a suitable input $u(\cdot)$ that will take the system to any desired initial state in a finite (often very "small") time.†

†Although we pose this problem in the context of analog-computer simulations, it also applies to any physical system whose state we would like to change at a given time, e.g., as in a *midcourse* correction for a rocket.

To see how this might be done, consider first the simplest case of a scalar (single-integrator) system

$$\dot{x}(t) = ax(t) + bu(t), \qquad t > 0-$$

and observe that if

$$u(t) = g\delta(t)$$

then

$$x(t) = bg \cdot 1(t) + (\text{a continuous function}), \qquad t > 0-$$

and

$$x(0+) - x(0-) = bg$$

Therefore the system can be taken to any desired initial condition $x_0$ by suitably choosing $g$. Moreover, the desired condition can be set up in "zero" time because of the impulsive nature of the input. This is impractical, of course, though it may be noted that in many problems "approximate" impulsive functions can be satisfactorily generated and used—it all depends on the "time constants" of the system under study. State-controllability problems with nonimpulsive inputs will be studied in Sec. 9.2. Here our goal is to pursue an analysis *dual* to that of Sec. 2.3.1.

To generalize the result just obtained, let us (again partly motivated by duality) consider next the controllability canonical form of Fig. 2.1-10. We notice from that figure that if

$$u(t) = g_1\delta(t)$$

then

$$\begin{aligned}\dot{x}_1(t) &= g_1\delta(t) + [-a_3x_3(t)] \\ &= g_1\delta(t) + (\text{nonimpulsive functions})\end{aligned}$$

and

$$\begin{aligned}x_1(t) &= g_1 \cdot 1(t) + (\text{continuous function}) \\ &\quad + [\text{response to } x_1(0-)]\end{aligned}$$

Therefore

$$x_1(0+) = g_1 + x_1(0-)$$

and $x_1(0+)$ can be arbitrarily set by properly choosing $g_1$. It is perhaps not so easy to see how to set $x_2(0+)$ and $x_3(0+)$. However, note that if

$$u(t) = g_1\delta(t) + g_2\delta^{(1)}(t)$$

then

$$\dot{x}_2(t) = g_2\delta(t) + (\text{nonimpulsive functions})$$

and

$$x_2(t) = g_2 \cdot 1(t) + (\text{continuous functions}) + [\text{response to } x_2(0-)]$$

so that

$$x_2(0+) = g_2 + x_2(0-)$$

which can be set arbitrarily by proper choice of $g_2$.

But what will happen to $x_1(0+)$ when $u(t) = [g_1\delta(t) + g_2\delta^{(1)}(t)]$ rather than just $g_1\delta(t)$? It is easy to see that $x_3(t)$ is still nonimpulsive so

$$\dot{x}_1(t) = g_1\delta(t) + g_2\delta^{(1)}(t) + (\text{nonimpulsive functions})$$

Integrating both sides from 0− to 0+ gives

$$\begin{aligned} x_1(0+) - x_1(0-) = \int_{0-}^{0+} \dot{x}_1(t)\,dt &= g_1 \int_{0-}^{0+} \delta(t)\,dt + g_2 \int_{0-}^{0+} \delta^{(1)}(t)\,dt \\ &\quad + \int_{0-}^{0+} (\text{nonimpulsive function})\,dt \\ &= g_1 + 0 + 0 = g_1 \end{aligned}$$

as before. This shows how to proceed given a general $n$th-order *controllability canonical form*: Let

$$u(t) = g_1\delta(t) + g_2\delta^{(1)}(t) + \cdots + g_n\delta^{(n-1)}(t)$$

Then arguments similar to those just used will show that

$$x_i(0+) = g_i + x_i(0-)$$

Thus, for the controllability form, it is easy to *set up* arbitrary initial conditions, just as it was easy to *determine* initial conditions on the dual observability form. But now, as before, we have the question of whether we can find inputs to set up arbitrary initial conditions for a general realization $\{A, b, c\}$. *By duality*, we would expect the answer to be [cf. (3)] yes if and only if the matrix

$$\mathcal{C} = [b \quad Ab \quad \cdots \quad A^{n-1}b] \quad \text{has full rank} \tag{7}$$

i.e., is nonsingular ($\mathcal{C}$ is a square matrix). $\mathcal{C}$ is called the *controllability matrix*† of the particular realization, and a realization for which $\mathcal{C}$ is non-

†For reasons that will appear later (see Sec. 2.3.3) this matrix is perhaps better called the *reachability matrix* or perhaps (see Sec. 3.2) the *modal controllability* matrix, but the above terminology is by now well entrenched.

singular is said to be *controllable*. Moreover, as we would expect by duality [cf. (5)], we can show that for the controllability canonical form

$$\mathcal{C}_{co} = I_{nxn} \tag{8}$$

and for the controller form

$$\mathcal{C}_c^{-1} = \mathcal{A}'_- \tag{9}$$

where [cf. (6)]

$$\mathcal{A}'_- = \text{an upper triangular Toeplitz matrix with first row } [1 \quad a_1 \quad \cdots \quad a_{n-1}]$$

and the $\{a_i\}$ are the coefficients of the characteristic polynomial of $A$.

Although the arguments via duality can be made quite completely, it will be useful to have a general proof of our claim that condition (7) is necessary and sufficient to solve our problem. For this, we introduce the following important result.

**A General Formula for x(0+) with Impulse Inputs.** If

$$\dot{x}(t) = Ax(t) + bu(t)$$

and

$$u(t) = g_1\delta(t) + \cdots + g_k\delta^{(k-1)}(t), \qquad k \geq 1 \tag{10a}$$

then

$$x(0+) = x(0-) + [b \quad Ab \quad \cdots \quad A^{k-1}b] \cdot [g_1 \quad g_2 \quad \cdots \quad g_k]' \tag{10b}$$

This important formula can be proved in several ways.† Here we give a proof that exploits the concept of superposition. Let

$$h(\cdot) = \text{the impulse response of } \dot{x} = Ax + bu$$
$$= \text{the response to } u(t) = \delta(t) \text{ with } x(0-) = 0$$

By linearity it follows that the response to the input

$$u(t) = g_1\delta(t) + g_2\delta^{(1)}(t) + \cdots + g_k\delta^{(k-1)}(t)$$

is

$$x(t) = g_1h(t) + g_2h^{(1)}(t) + \cdots + g_kh^{(k-1)}(t) + (\text{responses to zero input but nonzero initial conditions})$$

†The simplest uses the fact (easy to prove from first principles) that $x(t) = \int_0^t e^{A(t-\tau)}bu(\tau)d\tau$ if $x(0-) = 0$ (see also Sec. 2.5.1).

Therefore

$$x(0+) = g_1 h(0+) + g_2 h^{(1)}(0+) + \cdots + g_k h^{(k-1)}(0+) + x(0-)$$

Now

$$u(t) = 0, \qquad t \geq 0+$$

so that

$$\dot{h}(t) = Ah(t), \qquad t \geq 0+$$

and

$$h^{(i+1)}(t) = Ah^{(i)}(t), \qquad t \geq 0+$$

In particular, setting $t = 0+$ and recalling (from our initial arguments in this section) that $h(0+) = b$, we shall have

$$h^{(i)}(0+) = A^i h(0+) = A^i b$$

Therefore

$$\begin{aligned} x(0+) &= x(0-) + g_1 b + g_2 Ab + \cdots + g_k A^{k-1} b \\ &= x(0-) + [b \quad Ab \quad \cdots \quad A^{k-1}b][g_1 \quad g_2 \quad \cdots \quad g_k]' \end{aligned}$$

which is Eq. (10).

**State Controllability.** It is easy to see from (10), with $k = n$, that if $\mathcal{C}$ is nonsingular, the coefficients $\{g_i\}$ can be determined so as to (instantaneously) set up *arbitrary* initial conditions. On the other hand, if $\mathcal{C}$ is singular, then it is true that no matter what the $\{g_i\}$, certain vectors $x(0+) - x(0-)$ cannot be obtained from Eq. (10). The columns of $\mathcal{C}$ will be linearly dependent, and we shall not be able to obtain an *arbitrary* $n$-vector as a linear combination of these columns.† Moreover, as in the case of observability, the Cayley-Hamilton theorem shows that if $\mathcal{C}$ is singular, then higher orders of impulsive inputs will not provide any more linearly independent equations that can serve to determine an appropriate input.

Therefore we have proven that the nonsingularity of $\mathcal{C}$ is necessary and sufficient for the existence of an impulsive input that will change the state from a given value $x(t-)$ (the choice $t = 0$ is just a special case) to an arbitrary desired value at $t+$. This condition is called *state controllability*, and, like observability, it will be seen to be a fundamental property of a realization. As we might expect from the fact that arbitrary inputs can be regarded as an appropriate collection of impulsive inputs, the nonsingularity of $\mathcal{C}$ is also a necessary and sufficient condition for being able to change the

†If we have $x(0+) - x(0-)$ in the range of $\mathcal{C}$, then of course a solution always exists by definition (cf. Sec. 4 in the Appendix) whether or not $\mathcal{C}$ is nonsingular.

state arbitrarily in a finite time (i.e., with nonimpulsive inputs). To prove this, we shall need to know how to solve the state equations, and therefore a formal proof is postponed to Example 2.5-1.

A natural question raised by the results of Secs. 2.3.1 and 2.3.2 is whether there are realizations that are both controllable and observable, or, perhaps more fundamentally, why there should be any realizations that are not controllable, or not observable, or even neither? In fact, the reader may wonder why, if the above concepts are so important and if analog-computer realizations are so old (going back about 100 years to Kelvin and about 30 or 40 years to Bush), why it has taken so long for these concepts to be uncovered (basically in the early 1960s).

Insofar as *controllability* goes, one might say that the problem of setting up initial conditions by choice of a suitable system input $u(\cdot)$ is not relevant to analog-computer simulations. In such simulations, all integrators are generally directly accessible, and any desired initial conditions can be individually placed on each integrator (by a simple *charging* circuit). The problem we discussed is relevant for systems or system realizations where such independent access is unfeasible, e.g., because the system is at a remote location. Actually the concept was first encountered as a technical condition for certain optimal control problems and then in a slightly different form in a so-called *finite-settling-time* design problem [12], which we shall discuss in Sec. 2.3.3.

As for *observability*, the problem of determining the initial conditions seems to have been sidestepped by always choosing realizations in which proper initial conditions are easy to calculate, as in the observability or observer forms, or by always implicitly arranging things so that observability was ensured.† Thus it was apparently not necessary to formalize these concepts, though various difficulties, especially with multi-input, multi-output analog-computer simulations, had begun to show the need for a more fundamental analysis.

The present form of the definitions of controllability and observability and the recognition of the simple duality between them were worked out by R. E. Kalman in 1959–1960 (see the historical remarks in [18]). They were very clearly presented in a path-breaking paper [19], which has had a wide influence. It inspired many later researches, including some on the significance of jointly controllable and observable realizations by Gilbert [16], Kalman [17], and Popov [20]; these results will be studied in Sec. 2.4.

Here we shall first pause to show how our discussions so far can be carried over fairly readily to discrete-time [and to some extent also discrete-amplitude (cf. Example 2.3-6)] systems.

†For example (cf. Sec. 2.4.1), by working only with transfer functions in reduced form, i.e., without common factors between numerator and denominator.

**Reprise.** We have defined a realization $\{A, b, c\}$, or just a pair $\{A, b\}$, as being *controllable* if its controllability matrix $\mathcal{C}(A, b)$ has full rank. We showed that an equivalent characterization was being able to change the state arbitrarily by means of impulsive inputs. Change of state with bounded functions will be discussed in Sec. 9.2. One should note carefully the duality between the characterizations of observability and controllability. For example, make sure you understand completely the statement that a realization $\{A, b, c\}$ is observable (controllable) if and only if the dual realization $\{A', c', b'\}$ is controllable (observable). Note how controllability involves statements about inputs and states, while the dual statements for observability involve outputs and states. With these remarks in mind, reexamine the block diagrams for the four special realizations of Sec. 2.1.2.

One of the important points we wish to make as we proceed through this book is that the different state-space descriptions provide convenient coordinate frames, we might say, for better understanding system behavior. We could always work with a given realization, but different things stand out more clearly in different realizations. We should also repeat here that our use of different realizations is almost purely for conceptual purposes and not for actual numerical implementation. The numerical aspects have to be studied separately, often in each particular problem, though as a general rule, diagonal realizations (when feasible) appear to have the best numerical properties (see the remarks and references on singular-value decompositions in Sec. 15 of the Appendix).

### 2.3.3 *Discrete-Time Systems; Reachability and Constructibility*

In the previous sections, we have studied some aspects of state-space realizations of continuous-time systems. Very similar analyses and results can be obtained for *discrete-time* systems, where it is assumed that the inputs and outputs are known (or are of interest) only at discrete time instants $t_0$, $t_1$, $t_2, \ldots$. It can be arranged that these instants are all integral multiples of some basic unit $\Delta$, say,

$$t_0 = 0, t_1 = \Delta, t_2 = 2\Delta, \ldots$$

in which case $\Delta$ is often not explicitly shown and we assume that the time parameter, denoted by $k$ (or sometimes still by $t$), takes integral values, $k = 0, \pm 1, \pm 2, \ldots$.

We shall see that there are great similarities between the results for discrete- and continuous-time systems, and therefore most of the discussions in this book are for systems of the latter type. However, there are many problems where the discrete-time language is more intuitive, especially when the Markov parameters enter (cf. Chapter 5). There are also some differences from continuous-time systems (cf. Sec. 3.5.2). Many of these questions can be illuminated by using the discrete-time analogs of the canonical realizations

developed in Sec. 2.1.1, and some new results and insights will be developed in this way.

The discrete-time state-space equations are usually written in the form

$$\begin{aligned} x(k+1) &= Ax(k) + bu(k), & x(0) &= x_0 \\ y(k) &= cx(k) & k &\geq 0 \end{aligned} \tag{11}$$

In discrete time, the transfer function is defined via zee transforms [21] rather than Laplace transforms. We write

$$X(z) = \sum_0^\infty x(k)z^{-k} \tag{12}$$

$$\begin{aligned} \mathcal{Z}\{x(k+1)\} &= \sum_0^\infty x(k+1)z^{-k} \\ &= x(1) + x(2)z^{-1} + \cdots \\ &= z\sum_0^\infty x(k)z^{-k} - zx(0) \\ &= zX(z) - zx(0) \end{aligned} \tag{13}$$

Therefore, from the state-space equations we can write

$$Y(z) = c(zI - A)^{-1}b\, U(z) + cz(zI - A)^{-1}x_0 \tag{14}$$†

The transfer function, which is calculated with $x_0 = 0$, is

$$H(z) = \frac{Y(z)}{U(z)} = c(zI - A)^{-1}b \tag{15}$$

Now we notice the important fact that the expression for $H(z)$ is the same as that for the transfer function

$$H(s) = c(sI - A)^{-1}b$$

of the continuous-time system

$$\dot{x}(t) = Ax(t) + bu(t), \qquad y(t) = cx(t)$$

Therefore, modulo certain obvious changes, all results derived by algebraic manipulation of $H(s)$ can be immediately carried over to $H(z)$. As a first example, we note the analogs of the results of Sec. 2.1.

†Note the difference in the last term from its continuous-time analog; this difference could be avoided by taking the sum in (12) from 1 to $\infty$ rather than 0 to $\infty$—however, the latter convention is more traditional.

**Computer Simulations.** An analog of the continuous-time high-order differential equation specification of a system

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = b_1 u^{(n-1)}(t) + \cdots + b_n u(t) \tag{16}$$

is the difference equation

$$\begin{aligned} y(k+n) + a_1 y(k+n-1) + \cdots + a_n y(k) \\ = b_1 u(k+n-1) + \cdots + b_n u(k), \qquad k \geq 0 \end{aligned} \tag{17}$$

with initial conditions being given values of $\{y(0), \ldots, y(n-1)\}$. The zee transform, with zero initial conditions, is

$$z^n Y(z) + \cdots + a_n Y(z) = b_1 z^{n-1} U(z) + \cdots + b_n U(z) \tag{18}$$

so that

$$\frac{Y(z)}{U(z)} = \frac{b_1 z^{n-1} + \cdots + b_n}{z^n + a_1 z^{n-1} + \cdots + a_n} \tag{19}$$

which is the same, except that $z$ replaces $s$, as would be obtained by Laplace transformation of the differential equation (16). Therefore, with the obvious change that

$$z^{-1} \text{ can be implemented by a unit delay}$$

(while $s^{-1}$ is implementable by an integrator), all the arguments and results used in Sec. 2.1 can be repeated here. That $z^{-1}$ represents unit delay is evident from the fact that if $y$ is a unit-delayed version of $u$, i.e., $y(k) = u(k-1)$, then, from (17) and (18), $Y(z) = z^{-1}U(z)$.

We may perhaps also note explicitly that the powers of $z^{-1}$ serve to indicate the position in time at which the associated coefficient occurs. Thus, in particular, polynomials in $z$ correspond to discrete-time functions that start and end at finite "past" times $t$, $t \leq 0$.

**The Markov Parameters Define the Impulse Response.** As a matter of fact, several things turn out somewhat more nicely in discrete time. For example, the Markov parameters turn out to be just the values of the discrete-time impulse response. That is, if $\{h_i\}$ are the Markov parameters defined by [cf. (2.2-48)]

$$H(z) = \sum_1^\infty h_i z^{-i} = \sum_1^\infty (cA^{i-1}b) z^{-i} \tag{20}$$

then the inverse zee transform of $H(z)$ is the discrete-time function

$$h(k) = h_k = cA^{k-1}b, \qquad k \geq 1 \tag{21}$$

which is also the impulse response of the system; i.e., its response to the discrete-time impulse function

$$\delta(k) = \begin{cases} 1, & k = 0 \\ 0, & k \neq 0 \end{cases} \tag{22}$$

Thus, in discrete time the Markov parameters are much easier to interpret and obtain than they are in continuous time; however, they have the same algebraic properties in both cases, which is certainly a convenience.

**Specification of the Controllability and Observability Forms.** As an example, let us consider the controllability canonical form in discrete time, which we show in Fig. 2.3-1 (this is Fig. 2.1-10 drawn with delays in place of the integrators). In this form, the coefficients $\{\beta_1, \beta_2, \beta_3\}$ cannot be written down directly from the transfer function $b(z)/a(z)$. However, by putting an impulse $\delta(k)$ into the realization we see immediately from the realization that

$$\begin{aligned} \beta_1 &= \text{the response } y(\cdot) \text{ at } k = 1 \\ &= h_1, \qquad \text{by definition} \end{aligned}$$

and similarly we see that $\beta_2 = h_2$, $\beta_3 = h_3$ (cf. Eq. (2.1-8d) and Exercise 2.2-15). A similar derivation in continuous time is less direct but will be an interesting exercise for the reader.

Examination of the block diagram of the observability form (Fig. 2.3-2) shows similarly that the input vector is given by

$$b_{ob} = [h_1 \quad h_2 \quad h_3]'$$

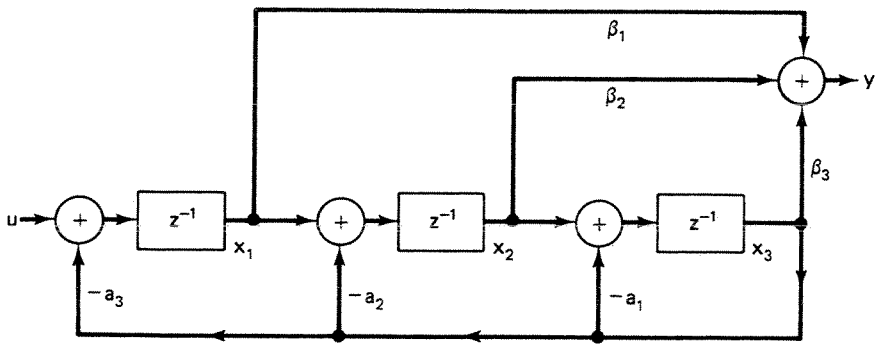



**Figure 2.3-1.** Controllability canonical form. It will be shown that this should perhaps better be called the *reachability canonical form* because it is easy to determine an input to take the state from zero at $t = 0$ to any value at $t \geq 3$.

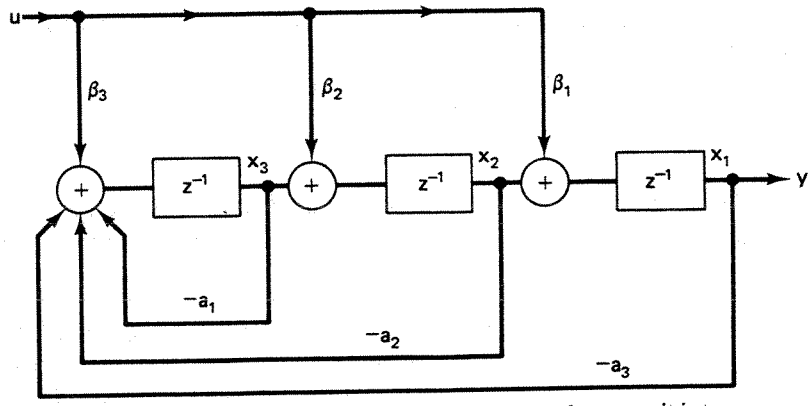



**Figure 2.3-2.** Observability canonical form. Note how easy it is to "read out" the values of the initial state [especially when $u(\cdot) \equiv 0$].

The several results of Sec. 2.3 on observability and controllability also have simpler discrete-time versions.

**Observability.** Thus the analog of the continuous-time equation (2) is

$$\begin{bmatrix} y(k) \\ y(k+1) \\ \cdot \\ \cdot \\ \cdot \\ y(k+n-1) \end{bmatrix} = \mathcal{O}x(k) + \mathbf{T} \begin{bmatrix} u(k) \\ u(k+1) \\ \cdot \\ \cdot \\ \cdot \\ u(k+n-1) \end{bmatrix} \tag{23}$$

where

$$\begin{aligned} \mathcal{O}' &= [c' \quad A'c' \quad \cdots \quad (A')^{n-1}c'], \text{ the (transposed) observability} \\ &\quad \text{matrix of } (c, A) \\ \mathbf{T} &= \text{the impulse response matrix, a lower triangular Toeplitz} \\ &\quad \text{matrix with first column } [0 \quad h_1 \quad \cdots \quad h_{n-1}]. \end{aligned} \tag{24}$$

Equation (23) shows that the state at any time $k$ can be uniquely recovered from known inputs and outputs if and only if the observability matrix $\mathcal{O}$ is nonsingular.

Moreover, the convenience of the observability canonical form is immediately evident from the realization (cf. Fig. 2.3-2): With no input, we can

just "read out" the states as (for $n = 3$)

$$y(k) = x_1(k), \qquad y(k+1) = x_2(k), \qquad y(k+2) = x_3(k)$$

We note also that in all cases relation (23) provides a viable calculation method, unlike in continuous time where the analogous method requires differentiation of $y(\cdot)$. (Note, though, that if the discrete-time system is obtained by *sampling* a continuous-time system, then in the limit as the sampling rate gets very high, it can be shown that the discrete-time solution will converge to the continuous-time solution based on derivatives.)

**Controllability.** Similar simplifications arise when we study the controllability problem in discrete time. Thus we can directly write from the state equation (11) that

$$x(n) = A^n x_0 + [b \quad Ab \quad \cdots \quad A^{n-1}b][u(n-1) \quad \cdots \quad u(0)]' \qquad (25)$$

the continuous analog of which was not so readily obtained [cf. Eq. (10) in Sec. 2.3.2]. We see from (25) that we can transfer *any* initial state to an *arbitrary* state (in not more than $n$ steps) if and only if the matrix $\mathcal{C}(A, b)$ is nonsingular.

Again, note that the discrete problem is somewhat easier and its solution more realistic, at least in that we do not use impulsive inputs. (Note again, though, that in the continuous-time limit the present solution will converge to the impulsive function solution of Sec. 2.3.2—it is in fact an interesting analytical exercise to prove this, as a reader with enough time might wish to show.) In fact, a special case of the above discrete-time problem, viz. the so-called *finite-settling-time problem* [12] in which it is required to find an input to return a perturbed ($x_0 \not\equiv 0$) system to the origin as quickly as possible, was apparently the one in which the notion of state controllability was first explicitly introduced, though the matrix $\mathcal{C}$ had arisen earlier in other control problems (cf. the historical remarks in [18]).

***Controllability to and from the Origin; Reachability.** This *controllability-to-the-origin* problem has several interesting aspects, which we shall discuss further in later chapters (e.g., Sec. 3.5.2). Let us consider one special aspect here, because it has been the source of some terminological problems. It arises from the fact that while, as (25) shows, the nonsingularity of $\mathcal{C}$ is *sufficient* to ensure that any arbitrary initial state $x_0$ can be driven to the origin in a finite time, this condition is not *necessary*. In other words, it may be possible to take any state to the *origin* even if $\mathcal{C}$ is singular. For example, this can happen if the $n \times n$ matrix $A$ has the property that $A^k = 0$ for some $k$; for such so-called *nilpotent* matrices, any $x_0$ can be driven to zero with the "zero" input $u(k) \equiv 0$. Another example will be given soon.

Therefore, *controllability p.s.t.o.* (pointwise state to the origin, to use Rosenbrock's terminology [22]) is not always equivalent to the nonsingularity of the matrix

$$\mathcal{C} = [b \quad Ab \quad \cdots \quad A^{n-1}b]$$

On the other hand, *controllability p.s.f.o.* (pointwise state from the origin), or *reachability* as it is often called, is easily seen from (25) to always be equivalent to the nonsingularity of the above matrix, which some people therefore prefer to call the *reachability* matrix.

A necessary and sufficient condition for controllability p.s.t.o. is also clear from (25):

$$A^n x_0 \in \mathcal{R}[b \quad Ab \quad \cdots \quad A^{n-1}b] \tag{26}$$

where $\mathcal{R}[\cdots]$ denotes the range space, or the collection of all linear combinations, of the columns $\{b, Ab, \ldots, A^{n-1}b\}$. It is easy to see that controllability p.s.t.o. is in fact completely equivalent to the nonsingularity of $\mathcal{C}$ when the matrix $A$ is nonsingular. For then, from (25), we can write, with $x(n) = 0$,

$$-x_0' = [u(n-1) \quad \cdots \quad u(0)]\mathcal{C}'(A^{-n})' \tag{27}$$

which shows that a suitable input sequence will exist (if and) only if $\mathcal{C}$ is nonsingular. With nonsingular $A$ it seems therefore more appropriate to consider the matrix (rather than $\mathcal{C}$)

$$A^{-n}\mathcal{C} = [A^{-n}b \quad \cdots \quad A^{-1}b] \tag{28}$$

when examining controllability to the origin.

**Example 2.3-A. Controllability Need Not Imply Reachability**

Let

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

The matrix

$$[b \quad Ab] = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

is singular, so that an arbitrary state, e.g., $[1 \ 1]'$, cannot be reached from the zero state no matter what the input is. On the other hand, any initial state can be returned to zero, for if $x_1(0) = \alpha$, $x_2(0) = \beta$, let $u(0) = -(\alpha + \beta)$, and then

$$x(1) = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}(-\alpha - \beta) \equiv 0$$

(Note that $A^k = A \neq 0$ for all $k$.)

However, controllability (to the origin) does imply reachability when $A$ is nonsingular. ∎

**Reachability and Controllability Canonical Forms.** It is also useful to explore the difference between reachability (controllability from the origin) and controllability to the origin in terms of (block diagrams of) canonical realizations. Thus if we ask for a realization in which the reachability problem is to be trivial, the realization of Fig. 2.3-1 immediately offers itself; it is clear from the figure that to set up any

state $\theta = [\theta_1 \ \theta_2 \ \theta_3]$ we just feed in the inputs

$$u(0) = \theta_3, \qquad u(1) = \theta_2, \qquad u(2) = \theta_1$$

and then we shall have

$$x(3) = \theta$$

Therefore the realization of Fig. 2.3-1 is in fact the *reachability canonical form.*

On the other hand, suppose we wish to take arbitrary states to the origin in the most natural way. If we assume that $A$ is nonsingular,† then formulas (27)–(28) show that we should seek a realization such that

$$[A^{-1}b \quad \cdots \quad A^{-n}b] = I \tag{29}$$

Some analysis‡ shows that the realization (written for $n = 3$)

$$A = \begin{bmatrix} -a_1 & 1 & 0 \\ -a_2 & 0 & 1 \\ -a_3 & 0 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} -a_1 \\ -a_2 \\ -a_3 \end{bmatrix} \tag{30}$$

has the above property. The output matrix $c$ is not needed here but can be found from the equation ($c = [\gamma_1 \ \gamma_2 \ \gamma_3]$)

$$-\begin{bmatrix} a_1 & a_2 & a_3 \\ a_2 & a_3 & 0 \\ a_3 & 0 & 0 \end{bmatrix}\begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} = \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} \tag{31}$$

Formula (31) can be obtained by noting that if the input sequence is $\{1, a_1, a_2, a_3\}$, then the output sequence [with $x(0) = 0$] should be $\{0, b_1, b_2, b_3\}$. This calculation is also very evident from the realization, which is shown in Fig. 2.3-3. From the realization, it is also readily evident that with the input

$$u(0) = -x_1(0), \qquad u(1) = -x_2(0), \qquad u(2) = -x_3(0)$$

the state will be zero at $k = 3$:

$$\begin{aligned} x(0) &= [x_1(0) \quad x_2(0) \quad x_3(0)]' \\ x(1) &= [x_2(0) \quad x_3(0) \quad 0 \ ]' \\ x(2) &= [x_3(0) \quad 0 \quad 0 \ ]' \\ x(3) &= [\ 0 \quad 0 \quad 0 \ ]' \end{aligned}$$

Therefore the realization in Fig. 2.3-3 should really have the name controllability (p.s.t.o.) canonical form, rather than the one in Fig. 2.3-1.

†The case of singular $A$ can also be studied by introducing some notions of *generalized inverses*; we shall not do so here.

‡*Hint:* Use Exercise A.32 on the inverse of a companion matrix.

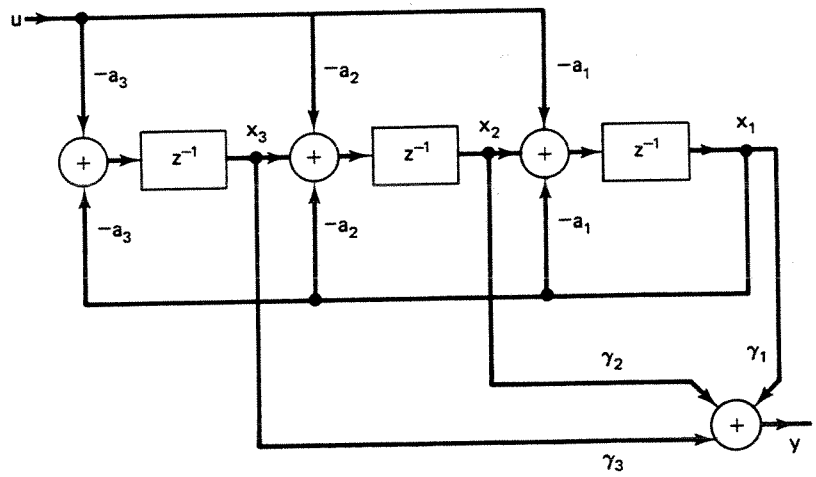



**Figure 2.3-3.** A form that perhaps better deserves the name controllability (to the origin) canonical form. However, it is necessary here that $A$ be nonsingular. Note how easy it is to determine an input that will take an arbitrary initial state to zero.

However, we shall not insist upon this distinction in this book, chiefly because the two concepts are always equivalent when $A$ is nonsingular and also in continuous time (where the nonsingular matrix $\exp A$ takes the place of the matrix $A$; cf. Sec. 2.5.1). Let us, however, briefly consider the corresponding questions for the observability problem.

**Observability and Constructibility.** There must clearly be dual distinctions in the observability problem, and it is useful to examine them. The discrete-time analog of what we called the observability canonical form in Sec. 2.1 (cf. Fig. 2.1-7) was already presented as Fig. 2.3-2. There is no question here that this name is well deserved, because it is clear from the figure that, with the input $u(\cdot) \equiv 0$, we can "read out" the state at any time $k$ as (for $n = 3$)

$$y(k) = x_1(k), \qquad y(k+1) = x_2(k), \qquad y(k+2) = x_3(k)$$

Then what corresponds in the observability problem to the distinction between controllability (to the origin) and reachability (from the origin)? Some reflection shows that the dual of reachability of a state (from the origin using past inputs) is observability of a state from future outputs. The dual of controllability of a state (to the origin using future inputs) should then be observability of a state using past outputs. This is called *constructibility* and is equivalent to observability (in the original sense) when $A$ is nonsingular.

**Example 2.3-B. Constructibility Need Not Imply Observability**

Let

$$c = [1 \quad 1], \qquad A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Then the observability matrix is

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

which is singular. Therefore the realization is unobservable, as is shown by the equations

$$y(0) = x_1(0) + x_2(0)$$
$$y(1) = 2x_1(0) + 2x_2(0)$$

so that we can only observe the sum $x_1(0) + x_2(0)$.

On the other hand, we see that

$$\begin{bmatrix} x_1(0) \\ x_2(0) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x_1(-1) \\ x_2(-1) \end{bmatrix} = \begin{bmatrix} x_1(-1) + x_2(-1) \\ x_1(-1) + x_2(-1) \end{bmatrix}$$

so that if we have available

$$y(-1) = x_1(-1) + x_2(-1)$$

then we can construct

$$x_1(0) = y(-1) \quad \text{and} \quad x_2(0) = y(-1)$$

Equivalently, if our observations can only occur for $k \geq 0$, note that we can construct

$$x_1(k+1) = y(k) = x_2(k+1)$$ ∎

Clearly, observability implies constructibility but not vice versa, unless $A$ is nonsingular, in which case the two concepts are completely equivalent. In the latter case, we can still ask for a canonical form in which constructibility is trivial. We leave it to the reader to show that such a form is (for $n = 3$)

$$A_{cb} = \begin{bmatrix} -a_1 & -a_2 & -a_3 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad b_{cb} = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix} \tag{32a}$$

$$c_{cb} = [-a_1 \quad -a_2 \quad -a_3] \tag{32b}$$

which yields

$$\begin{bmatrix} c_{cb}A_{cb}^{-1} \\ c_{cb}A_{cb}^{-2} \\ c_{cb}A_{cb}^{-3} \end{bmatrix} = I \tag{32c}$$

The $\{\gamma_i\}$ are determined by relation (31), and the constructibility canonical form is shown in Fig. 2.3-4.

Finally, in Fig. 2.3-5 we have summarized the relations among the concepts described above.

**Expanded Transfer Relations and Some Useful Identities.**† By definition, feeding $a(z)$ into a system with the transfer function $H(z) = b(z)/a(z)$ (and zero initial condi-

†The results of this section will be used only in Sec. 2.4.4.

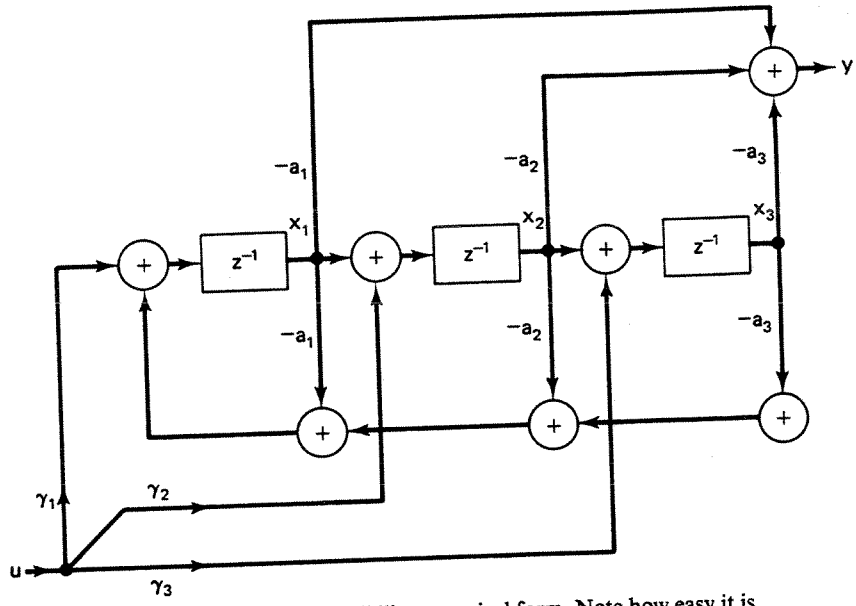



**Figure 2.3-4.** Constructibility canonical form. Note how easy it is to determine the state at a given time ($t \geq 3$) from knowledge of past $y(\cdot)$.

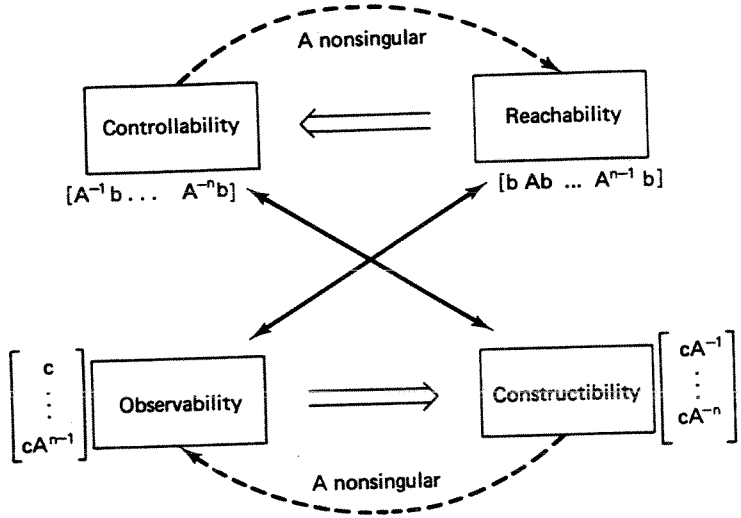



**Figure 2.3-5.** Relations among four basic concepts in discrete time. For continuous-time constant realizations, the distinction between reachability and controllability and between observability and constructibility disappear (because the nonsingular matrix $e^{At}$ takes the place of $A$).

tions) will give the output sequence

$$H(z)a(z) = b(z) \tag{33}$$

Comparing powers of $z$ on both sides of (33) will give us the matrix equation

$$\begin{bmatrix} 0 & & & & \\ h_1 & 0 & & & \\ h_2 & h_1 & \cdot & & \\ \cdot & & \cdot & & \\ \cdot & & & \cdot & \\ \cdot & & & & 0 \\ h_n & & \cdots & & h_1 \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_{n-1} \end{bmatrix} = \begin{bmatrix} 0 \\ b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_n \\ 0 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \tag{34}$$

Now note that by linearity and time invariance

$$H(z)a(z)z^k = z^k b(z), \qquad k = \text{any integer} \tag{35}$$

In matrix terms, if the input is shifted up or down, the Toeplitz nature of the impulse response matrix in (34) implies a corresponding shift in the output sequence. In particular, therefore, if we consider the $n$ shifts $\{k = 0, -1, \ldots\}$, we can write

$$\begin{bmatrix} 0 & & & & & & \\ h_1 & 0 & & & & & \\ \cdot & \cdot & & & & & \\ \cdot & & \cdot & & & & \\ \cdot & & & \cdot & & & \\ h_{n-1} & & & h_1 & 0 & & \\ \hline h_n & & \cdots & h_1 & 0 & & \\ & & & & h_1 & & \\ \cdot & & & & \cdot & \cdot & \\ \cdot & & & & \cdot & & \cdot \\ \cdot & & & & \cdot & & & \cdot \\ h_{2n-1} & & \cdots & h_n & h_{n-1} & \cdots & h_1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & & 0 \\ a_1 & 1 & & \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ a_{n-1} & a_{n-2} & & 1 \\ \hline a_n & a_{n-1} & & a_1 \\ 0 & a_n & & \cdot \\ \cdot & & \cdot & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & & \cdots & a_n \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & \cdots & 0 \\ b_1 & 0 & & \\ \cdot & & & \\ \cdot & b_1 & & \cdot \\ \cdot & & \cdot & \cdot \\ b_{n-1} & & & 0 \\ \hline b_n & b_{n-1} & & b_1 \\ 0 & b_n & & b_2 \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ 0 & & \cdots & b_n \end{bmatrix} \tag{36}$$

or more compactly as [cf. (2.3-24) and (2.2-49)], say,

$$\begin{bmatrix} \mathbf{T} & \bigcirc \\ M[1, n-1]\tilde{I} & \mathbf{T} \end{bmatrix} \begin{bmatrix} \mathfrak{A}_- \\ \mathfrak{A}_+ \end{bmatrix} = \begin{bmatrix} \mathfrak{B}_- \\ \mathfrak{B}_+ \end{bmatrix} \tag{37}$$

where $\tilde{I}$ is the *reversed* identity matrix,

$$\tilde{I} = \begin{bmatrix} & & & 1 \\ & \bigcirc & 1 & \\ & \cdot & & \\ & \cdot & & \\ 1 & & \bigcirc & \end{bmatrix}, \qquad \tilde{I}^2 = I \tag{38}$$

These equations yield some interesting algebraic identities (useful, e.g., in Sec. 2.4.4). Thus, from (37) we immediately obtain a nice factorization of the impulse response matrix,

$$\mathbf{T} = \mathfrak{B}_-\mathfrak{A}_-^{-1} \tag{39a}$$

Moreover, since (lower or upper) triangular Toeplitz matrices commute (cf. Exercise A.6), we can also write

$$\mathbf{T} = \mathfrak{A}_-^{-1}\mathfrak{B}_- \tag{39b}$$

From (38) and (39), we also obtain a formula for the Hankel matrix,

$$M[1, n-1] = [\mathfrak{B}_+ - \mathfrak{B}_-\mathfrak{A}_-^{-1}\mathfrak{A}_+]\mathfrak{A}_-^{-1}\tilde{I} \tag{40}$$

The matrix $\mathfrak{A}_-^{-1}$ has a nice interpretation, which can be obtained by comparing the definition of $\mathfrak{A}_-$ [cf. (36) and (37)] with formula (9) $\mathcal{C}_c^{-1}$, where $\mathcal{C}_c$ is the controllability matrix of the controller form $\{A_c, b_c, c_c\}$. We can identify

$$\mathfrak{A}_-^{-T} = \tilde{I}\mathfrak{A}_-^{-1}\tilde{I} = \mathcal{C}_c, \qquad \mathfrak{A}_- = \tilde{I}\mathcal{C}_c^{-1}\tilde{I} \tag{41}$$

Furthermore, it is easy to verify from the definition (2.2-49) of $M[1, n-1]$ that for any realization $\{A, b, c\}$

$$M[1, n-1] = \mathcal{O}(c, A)\mathcal{C}(A, b) \tag{42}$$

But then by combining (40)–(42), we obtain the nonobvious formula

$$\mathcal{O}_c = \mathcal{O}(c_c, A_c) = [\mathfrak{B}_+ - \mathfrak{B}_-\mathfrak{A}_-^{-1}\mathfrak{A}_+]\tilde{I} \tag{43}$$

For completeness we note here another nonobvious formula for $\mathcal{O}_c$,

$$\mathcal{O}_c = \tilde{I}b(A_c) = \tilde{I}[b_1A_c^{n-1} + \cdots + b_nI] \tag{44}$$

This formula, apparently first given by Wonham and Stuelpnagel [17, p. 178], can

be derived by an easy application† of the result of Exercise A.33, part 2. Note that it yields the unexpected matrix identity

$$\tilde{I}b(A_c)\tilde{I} = [\mathcal{B}_+ - \mathcal{B}_-\mathcal{A}_-^{-1}\mathcal{A}_+] \tag{45}$$

**Reprise.** In this section we have shown that many algebraic properties of a realization $\{A, b, c\}$ are independent of whether it is a continuous- or discrete-time realization. However, there are some differences in the physical characterizations, with the discrete-time problem often leading to more obvious and more elementary proofs (without derivatives or impulses). In this section we also provided some nice illustrations of the power of using the special state-space realizations of Sec. 2.1.2 in various calculations and explorations.

### *2.3.4 Some Worked Examples

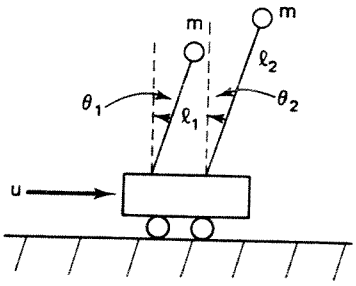
We shall provide several examples to illustrate some of the concepts introduced in the previous sections. The first two examples basically just show some of the algebra involved in many problems. Example 2.3-2 shows that the algebra can often be replaced, or at least illuminated, by returning to basic definitions; it also points the way to some of the results in the next section. The next three examples illustrate the value of the concepts of controllability and observability in answering some simple questions that arise in just manipulating the many possible state-space realizations of a given system. Finally, in Example 2.3-6, we show how many of the things we have said for systems over the real numbers can be carried over to digital systems operating over the binary field.

**Example 2.3-1. Cart with Inverted Pendulums**

A cart of mass $M$ has two inverted pendulums on it of lengths $l_1$ and $l_2$, both with bobs of mass $m$. For small $|\theta_1|$ and $|\theta_2|$, the equations of motion can be seen to be

$$M\dot{v} = -mg\theta_1 - mg\theta_2 + u$$
$$m(\dot{v} + l_i\ddot{\theta}_i) = mg\theta_i, \qquad i = 1, 2$$

where $v$ is the velocity of the cart and $u$ is an external force applied to the cart (see the figure).




†A detailed derivation is given in Example 2.3-5 in the next section.

1. Is it always possible to "control" both pendulums, i.e., keep them both vertical, by using the input $u(\cdot)$?

2. Is the system observable with output $y = \theta_1$?

**Solution.** Let

$$x_1 = \theta_1, \qquad x_2 = \theta_2, \qquad x_3 = \dot{\theta}_1, \qquad x_4 = \dot{\theta}_2$$

Then we obtain (after eliminating $\dot{v}$ from the equations of motion) the state equations $\dot{x} = Ax + bu$ with

$$A = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ a_1 & a_2 & 0 & 0 \\ a_3 & a_4 & 0 & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 0 \\ 0 \\ -1/Ml_1 \\ -1/Ml_2 \end{bmatrix}$$

where

$$a_1 = \frac{(M+m)g}{Ml_1}, \qquad a_2 = \frac{mg}{Ml_1}$$

$$a_3 = \frac{mg}{Ml_2} \qquad a_4 = \frac{(M+m)g}{Ml_2}$$

1. To keep the pendulums vertical, i.e., to have $\theta_1 = 0 = \theta_2$, the realization must be controllable from the input. Therefore we have to check $\det \mathcal{C}(A, b)$. After some algebra, we can obtain

$$\det \mathcal{C}(A, b) = \frac{2pq}{M^2 l_1 l_2} - \frac{q^2}{M^2 l_1^2} - \frac{p^2}{M^2 l_2^2}$$

where

$$p = \frac{a_1}{Ml_1} + \frac{a_2}{Ml_2}, \qquad q = \frac{a_3}{Ml_1} + \frac{a_4}{Ml_2}$$

Therefore $\mathcal{C}(A, b)$ will be singular if and only if

$$2l_1 l_2 pq - q^2 l_2^2 - p^2 l_1^2 = 0$$

i.e., after some more algebra, if and only if

$$M^2 g^2 l_1^2 l_2^2 (l_1 - l_2)^2 = 0$$

or,

$$l_1 = l_2$$

Therefore, if $l_1 \neq l_2$, the realization is controllable, and we can have $\theta_1 = 0 = \theta_2$ by suitable choice of the input.

2. If $y = \theta_1$, then $c = [1\ 0\ 0\ 0]$, and we can check that

$$\det \Theta(c, A) = -a_2^2 = -\left(\frac{mg}{Ml_1}\right)^2 \neq 0$$

so we shall have observability in all cases (even when $l_1 = l_2$).

*Remark:* Try to deduce these results by physical arguments. ▮

**Example 2.3-2. Dynamics of a Hot Air Balloon**

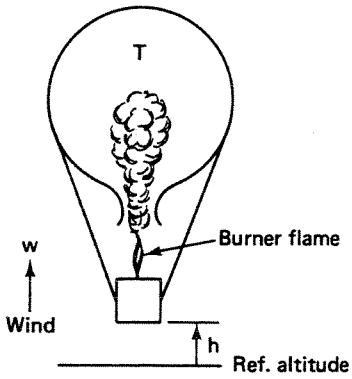
Approximate equations of motion for a hot air balloon (see the figure) are

$$\dot{\theta} = -\frac{1}{\tau_1}\theta + u$$

$$\dot{v} = -\frac{1}{\tau_2}v + \sigma\theta + \frac{1}{\tau_2}w$$

$$\dot{h} = v$$

where $\theta$ = temperature change of air in balloon away from equilibrium temperature, $u$ is proportional to change in heat added to air in balloon (control), $v$ = vertical velocity, $h$ = change in altitude from equilibrium altitude, and $w$ = vertical wind velocity (disturbance).




1. Can the temperature change $\theta(\cdot)$ and a constant wind velocity $w$ be observed by a continuous measurement of altitude change $h$? (Assume, as usual, that $u$ is known.)

2. Determine the transfer function from $u$ to $h$ and from $w$ to $h$. Is the system completely controllable by $u$? Is it completely controllable by $w$?

**Solution**

1. To the three state equations given, we must add a fourth for $w$ in order to decide whether a constant $w$ can be observed by measuring $h$. The required equation is evidently

$$\dot{w} = 0$$

Proceeding from fundamentals, to determine the observability of $\theta$ and $w$ we must see if we can solve for them from knowledge of $h$ and its derivatives. Now note that

$$\dot{y} = \dot{h} = v$$

which implies that $v$ is observable. We also have

$$\ddot{y} = \ddot{h} = \dot{v} = -\frac{1}{\tau_2}v + \sigma\theta + \frac{1}{\tau_2}w \qquad (*)$$

which implies that $\sigma\theta + (1/\tau_2)w$ is observable. Finally,

$$\dddot{y} = \dddot{h} = \ddot{v} = -\frac{1}{\tau_2}\dot{v} + \sigma\dot{\theta}$$

from which $\sigma\dot{\theta}$ is observable, whence $\theta$ is observable (if $\sigma \neq 0$). It then follows from (*) that $w$ also is observable.

To proceed more formally and determine the observability of the whole system, note that

$$\begin{bmatrix} \dot{\theta} \\ \dot{v} \\ \dot{h} \\ \dot{w} \end{bmatrix} = \begin{bmatrix} -1/\tau_1 & 0 & 0 & 0 \\ \sigma & -1/\tau_2 & 0 & 1/\tau_2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \theta \\ v \\ h \\ w \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} u$$

$$y = [0 \quad 0 \quad 1 \quad 0][\theta \quad v \quad h \quad w]'$$

Then it can be checked that $\mathcal{O}$ will be nonsingular when $\sigma \neq 0$.

2. We now consider $w$ as a *second input*. Our state equations are then

$$\begin{bmatrix} \dot{\theta} \\ \dot{v} \\ \dot{h} \end{bmatrix} = \begin{bmatrix} -\tau_1^{-1} & 0 & 0 \\ \sigma & -\tau_2^{-1} & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \theta \\ v \\ h \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0 & \tau_2^{-1} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} u \\ w \end{bmatrix}; \qquad y = [0 \quad 0 \quad 1] \begin{bmatrix} \theta \\ v \\ h \end{bmatrix}$$

Note that the eigenvalues are $-1/\tau_1$, $-1/\tau_2$, and 0.

The transfer function to the output from one input of a multi-input system is defined with all other inputs zero. Then when all inputs are present we merely use superposition.

Taking transforms with $w = 0$, we obtain, after some algebra,

$$\left.\frac{h(s)}{u(s)}\right|_{w=0} = \frac{h(s)}{v(s)}\frac{v(s)}{\theta(s)}\frac{\theta(s)}{u(s)} = \frac{\sigma}{s(s + 1/\tau_2)(s + 1/\tau_1)}$$

Similarly, with $u = 0$, we obtain

$$\left.\frac{h(s)}{w(s)}\right|_{u=0} = \frac{h(s)}{v(s)}\frac{v(s)}{w(s)} = \frac{1/\tau_2}{s(s + 1/\tau_2)} = \frac{1}{s(\tau_2 s + 1)}$$

The eigenvalue at $-1/\tau_1$ has evidently been cancelled by the numerator.

To determine controllabilities, we can calculate $\mathcal{C}_u$ and $\mathcal{C}_w$ to find that the system is controllable by $u$ (for $\sigma \neq 0$) but not by $w$. (The state equations show directly that, with $u = 0$, the equation for $\theta$ is undriven; hence $\theta$ is uncontrollable by $w$.) We shall see in Sec. 2.4 that this is related to the fact that all the system eigenvalues appear as poles of the transfer function from $u$, but the one at $-1/\tau_1$ is absent in the transfer function from $w$. We show there that no eigenvalues are cancelled out of the nominal transfer function from some input to some output if and only if the system is controllable by that input and observable from that output. Here the system is observable from $h$, so that any cancellations in a transfer function to $h$ imply uncontrollability by the corresponding input. ∎

**Example 2.3-3. Uniqueness of Realizations**

We have a realization $\{A, b, c\}$ with known characteristic polynomial $a(s) = \det(sI - A)$ and with $\mathcal{O}(c, A) = I$. Show that this information uniquely determines $\{A, b, c\}$.

**Solution.** Since $\mathcal{O} = I$, it is clear that $c = [1\ 0\ \cdots\ 0]$. Then $cA$ must equal the first row of $A$, and, since $\mathcal{O} = I$, this first row must be $[0\ 1\ 0\ \cdots\ 0]$. Proceeding thus, we see that the first $n - 1$ rows of $A$ have the form $[0\ I_{n-1}]$. Therefore $A$ must be a companion matrix, and it is easy to check that if

$$\det(sI - A) = a(s) = s^n + a_1 s^{n-1} + \cdots + a_n$$

then the last row of $A$ must be $-[a_n\ a_{n-1}\ \cdots\ a_1]$. Therefore the $c$ and $A$ matrices are uniquely determined. To check if there is a unique $b$ we must show that

$$c(sI - A)^{-1}b^{(1)} = c(sI - A)^{-1}b^{(2)} \qquad (*)$$

implies $b^{(1)} = b^{(2)}$. Now (*) implies that

$$cA^i b^{(1)} = cA^i b^{(2)}, \qquad i = 0, 1, \ldots$$

or

$$\mathcal{O}(b^{(1)} - b^{(2)}) = 0$$

or (since $\mathcal{O} = I$)

$$b^{(1)} - b^{(2)} = 0$$

Note that the fact that $\mathcal{O}(c, A) = I$ is not really necessary to obtain $b_1 = b_2$; it suffices for $\mathcal{O}(c, A)$ to be nonsingular. ∎

**Example 2.3-4. Similarity Transformation to Controller Form**

Let $\{A, b, c\}$ be a given realization function, and let $\{A_c, b_c, c_c\}$ be another realization in controller from such that $\det(sI - A) = \det(sI - A_c) = s^n + a_1 s^{n-1} + \cdots + a_n = a(s)$. Also let $c(sI - A)^{-1}b = c_c(sI - A_c)^{-1}b_c = b(s)/a(s)$.

Show that we can find a similarity transformation to take the realization $\{A, b, c\}$ to the controller form $\{A_c, b_c, c_c\}$ if and only if $\{A, b\}$ is controllable. *Hint:* Try to determine the transformation explicitly.

**Solution.** We have to try to find an invertible matrix $T$ such that

$$A_c = T^{-1}AT, \qquad b_c = T^{-1}b, \qquad c_c = cT$$

An explicit formula for $T$ can be found in various ways, but here let us start from scratch. Let

$$T = [t_1 \quad \cdots \quad t_n]$$

Then the fact that $b_c' = [1 \ 0 \ \cdots \ 0]$ gives

$$b = Tb_c = t_1$$

Now write $A_c = T^{-1}AT$ as

$$TA_c = AT = [At_1 \quad At_2 \quad \cdots \quad At_n]$$

Then, recalling the rule that multiplying a matrix by a column vector gives the weighted sum of the columns of the matrix, we shall have the equations

$$At_1 = -a_1t_1 + t_2 = -a_1b + t_2$$
$$At_2 = -a_2b + t_3, \ldots, At_n = -a_nb$$

which can be rearranged as

$$T = [b \quad Ab \quad \cdots \quad A^{n-1}b]\mathcal{A}_-'$$

where

$\mathcal{A}_-'$ = an upper triangular Toeplitz matrix with first row $[1 \quad a_1 \ \cdots \ a_{n-1}]$

This shows immediately that the matrix $T$ will be invertible if and only if $\mathcal{C}(A, b)$ is nonsingular, i.e., if $\{A, b\}$ is controllable.

We note from Eq. (2.3-9) that $\mathcal{A}_-' = \mathcal{C}_c^{-1}$, so that we have the explicit formula (worth remembering)

$$T = \mathcal{C}\mathcal{C}_c^{-1}, \qquad T^{-1} = \mathcal{C}_c\mathcal{C}^{-1} \quad \blacksquare$$

**Example 2.3-5. Observability of Controller Forms**

Show that the controller-form realization $\{A_c, b_c, c_c\}$ of a transfer function $b(s)/a(s)$, $a(s) = \det(sI - A_c)$, will be observable if and only if $b(s)$ and $a(s)$ are relatively prime, i.e., have no common roots.

**Solution.** We shall use the formula

$$\mathcal{O}_c = \tilde{I}b(A_c)$$

This is easily derived by application of the shifting property of companion matrices (see Exercise A.31):

$$e_i'A_c = e_{i-1}' \qquad \text{for } 2 \leq i \leq n$$

and

$$e_1'A_c = [-a_1 \quad -a_2 \quad \cdots \quad -a_n]$$

where $e_i'$ is the $i$th row of the identity matrix $I$. Now

$$\begin{aligned} e_n'b(A_c) &= b_1e_n'A_c^{n-1} + \cdots + b_ne_n' \\ &= b_1e_1' + b_2e_2' + \cdots + b_ne_n' \\ &= [b_1 \quad b_2 \quad \cdots \quad b_n] = c_c \end{aligned}$$

Then

$$e_{n-1}'b(A_c) = e_n'A_cb(A_c) = e_n'b(A_c)A_c = c_cA_c$$

since polynomial functions of a given matrix commute. Continuing similarly, we find

$$\mathcal{O}_c = [e_n \quad \cdots \quad e_1]'b(A_c) = \tilde{I}b(A_c)$$

This expression now shows that $\mathcal{O}_c$ will be nonsingular if and only if $\det b(A_c) \neq 0$. Now the determinant of a matrix is equal to the product of its eigenvalues; moreover, if $A_c$ has eigenvalues $\{\lambda_i, i = 1, \ldots, n\}$, then the eigenvalues of the polynomial function $b(A_c)$ will be $\{b(\lambda_i), i = 1, \ldots, n\}$ (cf. Exercise A.36). Therefore

$$\det b(A_c) = \prod_{i=1}^{n} b(\lambda_i)$$

and this will be zero if and only if one or more of the $b(\lambda_i)$ are zero. But, by definition, the $\{\lambda_i\}$ are such that $a(\lambda_i) = \det(\lambda_iI - A_c) = 0$. That is, $\det b(A_c)$ will be nonzero, and hence $\{A_c, b_c, c_c\}$ observable, if and only if $a(s)$ and $b(s)$ have no common roots. ∎

**Example 2.3-6. Systems Over the Binary Field**

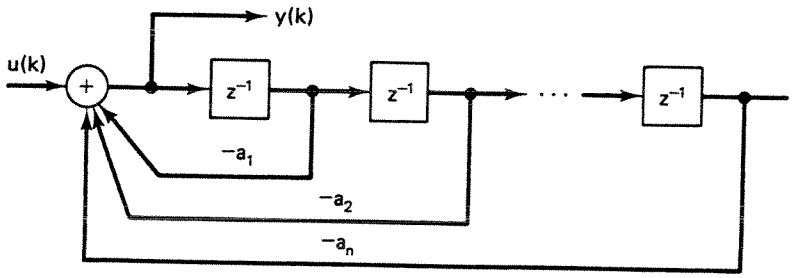
Consider a discrete-time system where the coefficients come not from the field of real numbers but from the finite field with elements {0, 1} with mod-2 addition, and multiplication only by 0 and 1. This is known as $GF(2)$, $G$ for Galois. We can define polynomials in $z^{-1}$ with coefficients in this field. These polynomials may be multiplied, for example,

$$(1 + z^{-1})^2 = 1 + 2z^{-1} + z^{-2} = 1 + z^{-2}$$

or factored, for example,

$$\begin{aligned} (1 + z^{-15}) = (1 + z^{-1})(1 + z^{-1} + z^{-2} + z^{-3} + z^{-4})(1 + z^{-3} + z^{-4}) \\ \cdot (1 + z^{-1} + z^{-4})(1 + z^{-1} + z^{-2}) \end{aligned} \qquad (*)$$

Consider the linear binary *feedback shift register* (FSR) in the figure.



1. Find the transfer function $H(z)$.

2. Let $n = 4$ (i.e., assume only four delay elements), and let $a_1 = a_4 = 1$, $a_2 = a_3 = 0$. Find the impulse response using mod-2 $z$ transforms.

**Solution**

1. The transfer function is determined from $y(k) = u(k) - a_1 y(k-1) - \cdots - a_n y(k-n)$ to be

$$H(z) = \frac{z^n}{z^n + a_1 z^{n-1} + \cdots + a_n}$$

Note that the numerator degree in this case equals the denominator degree since the present output depends on past *and* present inputs. State-space equations for this system can be written by including a direct feedthrough term in the output equation. Thus

$$H(z) = 1 - \frac{a_1 z^{n-1} + \cdots + a_n}{z^n + a_1 z^{n-1} + \cdots + a_n}$$

yields a realization $\{A, b, c, d\}$ where

$$A = \text{a top-companion matrix with first row } -[a_1 \quad \cdots \quad a_n]$$
$$b = [1 \quad 0 \quad \cdots \quad 0]'$$

2. The impulse response is the inverse $z$ transform of $H(z)$. It is possible to invert the $z$ transform by factoring the denominator polynomial and using partial fractions. However, the algebra satisfied by the roots of this polynomial may not be obvious. Another method (which we employ here) is to use the factorization of $1 + z^{-15}$ given in (*) above. Thus note that

$$\begin{aligned} H(z) &= \frac{1}{1 + z^{-1} + z^{-4}} \\ &= \frac{1}{1 + z^{-15}}(1 + z^{-1})(1 + z^{-1} + z^{-2} + z^{-3} + z^{-4})(1 + z^{-3} + z^{-4}) \\ &\quad \cdot (1 + z^{-1} + z^{-2}) \end{aligned}$$

Expanding the numerator, we get

$$H(z) = \frac{1 + z^{-1} + z^{-2} + z^{-3} + z^{-5} + z^{-7} + z^{-8} + z^{-11}}{1 + z^{-15}}$$

Note now that subtraction in $GF(2)$ is the same as addition ($0 - 0 = 0 + 0$, $1 - 0 = 1 + 0$, $1 - 1 = 1 + 1$, $-1 = +1$, etc.) and therefore

$$\frac{1}{1 + z^{-15}} = 1 + z^{-15} + z^{-30} + \cdots$$

Hence the impulse response is periodic, with period 15. The output during one period is determined by the numerator $1 + z^{-1} + z^{-2} + z^{-3} + z^{-5} + z^{-7} + z^{-8} + z^{-11}$ to be the sequence (1 1 1 1 0 1 0 1 1 0 0 1 0 0 0).

Exercises 2.3-33 and 2.3-34 provide further examples of such $GF(2)$ systems. References [23]–[25] provide detailed discussions, with interesting applications to coding and switching theory. ▮

***Exercises***

***2.3-1.***

a. Let $\mathcal{O}_i$ denote the observability matrices of two ($n$th-order) realizations $\{A_i, b_i, c_i\}$. If

$$\mathcal{O}_1 T = \mathcal{O}_2$$

for some invertible matrix $T$, by equating the first two rows we obtain

$$c_1 T = c_2, \qquad c_1 A_1 T = c_2 A_2 = c_1 T A_2$$

Can we conclude from this that $A_1 = T A_2 T^{-1}$?

b. We have two controllable realizations $\{A, b, c_1\}$ and $\{A, b, c_2\}$ of a given transfer function. Show that $c_1 = c_2$. Do this problem in as many different ways as you can.

***2.3-2.***

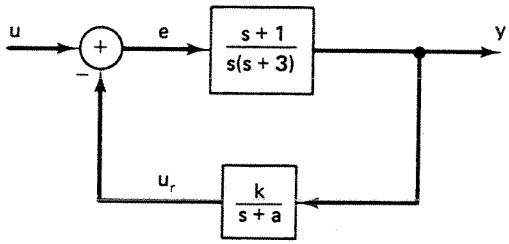
$\{A, b, c\}$ is an $n$th-order realization of a given transfer function $b(z)/a(z)$, $a(z) = z^n + a_1 z^{n-1} + \cdots + a_n$. Suppose that $\mathcal{C}(A, b) = I$. Show that this information completely determines $\{A, b\}$.

***2.3-3.***

Consider the system illustrated in the figure.

a. Give a state-variable realization of this system.

b. Is there any choice of parameters $k$ and/or $a$ for which this realization loses either controllability or observability or both?



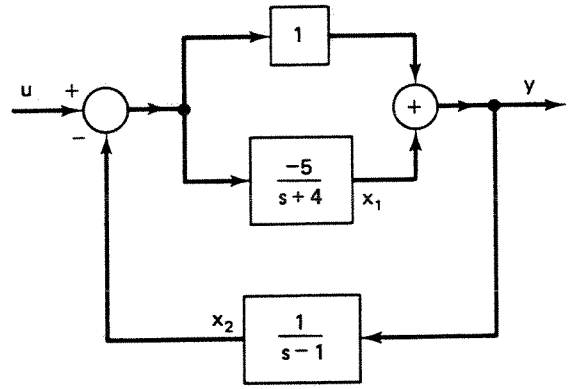

***2.3-4.***

Choose state variables as shown for the system shown in the figure.

a. Write the state equations.

b. Is this system realization controllable? Observable?

c. What is the transfer function from $u(\cdot)$ to $y(\cdot)$?




*2.3-5. Initial Conditions for Differential Equations*

Suppose that

$$y^{(n)}(t) + a_1 y^{(n-1)}(t) + \cdots + a_n y(t) = b_1 u^{(n-1)}(t) + \cdots + b_n u(t), \qquad t \geq 0+,\ u(t) = \delta(t)$$

Show that the initial-condition vector

$$Y'(0+) \triangleq [y(0+), \ldots, y^{(n-1)}(0+)], \qquad Y'(0-) \equiv 0$$

can be calculated as $Y(0+) = \mathfrak{A}_-^{-1} b$, where $b' = [b_1 \ \cdots \ b_n]$ and $\mathfrak{A}_-$ is a lower triangular Toeplitz matrix with first column $[1 \ a_1 \ \cdots \ a_{n-1}]'$.

*2.3-6.*

In Sec. 2.3.1, we noted that when $\mathcal{O}$ was singular, there was an undesirable loss of uniqueness in the solution of the state-determination problem. What is the corresponding phenomenon for the state-controllability problem, and how significant is it?

*2.3-7.*

Refer to Eq. (2.3-25) and suppose that $\mathcal{C}$ is singular. If $x(n) - A^n x_0$ is in the range of $\mathcal{C}$, then we can always find an input sequence to take $x_0$ to $x(n)$. In fact, if $\mathcal{C}$ is singular there will be many such sequences. Find the one with minimum energy, $\sum |u(i)|^2$.

*2.3-8. Alternative Proof of (2.3-10)*

Use the generalized initial-value theorem of Sec. 1.2 to give an alternative derivation of formula (2.3-10).

*2.3-9. The Moments of a System*

a. Given a transfer function $H(s) = b(s)/a(s)$, show that if $a_n \neq 0$ we can always obtain a realization $\{A_{cb}, b_{cb}, c_{cb}\}$, where $b'_{cb} = [\gamma_1, \ldots, \gamma_n]$ and

$$\gamma_k = c_{cb} A_{cb}^{-k} b_{cb}, \qquad k = 1, \ldots, n$$

Compare with formula (2.3-32c).

b. The $\{\gamma_i\}$ are called the *moments* of the system because they can be computed as

$$\int_0^\infty t^{i+1}h(t)\,dt$$

where $h(t) = ce^{At}b$ is the impulse response of the system. Try to prove this formula.

***2.3-10.*** *The Constructibility Problem in Continuous Time*

We recall that the Markov parameters $\{cA^{i-1}b,\ i \geq 1\}$ enter naturally into the observability problem of determining $x(t)$ from $y(t)$ and its derivatives. Show that when $A$ is nonsingular, the moments $\{cA^{-i}b,\ i < 0\}$ enter naturally into the problem of determining $x^{(n)}(t)$ from $y(t)$ and its derivatives. Compare the most natural canonical forms for these two problems. Compare also with the discrete-time results.

***2.3-11.*** *Another Observable Realization*

Given $H(s) = b(s)/a(s)$ as in Sec. 2.1, show that we can always obtain a realization in the form (shown for $n = 3$)

$$A = \begin{bmatrix} -d_1 & 1 & 0 \\ -d_2 & 0 & 1 \\ -d_3 - (b_3 + a_3) & -(b_2 + a_2) & -(b_1 + a_1) \end{bmatrix}, \quad b = \begin{bmatrix} -d_1 \\ -d_2 \\ -d_3 \end{bmatrix}$$

$$c = [1 \quad 0 \quad 0]$$

a. Draw an analog-computer simulation for this realization.
b. Show that this realization is always observable.
c. What statements can you make for the dual form?

***2.3-12.***

For the constant resistance networks of Exercises 2.2-22 and 2.2-23, determine what relations between $R$, $L$, and $C$ are required to make them uncontrollable and/or unobservable.

***2.3-13.***

Extend the formula for $\mathcal{O}_c$ in Example 2.3-5 to show that for any controllable realization $\{A, b, c\}$ of $H(s) = g(s)/a(s)$ we have

$$g(A) = [p_{n-1}(A')c' \quad \cdots \quad p_0(A')c']'$$

where the $p_i(s)$ are defined by $a(s)(sI - A)^{-1}b = [p_{n-1}(s), \ldots, p_0(s)]'$.

***2.3-14.***

Show that the pair

$$\left\{\begin{bmatrix} A & 0 \\ c & 0 \end{bmatrix}, \begin{bmatrix} b \\ 0 \end{bmatrix}\right\}$$

is controllable if and only if $\{A, b\}$ is controllable and

$$\begin{bmatrix} A & b \\ c & 0 \end{bmatrix}$$

has full rank.

***2.3-15.*** *Diagonal Forms*

Show that a pair $\{A, \beta\}$, where $A$ is diagonal with entries $\{\lambda_i\}$, is controllable if and only if (1) the $\lambda_i$ are distinct and (2) all components of $\beta$ are nonzero, with dual results for the observability of a pair $\{\gamma, A\}$. Also try to give simple physical explanations for why, when $A$ is diagonal, repeated eigenvalues cause a loss of controllability and observability. [*Warning:* When $A$ is not diagonal, or cannot be diagonalized, the simple conditions (1) and (2) do not apply; see Exercises 2.3-4 and 2.3-16.]

***2.3-16.*** *Controllability and Observability for Jordan Forms*

a. Let

$$J = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, \qquad \gamma' = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{bmatrix}$$

Find necessary and sufficient conditions on the $\{\beta_i\}$ and $\{\gamma_i\}$ for the nonsingularity of $\mathcal{O}(\gamma, J)$ and $\mathcal{C}(J, \beta)$.

b. Repeat for

$$J = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \mu \end{bmatrix}$$

c. Generalize to an arbitrary Jordan matrix $J$.

***2.3-17.***

Let $A$ be a companion matrix with $-[a_n \ a_{n-1} \ \cdots \ a_1]'$ as the last column. Show that $A' = H^{-1}AH$, where $H$ is an *upper Hankel* matrix with first row $[a_{n-1} \ \cdots \ a_1 \ 1]$.

***2.3-18.***

Let

$$\mathcal{C}_k = [b \quad Ab \quad \cdots \quad A^{k-1}b]$$

Show that if rank $\mathcal{C}_{k+1}$ = rank $\mathcal{C}_k$ for some $k$, then rank $\mathcal{C}_{k+i}$ = rank $\mathcal{C}_k$ for all $i \geq 1$.

***2.3-19.***

Given $\dot{x} = Ax + bu, y = cx$, with $\mathcal{C}$ of rank $r$, show that the transfer function can be written as a ratio of polynomials with denominator of degree $r$ and numerator of degree not greater than $r - 1$. Therefore there must be at least $n - r$ common roots between the numerator and denominator of $c(sI - A)^{-1}b$.

**2.3-20.**

With

$$(sI - A)^{-1}b = \frac{1}{a(s)}[p_{n-1}(s) \quad \cdots \quad p_1(s) \quad p_0(s)]'$$

show that $\{A, b\}$ is controllable if and only if the matrix $P$ defined by

$$[p_{n-1}(s) \quad \cdots \quad p_1(s) \quad p_0(s)]' = P[s^{n-1} \quad \cdots \quad s \quad 1]'$$

is nonsingular. *Hint:* What is the relation of $P$ to the matrix that transforms $\{A, b\}$ to controller form?

**2.3-21.** *Expanded State Equations*

Let

$$\mathcal{Y}_i = [y(0) \quad \cdots \quad y(i-1)]', \qquad \mathcal{U}_i = [u(0) \quad \cdots \quad u(i-1)]'$$
$$\mathcal{C}_i = [b \quad Ab \quad \cdots \quad A^{i-1}b], \qquad \mathcal{O}_i' = [c' \quad A'c' \quad \cdots \quad (A')^{i-1}c']$$

Show that we can write the *expanded* state equations

$$\begin{bmatrix} \mathcal{Y}_i \\ \hline x_i \end{bmatrix} = \left[\begin{array}{c|c} \mathcal{O}_i & \mathbf{T}_i \\ \hline A^i & \mathcal{C}_i \end{array}\right] \begin{bmatrix} x_0 \\ \hline \mathcal{U}_i \end{bmatrix}$$

where $\mathbf{T}_i$ is the Toeplitz matrix of the Markov parameters [cf. (2.3-2c)].

**2.3-22.** *Unknown-Input Observability*

We know that given the input and output (and their necessary derivatives) for an observable system we can determine its state. Prove that for an observable system we can determine the state *without* knowledge of the *input* if and only if the first $n - 1$ Markov parameters, $h_1$ to $h_{n-1}$, are zero. Show that this is equivalent to the fact that the transfer function has no (finite) zeros.

**2.3-23.** *Zero-Output Reachability*

Show that a state $x$ of a discrete-time system can be reached from the origin in $n$ steps while the output is maintained at zero if and only if

$$x = [A^{n-1}b \quad \cdots \quad Ab \quad b]p$$

for some vector $p$ such that

$$\mathbf{T}p = 0$$

where $\mathbf{T}$ is the Toeplitz matrix of Markov parameters defined in Eq. (2.3-24).

**2.3-24.** *Second-Order Vector Differential Equations*

To analyze systems with small damping it is often convenient to use sets of coupled second-order equations (e.g., in vibration and circuit analysis):

$$\ddot{x} + D\dot{x} + Kx = Gu$$

where $x$ = an $n$-vector of generalized coordinates and $u$ = an $n$-vector of control variables.

a. For a conservative system without gyroscopic coupling, $D = 0$, and $K$ is symmetric. Show that the eigenvalues of such a system occur in pairs $\pm\sigma$ or $\pm j\omega$, where $\sigma, \omega$ are real constants, and that the eigenvectors are orthogonal to each other.

b. For a conservative system with gyroscopic coupling, $D$ is antisymmetric (i.e., $D^T \equiv -D$), and $K$ is symmetric. Show that the eigenvalues of such a system are located symmetrically about both the real and imaginary axes.

c. A frictionless spinning top is an example of a conservative system with gyroscopic coupling. The equations of motion may be normalized to

$$\begin{bmatrix} \ddot{x}_1 \\ \ddot{x}_2 \end{bmatrix} + \begin{bmatrix} 0 & p \\ -p & 0 \end{bmatrix} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$

where $x_1, x_2$ are orthogonal lateral displacements from the vertical position and $p$ is proportional to the spin rate. What is the minimum value of $p$ for which the eigenvalues are pure imaginary?

**2.3-25.** *Some Zee-Transform Pairs*

Show that (all time functions are zero for $n < 0$)

a. If $x(n) = 1$, $X(z) = z(z-1)^{-1}$.

b. If $x(n) = e^{-\alpha n}$, $X(z) = z(z - e^{-\alpha})^{-1}$.

c. If $x(n) = n^k$, $X(z) = (-1)^k d^k z(z-1)^{-1}/dz^k$.

d. If $x(n) = (\alpha^n - \beta^n)/(\alpha - \beta)$, $X(z) = z(z-\alpha)^{-1}(z-\beta)^{-1}$.

e. If $x(n) = \sum_0^n f(k)g(n-k)$, $X(z) = F(z)G(z)$.

**2.3-26.** *The Fibonacci Sequence*

The Fibonacci sequence $\{0, 1, 1, 2, 3, 5, 8, 13, \ldots\}$ is generated by the equation

$$y_k = y_{k-1} + y_{k-2}, \qquad k \geq 2$$
$$y_0 = 0, \qquad y_1 = 1$$

a. Show that we can write

$$y_n = \frac{1}{\sqrt{5}}(\lambda_+^n - \lambda_-^n), \qquad \lambda_\pm = \frac{1 \pm \sqrt{5}}{2}$$

b. Show that

$$\lim_{n\to\infty} \frac{\ln y_n}{n} = \ln \frac{\sqrt{5}+1}{2}$$

**2.3-27.** *Realizations of Moving-Average and Autoregressive Models*

a. For a so-called moving-average (MA) model,

$$Y(z) = [b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}]U(z)$$

give, in matrix ($\{A, b, c, d\}$) and in block diagram form, two different types of realizations (controller and observer types).

b. For a so-called autoregressive (AR) model,

$$Y(z)(1 + a_1z^{-1} + \cdots + a_mz^{-m}) = U(z)$$

give, in matrix and in block diagram form, controller- and observer-type realizations. Use the results of Exercise 2.2-20 on inverse realizations to deduce realizations of AR models from MA models and vice versa.

c. Given a mixed or ARMA model,

$$\frac{Y(z)}{U(z)} = \frac{b_0 + b_1z^{-1} + \cdots + b_mz^{-m}}{1 + a_1z^{-1} + \cdots + a_nz^{-n}}$$
$$= \frac{b_0z^m + \cdots + b_m}{z^n + a_1z^{n-1} + \cdots + a_n} \cdot z^{n-m}$$

show how to "merge" the block diagrams of the MA and AR models in (a) and (b) so as to obtain realizations of the ARMA model. Which of these possible mergings will yield realizations with no more than $n$ integrators? Describe the merging procedure in matrix language as well.

***2.3-28.*** *Time-Variant Models*

Time-variant coefficients can be handled in discrete time much more easily than in continuous time. Thus suppose

$$x(i + 1) = A(i)x(i) + B(i)u(i), \qquad i = 0, 1, \ldots$$

a. Show that

$$x(k) = \Phi(k, j)x(j) + \sum_{i=j}^{k-1} \Phi(k, i + 1)B(i)u(i)$$

where

$$\Phi(k, j) \triangleq A(k - 1) \ldots A(j), \qquad \Phi(i, i) = I$$

b. Is it always true that $\Phi(i, j)\Phi(j, k) = \Phi(i, k)$, all $i, j, k \geq 0$?

***2.3-29.*** *An Alternative ["$u(k + 1)$"] Model*

In some problems, it turns out to be more natural to consider state-space models of the form

$$\xi(k + 1) = F\xi(k) + gu(k + 1), \qquad k \geq 0$$
$$y(k) = h\xi(k), \qquad \xi(0) = \xi_0$$

a. Show that the transfer function is

$$\mathcal{H}(z) = hz(zI - F)^{-1}g$$

Note the extra $z$ in the numerator corresponding to the "advanced" input $u(k + 1)$.

b. Prove that an arbitrary initial state $\xi_0$ can be transferred to some other

arbitrary state in a finite time if and only if the matrix $[g\ Fg\ \cdots\ F^{n-1}g]$ is nonsingular.

c. Consider how to translate results for the "$u(k)$" model to the present model. *Hint:* Define an augmented "state" $[x'(k), u(k)]'$.

***2.3-30.*** *Realization from the Impulse Response*

Suppose we have a discrete-time system with known impulse response $\{h_1, h_2, \ldots\}$. Assume also that somehow we know that the minimal order of any state-space realization of the system is $n$.

a. Show how to find a minimal realization $\{A, b, c\}$ of this system. *Hint:* Note the special Hankel matrices (cf. Exercise 2.2-14)

$$M[1, n-1] = \mathcal{O}(c, A)\mathcal{C}(A, b), \qquad M[2, n-1] = \mathcal{O}A\mathcal{C}$$

and assume that $\{A, b, c\}$ is in controllability canonical form.

b. Show that an irreducible transfer function

$$H(z) = \frac{b(z)}{a(z)} = \frac{b_1 z^{n-1} + \cdots + b_n}{z^n + a_1 z^{n-1} + \cdots + a_n}$$

can be computed via the equations

$$M[1, n-1][a_n, \ldots, a_1]' = -[h_{n+1}, \ldots, h_{2n}]'$$

and

$$[b_1, \ldots, b_n]' = \mathbf{T}(h)[1 \quad a_1 \quad \cdots \quad a_{n-1}]'$$

where $\mathbf{T}(h)$ is a lower triangular Toeplitz matrix with first column $[h_1, \ldots, h_n]'$. *Remark:* A basic result for such realization questions is the following result of Kronecker (1890): Let $\{h_1, h_2, \ldots\}$ be the impulse response of a discrete-time system. The system admits a finite-dimensional realization of order $n$ (not necessarily minimal) if and only if $\det M[1, n+i] = 0$, $i = 0, 1, 2, \ldots$.

***2.3-31.*** *Polynomial Inputs and the Cayley-Hamilton Theorem*

Note that inputs $u(z)$ that are polynomials in $z$ correspond to inputs occurring at times less than or equal to zero; e.g., $u(z) = z^2 - 1$ corresponds to an input that is 1 at $t = -2$, $-1$ at $t = 0$, and zero elsewhere. Show that the *response* at $t = 1$ of $x_{k+1} = Ax_k + bu_k$ to the polynomial input $u(z) = g(z) = g_0 + \ldots g_1 z + \ldots + g_m z^{m-1}$ is $x(1) = g(A)b$ (assuming that the system is at rest before the input is applied). This is the analog of Eq. (2.3-10) in the continuous-time case. *Remark:* The above result can be used to obtain a system-theoretical proof of the Cayley-Hamilton theorem. Thus given a matrix $A$ with characteristic polynomial $a(z) = \det(zI - A)$, choose arbitrary matrices $\{c, b\}$ to form a realization $\{A, b, c\}$. Denote $b(z) = c\ \text{Adj} \cdot (zI - A)b$. Now apply the input $u(z) = a(z)$ to this realization. By the above result, the response at $t = 1$ is $y(1) = ca(A)b$. But we also have $y(z) = [b(z)/a(z)]\, a(z) = b(z)$, a polynomial, so that $y(1) = 0$. Therefore $ca(A)b = 0$, which, since $\{c, b\}$ are arbitrary, implies that $a(A) = 0$, the

Cayley-Hamilton theorem. Note that this argument can also be carried out in continuous time using impulsive inputs and replacing $t = 1$ by $t = 0+$

**2.3-32.** *The Notion of State; Euclidean Division via the Controllability Form*

Continuing with the results of Exercise 2.3-31, recall that by the classical division theorem for polynomials we can write

$$u(z) = q(z)a(z) + r(z), \qquad \deg r(z) < n$$

a. Show that the response at $t = 1$ to the polynomial input $u(z)$ is

$$x(1) = r(A)b$$

This is a striking fact: The entire effect of any past input on the state at $t = 1$ is determined only by at most $n$ numbers—the coefficients of $r(z)$ or any nonsingular linear combination of these coefficients (cf. Sec. 2.2.2). Further consideration of this simple result will lead to an abstract definition of state space and state-space realizations; see Sec. 5.1.

b. Suppose the realization $\{A, b\}$ is in controllability form. Show then that $x'[1] = [r_n \cdots r_2 \; r_1]$, where $r(z) = r_1 z^{n-1} + \cdots + r_n$. Check your calculation by studying the state history $\{x(-2), x(-1), x(0), x(1)\}$ of the response of the controllability-form realization (as in Fig. 2.3-1) to an input $r(z) = r_1 z^2 + r_2 z + r_3$. *Remark:* In other words, we can find the remainder $r(z)$ in dividing a polynomial $u(z)$ by another polynomial $a(z)$ by feeding $u(z)$ into a controllability-form realization of $1/a(z)$ and reading out the components of the state vector at $t = 1$.

**2.3-33.** *Maximal Length Shift Register Sequences*

Consider a mod-2 system

$$x(k + 1) = Ax(k), \qquad y(k) = [1 \quad 0 \quad \cdots \quad 0]x(k)$$

If $A^k = I$ for some (smallest) $k$, then $y(\cdot)$ will clearly form a periodic sequence with period some divisor of $k$.

a. Show that if

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

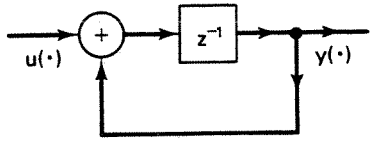
then the period is $15 = 2^4 - 1$.

b. Why is this the maximal-length cycle a four-stage shift register can have?

**2.3-34.** *Flip-Flops*

Consider a system consisting of a single delay element and a single mod-2 adder, as shown in the figure. With a 0 input, the delay element will retain its

state (either 0 or 1). An input 1 will cause the state to change, from 0 to 1 or from 1 to 0. Such a system is called a *trigger flip-flop* or *complementing flip-flop*. The flip-flop is a binary linear system with system function $H_F(z) = 1/(1 + z^{-1})$. Hence, if we have a system with system function $H(z)$ which can be written as $H_1[1/(1 + z^{-1})]$, this system may be constructed from flip-flops instead of delays. Use this fact to design a system using *only* flip-flops (no delays or adders) that is equivalent to the linear binary feedback shift register of Example 2.3-6.




## 2.4 FURTHER ASPECTS OF CONTROLLABILITY AND OBSERVABILITY

In this section, we shall explore further some of the important properties of the concepts of controllability and observability. In Sec. 2.4.1 we shall bring out their significance for the important problem of deciding when a given realization is minimal, i.e., realized with the smallest possible number of integrators (or delay elements). This is a nice application of some apparently rather special concepts in a general, nondynamical situation (we are not setting up or observing states). Furthermore, we can also now deduce an alternative characterization of minimality in terms of the irreducibility of the associated transfer function. This fact brings out some useful connections with the classical theory of polynomials and their resultants, a topic further explored in Sec. 2.4.4. In Sec. 2.4.2 we shall present some very useful standard forms for noncontrollable and/or nonobservable realizations, which shed further light on the deep relations between controllability and observability and system structure. A nice application of these results is to the development of some very powerful tests for controllability and observability, which we call the PBH tests (Sec. 2.4.3). In our opinion these tests are the most useful in many problems, and we shall apply them often in this book. Finally, in Sec. 2.4.5 we shall present several illustrative examples.

### *2.4.1 Joint Observability and Controllability: the Uses of Diagonal Forms*

The question of joint controllability and observability is not trivially resolvable for any of the four canonical forms of Sec. 2.1.2 because, for example, while the observer and observability forms have simple observability matrices, direct calculation of their controllability matrices does not seem to yield anything very transparent. In such situations, the diagonal form cor-

responding to the sum (parallel) realizations of Sec. 2.1.3 is often helpful in providing some insight into the situation. Thus, consider a realization with

$$A_d = \text{diag}\,\{\lambda_1, \ldots, \lambda_n\} \tag{1}$$

$$b_d = [\gamma_1 \quad \cdots \quad \gamma_n]', \qquad c_d = [\delta_1 \quad \cdots \quad \delta_n] \tag{2}$$

Then an easy calculation using the formula of Exercise A.7 for the determinant of a Vandermonde matrix shows that

$$\det \mathcal{C} = (\prod \gamma_i) \prod_{i<j} (\lambda_i - \lambda_j) \tag{3}$$

Therefore the realization will be controllable if and only if

$$\lambda_i \neq \lambda_j \tag{4a}$$

and

$$\gamma_i \neq 0, \qquad i, j = 1, \ldots, n \tag{4b}$$

Similarly, it can be proved that the realization will be observable if and only if

$$\lambda_i \neq \lambda_j \tag{5a}$$

and

$$\delta_i \neq 0, \qquad i, j = 1, \ldots, n \tag{5b}$$

It is not hard to see intuitively why controllability and/or observability are lost under the above conditions—we leave that to the reader. But we shall note here the implications for joint observability and controllability: If some of the eigenvalues $\{\lambda_i\}$ are repeated, or if some of the $\{\gamma_i\}$ or $\{\delta_i\}$ are zero, this means that the transfer function

$$H(s) = c_d(sI - A_d)^{-1}b_d = \sum_1^n \frac{\gamma_i \delta_i}{s - \lambda_i} \tag{6}$$

will in fact have less than $n$ terms in its partial fraction expansion. Therefore the "nominal" representation

$$H(s) = \frac{b(s)}{a(s)}, \qquad a(s) \triangleq \det\,(sI - A_d) = s^n + a_1 s^{n-1} + \cdots + a_n \tag{7}$$

will in fact have the "reduced" form

$$H(s) = \frac{\bar{b}_1 s^{r-1} + \cdots + \bar{b}_r}{s^r + \bar{a}_1 s^{r-1} + \cdots + \bar{a}_r}, \qquad r < n \tag{8}$$

In other words, we shall have *cancellations* in the transfer function. Reversing

this analysis, we see that if we set up an $n$th-order realization for a transfer function $H(s)$ given as (7) but which is in fact "reducible," then we would *expect* the realization to be either noncontrollable or nonobservable or both.

Why do we say "expect," rather than just assert the above? The reason is that we have only proved this for diagonal realizations. However, we might try to generalize the above arguments as follows: Given an arbitrary realization $\{A, b, c\}$, find an invertible transformation matrix $T$ such that

$$A_d = T^{-1}AT, \qquad b_d = T^{-1}b, \qquad c_d = cT \tag{9}$$

Because the transformation is invertible, we would expect that all conclusions reached for the transformed realization would also be valid for the original realization. *In particular the reader can easily verify that observability and controllability are preserved under such similarity transformations.*

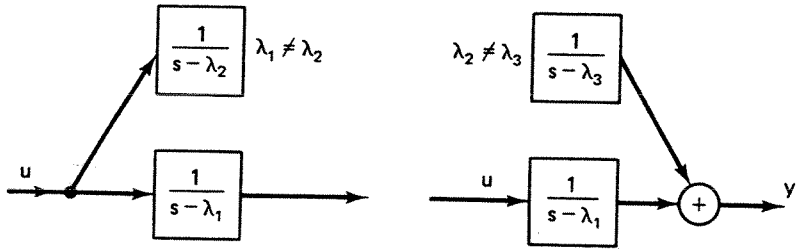
Unfortunately, while our conjectured result, that an $n$th-order realization of a transfer function $H(s)$ as in (7) will be both controllable and observable if and only if $b(s)/a(s)$ is *irreducible* [i.e., $b(s)$ and $a(s)$ have no common factors except constants], is indeed correct, the above arguments do not prove this, because it is not true that an arbitrary realization can be transformed to a diagonal realization, as we discussed in some detail in Example 2.2-A. Therefore, more complicated proofs must be used: Either one goes to the rather more complicated Jordan form (Sec. 13 in the Appendix), or one further examines the properties of the earlier canonical forms of Sec. 2.1.1. The latter route will be successfully pursued at the end of this section, though it is probably fair to say that the proofs given there were (as happens so often) stimulated by the fact that the diagonal form had suggested what had to be proved. Similarly, the reader may often find it useful to use the special (restricted) diagonal realization to probe problems for which the right questions or right answers do not seem obvious (see, e.g., the discussion of similarity transformations below).

**Potential General Validity of Results True for Diagonalizable Matrices.** It may be useful to make some remarks on when results for diagonalizable matrices might hold more generally. This will be the case when the results involve quantities that depend "continuously" on the elements of the matrices involved, i.e., if small changes in the elements produce only small changes in these quantities. For example, the coefficients of the characteristic polynomial, $a(s) = \det(sI - A)$, depend continuously on the values of the elements of $A$, and so do the roots of the characteristic polynomial (i.e., the eigenvalues of $A$). However, the rank of $A$ depends discontinuously on the values of $A$, as do the eigenvectors of $A$ [study, for example, a $2 \times 2$ lower triangular matrix with diagonal elements 1 and $\epsilon$ and $(2, 1)$-element 1]. Therefore results based on the coefficients of the characteristic polynomial or of the adjugate matrix generally extend to all matrices when they are true for diagonal

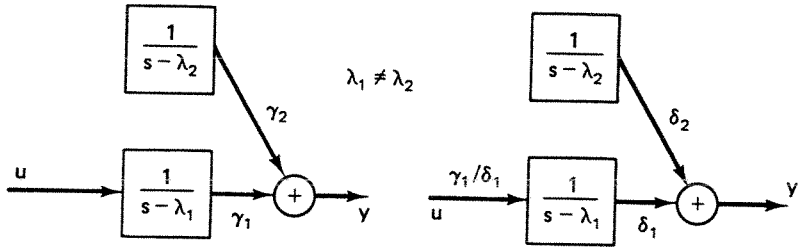
matrices. Examples are the Cayley-Hamilton theorem and the result noted above on the equivalence of joint controllability and observability to irreducibility of the transfer function. However, the result that a diagonalizable $n \times n$ matrix has $n$ linearly independent eigenvectors should not be expected to hold for arbitrary matrices (and it does not).

A firm mathematical foundation can be laid under these heuristic remarks by using a theorem of H. Weyl called the *principle of the irrelevance of algebraic inequalities*. This theorem states that if a polynomial equation in the entries of a matrix holds for all nonsingular matrices, then the equation holds for singular matrices as well. Several examples of the use of this theorem can be found in [26].

**Similarity Transformations Between Realizations.** It is obviously a useful thing to know when two realizations (of course, of a given transfer function) can be connected by a similarity transformation. The realizations must clearly have the same dimension, but is this enough? Using diagonal realizations shows that the answer is no—see Fig. 2.4-1(a). This example shows that




a. The two realizations have the same transfer function, but different eigenvalues and therefore cannot be similar (because similarity transformations do not change the eigenvalues or natural frequencies.




b. These two realizations can always be connected by a similarity transformation if $\lambda_1 \neq \lambda_2$ and $\gamma_i \neq 0$, $\delta_i \neq 0$, $i = 1, 2$. Note that both realizations are observable.

**Figure 2.4-1.** Diagonal realizations are useful in studying the problem of similar realizations.

a necessary condition is that the two realizations have the same natural frequencies. But the example in Fig. 2.4-1(a) shows that even this is not enough because, for example, there is no invertible matrix such that

$$1.\quad [1 \quad 0]T = [1 \quad 1] \qquad (cT = \bar{c})$$

and

$$2.\quad T^{-1}\begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}T = \begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix} \qquad (T^{-1}AT = \bar{A})$$

This is because, to satisfy the second condition, $T$ must be diagonal, and then the first condition can never be met (note that $\lambda_1 \neq \lambda_2$).

Reflection on this fact will lead us to consider the two observable realizations in Fig. 2.4-1(b). Now we can see that we can always achieve

$$[\gamma_1 \quad \gamma_2]T = [\delta_1 \quad \delta_2]$$

with a diagonal matrix, viz., $T = \text{diag}\,\{\delta_i/\lambda_i\}$. The input matrices are

$$b_1' = [1 \quad 0], \qquad b_2' = [\gamma_1/\delta_1 \quad 0]$$

which are related by $T^{-1}b_1 = b_2$.
Therefore the realizations in Fig. 2.4-1(b) are always similar.

These considerations lead us to conjecture the following general result, for not necessarily diagonal or even diagonalizable realizations.

**Lemma 2.4.1. Similarity of Scalar Realizations**

Any two realizations $\{A_i, b_i, c_i\}$ of the same order (and of the same transfer function) can be connected by a unique similarity transformation if

$$1.\qquad \det(sI - A_1) = \det(sI - A_2) \tag{10}$$

and

2. Both realizations are controllable or both are observable. The appropriate similarity transformation is given by

$$x_1(t) = Tx_2(t), \qquad T = \mathcal{C}(A_1, b_1)\mathcal{C}^{-1}(A_2, b_2) \tag{10a}$$

if both realizations are controllable, or by

$$x_1(t) = Tx_2(t), \qquad T = \mathcal{O}^{-1}(c_1, A_1)\mathcal{O}(c_2, A_2) \tag{10b}$$

if both realizations are observable. A general proof of this lemma will be given in Sec. 2.4.5 (Example 2.4-5). (In fact, Example 2.3-4 is already essentially a proof—show this.) ∎

The main aim of this discussion is to emphasize the important role of diagonal forms in testing and conjecturing results in system theory. (This is

apart from any potential advantages, e.g., in sensitivity, that might accrue from the actual use of diagonal realizations.) However, we must also emphasize the (true) *results* already guessed by this means, and we have summarized them in Fig. 2.4-2. There we have also shown a result that provides an important application of the concepts of controllability and observability—viz., that a controllable and observable realization $\{A, b, c\}$ is also *minimal* (or has *minimal* order). By this we mean that there can be no other realization $\{A_1, b_1, c_1\}$ with $A_1$ of smaller dimension than $A$. We recall, of course, that in saying "realization" we mean realization of a transfer function, and therefore what we are saying is that there can be no triple $\{A_1, b_1, c_1\}$ with $A_1$ "smaller" than $A$ in dimension and such that $c_1(sI - A_1)^{-1}b_1 = c(sI - A)^{-1}b$.

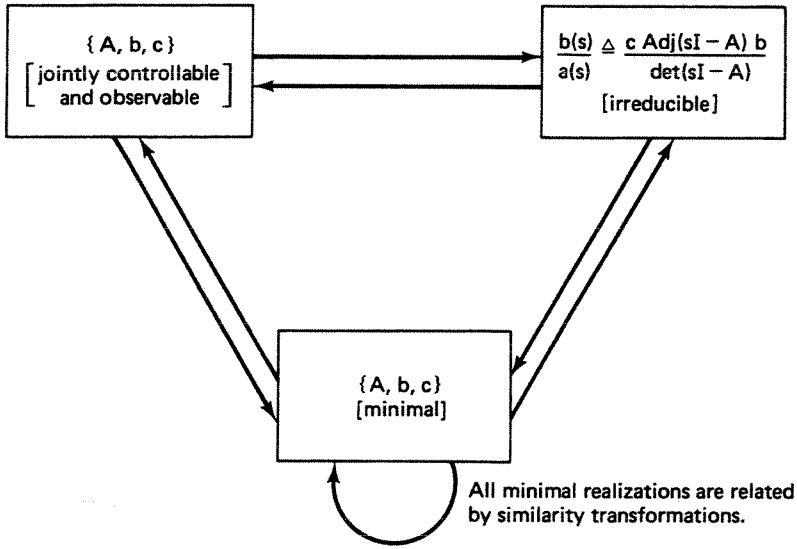



**Figure 2.4-2.** Some fundamental realizations.

Note that we now have an *explicit test* for minimality—it would be quite difficult to verify minimality by comparison with all other realizations! Moreover, in view of the relations previously indicated, another explicit test for the minimality of $\{A, b, c\}$ is to check that $a(s) \triangleq \det(sI - A)$ and $b(s) \triangleq a(s)c(sI - A)^{-1}b = c\,\text{Adj}\,(sI - A)b$ are *relatively prime*. Both these tests will be much used later—see especially Chapter 6 and also Exercise 2.4-5.

The results just stated on minimality can be proved quite directly from previous results, without assuming diagonal realizations, and we shall do so now. However, it will be valuable to gain as thorough an intuitive understanding as possible of the fundamental relations indicated in Fig. 2.4-2 before proceeding to the formal proofs.

**Proofs of Theorems Relating Controllability, Observability, Irreducibility, and Minimality.** Earlier in this section, we used the diagonal form to conjecture several important results on the characterization of simultaneously controllable and observable realizations. These results, which are summarized in Fig. 2.4-2, will be proven here. Proofs can be given in many different ways, depending on the order in which the different theorems are considered. The route we choose seems to be the most elementary in that, for example, it avoids the need for any results from linear algebra.

We begin with two preliminary results that are also of value in themselves.

**Lemma 2.4-2.**

If a transfer function

$$H(s) = \frac{b(s)}{a(s)} = \frac{b_1 s^{n-1} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n}$$

has one controllable and observable $n$th-order realization, then all $n$th-order realizations must also be controllable and observable.

**Proof.** We use the fact that if $\{A, b, c\}$ is a realization, then the Hankel matrix [cf. Eq. (2.2-49)]

$$M[1, n-1] = \begin{bmatrix} cb & cAb & \cdots & cA^{n-1}b \\ cAb & cA^2b & \cdots & cA^nb \\ \vdots & & & \\ cA^{n-1}b & \cdots & & cA^{2n-2}b \end{bmatrix}$$

depends only on the transfer function [$H(s) = \sum_1^\infty (cA^{i-1}b)s^{-i}$]. Furthermore, we can easily check that (cf. Exercise 2.2-14)

$$M[1, n-1] = \mathcal{O}(c, A)\mathcal{C}(A, b)$$

Therefore, if $\{A_1, b_1, c_1\}$ and $\{A_2, b_2, c_2\}$ are two $n$th-order realizations of $H(s)$,

$$\mathcal{O}(c_1, A_1)\mathcal{C}(A_1, b_1) = \mathcal{O}(c_2, A_2)\mathcal{C}(A_2, b_2)$$

Now since by hypothesis $\mathcal{O}(c_1, A_1)$ and $\mathcal{C}(A_1, b_1)$ are both nonsingular, so is their product, and thus also $\mathcal{O}(c_2, A_2)\mathcal{C}(A_2, b_2)$. But the product of two $n \times n$ matrices is nonsingular if and only if each matrix is nonsingular, so that $\mathcal{O}(c_2, A_2)$ and $\mathcal{C}(A_2, b_2)$ must be nonsingular for any $n$th-order realization $\{A_2, b_2, c_2\}$. ∎

In view of this result, it suffices to find one case in which joint controllability and observability can be assured. We shall see that the controller form will do.

**Lemma 2.4-3**

The $n$th-order controller form of $H(s) = b(s)/a(s)$, $n = \deg a(s)$, will be observable if and only if $b(s)$ and $a(s)$ are coprime, i.e., if $b(s)/a(s)$ is irreducible.

**Proof.** This result was proved in Example 2.3-5. ∎

We can now state the following theorem.

**Theorem 2.4-4**

A transfer function $H(s) = b(s)/a(s)$ is irreducible if and only if all $n$th-order realizations, $n = \deg a(s)$, are controllable and observable.

**Proof.** This is an immediate consequence of Lemmas 2.4-2 and 2.4-3. ∎

Next we recall that a realization $\{A, b, c\}$ is *minimal* if it has the smallest order (i.e., the smallest number of state variables) among all realizations having the same transfer function $c(sI - A)^{-1}b$. Clearly, there can be many minimal realizations, but they all share certain important properties.

**Theorem 2.4-5**

A realization $\{A, b, c\}$ is minimal if and only if $a(s) \triangleq \det(sI - A)$ and $b(s) \triangleq c \operatorname{Adj}(sI - A)b$ are relatively prime.

**Proof.** Let

$$H(s) = c(sI - A)^{-1}b = \frac{b(s)}{a(s)}$$

Suppose first that $\{A, b, c\}$ is minimal but that $b(s)/a(s)$ is not irreducible. Then using the *reduced* transfer function, we can obtain a realization with a lower-dimensional state vector. This is a contradiction. Similarly, to prove the converse, assume that $\{A, b, c\}$ is not minimal even though $b(s)/a(s)$ is irreducible. Then any minimal realization of $H(s)$ will have a transfer function with denominator of degree lower than $n$, the dimension of $A$. Therefore $b(s)/a(s)$ could not have been irreducible. ∎

We can combine Theorems 2.4-4 and 2.4-5 to obtain the following important result, which provides a direct test for minimality, instead of having to search over all possible realizations or check the nominal transfer function for irreducibility [17].

**Theorem 2.4-6**

A realization $\{A, b, c\}$ is minimal if and only if $\{A, b\}$ is controllable and $\{c, A\}$ is observable.

Finally we note that minimal realizations are very tightly related [17].

**Theorem 2.4-7**

Any two minimal realizations can be connected by a *unique* similarity transformation.

**Proof.** Minimal realizations satisfy the conditions of Lemma 2.4-1, and the theorem follows immediately. However, since a proof of Lemma 2.4-1 was deferred, we shall present a direct proof of the present theorem.

Since the two realizations are minimal, they are both observable and control-

lable. Therefore, if we define

$$T = \mathcal{O}^{-1}(c_1, A_1)\mathcal{O}(c_2, A_2)$$

from the previously noted identity (also easy to verify directly)

$$\mathcal{O}(c_1, A_1)\mathcal{C}(A_1, b_1) = \mathcal{O}(c_2, A_2)\mathcal{C}(A_2, b_2)$$

we can conclude that

$$T = \mathcal{C}(A_1, b_1)\mathcal{C}^{-1}(A_2, b_2)$$

Now we can easily see that

$$T^{-1}b_1 = b_2, \qquad c_1T = c_2$$

Finally, from the easily verified relation

$$\mathcal{O}(c_1, A_1)A_1\mathcal{C}(A_1, b_1) = \mathcal{O}(c_2, A_2)A_2\mathcal{C}(A_2, b_2)(= M[2, n-1]),$$

we obtain

$$A_2 = \mathcal{O}^{-1}(c_2, A_2)\mathcal{O}(c_1, A_1)A_1\mathcal{C}(A_1, b_1)\mathcal{C}^{-1}(A_2, b_2) = T^{-1}A_1T$$

Therefore this $T$ defines a similarity transformation relating the two realizations.

Moreover all matrices $\tilde{T}$ relating $\{A_i, b_i, c_i, i = 1, 2\}$ by similarity must be equal to $T$ as defined above. For if $\tilde{T}$ is any such transformation, we shall have

$$\mathcal{O}(c_1, A_1)(T - \tilde{T}) = 0$$

which implies that $T = \tilde{T}$, since $\mathcal{O}(c_1, A_1)$ is nonsingular. ∎

Example 2.4.6 in Sec. 2.4.5 will provide a nice illustration of the use of this theorem.

**Transformations Between the Canonical Realizations.** For various purposes, it will be useful to collect together here the special formulas that enable us to go between the four special realizations of Sec. 2.1.2 for an *irreducible* transfer function $H(s) = b(s)/a(s)$. We do this in Fig. 2.4-3, where the reader should be able to verify all the relations except that relating the observer form to the controller form—this will be treated in Sec. 2.4.4.

### *2.4.2 Standard Forms for Noncontrollable and/or Nonobservable Systems*

In previous sections we have discussed some canonical forms for controllable or observable realizations and have shown their usefulness in special applications. For other applications it will be useful to have standard forms in which noncontrollable and/or nonobservable systems can be represented. By using appropriate similarity transformations, we shall be able to find reali-

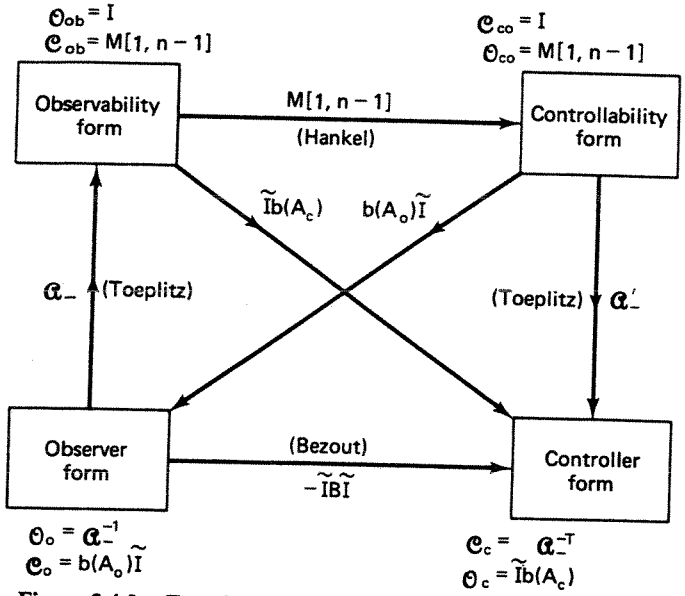



**Figure 2.4-3.** Transformations among four canonical realizations of an irreducible transfer function. Note that the (invertible) transformation matrix can be calculated as $T = \mathcal{O}_{old}^{-1}\mathcal{O}_{new} = \mathcal{C}_{old}\mathcal{C}_{new}^{-1}$. $\mathfrak{A}_-$ was defined in Eq. (2.3-6), $M[1, n-1]$ in Eq. (2.2-49), and $B$ will be defined in Sec. 2.4-4.

zations in which the noncontrollable and/or nonobservable state variables can be clearly separated out.

**Representation of Noncontrollable Realizations.** If we have a noncontrollable realization $\dot{x}(t) = Ax(t) + bu(t)$ and $A$ is diagonal with distinct eigenvalues, then we can readily say whether or not a particular state variable is controllable: It will be so if and only if the corresponding element of the input vector $b$ is not zero. However, it is clear that this simple separation will be lost after any change of variables (similarity transformation) that does not again yield a diagonal realization—every new state variable may now have both a controllable and noncontrollable part. In this section we shall show how any given (noncontrollable) realization $\{A, b\}$, with $A$ not necessarily diagonalizable, can be transformed into another realization where the controllable and noncontrollable state variables can be clearly identified.

Let $\{A, b, c\}$ be such that

$$\text{rank } \mathcal{C}(A, b) = r < n$$

Then we shall show that a transformation matrix $T$ can always be found such

that the realization

$$\{\bar{A} = T^{-1}AT,\ \bar{b} = T^{-1}b,\ \bar{c} = cT\}$$

has the form

$$\bar{A} = \left[\begin{array}{c|c} \bar{A}_c & \bar{A}_{c\bar{c}} \\ \hline 0 & \bar{A}_{\bar{c}} \end{array}\right]\begin{matrix} r \\ n-r \end{matrix}, \qquad \bar{b} = \left[\begin{array}{c} \bar{b}_c \\ \hline 0 \end{array}\right], \qquad \bar{c} = [\bar{c}_c \quad \bar{c}_{\bar{c}}] \tag{11}$$

Any realization in this form has the important properties that

1. The $r \times r$ subsystem $\{\bar{A}_c, \bar{b}_c, \bar{c}_c\}$ is controllable.
2. $\bar{c}(sI - \bar{A})^{-1}\bar{b} = \bar{c}_c(sI - \bar{A}_c)^{-1}\bar{b}_c$; i.e., the subsystem has the same transfer function as the original system.

If the state variables $\bar{x}$ are correspondingly partitioned as $x' = [\bar{x}_c' \quad \bar{x}_{\bar{c}}']'$, then the variables $\bar{x}_c$ can be said to be controllable and the variables $\bar{x}_{\bar{c}}$ noncontrollable. The realization $\{\bar{A}, \bar{b}, \bar{c}\}$ can be depicted as in Fig. 2.4-4, showing graphically the separation of states.

The above statements can be proved in several different ways. Here let us first check that for a system as in (11)†

$$\begin{aligned} \bar{c}(sI - \bar{A})^{-1}\bar{b} &= \bar{c}\begin{bmatrix} (sI - \bar{A}_c)^{-1} & \text{xx} \\ 0 & (sI - \bar{A}_{\bar{c}})^{-1} \end{bmatrix}\begin{bmatrix} \bar{b}_c \\ 0 \end{bmatrix} \\ &= [\bar{c}_c \quad \bar{c}_{\bar{c}}]\begin{bmatrix} (sI - \bar{A}_c)^{-1}\bar{b}_c \\ 0 \end{bmatrix} = \bar{c}_c(sI - \bar{A}_c)^{-1}\bar{b}_c \end{aligned}$$

[The xx denotes entries whose exact values are unimportant here, though it is easy to show that $\text{xx} = -(sI - \bar{A}_c)^{-1}A_{c\bar{c}}(sI - \bar{A}_{\bar{c}})^{-1}$.] Next we note that

$$\mathcal{C}(\bar{A}, \bar{b}) = \left[\begin{array}{cccc} \bar{b}_c & \bar{A}_c\bar{b}_c & \cdots & \bar{A}_c^{n-1}\bar{b}_c \\ \hline 0 & 0 & & 0 \end{array}\right]\begin{matrix} r \\ n-r \end{matrix} \tag{12}$$

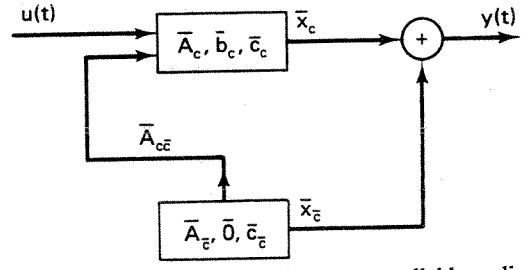



**Figure 2.4-4.** Decomposition of a noncontrollable realization.

†Another proof is to note that with zero initial conditions, as required when computing transfer functions, the states $x_{\bar{c}}$ will be identically zero.

Now since $\mathcal{C}(\bar{A}, \bar{b}) = T^{-1}\mathcal{C}(A, b)$, $\mathcal{C}(\bar{A}, \bar{b})$ has rank $r$, and therefore it can have only $r$ linearly independent rows and $r$ linearly independent columns. We shall show that if we start checking the columns from left to right, the first $r$ columns must be linearly independent. For suppose $\bar{A}_c^k\bar{b}_c$ is linearly dependent on $\{\bar{A}_c^i\bar{b}_c\}$, $i < k\}$. Then clearly $\bar{A}_c^{k+1}\bar{b}_c = \bar{A}_c\bar{A}_c^k\bar{b}_c$ will also be dependent on the set $\{\bar{A}_c^i\bar{b}_c, i < k\}$. In other words, in searching from left to right in (12), once we find a dependent vector, then all subsequent ones must be so too. But since the rank is $r$, the first $r$ columns $\{\bar{b}_c \cdots \bar{A}_c^{r-1}\bar{b}_c\}$ (and only these) must be linearly independent. This proves statement 1.

The above discussion also suggests a way of finding a suitable transformation matrix $T$. We require

$$T\mathcal{C}(\bar{A}, \bar{b}) = \mathcal{C}(A, b) \tag{13}$$

Partitioning these matrices appropriately, we get

$$[T_1 \quad T_2]\begin{bmatrix}\mathcal{C}(\bar{A}_c, \bar{b}_c) & \text{xx} \\ 0 & 0\end{bmatrix} = [b \quad Ab \quad \cdots \quad A^{r-1}b \quad \text{xx}] \tag{14}$$

from which

$$T_1 = [b \quad Ab \quad \cdots \quad A^{r-1}b]\mathcal{C}^{-1}(\bar{A}_c, \bar{b}_c) \tag{15}$$

If, in particular, we wish the new controllable subsystem to be in controllability form, then $\mathcal{C}(\bar{A}_c, \bar{b}_c) = I$, and the first $r$ (mutually independent) columns of $T$ are given by $T_1 = [b \; Ab \cdots A^{r-1}b]$. The remaining columns, those of $T_2$, need only to be linearly independent of each other and of the columns of $T_1$ [and hence of $\mathcal{C}(A, b)$] in order that $T$ be nonsingular. A particular choice for these columns is a set of $(n - r)$ vectors that is *orthogonal* to (and not merely independent of) the columns of $\mathcal{C}(A, b)$.

Examples of some different transformations to standard noncontrollable form are provided in Examples 2.4-1 and 2.4-2—the point is that instead of using a fixed method, the context should be exploited to find the appropriate transformation. For large systems, however, more systematic methods are desirable, and some efficient ones have been proposed by Rosenbrock [22, pp. 80–84] (see also Mayne [27] and Daly [28]). We stress also that in many analytical problems, explicit determination of the transformation is not necessary.

**Representation of Nonobservable Realizations.** It should be clear that similar (in fact, dual) statements can be made about nonobservable realizations. Thus if

$$\mathcal{O}(c, A) \text{ has rank } r < n$$

we can find a nonsingular matrix $T$ such that

$$\bar{A} = T^{-1}AT, \qquad \bar{b} = T^{-1}b, \qquad \bar{c} = cT$$

have the form

$$\bar{A} = \left[\begin{array}{c|c} \bar{A}_o & 0 \\ \hline \bar{A}_{\bar{o}o} & \bar{A}_{\bar{o}} \end{array}\right] \begin{matrix} r \\ n-r \end{matrix}, \qquad \bar{c}' = \left[\begin{array}{c} \bar{c}_o' \\ \hline 0 \end{array}\right], \qquad \bar{b} = \left[\begin{array}{c} \bar{b}_o \\ \hline \bar{b}_{\bar{o}} \end{array}\right] \tag{16}$$

and

$$\{\bar{c}_o, \bar{A}_o\} \text{ is observable}$$
$$c(sI - A)^{-1}b = \bar{c}_o(sI - \bar{A}_0)^{-1}\bar{b}_o$$

The separation into observable and nonobservable parts can be depicted as in Figure 2.4-5.

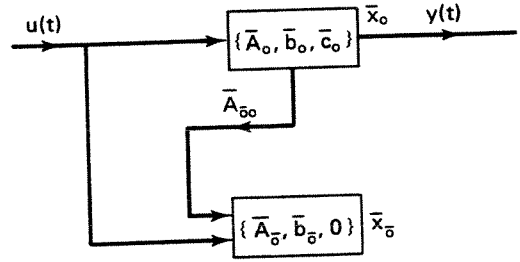



**Figure 2.4-5.** Decomposition of nonobservable realizations.

We can combine the above results for noncontrollable or nonobservable systems to obtain a decomposition of arbitrary realizations. Before stating the result, we point out that its content should first be anticipated by considering diagonal realizations, for which the decomposition shown in Fig. 2.4-6 is obvious. For nondiagonal realizations, there can be certain forms of state feedback between the blocks, as shown by the dashed lines in Fig. 2.4-6. The general theorem will be given at the end of this section, but we first interject another useful remark.

**Obtaining Minimal Realizations.** The above results suggest one way of obtaining a minimal realization from a given realization or transfer function. Thus, given a transfer function, we can always, for example, write down a controller-form realization by inspection and then separate out any nonobservable part if the realization is not minimal.

**General Decomposition Theorem.** We can always find an invertible state transformation that will allow us to rewrite the state equations

$$\dot{x} = Ax + bu, \qquad y = cx$$

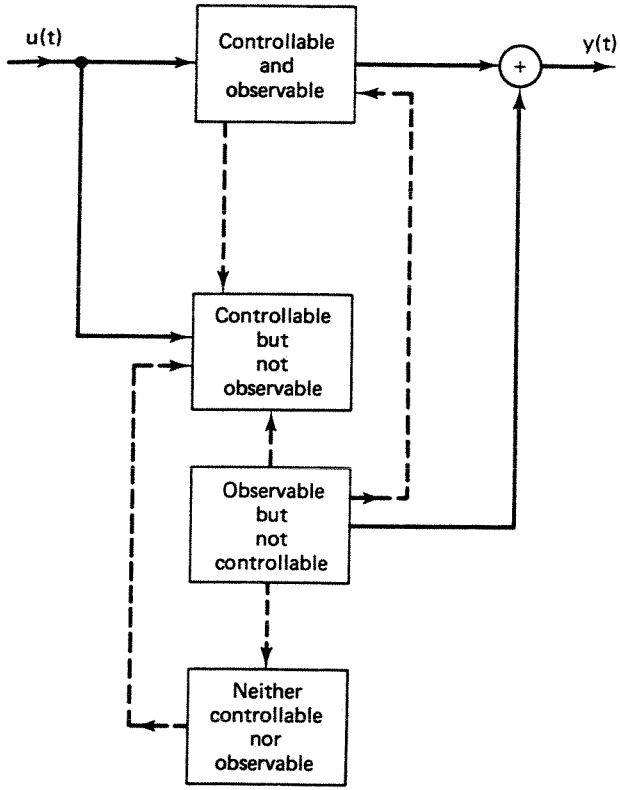



**Figure 2.4-6.** Canonical decomposition of diagonal realizations.

in the form

$$\dot{\bar{x}} = \bar{A}\bar{x} + \bar{b}u, \qquad y = \bar{c}\bar{x}$$

where

$$\bar{x}' = [\bar{x}'_{c,o} \quad \bar{x}'_{c,\bar{o}} \quad \bar{x}'_{\bar{c},o} \quad \bar{x}'_{\bar{c},\bar{o}}] \tag{17}$$

$$\bar{A} = \left[\begin{array}{c|c|c|c} \bar{A}_{c,o} & 0 & \bar{A}_{1,3} & 0 \\ \hline \bar{A}_{2,1} & \bar{A}_{c,\bar{o}} & \bar{A}_{2,3} & \bar{A}_{2,4} \\ \hline 0 & 0 & \bar{A}_{\bar{c},o} & 0 \\ \hline 0 & 0 & \bar{A}_{4,3} & \bar{A}_{\bar{c},\bar{o}} \end{array}\right], \qquad \bar{b} = \left[\begin{array}{c} \bar{b}_{c,o} \\ \hline \bar{b}_{c,\bar{o}} \\ \hline 0 \\ \hline 0 \end{array}\right] \tag{18}$$

$$\bar{c} = [\bar{c}_{c,o} \quad 0 \quad \bar{c}_{\bar{c},o} \quad 0]$$

and

1. The subsystem

$$\{\bar{A}_{c,o}, \bar{b}_{c,o}, \bar{c}_{c,o}\}$$

is controllable and observable, and

$$\bar{c}(sI - \bar{A})^{-1}\bar{b} = \bar{c}_{c,o}(sI - \bar{A}_{c,o})^{-1}\bar{b}_{c,o}$$

2. The subsystem

$$\left\{\left[\begin{array}{c|c} \bar{A}_{c,o} & 0 \\ \hline \bar{A}_{2,1} & \bar{A}_{c,\bar{o}} \end{array}\right], \left[\begin{array}{c} \bar{b}_{c,o} \\ \hline \bar{b}_{c,\bar{o}} \end{array}\right], [\bar{c}_{c,o} \quad 0]\right\}$$

is controllable;

3. The subsystem

$$\left\{\left[\begin{array}{c|c} \bar{A}_{c,o} & \bar{A}_{1,3} \\ \hline 0 & \bar{A}_{\bar{c},o} \end{array}\right], \left[\begin{array}{c} \bar{b}_{c,o} \\ \hline 0 \end{array}\right], [\bar{c}_{c,o} \quad \bar{c}_{\bar{c},o}]\right\}$$

is observable; and

4. The subsystem

$$\{\bar{A}_{\bar{c},\bar{o}}, [0], [0]\}$$

is neither controllable nor observable.

The actual values of the possible nonzero matrices $\{\bar{A}_{c,o}, \bar{A}_{2,1}, \ldots, \bar{b}_{c,o}, \ldots\}$ are not all unique, because they depend on the particular coordinate transformations used to identify the noncontrollable and nonobservable states, but the dimensions of the various blocks are unique (why?). Note, of course, that the subsystem in 1 has the same transfer function as the original system (prove this also by direct evaluation).

This general decomposition theorem, which was first enunciated by Gilbert [16] and Kalman [17], can be proved by combining the ideas used to analyze noncontrollable (or nonobservable) realizations—we shall omit the detailed construction (but try Exercise 2.4-16 for some physical insight).

**Reprise.** Given problems involving noncontrollable and/or nonobservable realizations, it is generally a good idea to first assume, as we can do without loss of generality, that they are in the standard forms described in this section—this often simplifies later analyses. It is important to realize that this assumption can be made without having to know a transformation matrix $T$ for taking a given realization to one or the other of these standard forms. Readers should familiarize themselves with the correspondence between Eq. (2.4-11) and Fig. 2.4-4, and Eq. (2.4-16) and Fig. 2.4-5, and should have a good understanding of the reasons for their special names—this again can be achieved without any knowledge of the explicit transformation to these forms.

### 2.4.3 The Popov-Belevitch-Hautus Tests for Controllability and Observability

The standard forms introduced in the previous section allow us to establish some extremely powerful criteria for testing the controllability and observability of realizations, especially when they arise by combining and arranging subsystems in special ways required by particular problems. These tests are especially useful for theoretical analysis and also in numerical problems whenever determination of matrix eigenvalues and eigenvectors is computationally feasible.

These criteria were introduced independently by several people, especially Popov [29], Belevitch [30, p. 413], Hautus [31], Rosenbrock [22], Hahn (cf. [18, p. 27]), Ford and Johnson ([32] and [33]), and no doubt others. In particular, for the special case where the $A$ matrix is diagonalizable, the test was first given by Gilbert [16]. Since Popov and Belevitch were perhaps the first to give a general statement and Hautus was perhaps the first to note their wide applicability, we shall call them the PBH tests. Several examples of the use of these tests will be given in this book.

**Theorem 2.4-8. PBH Eigenvector Tests**

1. A pair $\{A, b\}$ will be noncontrollable if and only if there exists a row vector $q \not\equiv 0$ such that

$$qA = \lambda q, \qquad qb = 0 \tag{19}$$

In other words, $\{A, b\}$ will be controllable if and only if there is no row (or left) eigenvector of $A$ that is orthogonal to $b$.

2. A pair $\{c, A\}$ will be nonobservable if and only if there exists a (column) vector $p \not\equiv 0$ such that

$$Ap = \lambda p, \qquad cp = 0 \tag{20}$$

in other words, if and only if some eigenvector of $A$ is orthogonal to $c$.

**Proof.**

1. *The "if" part.* If there exists $q \not\equiv 0$ such that

$$qA = \lambda q, \qquad qb = 0$$

then

$$qAb = \lambda qb = 0$$

and

$$qA^2b = \lambda qAb = 0$$

and so on, until we see that

$$q\mathcal{C}(A, b) = q[b \quad Ab \quad \cdots \quad A^{n-1}b] = 0$$

which means that the controllability matrix is singular, i.e., $\{A, b\}$ is not controllable.

*The "only if" part.* We have to show that $\{A, b\}$ noncontrollable implies the existence of a vector $q$ as in (19). Now whenever we have to show things about noncontrollable realizations, it is generally a good idea to begin by assuming that the realization has been put into the standard noncontrollable form (11) described in the previous section,

$$A = \begin{bmatrix} A_c & A_{c\bar{c}} \\ 0 & A_{\bar{c}} \end{bmatrix} \begin{matrix} r \\ n-r \end{matrix}, \qquad b = \begin{bmatrix} b_c \\ 0 \end{bmatrix}$$
$$r \quad n-r$$

where $r = \text{rank } \mathcal{C}(A, b) < n$. Now it is clear that a particular row vector $q$ that is orthogonal to $b$ has the form $q = [0 \mid z]$, and it is perhaps not hard to guess that we should choose $z$ as an eigenvector of $A_{\bar{c}}$,

$$zA_{\bar{c}} = \lambda z$$

for then

$$qA = [0 \quad z]A = [0 \quad \lambda z] = \lambda q$$

Therefore we have shown how to find a row vector $q$ satisfying (19), and this completes the proof of part 1.

2. This result is the "dual" of the one in part 1. Let us go through the details of showing this.

We first note that $\{c, A\}$ will be observable if and only if $\{A', c'\}$ is controllable. But, by part 1, this will be true if and only if there exists no row vector $p'$ such that

$$p'A' = \lambda p' \quad \text{and} \quad p'c' = 0$$

i.e., such that

$$Ap = \lambda p \quad \text{and} \quad cp = 0$$

which is the criterion stated in part 2. ∎

Another form of these tests is often useful.

**Theorem 2.4-9. PBH Rank Tests**

1. A pair $\{A, b\}$ will be controllable if and only if

$$\text{rank } [sI - A \quad b] = n \qquad \text{for } \textit{all } s \tag{21}$$

2. A pair $\{c, A\}$ will be observable if and only if

$$\text{rank} \begin{bmatrix} c \\ sI - A \end{bmatrix} = n \qquad \text{for } \textit{all } s \tag{22}$$

Here, of course, $n$ is the size of $A$. Note also that these conditions will clearly be met for all $s$ that are not eigenvalues of $A$, because $\det (sI - A) \neq 0$ for such $s$; the point

of the theorem is that the rank must be $n$ even when $s$ is an eigenvalue of $A$. We shall see in Sec. 6.3 that conditions (21)–(22) are ways of stating that the matrix polynomials $\{sI - A, b\}$ and $\{c, sI - A\}$ obey certain *relative primeness* (or *co-primeness*) conditions.

**Proof.**

1. If $[sI - A \;\; b]$ has rank $n$, there cannot be a nonzero row vector $q$ such that

$$q[sI - A \quad b] = 0 \qquad \text{for any } s$$

i.e., such that

$$qb = 0 \quad \text{and} \quad qA = sq$$

But then by Theorem 2.4-8 $\{A, b\}$ must be controllable. The converse follows easily by reversing the above arguments.

2. This can be proved in the same way or via duality. ∎

As noted before, several examples and applications of these tests will appear throughout this book. In fact, when faced with problems of checking for controllability and/or observability, it is a good heuristic rule to first try to apply the PBH tests.

**Diagonalizable Realizations** (Gilbert [16]). The special case of diagonalizable realizations provides a good insight into the nature of the PBH tests.

Thus, suppose we have a system

$$\dot{x}(t) = Ax(t) + bu(t)$$

where we can write

$$T^{-1}AT = \Lambda = \text{diag}\,\{\lambda_1, \ldots, \lambda_n\}$$

with

$$T = [p_1 \ldots p_n] \quad \text{and} \quad Ap_i = \lambda_i p_i$$

In other words, $\{\lambda_i\}$ and $\{p_i\}$ are the eigenvalues and eigenvectors of $A$. Also let

$$q_i = \text{the } i\text{th row of } T^{-1}$$

It is easy to see that

$$q_i p_j = \delta_{ij}$$

and that the $\{q_i\}$ are left eigenvectors of $A$; i.e., $q_i A = \lambda_i q_i$. Now the realization $\{A, b\}$ can be converted to the diagonal realization

$$\dot{\chi}(t) = \Lambda\chi(t) + \beta u(t), \qquad \beta = T^{-1}b$$

Then as we noted in Sec. 2.4.1, and as is directly evident, this realization will

be noncontrollable if any of the components of the input vector is zero. But

$$\beta_i = q_i b, \qquad i = 1, \ldots, n$$

so that we shall have noncontrollability if some left eigenvector of $A$ is orthogonal to $b$, as claimed by the PBH criterion.

Now we also know that even if all the $\{\beta_i\}$ are nonzero, we can lose controllability if the eigenvalues are repeated, e.g., if $\lambda_1 = \lambda_2$ (cf. Sec. 2.4.1). What happens here? We have, say,

$$q_1 b = \beta_1 \neq 0, \qquad q_2 b = \beta_2 \neq 0$$

with $q_1$ and $q_2$ linearly independent.

But because $q_1$ and $q_2$ are associated with the same eigenvalue $\lambda$, then clearly any linear combination of $q_1$ and $q_2$ will also be a left eigenvector associated with $\lambda$. Therefore we can readily find an eigenvector orthogonal to $b$, in fact, $(\beta_2 q_1 - \beta_1 q_2)b = 0$. Therefore for diagonalizable realizations, the PBH eigenvector test is relatively obvious.

For nondiagonalizable realizations, we can prove the test by using the Jordan canonical form, but the proof given above (Theorem 2.4-8) is more self-contained. We should remember, however, from the above argument that whenever more than one independent eigenvector can be associated with a single eigenvalue, we shall lose controllability (and observability).

**Example 2.4-A.**

Show that any pair $\{A_1, b\}$, where

$$A_1 = \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$$

is always noncontrollable.

**Solution.** There are three eigenvalues $\lambda$. As for eigenvectors, the equations

$$[\alpha_1 \quad \alpha_2 \quad \alpha_3]A = \lambda[\alpha_1 \quad \alpha_2 \quad \alpha_3]$$

yield

$$\alpha_1 \lambda = \lambda \alpha_1, \qquad \alpha_2 \lambda = \lambda \alpha_2, \qquad \alpha_2 + \lambda \alpha_3 = \lambda \alpha_3$$

which can be satisfied by

$$[1 \quad 0 \quad 1] \text{ or } [0 \quad 0 \quad 1]$$

and any linear combinations thereof. Clearly, we can always find a combination that will be orthogonal to any 3-vector $b$.

An alternative solution is provided by the PBH rank test, as we ask the reader to check. ∎

**Example 2.4-B.**

Find conditions on $c$ so that $\{c, A_2\}$ will always be observable when

$$A_2 = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}$$

**Solution.** By the PBH rank test, we must check the rank of

$$\begin{bmatrix} c_1 & c_2 & c_3 \\ s-\lambda & -1 & 0 \\ 0 & s-\lambda & -1 \\ 0 & 0 & s-\lambda \end{bmatrix}$$

When $s = \lambda$, we see that the rank will be 3 if and only if $c_1 \neq 0$. ∎

We note that one of the PBH tests might be simpler than the other in any particular problem.

**Uncontrollable and Controllable Modes and Eigenvalues.** The reader should know from earlier experience with linear systems that the eigenvalues of the matrix $A$ determine the so-called *modes* of a realization $\{A, b, c\}$ (cf. Sec. 2.5.2 for a more formal discussion). The PBH tests allow us to say that the modes associated with the eigenvalue $\lambda$ are *uncontrollable* if and only if some associated left eigenvector is orthogonal to $b$. (Otherwise the associated modes may be called *controllable*.)

The significance of this definition is very clear for diagonalizable realizations, where there will be no input to the subsystem associated with this mode. In Chapter 3, we shall see also that an uncontrollable mode corresponds to an eigenvalue of $A$ that cannot be changed to some other value by using state feedback. In this sense we can also talk about controllable and uncontrollable eigenvalues (or natural frequencies).

It is worth noting that every mode and every eigenvalue of a realization can be classed as completely controllable or completely uncontrollable, while this is not possible for any *state variable* in a realization, unless the realization happens to be in the standard form of Sec. 2.4.2.

It also should go without saying that all the statements we have made above about controllability properties go over, with obvious changes, to observability.

**Reprise.** The PBH eigenvector and rank tests for controllability and observability are very useful. Their intuitive meaning can be brought out by recalling how (diagonalizable) realizations are transformed to diagonal form. The tests also allow us to identify controllable and/or observable modes and natural frequencies (eigenvalues).

### *2.4.4 Some Tests for Relatively Prime Polynomials

In Sec. 2.4.1 we noted that the minimality of a realization $\{A, b, c\}$ could be determined either by checking the nonsingularity of the observability and controllability matrices $\mathcal{O}(c, A)$ and $\mathcal{C}(A, b)$ or by checking the relative primeness of $a(z) = \det(zI - A)$ and $b(z) = c \operatorname{Adj}(zI - A)b$. Now while controllability and observability may be relatively new concepts, that of relatively prime polynomials is a very old one, and there exist many tests for it. We shall give three of the better known ones, due, respectively, to Sylvester [34], Bezout [35], and MacDuffee [36]. There should of course be intimate relations between these criteria and the notions of controllability and observability because both provide tests for minimality; we shall in fact display some of these.

The question of checking the relative primeness of two polynomials $a(z)$ and $b(z)$ can be regarded as a special case of the problem of finding the greatest common divisor (gcd) of two polynomials. This can be done by using the celebrated Euclidean algorithm, which Knuth [37, p. 294 ff.] has remarked is the oldest nontrivial algorithm that has survived to the present day.

The Euclidean algorithm is based on the fact that given two polynomials

$$a(z) = a_0 z^n + a_1 z^{n-1} + \cdots + a_n, \qquad a_0 \neq 0 \tag{23a}$$

$$b(z) = b_0 z^m + \cdots + b_m, \qquad m \leq n \tag{23b}$$

there exists a unique quotient polynomial $q(z)$ and a unique remainder polynomial $r(z)$ such that

$$a(z) = q(z)b(z) + r(z), \qquad \deg r(z) < \deg b(z)$$

Suppose that

$$\deg b(z) \leq \deg a(z)$$

Then by successive use of the above polynomial division formula we can write

$$\begin{aligned}
a(z) &= q_1(z)b(z) + r_1(z), && \deg r_1 < \deg b \\
b(z) &= q_2(z)r_1(z) + r_2(z), && \deg r_2 < \deg r_1 \\
&\vdots \\
r_{p-3}(z) &= q_{p-1}(z)r_{p-2}(z) + r_{p-1}(z), && \deg r_{p-1} < \deg r_{p-2} \\
r_{p-2}(z) &= q_p(z)r_{p-1}(z) + 0
\end{aligned}$$

The algorithm stops when the remainder $r_p(z) = 0$ and then the gcd of $a(z)$ and $b(z)$ is $r_{p-1}(z)$.

The reader should check this for himself with simple numerical examples. A formal proof is not very difficult. Note first that the last equation shows that $r_{p-1}(z)$ divides $r_{p-2}(z)$ (written $r_{p-1} \mid r_{p-2}$). The last-but-one equation can be written

$$r_{p-3} = q_{p-1}(q_p r_{p-1}) + r_{p-1}$$

so we see that $r_{p-1} | r_{p-3}$. Proceeding in this way gives

$$r_{p-1}(z) | b(z), \qquad r_{p-1}(z) | a(z)$$

To show that this is the greatest common divisor, we have to show that any other divisor of $a(z)$ and $b(z)$ also divides $r_{p-1}(z)$. For this, we first note that successive substitution in the equations

$$r_{p-1} = r_{p-3} - q_{p-1}r_{p-2}, \qquad r_{p-2} = r_{p-4} - q_{p-2}r_{p-3}, \qquad \ldots$$

shows that there exist two polynomials $x(z)$ and $y(z)$ such that

$$r_{p-1}(z) = x(z)a(z) + y(z)b(z) \tag{24}$$

Now it is clear that any common factor of $a(z)$ and $b(z)$ will also be a factor of $r_{p-1}(z)$, which is thereby proved to be the gcd. The gcd is only unique up to a constant, but it can be made unique by requiring it to be monic, i.e., to have highest-order coefficient equal to unity.

The algebra involved in computing the gcd can be organized in several different ways, which we shall not pursue here. (One of the most efficient algorithms can be found in [38, Sec. 8.4]). Our main interest is in the case where the polynomials are relatively prime (or coprime, as we shall often say). For this case, we can state the following slight strengthening of the result (23).

**Lemma 2.4-10.**

The polynomials $a(z)$ and $b(z)$ in (24) will be coprime if and only if there exist two polynomials $\tilde{x}(z)$ and $\tilde{y}(z)$ such that

$$\tilde{x}(z)a(z) + \tilde{y}(z)b(z) = 1 \tag{25a}$$

and

$$\deg \tilde{x}(z) < m, \qquad \deg \tilde{y}(z) < n \tag{25b}$$

Moreover, if $a(z)$ and $b(z)$ are coprime, then $\tilde{x}(z)$ and $\tilde{y}(z)$ will be unique.

**Proof.** Everything has been proved already except for the possibility of the degree constraints on $x(z)$ and $y(z)$. That is, by (24) we know that $a(z)$ and $b(z)$ are coprime if and only if there exist $\{x(z), y(z)\}$ such that $x(z)a(z) + y(z)b(z) = 1$. But now let $y(z) = q(z)a(z) + r(z)$, $\deg r < n$, and define

$$\tilde{y}(z) = r(z) = y(z) - q(z)a(z), \qquad \tilde{x}(z) = x(z) + q(z)b(z)$$

Then

$$\begin{aligned}\tilde{x}(z)a(z) + \tilde{y}(z)b(z) &= x(z)a(z) + y(z)b(z) + \{q(z)b(z)a(z) - q(z)a(z)b(z)\} \\ &= 1 + 0 = 1\end{aligned}$$

Also note that

$$\deg [\tilde{x}(z)a(z)] = \deg [1 - \tilde{y}(z)b(z)] < m + n$$

so that

$$\deg \tilde{x}(z) < m + n - \deg a(z) = m$$

To prove uniqueness, suppose that $\{a(z), b(z)\}$ are coprime but that there exist two sets $\{\tilde{x}_i(z), \tilde{y}_i(z), i = 1, 2\}$ satisfying (25). Then

$$[\tilde{x}_1(z) - \tilde{x}_2(z)]a(z) + [\tilde{y}_1(z) - \tilde{y}_2(z)]b(z) = 0$$

which implies that

$$\frac{a(z)}{b(z)} = \frac{\tilde{y}_2(z) - \tilde{y}_1(z)}{\tilde{x}_1(z) - \tilde{x}_2(z)}$$

But since $\deg [\tilde{y}_2(z) - \tilde{y}_1(z)] < m$ and $\deg [\tilde{x}_1(z) - \tilde{x}_2(z)] < n$, this means that $b(z)$ and $a(z)$ must have a common factor, contradicting our initial assumption. Hence. . . . ∎

The result of Lemma 2.4-10 will be used in Sec. 4.5.3 on compensator design. In that connection, we should note that there are several other criteria and tests for coprimeness, and here we shall note three of the best known ones.

**Sylvester's Resultant (1840).** Given polynomials $\{a(z), b(z)\}$ as in (25), define a so-called *Sylvester matrix* (shown for $n = 3, m = 2$)

$$\tilde{S}(a, b) = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 & 0 \\ 0 & a_0 & a_1 & a_2 & a_3 \\ 0 & 0 & b_0 & b_1 & b_2 \\ 0 & b_0 & b_1 & b_2 & 0 \\ b_0 & b_1 & b_2 & 0 & 0 \end{bmatrix} \tag{26}$$

Sylvester showed that

$$\{a(z), b(z)\} \text{ are coprime if and only if } \det \tilde{S}(a, b) \neq 0 \tag{27}$$

The determinant of $\tilde{S}(a, b)$ is known as the *Sylvester resultant*. We note that since elementary row and column operations leave the determinant unchanged, we could rearrange $\tilde{S}(a, b)$ in many ways. In (26), we have shown a form favored by Jury ([39] and [40]), which displays a "left triangle" of zeros. Another useful form (especially when $m = n$, as shown below for $m = 3 = n$) is

$$S(a, b) = \left[\begin{array}{ccc:ccc} a_0 & 0 & 0 & b_0 & 0 & 0 \\ a_1 & a_0 & 0 & b_1 & b_0 & 0 \\ a_2 & a_1 & a_0 & b_2 & b_1 & b_0 \\ \hdashline a_3 & a_2 & a_1 & b_3 & b_2 & b_1 \\ 0 & a_3 & a_2 & 0 & b_3 & b_2 \\ 0 & 0 & a_3 & 0 & 0 & b_3 \end{array}\right] = \left[\begin{array}{c:c} \mathfrak{A}_- & \mathfrak{B}_- \\ \hdashline \mathfrak{A}_+ & \mathfrak{B}_+ \end{array}\right] \tag{28}$$

where the notation $\mathfrak{A}_-$, $\mathfrak{B}_-$ agrees with definitions introduced earlier in Eqs. (2.3-36 and 2.3-37). The Sylvester test (27) can be established in many ways, one of which is described in Exercise 2.4-18.

**MacDuffee's Resultant (1950).** MacDuffee [36] showed that polynomials $b(z)$ and $a(z)$ will be relatively prime if and only if

$$\det [b(A_c)] \neq 0 \tag{29a}$$

or equivalently

$$\det [a(B_c)] \neq 0 \tag{29b}$$

where $A_c$ and $B_c$ are companion matrices of $a(z)$ and $b(z)$, with the coefficients of $a(z)$ and $b(z)$ in the top rows. The determinants in (29) are often known as *MacDuffee resultants;* whether (29a) or (29b) is used will depend on whether $m$ is less than $n$ or greater than $n$.

A proof follows easily from the fact that if $\{\lambda_1, \ldots, \lambda_n\}$ are the eigenvalues of $A_c$, then the eigenvalues of the matrix polynomial $b(A_c)$ are (cf. Exercise A.36) $\{b(\lambda_1), \ldots, b(\lambda_n)\}$. Therefore

$$\det b(A_c) = (-1)^n \det [zI - b(A_c)]\Big|_{z=0} = \prod_1^n b(\lambda_i)$$

and clearly $\det b(A_c)$ will be zero if and only if at least one of the $\{b(\lambda_i)\}$ is zero, i.e., if and only if $b(z)$ and $a(z)$ have at least one common factor, and similarly for the test involving $a(B_c)$.

**Bezout's Resultant (1764).** It will be convenient here to assume that the polynomials $a(z)$ and $b(z)$ have equal degrees ($m = n$), which can of course always be arranged by using zero coefficients. With this assumption, we return to the Sylvester matrix in the form (28). The Bezout test becomes evident in trying to simplify $S(a, b)$ by reducing it to triangular form. Thus note that

$$\begin{bmatrix} \mathfrak{A}_- & \mathfrak{B}_- \\ \mathfrak{A}_+ & \mathfrak{B}_+ \end{bmatrix} \begin{bmatrix} I & \mathfrak{B}_- \\ 0 & -\mathfrak{A}_- \end{bmatrix} = \begin{bmatrix} \mathfrak{A}_- & 0 \\ \mathfrak{A}_+ & \tilde{B} \end{bmatrix} \tag{30}$$

where we have used the fact that lower triangular Toeplitz matrices commute (cf. Exercise A.6) and have defined

$$\tilde{B} = \mathfrak{A}_+\mathfrak{B}_- - \mathfrak{B}_+\mathfrak{A}_- \tag{31}$$

Since $\mathfrak{A}_-$ is always nonsingular ($a_0 \neq 0$ by assumption), it is clear that

$$\det S(a, b) = 0 \Longleftrightarrow \det \tilde{B} = 0 \tag{32}$$

and therefore $\det \tilde{B}$ can also be used as a resultant. It turns out to be somewhat more convenient to introduce the *Bezout matrix* or *Bezoutian*

$$B = \tilde{I}\tilde{B} = \tilde{I}[\mathfrak{A}_+\mathfrak{B}_- - \mathfrak{B}_+\mathfrak{A}_-] \tag{33}$$

because it will follow (as we ask the reader to show—cf. Exercise 2.4-21) that $B$ will be symmetric, unlike $\tilde{B}$. We shall define

$$\text{Bezout's resultant} \triangleq \det B$$

We may note that the Bezoutian can also be introduced via the bilinear form

$$B(s, \sigma) = \frac{a(s)b(\sigma) - b(s)a(\sigma)}{\sigma - s} \tag{34a}$$

$$= \sum_{i,j=1}^{n} B_{ij} s^{i-1} \sigma^{j-1} \tag{34b}$$

The matrix $[B_{ij}]$ in (34) must clearly be symmetric, and it can be verified that it equals the expression in (33).

**System-Theoretic Interpretations.** As stated in the introduction, because of the results of Sec. 2.4-1, we would expect some relationships between the above tests for relative primeness and the concepts of controllability and observability. In fact, for example, MacDuffee's test was rediscovered by Kalman ([17] and [41]) on the basis of the result [cf. (2.3-44)] that for strictly proper systems ($b_0 = 0, a_0 = 1$)

$$\tilde{I}b(A_c) = \mathcal{O}_c, \quad \begin{array}{l}\text{the observability matrix of the controller-form} \\ \text{realization } \{A_c, b_c, c_c\} \text{ of } b(z)/a(z)\end{array} \tag{35}$$

Now we know (cf. Fig. 2.4-2) that $\{A_c, b_c, c_c\}$ will be jointly controllable and observable if and only if $b(z)$ and $a(z)$ are coprime. Since the controller form is always controllable, this means that $b(z)$ and $a(z)$ will be coprime if and only if $\mathcal{O}_c$ is nonsingular, or, using (35), if and only if $b(A_c)$ is nonsingular, which is MacDuffee's test (29).

Next recall that in Sec. 2.3.3 we had obtained another expression for $\mathcal{O}_c$, namely [cf. Eq. (2.3-43)]

$$\mathcal{O}_c = [\mathcal{B}_+ - \mathcal{B}_-\mathcal{A}_-^{-1}\mathcal{A}_+]\tilde{I} = [\mathcal{B}_+ - \mathcal{A}_-^{-1}\mathcal{B}_-\mathcal{A}_+]\tilde{I}$$

Then some algebra yields [use (33)]

$$-\tilde{I}\mathcal{A}_-\mathcal{O}_c\tilde{I} = \tilde{I}[\mathcal{B}_-\mathcal{A}_+ - \mathcal{A}_-\mathcal{B}_+] = B' = B \tag{36}$$

which shows that, like MacDuffee's test, the Bezout test is just another way of saying that the controller form of $b(s)/a(s)$ will be observable if and only if $b(s)$ and $a(s)$ are coprime.

Furthermore, we now recall that [cf. (2.3-6)] $\mathcal{O}_o^{-1} = \mathcal{A}_-$, so (36) can be rewritten as (cf. [42, Theorem 4])

$$-\tilde{I}B\tilde{I} = \mathcal{O}_o^{-1}\mathcal{O}_c \tag{37}$$

which we can recognize [cf. (2.4-10)] as the transformation matrix from the observer-form realization of $b(s)/a(s)$ to its controller-form realization, a transformation that will only exist when $b(s)$ and $a(s)$ are coprime (why?).

There are numerous other variants and identities for the resultants, especially concerning transformations from a (minimal) realization to its dual (see Fig. 2.4-3) and they are closely connected to classical results on root location, Padé approximation, etc. (see, e.g., [42]–[48]). However, we do not have space to pursue such questions here.

### *2.4.5 Some Worked Examples

We present several examples to reinforce some of the results and tests given in Secs. 2.3 and 2.4.

The first two examples illustrate some computational as well as conceptual features. As with all worked examples, the reader will profit most by first trying to work out the problems for himself. In any case, however, it will be useful for the reader to make for himself a list of important facts and ideas gained by a close study of these examples. Here we draw attention only to the methods described for computing the standard noncontrollable form and to the facts that controllability and observability depend on the variables chosen for the state-space description of a given physical system. The remaining examples are more theoretical but also bring out several things worth remembering. Example 2.4-3 provides another example of a useful result that can be easily conjectured by using the diagonal form (cf. the discussion in Sec. 2.4.1). The results of Example 2.4-4 (and of the closely related Exercises 2.4-6 and 2.4-7) are very helpful in avoiding some perhaps tedious algebra in many problems. In Example 2.4-5 we give a previously promised proof of Lemma 2.4-1. Finally in Example 2.4-6, we show an application of the fact that a unique similarity transformation relates any two minimal realizations (Theorem 2.4-7)—this problem arose in the study of reciprocal networks and was in fact one of the motivations for the development of the theorem by Youla [49], independently of Kalman [17].

**Example 2.4-1. Radial and Tangential Control of a Satellite**

In Example 2.2-E, we noted that the linearized equations for a satellite in a circular equatorial orbit are given by

$$\dot{x}(t) = Ax(t) + Bu(t), \qquad x' = [r \quad \dot{r} \quad \theta \quad \dot{\theta}]$$

where (with $m = 1 = r_0$, $\omega_0 = \omega$, as compared to Example 2.2-E)

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 3\omega^2 & 0 & 0 & 2\omega \\ 0 & 0 & 0 & 1 \\ 0 & -2\omega & 0 & 0 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \qquad u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

and the control $u_1(\cdot)$ represents radial thrust and $u_2(\cdot)$ represents tangential thrust.

Determine if the system is controllable when

1. There is only radial thrust [$u_2(\cdot) \equiv 0$].
2. There is only tangential thrust [$u_1(\cdot) \equiv 0$].

Also transform the realization into standard noncontrollable form when appropriate.

**Solution.**

1. When $u_2(\cdot) \equiv 0$ (only radial thrust), $b_1' = [0\ 1\ 0\ 0]$, $B$ becomes a column matrix, and we can calculate

$$\mathcal{C}(A, b_1) = \begin{bmatrix} 0 & 1 & 0 & -\omega^2 \\ 1 & 0 & -\omega^2 & 0 \\ 0 & 0 & -2\omega & 0 \\ 0 & -2\omega & 0 & 2\omega^3 \end{bmatrix}$$

We may note that the last column is $\omega^2$ times the second one, so that $\mathcal{C}(A, b_1)$ is singular and the system is not controllable using only a radial force. The PBH tests could also be used—thus one can check that $[sI - A \ \ b_1]$ loses rank at $s = 0$ or equivalently that $[2\omega\ 0\ 0\ 1]$ is a left eigenvector of $A$ (associated with the eigenvalue 0), which is orthogonal to $b_1$.

Although the state equations in this problem are fairly simple, it is not quite obvious from them which variables (or, rather, combinations of variables) are noncontrollable. To determine these, we should transform the given equations to a standard noncontrollable form, as described in the discussion following Eq. (13). *One* way to choose the transformation matrix $T$ is

$$T = [b_1 \quad Ab_1 \quad A^2b_1 \quad t]$$

where $t$ is independent of the preceding vectors (and may be chosen to be orthogonal to them). For example, we may choose

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

Then some algebra yields†

$$T^{-1}AT = \bar{A} = \left[\begin{array}{ccc:c} 0 & 0 & 0 & 6\omega^3 + \frac{3\omega}{2} \\ 1 & 0 & -\omega^2 & 0 \\ 0 & 1 & 0 & -(1/2\omega) \\ \hdashline 0 & 0 & 0 & 0 \end{array}\right]$$

Also, since $T\bar{b} = b$, we must have $\bar{b}' = [1\ 0\ 0 \mid 0]$. $\bar{A}$ and $\bar{b}$ are now in the standard form for displaying noncontrollability. The uncontrollable part has characteristic polynomial $s$, as already indicated by the PBH tests. Note that from $\bar{A}$ it is easy to

†It is worth remarking here that the algebraic labor involved in the above calculations can be conceptually illuminated by using some basic linear algebra. It would therefore be useful to read Secs. 5.2.1 and 5.2.2 at this point.

determine the overall characteristic polynomial as $s\cdot(s^2+\omega^2)s$, so that the eigenvalues are $\{0, 0, \pm j\omega\}$.

Another route to obtaining a suitable $T$ for transforming to the standard form is to note that we require

$$T^{-1}A = \left[\begin{array}{c|c} A_c & A_{12} \\ \hline 0 & \lambda \end{array}\right] T^{-1}, \qquad T^{-1}b = \left[\begin{array}{c} b_c \\ \hline 0 \end{array}\right]$$

where $\lambda$ is an uncontrollable eigenvalue.

If the last row of $T^{-1}$ is denoted by $t_n$, we require

$$t_n A = \lambda t_n, \qquad t_n b = 0$$

It is clear from the PBH test that a suitable $t_n$ is then

$$t_n = [2\omega \quad 0 \quad 0 \quad 1]$$

The remaining rows of $T^{-1}$ may be arbitrarily chosen to form an independent set (so that the transformation matrix is invertible). Let us take

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

We then obtain,

$$\bar{A}_1 = T^{-1}AT = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ -\omega^2 & 0 & 0 & 2\omega \\ -2\omega & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 0 \end{array}\right], \qquad \bar{b}_1 = T^{-1}b = \left[\begin{array}{c} 0 \\ 1 \\ 0 \\ \hline 0 \end{array}\right]$$

which again immediately displays the noncontrollability of the eigenvalue 0. We should also note the nonuniqueness of the standard forms: $\{\bar{A}, \bar{b}\}$ is not the same as $\{\bar{A}_1, \bar{b}_1\}$. But in both forms we can see that the uncontrollable variable is $2\omega x_1 + x_4 \triangleq 2\omega r + \dot{\theta}$ (a fact also evident from the PBH eigenvector test). There is actually a nice physical interpretation for this: Radial thrust cannot change the angular momentum, and the interested reader can check (referring back to Example 2.2-E) that if the angular momentum is to be the same after a small perturbation from the nominal orbit, then we must have $2\omega r + \dot{\theta} = 0$.

2. When $u_1(\cdot) \equiv 0$ (tangential thrust only), $b_2' = [0\ 0\ 0\ 1]$, and we can see readily that $\mathcal{C}(A, b_2)$ is nonsingular [e.g., its determinant is $-2\omega(-8\omega^3 + 2\omega^3) = 12\omega^4$]. Therefore tangential thrust is enough to provide controllability.

The PBH tests are not quite so simple now, since we must show that there is *no*

$s$ for which rank $[sI - A \;\; b_2]$ is less than $n$ or that there is *no* left eigenvector of $A$ orthogonal to $b_2$. We shall not go through the labor here.

It is worth mentioning here that in a multi-input system, as in this example, we do not usually require the system to be controllable by each input acting *separately*; the system is controllable if the inputs acting together in combination can set up arbitrary states. It will be shown in Sec. 6.2 that the test for controllability is then that the matrix $\mathcal{C} = [B \;\; AB \;\; \cdots \;\; A^{n-1}B]$ have full rank, and *not* that the $\mathcal{C}_i = [b_i \;\; Ab_i \;\; \cdots \;\; A^{n-1}b_i]$ all have full rank. ∎

**Example 2.4-2. An Inverted Pendulum on a Cart**

An important idealized control problem is that of a pendulum mounted on a moving carriage. After linearization, the equations of motion can be written as

$$\dot{x}(t) = Ax(t) + bu(t)$$

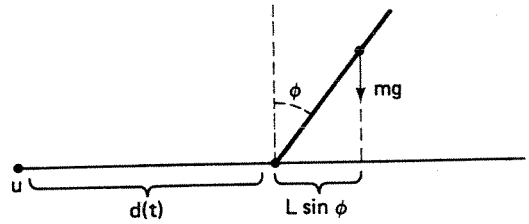
where

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -F/M & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -g/L & 0 & g/L & 0 \end{bmatrix}, \qquad b = \begin{bmatrix} 0 \\ 1/M \\ 0 \\ 0 \end{bmatrix}$$

and $M$ is the mass of the carriage (assumed much greater than the mass $m$ of the pendulum), $F$ is the friction coefficient for the motion of the carriage, $g$ is the gravitational constant, and $L = (J + ml^2)/ml$, where $l$ is the distance of the center of gravity from the pivot point and $J$ is the moment of inertia about this point. The input $u(\cdot)$ is the force on the cart. The state variables are

$$x'(t) = [d(t) \quad \dot{d}(t) \quad d(t) + L\phi(t) \quad \dot{d}(t) + L\dot{\phi}(t)]$$

where $d(\cdot)$ is the displacement of the cart from the origin and $\phi$ is the deviation of the pendulum from the vertical (see the figure).




1. Find the eigenvalues and left and right eigenvectors of $A$.
2. Determine the controllability properties of the system.
3. Assume that the output equation is $y(\cdot) = \phi(\cdot)$. Show that the system is unobservable and determine the observable and unobservable variables. With this knowledge, what observations would be needed to observe all the states?

**Solution.**

1. Because $A$ is *block triangular*, we can see immediately that the eigenvalues are $\{0, -F/M, \pm\sqrt{g/L}\}$.

We note that the corresponding right eigenvectors are

$$p_1' = [1 \quad 0 \quad 1 \quad 0], \qquad p_2' = [1 \quad -F/M \quad \alpha \quad -\alpha F/M]$$
$$p_3' = [0 \quad 0 \quad 1 \quad \sqrt{g/L}], \qquad p_4' = [0 \quad 0 \quad 1 \quad -\sqrt{g/L}]$$

where $\alpha = (g/L)/(gL^{-1} - F^2M^{-2})$. The left eigenvectors can be found as

$$q_1 = [F/M \quad 1 \quad 0 \quad 0], \qquad q_2 = [0 \quad 1 \quad 0 \quad 0]$$
$$q_3 = [-\sqrt{g/L} \quad \beta_+ \quad \sqrt{g/L} \quad 1], \qquad q_4 = [\sqrt{g/L} \quad \beta_- \quad -\sqrt{g/L} \quad 1]$$

where $\beta_\pm = -\sqrt{g/L}/[(F/M) \pm \sqrt{g/L}]$. As a check, notice that $q_i p_j = 0$, $i \neq j$, while $q_i p_i \neq 0$.

2. We see that $q_i b \neq 0$, and so by the PBH test the realization is controllable. (This can also be seen quite easily from the PBH rank test or by direct calculation of the controllability matrix.)

3. Since $\phi(t) = [x_3(t) - x_1(t)]/L$, the output equation is

$$y(t) = [-1/L \quad 0 \quad 1/L \quad 0]x(t)$$

By the PBH eigenvector test we see that

$$cp_1 = 0, \qquad cp_i \neq 0, \qquad i = 2, 3, 4$$

Therefore the unstable mode corresponding to $s = 0$ is unobservable.

This can also be seen by computing the transfer function from $u(\cdot)$ to $y(\cdot)$, which turns out to be (e.g., use Exercise A.13, part 2) $H(s) = -s/[LM(s - F/M)(s^2 - g/L)]$. The pole at $s = 0$ evidently has been cancelled out of the transfer function. Since the system is controllable, the cancellation must correspond to an unobservable natural frequency or unobservable mode of oscillation.

The unobservability of the system using observations of $\phi$ alone may have been suspected on purely physical grounds. It would seem that knowledge of $\phi$ can never tell us what is $d$, the displacement of the cart; a different initial $d$ could still give us the same $\phi$ history.

We attempt therefore to rewrite the state equations in such a way as to make the conjectured unobservability of $d$ obvious. Choosing a new set of state variables as, for example,

$$\bar{x} = [\dot{d} \quad L\phi \quad L\dot{\phi} \quad d]'$$

we get [cf. (2.4-16)]

$$\bar{A} = \left[\begin{array}{ccc:c} -F/M & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & g/L & 0 & 0 \\ \hdashline 1 & 0 & 0 & 0 \end{array}\right] = \begin{bmatrix} A_o & 0 \\ A_{21} & A_{\bar{o}} \end{bmatrix}$$

and

$$\bar{c} = [0 \quad 1/L \quad 0 \;\vdots\; 0] = [c_o \quad 0]$$

It is immediately seen that $d$ (and its associated eigenvalue of zero) is unobservable. The zero eigenvalue is a consequence of the fact that, with no input and no initial condition on any of the other state variables, any initial condition on $d$ remains constant.

To make the original realization observable, we could measure $d(\cdot)$ as well, giving a two-output system, say,

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} -1/L & 0 & 1/L & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} x(t)$$
$$= Cx(t)$$

This system is now clearly observable, and though we have not studied multiple-output systems as yet, the reader can verify that the appropriate algebraic criteria for observability are that either (1) rank $[C' \; A'C' \; \cdots \; (A')^{n-1}C'] = n$, or (2) no right eigenvector of $A$ is orthogonal to all the rows of $C$, or that (3) $[sI - A' \; C']$ has full rank for all $s$.

We can, however, make the system observable without converting it to a two-output system. The reader can verify that with the single measurement $y = d + L\phi = [0 \; 0 \; 1 \; 0]x$, the system is observable. ∎

**Example 2.4-3.**

Let

$$(sI - A)^{-1}b = \frac{[p_1(s), p_2(s), \ldots, p_n(s)]'}{a(s)}, \qquad a(s) = \det(sI - A)$$

Use the diagonal form to guess the relationship between the controllability of $\{A, b\}$ and the fact that the $n + 1$ polynomials $\{p_1(s), \ldots, p_n(s), a(s)\}$ have no non-trivial common factors.

Then give general proofs (in both directions) of your conjectured relation.

**Solution.** Let $n = 3$ and $A = \text{diag}\{\lambda_1, \lambda_2, \lambda_3\}$, $b' = [b_1 \; b_2 \; b_3]$. Then

$$(sI - A)^{-1}b = [b_1(s - \lambda_1)^{-1} \; b_2(s - \lambda_2)^{-1} \; b_3(s - \lambda_3)^{-1}]'$$

and $p_1(s) = b_1(s - \lambda_2)(s - \lambda_3)$, $p_2(s) = b_2(s - \lambda_1)(s - \lambda_3)$, and $p_3(s) = b_3(s - \lambda_1)(s - \lambda_2)$, while $a(s) = (s - \lambda_1)(s - \lambda_2)(s - \lambda_3)$. The controllability of $\{A, b\}$ is equivalent to the conditions $b_i \neq 0$, $\lambda_1 \neq \lambda_2 \neq \lambda_3$. In this case the $\{p_i(s), a(s)\}$ will have no common roots, whereas they clearly will if one of the $\{b_i\}$ is zero or if the $\{\lambda_i\}$ are not all distinct. Thus we may make the conjecture that

$$\{A, b\} \text{ controllable} \iff \{p_i(s), a(s)\} \text{ have no common roots}$$

We can in fact strengthen the conjecture to

$$\{A, b\} \text{ controllable} \iff \{p_i(s)\} \text{ have no common roots}$$

We now can try to prove the first conjecture directly; the second we leave to the reader.

First, if $\{A, b\}$ is controllable, then without loss of generality we can assume that we are in controller form $\{A_c, b_c\}$. In this case, it is easy to see (cf. Exercise A.33) that

$$p_{ic}(s) = s^{n-i}, \qquad i = 1, \ldots, n$$

so that $\{p_{ic}(s), a(s)\}$ will have no common factors (except the trivial factor unity). This result must also hold for the original pair $\{A, b\}$, because we have made an invertible transformation to get to $\{A_c, b_c\}$. But to get more confidence with such arguments, let us confirm this by a direct calculation. Let

$$A_c = T^{-1}AT, \qquad b_c = T^{-1}b$$

Then

$$(sI - A)^{-1}b = T(sI - A_c)^{-1}b_c$$

which shows that

$$[p_1(s), \ldots, p_n(s)]' = T[p_{1c}(s), \ldots, p_{nc}(s)]'$$

Let $\lambda$ be a root of $a(s) = 0$. Then it is clear that $[p_{1c}(\lambda), \ldots, p_{nc}(\lambda)] \not\equiv 0$ (because $p_{nc} = 1$). But since $T$ is invertible, it must also hold that $[p_1(\lambda), \ldots, p_n(\lambda)] \not\equiv 0$, which is the desired result. However, we should emphasize that such confirmation is not necessary—it is enough to know that the transformation involved is invertible. Note also that we do not need to know the explicit form of the transformation.

We shall use similar arguments to prove the converse. Thus, suppose now that $\{A, b\}$ is not controllable. We can without loss of generality assume that we have the standard noncontrollable form (2.4-11), so that $[(sI - \bar{A})^{-1}\bar{b}]' = [((sI - \bar{A}_c)^{-1}b_c)' \; 0]$, while $a(s) = \det(sI - \bar{A}_c) \det(sI - \bar{A}_{\bar{c}}) = a_c(s)a_{\bar{c}}(s)$, say. Therefore

$$[\bar{p}_1(s), \ldots, \bar{p}_n(s)] = [p_{1c}(s)a_{\bar{c}}(s), \ldots, p_{rc}(s)a_{\bar{c}}(s), 0, \ldots, 0]$$

so that $\{\bar{p}_i(s), a(s)\}$ have the common factor $a_{\bar{c}}(s)$. ∎

**Example 2.4-4. Observability of Series Realizations**

Let $\{A_i, b_i, c_i\}$ be observable realizations of order $n_i$ of the transfer functions $H_i(s) = g_i(s)/a_i(s)$, $i = 1, 2$, where $\deg a_i(s) = n_i$.

1. Write a natural set of state equations for the series (or cascade) connection of $H_1(s)$ followed by $H_2(s)$.

2. Show that these equations will be observable if and only if $a_1(s)$ and $g_2(s)$ are coprime.

**Solution.**

1. From the equations

$$\dot{x}_1 = A_1x_1 + b_1u, \qquad \dot{x}_2 = A_2x_2 + b_2y_1$$
$$y_1 = c_1x_1, \qquad y_2 = c_2x_2$$

we can write

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}A_1 & 0\\ b_2c_1 & A_2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}b_1\\ 0\end{bmatrix}u$$
$$y = [0 \quad c_2]\,[x_1' \quad x_2']'$$

2. This problem can be approached in many ways, but here we shall use the PBH rank test. A preliminary result, which is useful in its own right, is first stated.

The matrix $sI - A$, where $A$ is a left-companion matrix with $-[a_1 \cdots a_n]'$ as the first column, can be reduced to the form

$$\begin{bmatrix} & -1 & \bigcirc\\ & & \ddots\\ \bigcirc & & \\ & \bigcirc & -1\\ \hline a(s) & \bigcirc & \end{bmatrix}, \qquad a(s) = s^n + a_1 s^{n-1} + \cdots + a_n$$

by elementary row and column operations corresponding to pre- and postmultiplication by certain well-structured matrices that we ask the reader to determine.

Now assume without loss of generality that the two given observable realizations are in observer form. Noting that $c_1 = [1\ 0\ \cdots\ 0] = c_2$, we see that

$$\begin{bmatrix}sI - A\\ c\end{bmatrix} = \left[\begin{array}{c|c} sI - A_1 & 0\\ -b_2 \quad 0 \cdots 0 & sI - A_2\\ \hline 0 \cdots 0 & 1 \quad 0 \cdots 0\end{array}\right]$$

Using the elementary row and column operations mentioned above, this matrix can be reduced to the form

$$\left[\begin{array}{c|c|c|c} & -1 & & \\ 0 & \ddots & & \bigcirc\\ & -1 & & \\ \hline a_1(s) & 0 & & \\ \hline & & & -1\\ 0 & \bigcirc & 0 & \ddots\\ & & & -1\\ \hline -g_2(s) & 0 & a_2(s) & 0\\ \hline 0 & 0 & 1 & 0\end{array}\right]$$

The composite system is observable if and only if this $(n_1 + n_2 + 1) \times (n_1 + n_2)$

matrix has rank $n_1 + n_2$ for all values of $s$, i.e., if it has all its columns independent for all $s$. This will clearly be so if and only if the first column is never zero, i.e., if and only if there is *no* $\lambda$ for which $a_1(\lambda) = g_2(\lambda) = 0$.

In Exercise 2.4-9, we suggest that another simple proof can be obtained by using the (dual of the) criterion of Example 2.4-3. ▮

**Example 2.4-5. Transformation Between Controllable Realizations (Lemma 2.4-1)**

Show that any two controllable realizations (of the same transfer function) with the same characteristic polynomial can be related by a similarity transformation. (This will essentially prove Lemma 2.4-1.)

**Solution.** One method is to note that there exists an invertible transformation from any controllable realization to a controller-form realization with the same characteristic polynomial (see Example 2.3-4). Here we shall give a somewhat more direct proof.

If $x_1(t) = Tx_2(t)$, we know $T^{-1}A_1T = A_2, T^{-1}b_1 = b_2$. Therefore $\mathcal{C}_2 = T^{-1}\mathcal{C}_1$ or $T = \mathcal{C}_1\mathcal{C}_2^{-1}$ if both realizations are controllable. Now let us try to reverse this argument. Since we are given $\{A_i, b_i, c_i, i = 1, 2\}$ both controllable, we can certainly define $T = \mathcal{C}_1\mathcal{C}_2^{-1}$, but we have to see if this works, i.e., if with this $T$, $T^{-1}A_1T = A_2$, $T^{-1}b_1 = b_2$, $c_1T = c_2$. Now

$$T^{-1}b_1 = \mathcal{C}_2\mathcal{C}_1^{-1}b_1 = \mathcal{C}_2[b_1 \quad Ab_1 \quad \cdots \quad A^{n-1}b_1]^{-1}b_1 = b_2$$

Next we examine

$$T^{-1}A_1T = \mathcal{C}_2\mathcal{C}_1^{-1}A_1\mathcal{C}_1\mathcal{C}_2^{-1}$$

Consider for simplicity $n = 3$, and

$$\mathcal{C}_1^{-1}A_1\mathcal{C}_1 = [b_1 \quad A_1b_1 \quad A_1^2b_1]^{-1}[A_1b_1 \quad A_1^2b_1 \quad A_1^3b_1]$$

Now, by the Cayley-Hamilton theorem,

$$(A_1^3 + a_1A_1^2 + a_2A_1 + a_3I)b = 0$$

where the $\{a_i\}$ have no subscript because by assumption they are the same for both realizations. Therefore

$$\mathcal{C}_1^{-1}A_1\mathcal{C}_1 = \begin{bmatrix} 0 & 0 & -a_1 \\ 1 & 0 & -a_2 \\ 0 & 1 & -a_3 \end{bmatrix}$$

which is the *same* for both realizations. Therefore

$$\mathcal{C}_1^{-1}A_1\mathcal{C}_1 = \mathcal{C}_2^{-1}A_2\mathcal{C}_2$$

and

$$T^{-1}A_1T = \mathcal{C}_2\mathcal{C}_1^{-1}A_1\mathcal{C}_1\mathcal{C}_2^{-1} = \mathcal{C}_2\mathcal{C}_2^{-1}A_2\mathcal{C}_2\mathcal{C}_2^{-1} = A_2$$

Finally, to bring in the $\{c_i\}$, we have to refer to the transfer function or to the Markov

parameters, which yield

$$c_1 b_1 = c_2 b_2, \qquad c_1 A_1 b_1 = c_2 A_2 b_2, \qquad \cdots$$

or

$$c_1 \mathcal{C}_1 = c_2 \mathcal{C}_2$$

or

$$c_2 = c_1 \mathcal{C}_1 \mathcal{C}_2^{-1} = c_1 T, \qquad \text{as desired}$$

Note that the assumption that the $\{c_i, A_i\}$ are observable is not necessary. ∎

**Example 2.4-6. Transformation Between Dual Realizations**

Let $\{A, b, c\}$ be a minimal realization of a scalar transfer function $H(s)$. Show that there exists a unique *symmetric* matrix $T$ satisfying the relations

$$TA' = AT \quad \text{and} \quad cT = b'$$

**Solution.** Since $H(s)$ is scalar, we have $H(s) = H'(s)$, so that

$$c(sI - A)^{-1} b = b'(sI - A')^{-1} c'$$

Therefore if $\{A, b, c\}$ is a minimal realization, so is $\{A', c', b'\}$ (by Theorem 2.4-6 and Lemma 2.4-2). Therefore by Theorem 2.4-7 there must be a *unique* invertible matrix such that

$$TA' = AT, \qquad Tc' = b, \qquad b' = cT$$

Taking transposes, we obtain the equalities

$$T'A' = AT', \qquad cT' = b', \qquad T'c' = b$$

which show that $T'$ is also a similarity transformation between the two realizations. But by Theorem 2.4-7 there can be only one such transformation. Therefore $T = T'$, which completes the proof.

This result was discovered by Youla in his studies of the properties of *reciprocal networks*—see [49]; in fact, it appears that the above problem led Youla to an independent proof of Theorem 2.4-7. ∎

***Exercises***

At this stage the reader will profit from another look at the exercises of the previous sections, many of which will be simpler to do now.

***2.4-1.***

Consider a realization with

$$A = \text{block diag}\left\{\begin{bmatrix} -2 & 1 \\ 0 & -2 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right\}$$

$$b' = [1 \quad 0 \quad 1 \quad 1], \qquad c = [1 \quad 0 \quad 0 \quad 0]$$

Draw a block diagram of the realization and check its controllability in as many ways as you can.

***2.4-2.***

Consider a realization $\{A, b, c\}$ with

$$A = \begin{bmatrix} -0.5 & 1 & 0 \\ -1 & -0.5 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}, \quad c' = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

a. Is the system observable? If not completely observable, what quantities are unobservable?

b. Is the system controllable? If not completely controllable, what quantities are uncontrollable?

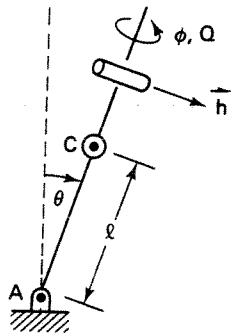
***2.4-3.*** *(DeBra)*

An inverted pendulum, of mass $m$, is hinged at $A$. A gyro with spin angular momentum, $h$, is attached to the pendulum but is free to rotate about the pendulum axis (angle $\phi$) as shown in the figure. A control torque, $Q$, can be applied to the gyro from the pendulum. The equations of motion are $I\ddot{\theta} = mgl\theta - h\dot{\phi}$ and $J\ddot{\phi} = h\dot{\theta} + Q$, where $I$ = the moment of inertia of the pendulum plus gyro about $A$, $J$ = the moment of inertia of the gyro about axis $AC$, and $C$ = the mass center of the pendulum plus gyro.

a. Compute the transfer functions from $u(\cdot)$ to $\phi(\cdot)$ and $u(\cdot)$ to $\theta(\cdot)$.

b. Show that the system is controllable by $Q$, observable with $\phi$, and unobservable with $\theta$.

c. Show that the system is always unstable.




***2.4-4.*** *A Model Common in System Identification Problems*

Consider the equation

$$y(t) + a_1 y(t-1) + \cdots + a_n y(t-n) = b_1 u(t-1) + \cdots + b_n u(t-n)$$

Assume that the associated polynomials $a(z)$ and $b(z)$ are relatively prime, and

introduce as a "state" vector

$$\theta'(t) = [y(t) \quad y(t-1) \quad \cdots \quad y(t-n+1) \quad u(t-1) \quad \cdots \quad u(t-n+1)]$$

a. Write the associated state equations for the system.

b. Show that this state-space realization is not minimal and determine the eigenvalues corresponding to the "hidden" modes.

***2.4-5.***

a. Consider the cascade connections of minimal realizations of $H_1(s)$ and $H_2(s)$ as $H_1(s)H_2(s)$ and $H_2(s)H_1(s)$, where $H_1(s) = 1/(s+1)$ and $H_2(s) = (s+1)/(s+2)(s+3)$. For each connection, determine the uncontrollable and unobservable modes, if any.

b. Repeat for the realizations connected in feedback form, first with $H_1(s)$ in the feedforward path and $H_2(s)$ in the feedback path, and then vice versa.

c. In Sec. 2.0, we noted that the behavior of the cascade connection of systems with $H_1(s) = 1/(s-1)$ and $H_2(s) = (s-1)/(s+1)$ depended on the order in which they were connected. Can you give a simple explanation of the differences?

***2.4-6.*** *Controllability and Observability of Interconnected Subsystems*

Let $\{A_i, b_i, c_i, i = 1, 2\}$ be realizations of order $n_i$ of the transfer functions $H_i(s) = g_i(s)/a_i(s)$, also of order $n_i$.

a. Show that if the realizations are controllable, then the *series* combination of system 1 followed by 2 is controllable if and only if $g_1(s)$ and $a_2(s)$ are coprime.

b. Show that if the realizations are observable (controllable), then the *parallel* combination is observable (controllable) if and only if $a_1(s)$ and $a_2(s)$ are coprime.

c. Show that if the realizations are observable (controllable), then the *feedback* configuration with system 1 in the forward path and 2 in the feedback path is observable (controllable) if and only if $g_1(s)$ and $a_2(s)$ are coprime. *Note:* One way to prove part c is to show, using general arguments, that the feedback system is equivalent, insofar as observability (controllability) goes, to the series combination of system 2 followed by system 1 (system 1 followed by system 2).

d. Extend the results to the case of systems with direct feedthrough from input to output.

***2.4-7.*** *Characteristic Polynomial of Interconnected Systems*

Let $\{A_i, b_i, c_i, i = 1, 2\}$ be two minimal realizations with characteristic polynomials $a_i(s) = \det(sI - A_i)$.

a. Show that the characteristic polynomial of the
   (1) Series connection (of the two systems) is $a_1(s)a_2(s)$.
   (2) Parallel connection is $a_1(s)a_2(s)$.
   (3) Feedback connection, with $\{A_1, b_1, c_1\}$ in the forward path and $\{A_2, b_2, c_2\}$ in the feedback path, is $a_1(s)a_2(s) + b_1(s)b_2(s)$.

b. Let $H_i(s) = b_i(s)/a_i(s)$, $i = 1, 2$. Show that the denominators of the

nominal (i.e., without cancellation of common factors) transfer functions of the series and parallel combinations are just the characteristic polynomials found in parts a(1) and a(2). For the feedback connection, we can write

$$\begin{aligned} H_f(s) &= \frac{H_1(s)}{1 + H_1(s)H_2(s)} \\ &= \frac{b_1(s)}{a_1(s)[a_1(s)a_2(s) + b_1(s)b_2(s)]} a_1(s)a_2(s) \\ &= \frac{b_1(s)a_2(s)}{a_1(s)a_2(s) + b_1(s)b_2(s)} \end{aligned}$$

where the denominator is again the characteristic polynomial of the feedback combination [cf. part a(3)]. Yet there was a cancellation of $a_1(s)$ in forming the overall transfer function $H_f(s)$. How do you reconcile these two facts?

***2.4-8.***

a. Use the PBH tests to show that the controller-form realization of $b(s)/a(s)$ will be observable if and only if $\{b(s), a(s)\}$ are coprime. (Another proof was given in Example 2.3-5.) (*Hint:* See Exercise A.35.)

b. The differential equation $\dot{x}(t) = Ax(t) + bu(t)$ can be approximated by the equations $x_{k+1} = (I + A\Delta)x_k + \Delta b u_k$. If $\{A, b\}$ is controllable, what can you say about the controllability of $\{I + A\Delta, \Delta b\}$?

***2.4-9.*** *Alternative Tests for Observability*

a. Show that $\{c, A\}$ is observable if and only if the $n + 1$ polynomials $\{q_i(s), a(s)\}$ defined by $c(sI - A)^{-1} = [q_1(s), \ldots, q_n(s)]/a(s)$ have no nontrivial common factors.

b. Use this result to obtain an alternative proof of the condition of Example 2.4-4 for observability of a series combination of two observable subsystems.

***2.4-10.*** *Alternative Characterizations of Controllability*

a. Prove that $\{A, b\}$ is controllable if and only if $\{A - bk, b\}$ is controllable for all $k$.

b. Show that $\{A, b\}$ is controllable if and only if the only $n \times n$ matrix $X$ such that $AX = XA$ and $Xb = 0$ is the matrix $X \equiv 0$.

***2.4-11.*** *Simple Multivariable Systems*

a. Let

$$Y(s) = \begin{bmatrix} Y_1(s) \\ Y_2(s) \end{bmatrix} = \begin{bmatrix} \frac{1}{s+1} & \frac{2}{s+1} \\ \frac{-1}{(s+1)(s+2)} & \frac{1}{s+2} \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \end{bmatrix} = H(s)U(s)$$

be the transfer function of a two-input, two-output linear system. Find a realization $\{A, B, C\}$ of order not greater than 4 and another realization of order 3. Prove that 3 is the minimal order by using the theorem that minimality is equivalent to simultaneous controllability and observability (cf. Example 2.4-1).

b. A two-input, two-output system is described by the equations

$$\dot{y}_1(t) + y_2(t) = u_1(t) + u_2(t)$$
$$\dot{y}_2(t) + y_1(t) = u_2(t)$$

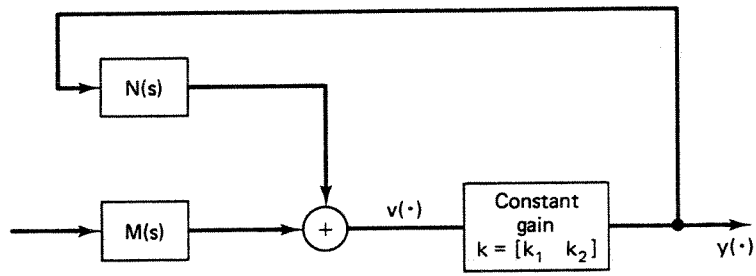
Calculate the transfer function matrix. Try to give at least two different analog-computer simulations, say one using three integrators and one using two integrators.

*2.4-12. A Nice Form for System Identification*

Let $\{y(\cdot), u(\cdot)\}$ be the input and output of a system with a strictly proper irreducible transfer function, with denominator of degree $n$. For certain *adaptive identification* schemes, it is useful to reparametrize the system as shown in the figure. Here we choose

$$N(s) = \begin{bmatrix} (sI - F)^{-1}g \\ 0 \end{bmatrix}, \qquad M(s) = \begin{bmatrix} 0 \\ (sI - F)^{-1}g \end{bmatrix}$$

where $\{F, g\}$ is an *arbitrary* controllable pair, $F$ being $n \times n$ and $g$ $n \times 1$, where $n$ is the order (assumed known) of the unknown system.




a. Show that the transfer function of the system shown can be written as

$$W(s) = \frac{Y(s)}{U(s)} = [1 - k_1(sI - F)^{-1}g]^{-1}k_2(sI - F)^{-1}g$$
$$= k_2[sI - (F + gk_1)]^{-1}g$$

b. Show that if $\{F, g\}$ is controllable, then by proper choice of $[k_1 \;\; k_2]$ we can make $W(s)$ have an arbitrary $n$th-degree denominator polynomial and an arbitrary numerator polynomial of degree less than or equal to $n - 1$; i.e., we have a "model" of the given system.

c. What will happen if $n$ happens to be larger than the order of the given system? *Remark:* For the application of this model to the so-called *model reference adaptive identification* techniques, see a recent account by B.D.O. Anderson (*Automatica*, Vol. 13, pp. 401–408, 1977).

*2.4-13.*

Suppose $\{A, b, c\}$ is minimal and $a(s) = \det(sI - A)$ has a repeated root. Prove that $A$ cannot be diagonalized by a similarity transformation.

*2.4-14.* (*Brockett*)

Suppose that $\{A, b, c\}$ is minimal. Prove that $A$ and $bc$ cannot commute if $n \geq 2$ ($A$ is $n \times n$).

*2.4-15.*

a. If $\{A, b\}$ is given and *not* controllable, is it always possible to choose $c$ so that $[c, A]$ is observable? A full explanation or a counterexample will suffice.

b. If $\{A, b\}$ is given and is controllable, can we always choose $c$ so that $\{c, A\}$ is observable?

*2.4-16.*

Use the state equations to give a physical explanation of why the variables $\bar{x}_{\bar{c}}$ in Eq. (2.4-11) are always noncontrollable. Give the dual argument for the states $\bar{x}_0$ in Eq. (2.4-16).

*2.4-17. A Classical Resultant Formula*

Suppose $a(s) = a_0 \prod_1^n (s - \alpha_i)$, $a_0 \neq 0$, and $b(s) = b_0 \prod_1^m (s - \beta_i)$, $b_0 \neq 0$. Show that

$$\det \tilde{S}(a, b) = a_0^m b_0^n \prod (\alpha_i - \beta_j)$$

where $\tilde{S}(a, b)$ is the Sylvester matrix of $\{a(s), b(s)\}$.

*2.4-18. Alternative Derivation of the Sylvester Test*

Let $\tilde{S}(a, b)$ be the Sylvester matrix for the polynomials $a(z)$ and $b(z)$ of Eq. (2.4-23) (with $a_n \neq 0 \neq b_m$). Define a matrix $R$ with first column $[z^{n+m-1} \cdots z\ 1]$, ones on the diagonal [except in the (1, 1) location], and zeros everywhere else. Show by evaluating the determinant in two different ways that

$$\det [\tilde{S}(a, b)R] = [\det \tilde{S}(a, b)]z^{n+m-1} = a(z)f(z) + b(z)g(z)$$

where $\{f(z), g(z)\}$ are polynomials of degree *at most* $n - 1$ and $m - 1$, respectively. Show from this relation that $\{a(z), b(z)\}$ have a nontrivial common factor if and only if $\det \tilde{S}(a, b) = 0$. (Reference: H. Skala, *Am. Math. Mon.*, Vol. 78, pp. 889–890, 1971.)

*2.4-19. Another Test for Coprimeness*

a. From a systems point of view another test for relative primeness is that the controllability-form realization of $b(z)/a(z)$ be observable. Show that this fact leads to a Hankel matrix test for coprimeness.

b. Deduce this test in another way by proving the identity $\beta = -\mathcal{C}'_- M[1, n - 1]\mathcal{C}_-$. (*Hint:* See Fig. 2.4-3.)

*2.4-20. The Sylvester Matrix and an Observability Matrix*

Let $a(z) = z^n + a_1 z^{n-1} + \cdots + a_n$ and $b(z) = b_1 z^{n-1} + \cdots + b_n$. Show that by elementary row *and* column operations we can reduce the Sylvester

matrix to the form block diag $\{I_{n-1}, \mathcal{O}(b, A_c)\}$, where $b = [b_1 \ \cdots \ b_n]$ and $A_c$ is a companion matrix with $-[a_1 \ \cdots \ a_n]$ as the first row. Use this to give another proof of the Sylvester test for coprimeness.

*2.4-21. Symmetry of the Bezoutian*

Prove that the matrix $B$ defined by (2.4-33) is symmetric. *Hint:* Use the matrix identities corresponding to the trivial polynomial identity $a(z)b(z) = b(z)a(z)$.

## *2.5 SOLUTIONS OF STATE EQUATIONS AND MODAL DECOMPOSITIONS

So far in this chapter we have deliberately said very little about actually solving the state equation, because, as we have seen already and shall see even more as we proceed, one does not need explicit solutions in order to gain useful information about the system. This is fortunate, because often the solutions may be impossible to get in any convenient analytical form, especially for systems with time-variant coefficients. And furthermore, in many studies it is more important to have a nice "representation" of the solution rather than an actual "solution." Long usage often tends to blur this distinction, especially with scalar equations. For example, the solution of the scalar equation

$$\dot{x}(t) = ax(t), \qquad x(0) = 1$$

is

$$x(t) = e^{at} \quad \text{or} \quad \exp at$$

Now we can readily plot the solution using a table of exponentials or by some direct method of evaluating exponentials. However, often we do not need any explicit values of exp $at$ but only the "defining" property

$$\frac{de^{at}}{dt} = ae^{at}$$

In this case exp $at$ is more a useful "representation" of a solution than a solution itself. The significance of this statement will become clearer as the reader studies matrix exponentials in Sec. 2.5.1 and their time-variant generalizations in Chapter 9.

In Sec. 2.5.2, we shall give a brief discussion of the concept of modes and modal decompositions of a realization, a topic that has both conceptual and numerical value.

* This section may be omitted without loss of continuity, especially since, to get an exposure to some more significant applications, the reader could at this point proceed directly to Chapter 3. Section 2.5.1 can be taken up just before the study of time-variant systems in Chapter 9.

# ASYMPTOTIC OBSERVERS AND COMPENSATOR DESIGN

# 4

## 4.0 INTRODUCTION

In Chapter 3, we showed that if a realization $\{A, b, c\}$ is controllable, then state-variable feedback can modify the eigenvalues of $A$ at will. We shall now discuss the problem of actually obtaining the states of the realization from knowledge only of the system input $u(\cdot)$ and system output $y(\cdot)$. We have already seen in Sec. 2.3 that if the realization $\{A, b, c\}$ is also observable, then a differentiation technique can be used to calculate the states. However, this technique is clearly impractical, and therefore in Sec. 4.1 we shall develop a more realistic state estimator, which is usually known as an asymptotic observer. The name arises from the fact that the states can only be obtained with an error, but one that can be made to go to zero at any specified exponential rate.

In Sec. 4.2 we shall discuss the use of such state estimates in place of the possibly unavailable true states. We shall find that the design equations for the controller are not affected by the fact that approximate states are being used instead of the true states. More crucially, however, we shall find the important and a priori nonobvious fact that the overall observer-controller configuration is *internally stable*, which was an issue not completely faced by classical design methods. However, use of the estimated instead of the true states, for feedback, may lead in general to a deterioration of the transient response (see Example 4.2-1).

The observer studied in Sec. 4.1 was apparently first introduced in un-

published work by Bertram in 1961 (see the comments in Astrom [1, p. 158]) and by Bass in 1963 (see the comment of Kalman et al. [2, p. 55]). An independent and rather different approach was published by Luenberger in 1964 [3] (see also his Ph.D. thesis, Stanford University, Stanford, Calif., 1963). Luenberger treats a somewhat more general estimation problem which, when specialized to our situation, actually yields an observer with one less state variable than the observer of Sec. 4.1. This so-called reduced-order or Luenberger observer is discussed in Sec. 4.3 by a method due to Gopinath [4] and Cumming [5] (see also the survey papers [6] and [7]).

In Sec. 4.4, we shall briefly discuss the question of optimum observer pole locations, which will lead us to some results *dual* to those developed in Sec. 3.4.1 for the optimal quadratic regulator. Analogs of the results in the other sections of Secs. 3.4 and 3.5 can be developed, though we shall not do so here. Instead, in our concluding Sec. 4.5, we shall explore at some length the fact that it is possible to obtain the combined observer-controller configuration directly by transfer function analysis, thus obviating the need for the notions of state, controllability, and observability. This was shown by Chen ([8]) although, as we shall explain in Sec. 4.5.1, without the insights provided by the state-variable design, certain judicious choices and assumptions in the arguments of Reference [8] might not have been apparent. However, once the transfer function method has been introduced, we shall find that this suggests certain extended results that are not immediately obvious in the state-space approach; moreover, this method will extend nicely to the multivariable case (Sec. 7.5).

## 4.1 ASYMPTOTIC OBSERVERS FOR STATE MEASUREMENT

We shall now begin to explore the question of methods of actually determining the states of a realization

$$\begin{aligned} \dot{x}(t) &= Ax(t) + bu(t), \qquad x(0-) = x_0 \\ y(t) &= cx(t) \qquad\qquad\quad t > 0- \end{aligned} \tag{1}$$

given knowledge only of $y(\cdot)$ and $u(\cdot)$.

In Sec. 2.3. we have already described one method of determining the states at any time $t$. But this method involves differentiation and is therefore impractical.

When we reflect on the fact that we know $A$, $b$, $c$, $y(\cdot)$, and $u(\cdot)$, which is really quite a lot, we wonder why $x(\cdot)$ cannot be reconstructed by forming a *dummy* system $\{A, b\}$ and driving it with $u(\cdot)$. The problem,† of course, is

†It is important to note that the dummy system need not use the same physical components as the original system—it can be a miniaturized electronic analog computer simulation or a special-purpose digital computation package. For the present purpose, quite large systems—chemical plants, satellites, etc.—may be quite cheaply simulated in this way.

that we do not know the initial condition $x(0-) = x_0$. This again has to be found by the differentiation technique, but we have a slight simplification since it is reasonable to assume that $u(t) = 0$, $t \leq 0-$. Then some calculation shows that $x(0-)$ can be obtained from the equation

$$\mathcal{O}x(0-) = [y(0-), \ldots, y^{(n-1)}(0-)]' \tag{2}$$

and in many problems the values $\{y(0-), \ldots, y^{(n-1)}(0-)\}$ are given to us as part of the problem statement.

**An Open-Loop Observer.** Thus, by setting up a dummy system $\{A, b, c\}$ with the proper initial conditions and driving it with the known input $\{u(t), t > 0-\}$, we can obtain $\{x(t), t > 0-\}$. Note that the values of the output $\{y(t), t > 0-\}$ are no longer required. Unfortunately, however, this last fact is really more of a disadvantage than a help. The point is that strategies that are the same no matter what the output is (called *open-loop* strategies as opposed to *feedback* or *closed-loop* methods, which adjust parameters according to the current value of the system output or system state) are obviously susceptible to disturbances that may arise during system operation, and, furthermore, they allow no means of compensating for any errors, possibly small but still almost inevitable, in the predetermined strategy.

Thus, suppose that the initial condition that we use for our dummy system is slightly in error—say we have not $x_0$ but

$$\hat{x}_0 \triangleq x_0 - \epsilon, \qquad \|\epsilon\| \ll \|x_0\| \tag{3}$$

where we shall use, both here and elsewhere, the caret ($\wedge$) to denote *estimate of*. What is the effect of this small error on the states calculated from our dummy system? The states will no longer be $x(t)$ but will be a different function, say $\hat{x}(t)$, which satisfies the equation

$$\dot{\hat{x}}(t) = A\hat{x}(t) + bu(t), \qquad \hat{x}(0-) = \hat{x}_0 = x_0 - \epsilon \tag{4}$$

Clearly, there will be an error

$$\tilde{x}(t) \triangleq x(t) - \hat{x}(t)$$

which will satisfy the differential equation

$$\dot{\tilde{x}}(t) = A\tilde{x}(t), \qquad \tilde{x}(0-) = \epsilon \tag{5}$$

That is, the error $\epsilon$ in the initial condition will produce an error, $\tilde{x}(\cdot)$, at all later times. Now, as can be seen from a partial fraction expansion of

$$\tilde{X}(s) = (sI - A)^{-1}\epsilon$$

$\tilde{x}(t)$ will be a sum of terms of the form $\{e^{\lambda_i t}, t^j e^{\lambda_i t}, \ldots\}$, where the $\{\lambda_i\}$ are the eigenvalues of $A$. Clearly, if the system is unstable (recall that we are interested in determining states to be fed back to achieve stabilization), then the error $\tilde{x}(t)$ will become *arbitrarily large* as $t \longrightarrow \infty$, no matter *how small* the initial error is. Less dramatically, even if the system is stable but some eigenvalues have real parts that are very small, the effects of errors in the initial estimates will take a long time to die out.

**A Closed-Loop Observer.** There may be situations in which open-loop estimators are fairly satisfactory, especially if periodic *resetting* is feasible so as to eliminate, or at least suitably control, the effects of initial errors and later disturbances.

However, the classical way of overcoming the potential difficulties of open-loop systems is to use *feedback*, viz., to try to *zero the error* by driving the system with a term proportional to the error in the estimate. But how can we determine this "error?" In classical feedback problems, this is provided by the *reference* signal.† In our problem, the error is $x(\cdot) - \hat{x}(\cdot) = \tilde{x}(\cdot)$, but of course $x(\cdot)$ is not available! Therefore, we have to obtain a reference signal in some other way. The system output $y(\cdot)$, which was not used in the open-loop solution, comes in now because it is related to the quantity $x(\cdot)$ we are interested in: $y(\cdot) = cx(\cdot)$, and $y(\cdot)$ is clearly available. Therefore, an *error* signal can be generated as

$$y(t) - \hat{y}(t) = y(t) - c\hat{x}(t) = c[x(t) - \hat{x}(t)] = c\tilde{x}(t)$$

and it can be used to drive the estimator equation. These considerations lead us‡ to consider an estimator for $x(\cdot)$ of the form (see Fig. 4.1-1)

$$\dot{\hat{x}}(t) = A\hat{x}(t) + bu(t) + l[y(t) - c\hat{x}(t)], \qquad \hat{x}(0) = \hat{x}_0 \tag{6}$$

where

$\hat{x}_0$ = an estimated initial state vector

and

$l$ = a *feedback gain vector*, to be suitably chosen

The system yielding $\hat{x}(\cdot)$ is called an observer, or, actually, for reasons that will become clearer presently, an *asymptotic observer*.

We should, of course, choose $l$ so as to properly control the error $\tilde{x}(\cdot)$. Now the error $\tilde{x}(\cdot)$ obeys the differential equation

$$\begin{aligned} \dot{\tilde{x}}(t) &= \dot{x}(t) - \dot{\hat{x}}(t) \\ &= (A - lc)\tilde{x}(t), \qquad \tilde{x}(0) = x_0 - \hat{x}_0 \end{aligned} \tag{7}$$

†For example, in a thermostat, the reference signal is the desired ambient temperature.

‡See also Exercise 4.1-2.

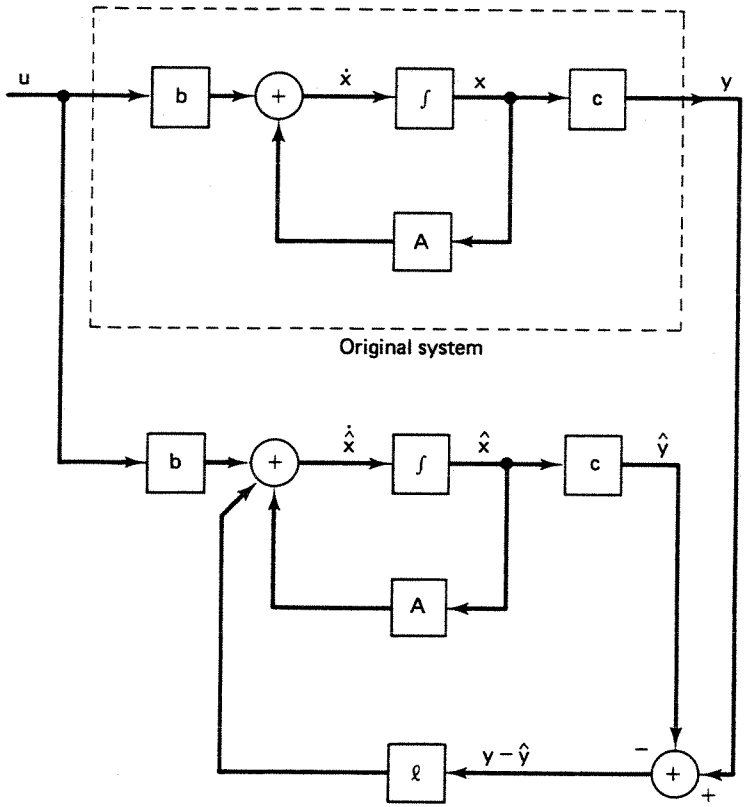



**Figure 4.1-1.** Block diagram of asymptotic observer. Access to the original system is assumed to be possible at the input and output terminals, while this restriction is not necessary for the observer.

Notice that when $l = 0$ this equation reduces to the open-loop error equation (5). The effect of the *error feedback* $l[y(\cdot) - \hat{y}(\cdot)] = lc[x(\cdot) - \hat{x}(\cdot)]$ is to give us some control over the behavior of the error $\tilde{x}(\cdot)$. In fact, the natural frequencies will be the eigenvalues of $A - lc$, and what we can try to do is to choose $l$ so that the error will die out as rapidly as we deem suitable. Notice that the actual value of the initial estimate $\hat{x}_0$ is unimportant —if we have no special information, we often take $\hat{x}_0 = 0$. The reason for the name *asymptotic observer* should now be clear.

But can we always find a suitable $l$? We shall prove that if $\{c, A\}$ is observable, then we can choose $l$ so that $A - lc$ has arbitrary eigenvalues, or equivalently

$$\det (sI - A + lc) = \alpha(s) = s^n + \alpha_1 s^{n-1} + \cdots + \alpha_n \tag{8}$$

where the $\{\alpha_i\}$ are completely arbitrary. The reader will recognize that this problem is essentially the same as the one we discussed in Sec. 3.2 in determining the feedback vector $k$ required to give a controllable realization arbitrary dynamics. There is in fact a very close connection, and we shall actually solve our problem by *dualizing* the results of Sec. 3.2.

**Formulas for the Observer Gain.** In Sec. 3.2 we showed that given any realization $\{A, b, c\}$ with $\{A, b\}$ controllable we could in several ways find $k$ so that

$$\det (sI - A + bk) = s^n + \alpha_1 s^{n-1} + \cdots + \alpha_n \tag{9}$$

for any $\{\alpha_i\}$. For example, one formula is [cf. (3.2-13)]

$$k = (\alpha - a)\mathfrak{A}^{-1}\mathfrak{C}^{-T}(A, b) \tag{10}$$

where $a$ is the vector of coefficients of $a(s) = \det (sI - A)$ and $\mathfrak{A}^T$ is an upper triangular Toeplitz matrix with $[1 \;\; a_1 \;\; \cdots \;\; a_{n-1}]$ as the first row.

To recast our observer problem into this form we merely have to note that

$$\det (sI - A + lc) = \det (sI - A' + c'l')$$

so that if we let

$$A \longrightarrow A', \qquad b \longrightarrow c', \qquad k \longrightarrow l'$$

in the solution of the controller problem, we deduce that we shall be able to find $l$ if and only if $\mathfrak{C}(A', c')$ is nonsingular. Then

$$l' = (\alpha - a)\mathfrak{A}^{-T}\mathfrak{C}^{-1}(A', c')$$

But

$$\begin{aligned} \mathfrak{C}(A', c') &= [c' \quad A'c' \quad \cdots \quad A'^{(n-1)}c'] \\ &= \mathfrak{O}'(c,A), \qquad \text{the transpose of the observability matrix of } \{A, b, c\} \end{aligned} \tag{11}$$

Hence the observer gain vector $l$ can be calculated as

$$l = \mathfrak{O}^{-1}(c,A)\mathfrak{A}^{-1}(\alpha - a)' \tag{12}$$

It is perhaps somewhat unexpected that asymptotic observability requires the same condition—the nonsingularity of $\mathfrak{O}(A, c)$—as exact observability does. But some reflection on what *unobservable states* are (cf. Sec. 2.4.2) will provide an explanation.

The duality between the asymptotic observer problem and the modal controller problem is quite striking and useful and can be used as above to

translate almost all the results of Chapter 3 to the present context. We shall leave these extensions to the reader (and partly to the exercises).

**Example 4.1-1. The Pointer-Balancing Problem**

For the pointer-balancing problem of Example 3.3-1, design an observer with modes at $-10 \pm j10$.

For $\varphi(0) = 0.1$, $\dot{\varphi}(0) = 0$, calculate and plot the observer errors $\tilde{\varphi}(\cdot)$ and $\dot{\tilde{\varphi}}(\cdot)$.

**Solution.** Recall that with $g/L = 9$ and $z = [\varphi \ \dot{\varphi}]'$ we have

$$\dot{z} = \begin{bmatrix} 0 & 1 \\ 9 & 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ -1 \end{bmatrix} u$$

$$y = [1 \quad 0]z$$

We want $\det(sI - A + lc) = (s + 10)^2 + 10^2 = s^2 + 20s + 200 = \alpha_0(s)$, and a simple calculation yields $l' = [20 \quad 209]$.
Let $\hat{z}(0) = 0$; then $\tilde{z}(0) = z(0) = [0.1 \ 0]'$ and

$$\tilde{Z}(s) = (sI - A + lc)^{-1}\tilde{z}(0)$$

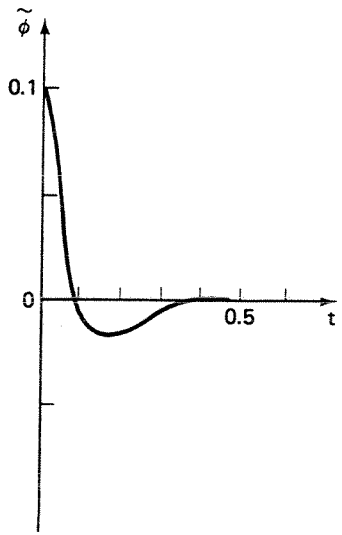
$$= \frac{1}{\alpha_0(s)} \begin{bmatrix} s & 1 \\ -200 & s + 20 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0 \end{bmatrix}$$

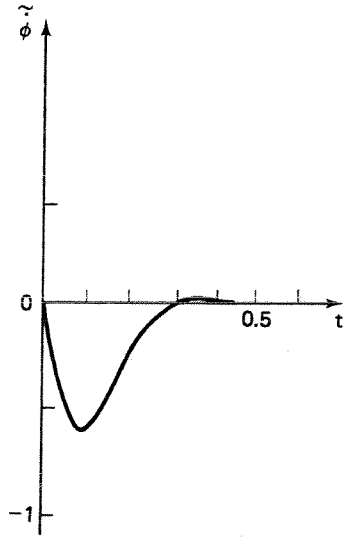
Therefore

$$\tilde{z}_1(t) = \tilde{\varphi}(t) = \mathcal{L}^{-1}\left[\frac{0.1s}{\alpha_0(s)}\right]$$

$$= 0.1\mathcal{L}^{-1}\left[\frac{(s + 10)}{(s + 10)^2 + 10^2} - \frac{10}{(s + 10)^2 + 10^2}\right]$$

$$= 0.1e^{-10t}(\cos 10t - \sin 10t)$$



Similarly, $\tilde{\varphi}(t) = -2e^{-10t}\sin 10t$. These are sketched in the figure. Note that in about three time constants ($3 \times \frac{1}{10}$) the observer error is down to a small fraction of its maximum value.

**Example 4.1-2. The Need for Proper Modeling**

Discuss the design, as $\epsilon \longrightarrow 0$, of observers for a realization

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

Note that $x_1(\cdot)$ does not depend on $x_2(\cdot)$, and therefore as $\epsilon \longrightarrow 0$, $x_2(\cdot)$ tends to disappear completely from the observation $y(\cdot)$. Therefore $x_2(\cdot)$ becomes unobservable as $\epsilon \longrightarrow 0$, and we should expect some difficulties in trying to estimate it.

**Solution.** Since this is an observer problem, we can, without loss of generality, assume that $u(\cdot) \equiv 0$. We can now readily calculate the observer gain required to obtain any set of observer poles, but we shall see that the parameter $\epsilon$ will give us indications of potential difficulties. For example,

$$\Theta = \begin{bmatrix} 1 & \epsilon \\ -1 + 2\epsilon & -\epsilon \end{bmatrix}, \qquad \det \Theta = -2\epsilon^2$$

so that $\Theta$ is almost singular for small $\epsilon$. Therefore we would expect trouble even if perfect differentiation were feasible. Thus note that

$$\Theta^{-1} = -\frac{1}{2\epsilon^2} \begin{bmatrix} -\epsilon & -\epsilon \\ 1 - 2\epsilon & 1 \end{bmatrix}$$

Therefore, if, for example, $\epsilon = 0.01$, we see that the ideal observer will give

$$\hat{x}_1(t) = 50[y(t) + \dot{y}(t)]$$
$$\hat{x}_2(t) = -4900y(t) - 5000\dot{y}(t)$$

while if $\epsilon = 0.02$, the estimates change quite a bit to

$$\hat{x}_1(t) = 25[y(t) + \dot{y}(t)]$$
$$\hat{x}_2(t) = -1200y(t) - 1250\dot{y}(t)$$

Such difficulties will persist no matter what method of state estimation is used. In this problem, when $\epsilon$ is small we see that $y(\cdot)$ depends almost completely on $x_1(\cdot)$, and it is better to represent the model by the equations

$$\dot{x}_1(t) = -x_1(t) + u(t)$$
$$y(t) = x_1(t)$$

for which no observer is needed. Of course this is because the $x_2(\cdot)$ equation is

stable so that $x_2(\cdot)$ will tend to die out with time. If $x_2(\cdot)$ is unstable, then we shall have trouble, and we should reexamine the original model to see if it can be modified.

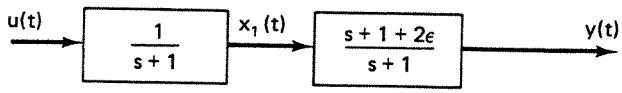
The point is that one should not just blindly plug into a mathematical formula but should also examine the solution and the model for sensitivity to parameter changes, *singular* phenomena, etc.

In this problem, some other clues to the potential difficulties might have been recognized from the transfer function,

$$H(s) = \frac{s + 1 + 2\epsilon}{(s + 1)^2}$$

which will show that, for small $\epsilon$, we really have a first-order system. The pole-zero cancellation for $\epsilon = 0$ shows that we shall lose observability or controllability of the realization. A block diagram of the realization (see the figure) shows that as $\epsilon \longrightarrow 0$, a pole of the first subsystem cancels a zero of the second subsystem, showing (according to Example 2.4-4) that we have unobservability of the overall realization.

The active reader should try to develop the *duals* for controllability of the above problem and remarks.




***Exercises***

***4.1-1.*** *Ackermann's Formula*

Obtain a formula for $l$ in terms of $\{A, c\}$ and the coefficients $\{\alpha_i\}$ of the desired characteristic polynomial, $\alpha(s)$.

***4.1-2.*** *Why Not Feedback to the Input?*

In Chapter 3, we used feedback to the input according to $u(t) = v(t) - kx(t)$. We could do this for the observer $u(t) = v(t) + l[y(t) - \hat{y}(t)]$, where, of course, $l$ is now a scalar. What can be achieved with such feedback? Why is it that we use feedback to the states in the observer problem but not in the controller problem?

***4.1-3.***

Is the asymptotic observer of an observable system itself observable for all possible $l$? Give a proof.

***4.1-4.*** *Feedback and Observability*

Show that the observability of a realization $\{A, b, c\}$ is not invariant under general state feedback $(u \longrightarrow v - kx)$ but is invariant under linear or nonlinear output feedback $\{u(t) \longrightarrow v(t) - f[y(s)], s \leq t\}$.

***4.1-5.*** *Another Approach to Observers* [3]

If $\dot{x}(t) = Ax(t) + bu(t)$, $y(t) = cx(t)$, $x(t_0) = x_0$, let $\hat{x}(\cdot)$ obey $\dot{\hat{x}}(t) = F\hat{x}(t) + gu(t) + hy(t)$, $\hat{x}(t_0) = \hat{x}_0$. The second equation can be said to define an *observer* for the first if $x_0 = \hat{x}_0 \Longrightarrow x(t) = \hat{x}(t)$, $t \geq t_0$. Show that a necessary and sufficient condition for this is that $F = A - kc$, $h = k$, $g = b$, where $k$ is an arbitrary $n \times 1$ vector.

*4.1-6. Effects of Mismatched Models*

Given a realization $\dot{x}(t) = Ax(t) + bu(t), y(t) = cx(t)$, consider an *observer* $\dot{\chi}(t) = \hat{A}\chi(t) + \hat{b}u(t) + v(t)$, $\hat{y}(t) = \hat{c}\chi(t)$, $v(t) = \hat{b}l(y - \hat{y})$, where the $\{\hat{A}, \hat{b}, \hat{c}\}$ are *estimates* (approximations to) of $\{A, b, c\}$. Show that $\dot{\epsilon}(t) = [\hat{A} - \hat{b}l\hat{c}]\epsilon(t) + [\delta A - \hat{b}l(\delta c)]x(t) + (\delta b)u(t)$, $\epsilon(t) = x(t) - \chi(t)$, where $\delta A = A - \hat{A}$, $\delta b = b - \hat{b}$, and $\delta c = c - \hat{c}$. *Note:* Further analysis allows us to relate these results to questions of system sensitivity (as originally conceived by Bode (1945): See W. A. Porter, *IEEE Trans. Autom. Control*, **AC-22**, pp. 144–146, Feb. 1977.

*4.1-7.*

Design an observer for the oscillatory system $\dot{x}(t) = v(t)$, $\dot{v}(t) = -\omega_0^2 x(t)$ using measurements of the velocity $v(\cdot)$. Place both observer poles at $s = -\omega_0$.

*4.1-8. Station-Keeping Satellite*

For the station-keeping satellite of Example 3.3-2, design an observer using measurements $y(\cdot)$ of azimuthal position perturbation. Place the observer poles at $s = -2\omega$, $s = -3\omega$, $s = -3\omega \pm j3\omega$, which means that the estimate errors will decay in about $2\frac{1}{2}$ days ($\doteq \frac{1}{2}\omega$, where $\omega = 2\pi/29.3$ rad/day).

*4.1-9. Deadbeat Observers in Discrete-Time*

Develop observer designs for discrete-time systems and show how to design an observer for which the error will go to zero in no more than $n$ steps, where $n$ is the number of states.

## 4.2 COMBINED OBSERVER-CONTROLLER COMPENSATORS

We were led to the observer problem by the need to obtain the states for use in the controller. Now we only have asymptotically correct estimates of the states rather than the states themselves, and a natural question is whether our previous result on arbitrary pole placement via state-variable feedback will continue to hold when only such estimates of the actual states are available. We have no option but to see what happens with the estimators that we have available.

In steady state there should clearly be no loss in using the asymptotic observer, since the error in the estimates will be zero. Therefore, as we shall soon confirm, the transfer function of the combined observer-controller will be just that of a pure controller (with perfect state feedback). However, this is not our major worry since, as noted in Sec. 3.1.1, there are many ways of arranging for a desired transfer function. The real question for the present scheme is whether the incorporation of the observer dynamical system into the feedback loop will affect the *stability* of the overall system—the point being that interconnections of stable subsystems may lead to unstable overall systems (cf. Sec. 3.1). However, we shall now prove the nice result that *incorporation of a stable asymptotic observer does not impair stability.*

We begin our analysis by setting up a joint observer-controller system. This is provided by Fig. 4.2-1, which the reader should be able by now to set up on his own. However, some brief words of explanation might be helpful. The original system may be described by its transfer function $H(s)$ or by a realization $\{A, b, c\}$. If we have $H(s)$, we must set up a realization in any form convenient for us (e.g., controller form) so as to be able to use state-variable feedback. If the states are directly measurable, we can close the loop through the feedback gain vector $k$. Otherwise we have to use state estimates $\hat{x}(\cdot)$, which we obtain by setting up a dummy system driven by the *error* term $l[y(\cdot) - c\hat{x}(\cdot)]$ and *also* by the same input $[u(\cdot) = v(\cdot) - k\hat{x}(\cdot)]$ as enters the state equations of the original system. This last fact should be well understood: If this input is not fed back to the observer, its states $\hat{x}(\cdot)$ will certainly not follow the states $x(\cdot)$. We can see this mathematically as well. Thus, for the scheme shown in Fig. 4.2-1, the equations for the overall system can be written down by inspection as

$$\dot{x}(t) = Ax(t) - bk\hat{x}(t) + bv(t), \qquad x(0-) = x_0$$
$$\dot{\hat{x}}(t) = lcx(t) + (A - lc - bk)\hat{x}(t) + bv(t), \qquad \hat{x}(0-) = \hat{x}_0$$

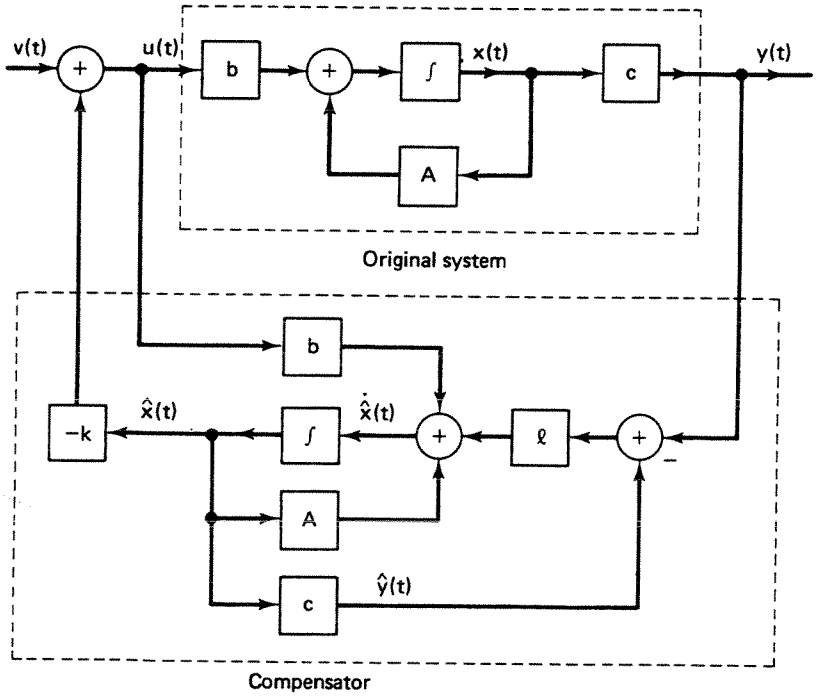



**Figure 4.2-1.** Combined observer-controller. Note that the observer is driven by both the output $y(\cdot)$ and the input $u(\cdot)$ $[= v(\cdot) - kx(\cdot)]$ of the original system.

The error in the estimates obeys the equation

$$\begin{aligned}\dot{\tilde{x}}(t) &= \dot{x}(t) - \dot{\hat{x}}(t)\\ &= Ax(t) - A\hat{x}(t) - lcx(t) + lc\hat{x}(t)\\ &= (A - lc)\tilde{x}(t), \qquad \tilde{x}(0-) = x_0 - \hat{x}_0\end{aligned} \tag{1}$$

and the effect of the input, $u(\cdot) = v(\cdot) - k\hat{x}(\cdot)$, is cancelled out. To study the overall system we arrange the equations for $x(t)$ and $\hat{x}(t)$ in matrix form as

$$\begin{bmatrix}\dot{x}(t)\\ \dot{\hat{x}}(t)\end{bmatrix} = \begin{bmatrix}A & -bk\\ lc & A - lc - bk\end{bmatrix}\begin{bmatrix}x(t)\\ \hat{x}(t)\end{bmatrix} + \begin{bmatrix}b\\ b\end{bmatrix}v(t), \qquad \begin{bmatrix}x(0-)\\ \hat{x}(0-)\end{bmatrix} = \begin{bmatrix}x_0\\ \hat{x}_0\end{bmatrix} \tag{2a}$$

$$y(t) = cx(t) \tag{2b}$$

**Calculation of the Transfer Function.** When computing transfer functions we have to assume zero initial conditions. Doing this and then taking Laplace transforms, we obtain

$$sX(s) = AX(s) - bk\hat{X}(s) + bV(s) \tag{3a}$$

$$s\hat{X}(s) = lcX(s) + (A - lc - bk)\hat{X}(s) + bV(s) \tag{3b}$$

Then eliminating $\hat{X}(s)$, we get

$$\begin{aligned}&[sI - A + bk(sI - A + bk + lc)^{-1}lc]X(s)\\ &\qquad = [I - bk(sI - A + bk + lc)^{-1}]bV(s)\end{aligned} \tag{4}$$

It seems difficult to proceed, but if we use the important matrix-inversion identity (Exercise A-21)

$$[I + C(sI - A)^{-1}B]^{-1} = I - C(sI - A + BC)^{-1}B$$

we can rewrite (4) as

$$\begin{aligned}bV(s) &= [I + bk(sI - A + lc)^{-1}]\{sI - A + lc\\ &\qquad - [I - bk(sI - A + bk + lc)^{-1}]lc\}X(s)\\ &= (sI - A + lc + bk - lc)X(s)\\ &= (sI - A + bk)X(s)\end{aligned}$$

Therefore

$$X(s) = (sI - A + bk)^{-1}bV(s)$$

so that the transfer function of the overall system is

$$H_{o\text{-}c}(s) = c(sI - A + bk)^{-1}b \tag{5}$$

(where the subscript *o-c* stands for observer-controller). These are the same equations we would have had with perfect state feedback; i.e., the overall

transfer function is just that of the controlled system and does not depend on the dynamics of the observer. In other words, the transfer function does not depend on how quickly the error in the state estimates goes to zero.

The explanation is that when the initial conditions for $x(\cdot)$ and $\hat{x}(\cdot)$ are both zero, and therefore the *same*, then of course the observer, when driven with the same inputs as the original system, will have the same outputs as the original system (see our earlier discussions in Sec. 4.1). That is, $x(t) = \hat{x}(t)$ when $x(0) = 0 = \hat{x}(0)$. Therefore, when making transfer function calculations, the asymptotic observer is the same as the perfect observer!

This is a nice result, but our major concern of course is with the modes of the overall realization. These will be the roots of the characteristic equation of realization (2), namely

$$a_{o\text{-}c}(s) = \det \begin{bmatrix} sI - A & bk \\ -lc & sI - A + lc + bk \end{bmatrix} \tag{6}$$

We can simplify this determinant by obvious row and column transformations to obtain,

$$a_{o\text{-}c}(s) = \det (sI - A + bk) \det (sI - A + lc) \tag{7}$$

$$= a_{\text{cont}}(s) a_{\text{obs}}(s), \text{ say.} \tag{8}$$

That is, the characteristic polynomial of the overall system is just the product of the characteristic polynomial of the observer and the characteristic polynomial of the controlled system assuming perfect knowledge of the states. This is nice, because it means that the natural frequencies or modes of the overall system can always be arranged to be stable. In fact, they can be chosen completely arbitrarily (if the original realization is controllable and observable). For by choosing $k$ as in Sec. 3.2, we can make $a_{\text{cont}}(s) = \det (sI - A + bk)$ arbitrary. Therefore, if $a_{\text{cont}}(s)$ is stable (i.e., has no roots with zero or positive real parts) and $a_{\text{obs}}(s)$ is stable, then $a_{o\text{-}c}(s)$ is stable. This is a fundamental result and one that is not a priori obvious because, as noted earlier, there are many situations where the interconnection of stable systems leads to an unstable system.

Another useful consequence of (7) and (8) is that the *controller and observer can be designed independently of each other*. Whether the true states are available, or only asymptotically correct estimates of the states, is immaterial to the calculation of the feedback gains $k$; similarly, the dynamics of the asymptotic observer can be calculated from knowledge of $A$ and $c$ without caring if the observer is to be combined with a feedback controller or not. This is the so-called *separation property* of the observer-controller design procedure. However there will generally be a deterioration in the transient response of the combined system—see Example 4.2-1 below.

**Implementation of the Observer.** The fact that the observer has as many states as the original system might be disturbing—does this mean that we essentially have to replicate the original system in order to achieve the above nice results? The original system might be a rather complicated power plant, chemical processor, etc. Fortunately, we recognize that except under very special circumstances the observer need not be constructed in the same way as the original system—in particular, we can build the observer with (miniaturized) electronic circuitry, with great advantages of size, ruggedness, cost, etc. In fact, with modern developments in integrated electronics, we can readily envisage the use of very high-order observers for help in the control of quite complicated physical systems. Thus, in several applications, it will be feasible to implement the observer via a microprocessor specially designed to integrate the observer state equations.

**Summary.** We have now obtained some clear and definite answers to several of the questions we raised in Sec. 3.1.1. Briefly, by using feedback of the states of a completely controllable and completely observable realization of the original transfer function, we can obtain a new internally stable realization whose natural frequencies are completely under our control. While we have obtained this result only for scalar (single-input, single-output) systems, it also has a fairly natural extension to the multivariable case (Sec. 7.4) and in this sense represents perhaps the major triumph in the state-space approach to linear systems. These points are worth further discussion, which we shall begin to pursue further in Example 4.2-3.

**Example 4.2-1.**

In this example we shall illustrate the sort of response obtained by using an observer to generate state estimates for feedback. The results were obtained by simulation on an analog computer.

The system we shall consider corresponds to the pointer-balancing problem of Example 3.3-1, except that we now take $g/L = 1$ for convenience of simulation. We then have

$$\dot{z} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ -1 \end{bmatrix} u$$

$$y = [1 \quad 0]z$$

and the system modes are at $\lambda = \pm 1$. An analog simulation of the system is shown in Fig. a.

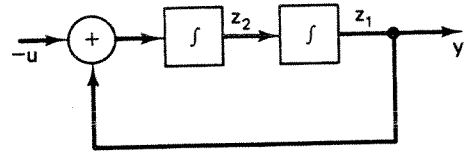



**Figure a.** Simulation of system.

The system is clearly unstable, as indicated in Fig. b by the response to a nonzero initial condition.

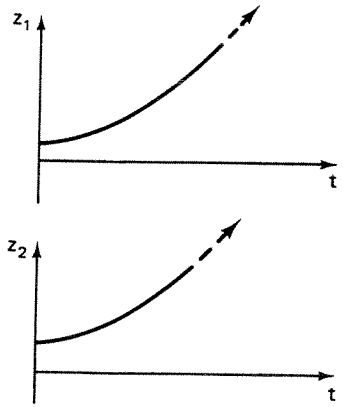



**Figure b.** Unstabilized response.

We use state feedback $u = -kz$ to move the system poles to $-0.5 \pm j0.5$. The required $k$ is easily obtained as $k = [-1.5 \quad -1]$. The decay of the system from an initial state to the origin, using this direct state feedback, is shown in Fig. c.

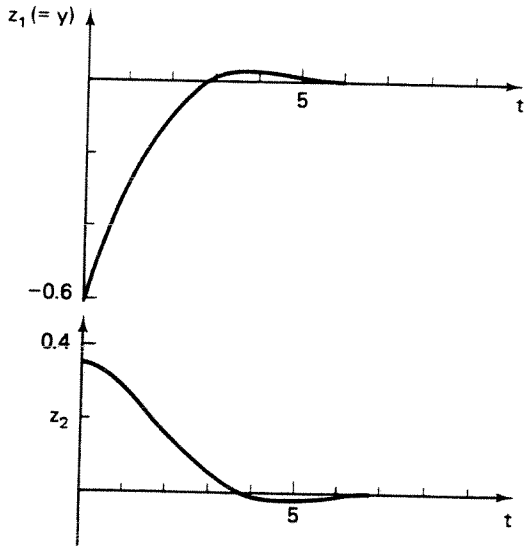



**Figure c.** Response with direct state feedback.

We now construct an observer for the system. The observer modes are chosen to lie at $-1 \pm j1$. (They are faster than the desired closed-loop poles but slow enough for us to see clearly their effect on the system response. A brief discussion of the factors involved in selecting reasonable observer modes is presented in Sec. 4.4.)

The observer gain is readily found as $l = [2 \quad 3]'$, and the observer is then easily constructed.

We now use the observer states to generate the required control, $u = -k\hat{z}$, and the resulting compensator is shown in Fig. d.

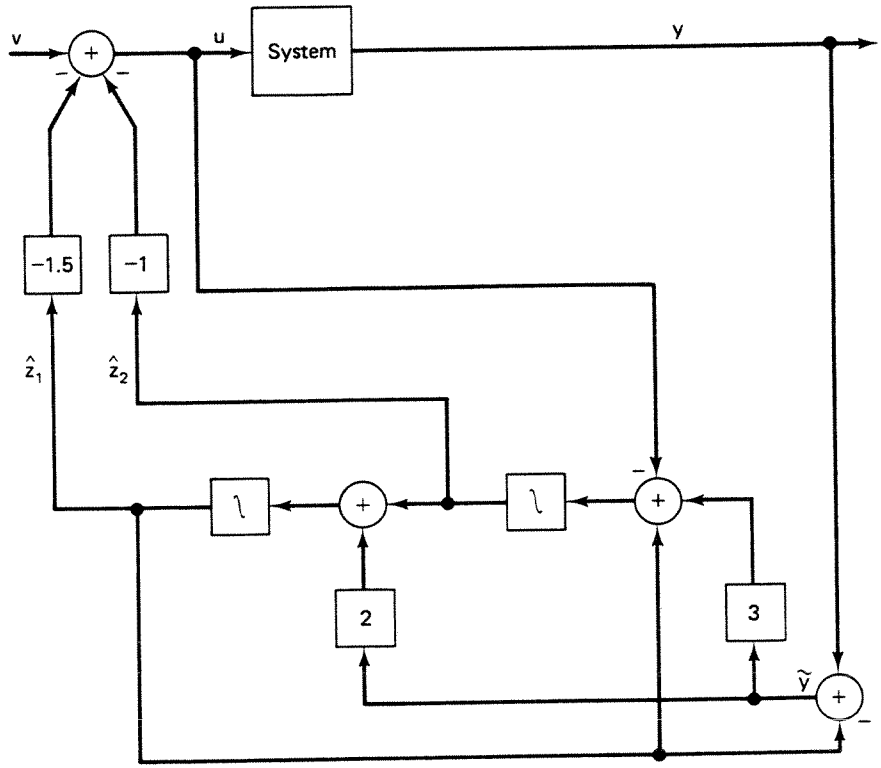



**Figure d.** Compensator structure.

The response of this closed-loop system is indicated in Fig. e. Note the manner in which the observer error decays; it is clear that the observer error goes to zero at about twice the rate at which the system settles to zero, as is to be expected from the fact that the error poles are twice the closed-loop poles. Comparison of Figs. c and e indicates the deterioration in response due to the use of estimated rather than actual states for feedback. In Fig. f we compare the controls resulting from direct state feedback and from use of the observer-controller compensator structure; the initial large-amplitude oscillation of the control in the latter case is again a result of the large initial uncertainty as to the true state of the system.

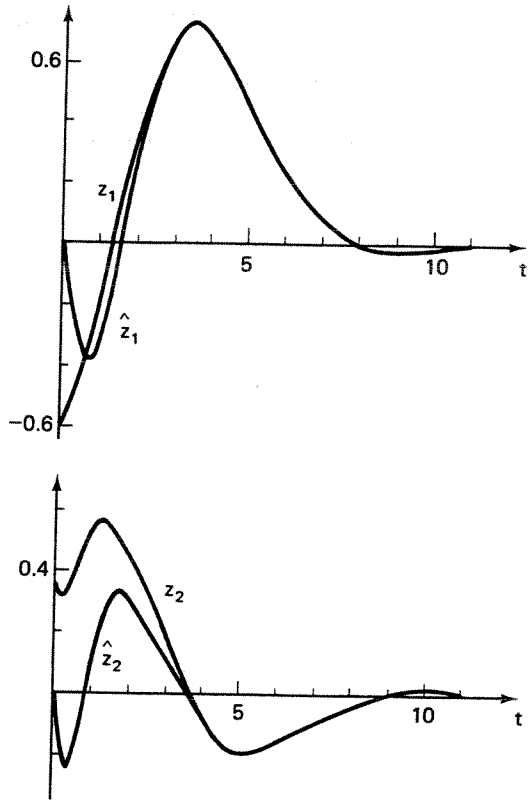



**Figure e.** Response using compensator of Fig. d.

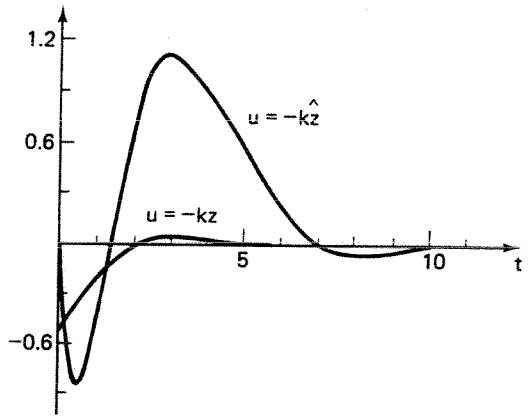



**Figure f.** Comparison of controls generated by direct state feedback and using the compensator.

**Example 4.2-2. Constant Disturbances and Integral Feedback**

For a system driven by a *constant* unknown disturbance $w$, design an observer to estimate $w$, and use this to compensate for the disturbance.

**Solution.** We have

$$\dot{x} = Ax + bu + bw$$
$$\dot{w} = 0$$
$$y = cx$$

where $u$ is the control input and $y$ the observed output. The constant disturbance $w$ is modeled as the output of an undriven integrator. We then have the *augmented* system shown in Fig. a.

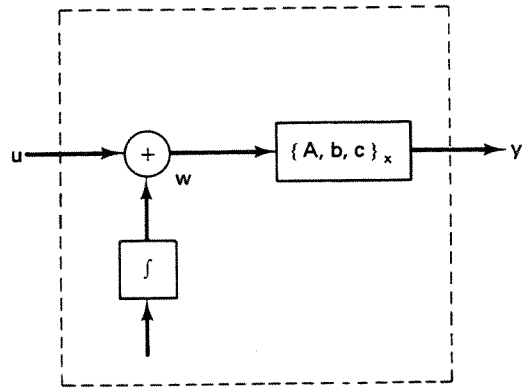



**Figure a.**

If now we had an estimate $\hat{w}$ of $w$, we could set $u = -\hat{w}$ to attempt to cancel out the disturbance. This motivates us to set up an observer to estimate $w$.

An observer for the augmented system is given by

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{w}} \end{bmatrix} = \begin{bmatrix} A & b \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{w} \end{bmatrix} + \begin{bmatrix} b \\ 0 \end{bmatrix} u + l(y - c\hat{x})$$
$$\hat{x}(0) = 0, \qquad \hat{w}(0) = 0$$

where $l$ is an $(n + 1) \times 1$ vector. Partitioning $l$ as $[l_1' \;\; l_2']'$, with $l_2$ a scalar, we get

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{w}} \end{bmatrix} = \begin{bmatrix} A - l_1 c & b \\ -l_2 c & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{w} \end{bmatrix} + \begin{bmatrix} b \\ 0 \end{bmatrix} u + \begin{bmatrix} l_1 \\ l_2 \end{bmatrix} y$$

The observer structure is then as shown in Fig. b. Now *if* the augmented system is observable, we can choose $l$ so as to obtain arbitrary error decay modes and thus ensure that $\hat{w}$ approaches $w$ asymptotically. Let us temporarily ignore the question

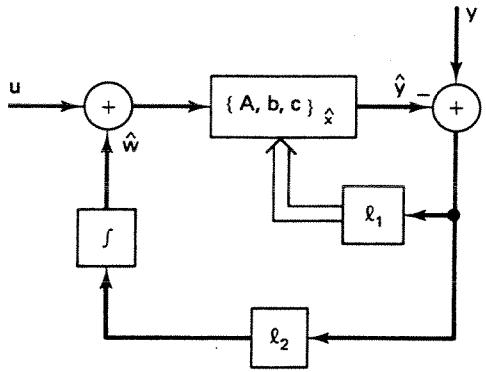



**Figure b.**

of observability. We also make a particular choice of $l$ as $[0 \;\; l_2]$, which will simplify our observer considerably; for the moment we shall not worry about whether this can still ensure that the error-decay modes are stable. Now, on setting $u = -\hat{w}$, our observer equation reduces to

$$\begin{bmatrix} \dot{\hat{x}} \\ \dot{\hat{w}} \end{bmatrix} = \begin{bmatrix} A & 0 \\ -l_2 c & 0 \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{w} \end{bmatrix} + \begin{bmatrix} 0 \\ l_2 \end{bmatrix} y, \qquad \hat{x}(0) = 0, \qquad \hat{w}(0) = 0$$

Since the equation for $\hat{x}$ is undriven and the initial condition is zero, we have $\hat{x} \equiv 0$, and our observer is simply

$$\dot{\hat{w}} = l_2 y, \qquad \hat{w}(0) = 0$$

The resulting overall compensation scheme is shown in Fig. c (where the dashed lines indicate parts that drop out of the compensator).

The result of the above procedure is thus precisely the technique that was presented in Sec. 3.2.3 for compensation of constant unknown disturbances, namely *integral feedback*. It arises here in a more natural and motivated manner.

The question we have so far avoided is whether proper choice of $l_2$ can ensure that $\hat{w}$ approaches $w$.

Our earlier observer equation shows that the observer error behavior is determined by the roots of (see Exercise A-11)

$$\alpha(s) = \det \begin{bmatrix} sI - A & -b \\ l_2 c & s \end{bmatrix} = \det (sI - A) \det [s + l_2 c(sI - A)^{-1} b]$$

$$= sa(s) + l_2 b(s) = 0$$

We assume now that the original system $\{A, b, c\}$ was stable (or stabilized) and hence that $a(s)$ is stable, i.e., has roots with strictly negative real parts. It can then be shown that proper choice of $l_2$ can give stable $\alpha(s)$ if and only if $b(s)$ has no root

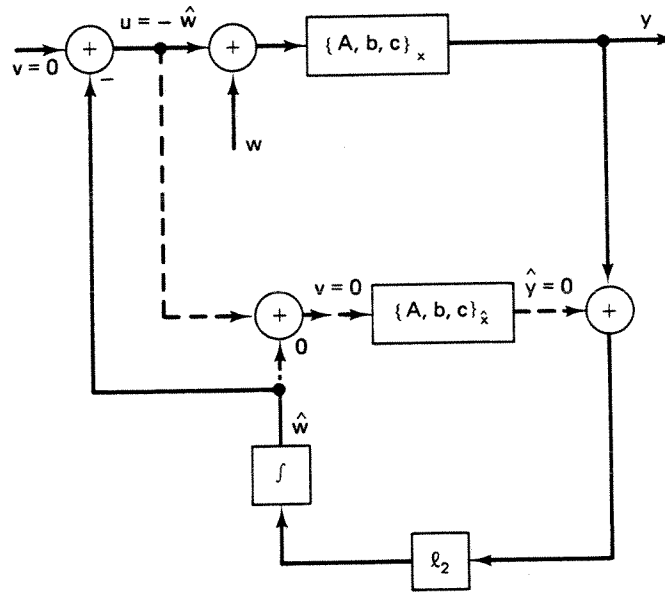



**Figure c.**

at the origin. [Necessity is obvious, since if $b(s)$ contained $s$ as a factor, then $\alpha(s)$ would also contain this unstable factor; initial observer error would then not decay to zero. Sufficiency is easily proved using root-locus-type arguments, but we shall omit these here.]

The case where $b(s)$ has a root at the origin (i.e., where the original system has a zero at the origin) is, however, rather trivial, since the presence of a zero at the origin makes the output of the system insensitive to a constant input anyway (see Sec. 3.2.3). Another way of understanding the above fact is to note (see Example 2.4-4) that if $b(s)$ contained $s$ as a factor then the augmented system (from the disturbance initial conditions to $y(\cdot)$) would be unobservable due to the cancellation of the term $s$.

This is a simple example of the so-called "internal model" principle for the rejection of disturbances.

**Example 4.2-3. A Review of the Overall Observer-Controller Design Problem**

We are given a system with transfer function $H(s)$. The *system* may be a process, power plant, machine, market, bacterial colony, etc. We have control over one variable, the *input*, and are interested in controlling the behavior (evolution, response) of some other system variable, the *output*. The system given to us is characterized by a certain input-output behavior, say,

$$H(s) = \frac{(s-2)(s-3)}{s(s-1)(s-4)} = \frac{s^2 - 5s + 6}{s^3 - 5s^2 + 4s}$$

The system may be unstable (as above), or its response to a given input may not be what we wish it to be. Suppose we wish to relocate the poles at $-1 \pm j$ and $-5$ (guided perhaps by the optimum root locus—draw this!).

Cascade compensation, which cancels the unstable poles, is unsatisfactory.

Output feedback, $u \longrightarrow v - ky$, will not suffice; we have one free parameter and three quantities to be independently adjusted.

Position, velocity, and acceleration feedback, $(k_2 + k_1 s + k_0 s^2)y$, gives us enough parameters, but the system order increases to 4,

$$H(s) = \frac{(s-2)(s-3)}{s(s-1)(s-4) + (k_2 + k_1 s + k_0 s^2)(s-2)(s-3)}$$

Other strategies can be tried, as in Sec. 3.1.1, without leading to any obvious clear answer, and therefore we shall try to approach the problem with state-space ideas. We know that to modify the system arbitrarily we must know what the system is doing. And this (by definition, or by construction, . . .) is told us by the state variables. Suppose then that someone gives us these state variables (or states, for short). This statement by itself is meaningless. One cannot talk of states apart from the *realization* they correspond to. To make sense, we have to be given measurements (terminals, . . .) that define the state variables $x(\cdot)$ *of a realization* $\{A, b, c\}$. How was the realization obtained by the person who gives us the states? Again, it is meaningless to talk of *the* realization. What the person had was a set of equations describing the system; these were transformed (and perhaps linearized) to *a* set of first-order differential equations in the variables $x(\cdot)$, which are now made available. If the realization is controllable, we can use linear state feedback, $kx(\cdot)$, with $k$ chosen to obtain any desired free response (natural frequencies).

Assume that we have available the states of a controller-form realization of $H(s)$,

$$A_c = \begin{bmatrix} +5 & -4 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \qquad b_c = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$c_c = [1 \quad -5 \quad 6]$$

We would like a new characteristic polynomial

$$\begin{aligned} \alpha(s) &= (s+5)(s+1+j)(s+1-j) \\ &= s^3 + 7s^2 + 12s + 10 \end{aligned}$$

Now with state feedback,

$$A_c - b_c k_c = \begin{bmatrix} -7 & -12 & -10 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

which shows that

$$k_c = (\alpha - a) = [12 \quad 8 \quad 10]$$

Now how realistic is the assumption that the states are available? What is involved

in making the states available? Note that by making the states available we are in effect defining $n$ new outputs. It is then perhaps not so surprising that we can achieve so much more (namely arbitrary pole location) with state feedback than with feedback of the single original output. The obvious cost involved in making the states available is the cost of the sensors required to measure the states and the transducers needed to convert these measurements to useful quantities for control.

(There is a whole *dual* range of possibilities for system modification that arises from the definition of additional *inputs*, or, equivalently, from postulating access to the *inputs* of the *integrators* of the realization, but we shall not go into this here.)

So what do we do if the states are not available (or the cost of making them available is prohibitive)? If the system is observable, we can determine its state at any time by differentiation of the output and input. This is unrealistic because of noise considerations. If, however, we know the state of the system at some instant, we can construct a dummy system (realization) with this given initial state and drive it with the same input as the original system from that time on. (Note that this dummy system is only mathematically equivalent to the original—its actual implementation may be quite different, perhaps an analog-computer simulation, or often a *digital-computer algorithm* for integrating the state equations.) The (simulated) states are now available.

It is desirable to use an additional input to the dummy, namely $y - \hat{y} = \tilde{y}$ (the difference between the actual and dummy outputs). This allows asymptotic correction of errors in initial-state determination and in fact obviates the need for initial-state determination altogether. The system is still required to be observable, however, in order that the error dynamics may be arbitrarily chosen.

Returning to our example, we construct a dummy system $\{A_c, b_c, c_c\}$ with states $\hat{x}_c$, driven by an additional input $l_c(y - c_c\hat{x}_c)$. Suppose we want the observer poles at $-6 \pm j6$ and $-6$ (a "reasonable" choice for our given closed-loop pole locations). Then $l_c$ must be chosen such that the characteristic polynomial of $A_c - l_c c_c$ is

$$\alpha_l(s) = (s + 6)(s^2 + 12s + 72) = s^3 + 18s^2 + 144s + 432$$

Some algebra then shows that

$$l_c = [807.5 \quad 127.5 \quad -24.5]'$$

The combined observer-controller can now be implemented as in Figure 4.2-1. In practice, however, there are still problems to be overcome in the actual realization, one of the most important being *scaling*: We may need to transform the realization to a form where the required gains are of reasonable magnitude—see Exercise 2.2-19.

***Exercises***

***4.2-1.*** *Alternative State Equations*

The fact that $\tilde{X}(s) = X(s) - \hat{X}(s) = 0$ in (3a) and (3b) suggests that perhaps a more convenient set of state variables for the observer-controller system is $[x, \tilde{x}]$. Write state equations for these variables, and use them to calculate $H_{o\text{-}c}(s)$ and $a_{o\text{-}c}(s)$.

***4.2-2.*** *Controllability and Observability of the Observer-Controller Realization*

The realization in Fig. 4.2-1 is clearly not minimal.

a. Show that this realization is not controllable. What are the noncontrollable state variables?

b. Show that the realization will be nonobservable if and only if at least one of the following holds:

(1) $\{k, A - lc\}$ is not observable,

(2) $\{c, A - bk\}$ is not observable, or

(3) A pole of the observer cancels a zero of the original transfer function $H(s)$.

It is, of course, assumed that $\{A, b, c\}$ is minimal.

*4.2-3.*

In the combined controller-observer design we select $k$ and $l$ so that $a_c(s)$ and $a_o(s)$ both have poles in the left half plane. Is it true that the resulting design must be stable even if the loop is broken open at $y$, for instance? Explain your answer briefly.

*4.2-4.*

Consider the undamped harmonic oscillator $\dot{x}_1(t) = x_2(t)$, $\dot{x}_2(t) = -\omega_0^2 x_1(t) + u(t)$. Using an observation of velocity, $y(\cdot) = x_2(\cdot)$, design an observer/state-feedback compensator to control the position $x_1(\cdot)$. Place the state-feedback controller poles at $s = -\omega_0 \pm j\omega_0$ and both observer poles at $s = -\omega_0$.

## *4.3 REDUCED-ORDER OBSERVERS

The observer design method described in Sec. 4.1 is due, as noted earlier, to Bertram (1961) and Bass (1963). In his Ph.D. thesis (Stanford, 1963) Luenberger took a different approach to the observer problem. We shall not really discuss this approach here but shall focus instead on its most significant contribution. The observer obtained in Sec. 4.1 had $n$ states, where $n$ was the order of the realization whose states were being observed. Luenberger [3] pointed out that the order of the observer can actually be less than this because the observed output provides a linear relationship $y(t) = cx(t)$ between the state variables. Therefore it suffices to observe $n - 1$ of the states and then to calculate the final one from this linear relationship.

The reduction by one in observer dimension is not particularly significant, especially when the observer is implemented with (integrated-circuit) electronic logic. However, the question has some interesting theoretical aspects, which we shall explore here. Moreover, for multioutput systems, somewhat more substantial reductions can be obtained (cf. Sec. 7.3). In this section, we shall essentially follow the historical approach to the topic, which starts with the unreduced observer of Sec. 4.1. The reduced-order observer will arise in a more direct way in the frequency-domain design procedure to be presented in Sec. 4.5.1.

We shall first explain why reduction is possible and describe Luenberger's idea [3] for exploiting it. However for the detailed calculations we shall use a different technique ([4] and [5]).

Let us begin with a realization $\{A, b, c\}$ and make an easily found (and non-unique) similarity transformation such that the output vector has the form†

$$c = [0 \quad \cdots \quad 1] \tag{1}$$

and

$$y = [0 \quad \cdots \quad 1]x(t) = x_n(t) \tag{2}$$

Therefore we have clearly displayed the fact that one state is directly observable,

$$\hat{x}_n(t) = y(t) = x_n(t)$$

and we only need to estimate the states $x_r = [x_1 \ \cdots \ x_{n-1}]$, where the subscript $r$ is used to denote the states that remain to be estimated. It should be possible to estimate these $n - 1$ states by an $(n - 1)$-dimensional observer. This is true, but the design has to be approached with a little care. Thus, consider the partitioned state equations

$$\begin{bmatrix} \dot{x}_r(t) \\ \dot{x}_n(t) \end{bmatrix} = \begin{bmatrix} A_r & b_r \\ c_r & a_{nn} \end{bmatrix} \begin{bmatrix} x_r(t) \\ x_n(t) \end{bmatrix} + \begin{bmatrix} g_r \\ g_n \end{bmatrix} u(t), \qquad y(t) = x_n(t) \tag{3}$$

where the reasons for our notation will become clear later. Now we set up a dummy system for $\hat{x}_r(t)$,

$$\dot{\hat{x}}_r(t) = A_r\hat{x}_r(t) + b_r y(t) + g_r u(t) + \text{(feedback term)} \tag{4}$$

The feedback term should be derived from the error $x_r(t) - \hat{x}_r(t)$, but not only is this term inaccessible, just as $x(t) - \hat{x}(t)$ was in the $n$th-order observer of Sec. 4.1, but the accessible quantity $[y(t) - c\hat{x}(t)]$ used in Sec. 4.1 is identically zero [with our choice (2) for $c$] if we insist that $\hat{x}_n(t) = y(t)$. Therefore, it seems difficult to get a *feedback* estimator for $x_r(t)$. However, it is worthwhile to persist a bit.

Note that without feedback, the *open-loop* equation (4) for the error $\tilde{x}_r(t) = \hat{x}_r - x_r(t)$ is

$$\dot{\tilde{x}}_r(t) = \dot{x}_r(t) - \dot{\hat{x}}_r(t) = A_r\tilde{x}_r(t), \qquad \tilde{x}_r(0) = x_r(0) - \hat{x}_r(0) \tag{5}$$

[Note again that we might as well have assumed the *known* inputs $b_r y(\cdot)$ and $g_r u(\cdot)$ to be zero.] We had a similar equation in Sec. 4.1 for the $n$-dimensional open-loop observer,

$$\dot{\tilde{x}}(t) = A\tilde{x}(t), \qquad \tilde{x}(0) = x(0) - \hat{x}(0) \tag{6}$$

†One such realization is obtained from our usual observer form $\{A_0, b_0, c_0\}$ by merely labeling the states in the reverse order to our conventional order. In fact, the reader will find it helpful to interpret various results stated for observable realizations by assuming the realizations to be in observer (or relabeled observer) form. For actual computation, however, it may be generally simpler to avoid transformation all the way to observer form and to merely find some convenient transformation that gives $c = [0 \ \cdots \ 0 \ 1]$.

and we concluded that the error dynamics were beyond our control because the eigenvalues of $A$ could not be changed by similarity transformations taking $A$ to $T^{-1}AT$. However, it was at this point that Luenberger made the key observation that even though the eigenvalues of a matrix $A$ could not be changed by similarity transformations, this is *not* true of eigenvalues *of submatrices of* $A$. (Note that if $\bar{A} = T^{-1}AT$, it is unlikely that [cf. (3) for the notation] $\bar{A}_r$ is similar to $A_r$.) Luenberger showed in fact that if the original equations are observable, then state transformations can be found that will yield a *submatrix* with arbitrary eigenvalues (see Exercise 4.3-1).

Luenberger's technique is ingenious and makes interesting use of different state realizations. However, it is somewhat confusing (at least initially) to keep changing realizations, since after all we are basically interested in only one realization, viz., the one that we assumed for the original system that was to be controlled by state feedback.

**Another Approach (Gopinath [4] and Cumming [5]).** For another approach we can return to the original equations (3) and note that so far we have only really examined the subset of equations describing the evolution of $x_r(\cdot)$. A direct approach using this set did not succeed—see (4)—because we could not obtain a suitable *feedback term*. However, we did not use the remaining equation

$$\dot{x}_n(t) = a_{nn}x_n(t) + c_r x_r(t) + g_n u(t)$$

which certainly does provide some more information about $x_r(\cdot)$. Thus, note that if we define

$$y_r(t) = \dot{y}(t) - a_{nn}y(t) - g_n u(t) \tag{7}$$

then we can write

$$\dot{x}_r(t) = A_r x_r(t) + b_r y(t) + g_r u(t) \tag{8}$$

$$y_r(t) = c_r x_r(t) \tag{9}$$

The function $y_r(\cdot)$ is completely determined by $y(\cdot)$ and $u(\cdot)$ and can therefore be regarded as an observation process. The difficulty is that $y_r(\cdot)$ contains the derivative $\dot{y}(\cdot)$ and cannot therefore be realistically obtained from $y(\cdot)$. However, if we temporarily ignore this difficulty, we see that Eqs. (8) and (9) display the $n - 1$ states $x_r(\cdot)$ in a form to which the feedback design of Sec. 4.1 can be applied.

That is, we can set up an observer of the form

$$\dot{\hat{x}}_r(t) = A_r\hat{x}_r(t) + b_r y(t) + g_r u(t) + l_r[y_r(t) - c_r\hat{x}_r(t)] \tag{10}$$

where $l_r$ is an $(n - 1) \times 1$ matrix. Now

$$\dot{\tilde{x}}_r(t) = (A_r - l_r c_r)\tilde{x}_r(t), \qquad \tilde{x}_r(0) = x_r(0) - \hat{x}_r(0) \tag{11}$$

and we see that if

$$\{c_r, A_r\} \text{ is observable} \tag{12}$$

then $\tilde{x}_r(\cdot)$ can be made to go to zero arbitrarily fast (but, of course, not faster than *exponentially*).

The observability of $\{c_r, A_r\}$ is to be expected from that of $\{c, A\}$ because (3) shows that the only way $y(\cdot) = x_n(\cdot)$ gives information about the remaining states $x_r(\cdot)$ is through $y_r(\cdot)$. In any case, a simple proof of this fact can be obtained by using the PBH tests.

Therefore, we now have a reduced-order observer, except for the fact that $y_r(\cdot)$ contains a derivative of the observation $y(\cdot)$, which is generally unacceptable. However, our experience with analog-computer simulations in Sec. 2.1 shows how to avoid this difficulty. The method is clear from Figs. 4.3-1 and 4.3-2. [Note that removing the differentiator (cf. Fig. 4.3-2) changes the input to the integrator from $\hat{x}_r(\cdot)$ to $\hat{x}_r(\cdot) - l_r\dot{y}(\cdot)$ and changes the integrator output to $\hat{x}_r(\cdot) - l_r y(\cdot)$; therefore, $\hat{x}_r(\cdot)$ can be recovered by adding $l_r y(\cdot)$ to the output of the integrator.]

State equations can be written down from Fig. 4.3-2 to yield

$$\dot{\theta}(t) = (A_r - l_r c_r)\theta(t) + (b_r - l_r a_{nn} + A_r l_r - l_r c_r l_r)y(t) + (g_r - l_r g_n)u(t) \tag{13}$$

$$\hat{x}_r(t) = \theta(t) + l_r y(t) \tag{14a}$$

$$\hat{x}_n(t) = x_n(t) = y(t) \tag{14b}$$

as the equations for the observer.

Of course, there are other methods of implementing the equations we first obtained for the observer. Thus we could take Laplace transforms of (7)–(10) to

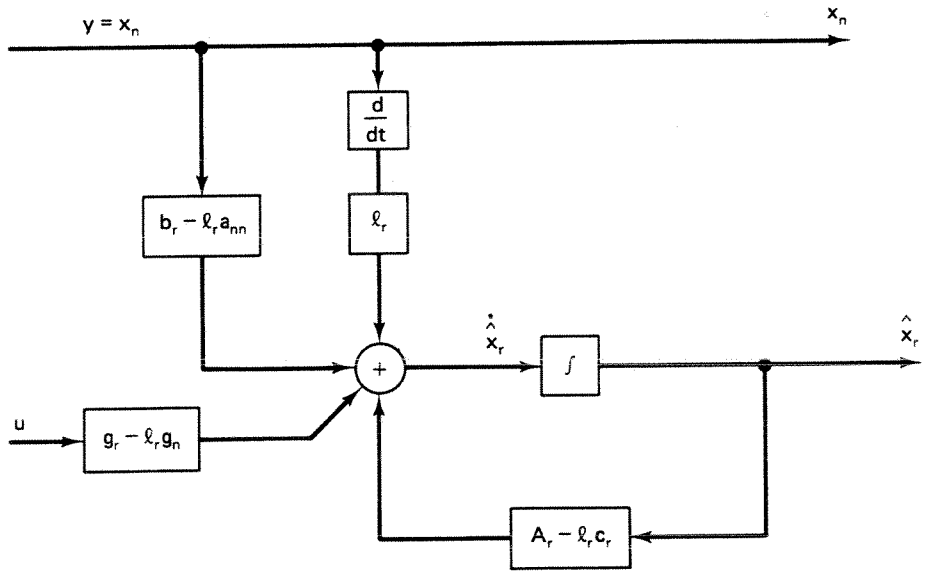



**Figure 4.3-1.** Implementation using differentiators.

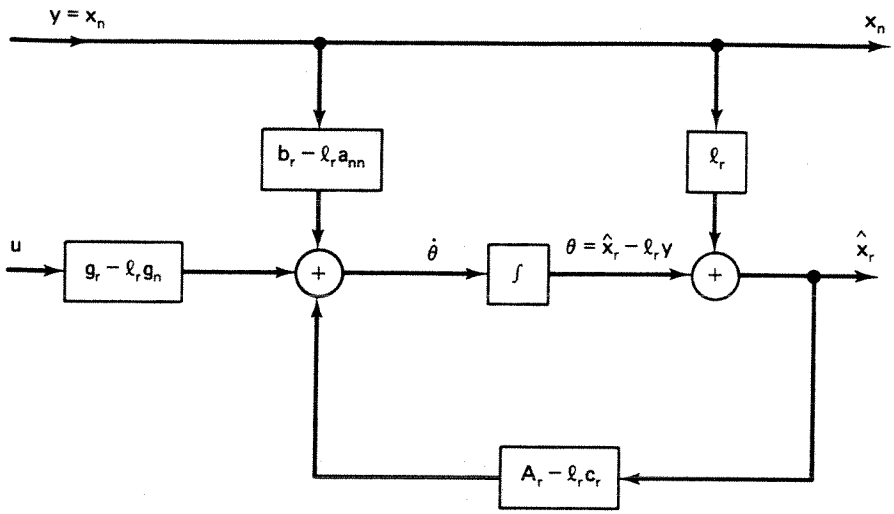



**Figure 4.3-2.** Equivalent implementation without differentiators.

obtain the transfer function

$$\hat{X}(s) = (sI - A_r + l_r c_r)^{-1}[(sl_r + b_r - l_r a_{nn})Y(s) + (g_r - l_r g_n)U(s)] \quad (15)$$

which can be simplified and then realized in several different ways. The advantage of the first technique is that the realization (Fig. 4.3-2) it yields contains the parameters $\{A_r, b_r, c_r, a_{nn}, l_r\}$ in a direct way, while if (15) is used, these various parameters may be recombined in a way that destroys their identity.

From Fig. 4.3-2 and (13)–(15) we see that the reduced-order observer can be straightforwardly incorporated into our compensator scheme.

**Example 4.3-1. Pointer-Balancing Problem**

We shall present here the results obtained by analog-computer simulation of the system of Example 4.2-1, now compensated with a reduced-order observer.

Recall that the system was described by

$$\dot{z} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} z + \begin{bmatrix} 0 \\ -1 \end{bmatrix} u, \qquad y = [1 \quad 0]z$$

and we had determined a state-feedback gain vector to move the poles from $\pm 1$ to $-0.5 \pm j0.5$, namely $k = [-1.5 \quad -1]$.

We design from first principles a reduced-order observer with pole at $-1.5$. Since $z_1$ is directly observed, we construct an observer for $z_2$. We thus have

$$\dot{\hat{z}}_2 = z_1 - u + l \times (\text{error signal})$$

Noting that $\dot{z}_1 = \dot{y} = z_2$, we see that a convenient error signal can be generated,

using only available measurements, as $\dot{y} - \hat{z}_2$, to give

$$\dot{\hat{z}}_2 = -l\hat{z}_2 + y - u + l\dot{y} \qquad \text{when } \dot{\tilde{z}}_2 = -l\tilde{z}_2$$

Choosing $l = 1.5$ gives the desired observer-error decay. To implement the observer without differentiation of $y$, we work with a new variable $\theta = \hat{z}_2 - ly$. Our observer equation is then

$$\begin{aligned}\dot{\theta} &= -l\theta - l^2 y + y - u \\ &= -1.5\theta - 1.25y - u\end{aligned}$$

and

$$\hat{z}_1 = y, \qquad \hat{z}_2 = \theta + 1.5y$$

Figure 4.3-3(a) shows the resulting compensator structure. In Fig. 4.3-3(b) we illustrate the closed-loop response using this compensator; the results shown are for $\theta(0) = +0.5$ (a value chosen for convenience of simulation), and, as before, $z(0) = [-0.6 \;\; 0.35]$. This response may be compared with that obtained by direct state feedback and by using a full-order observer in the compensator—see Example 4.2-1. Figure 4.3-3(c) shows the control generated by the reduced-order compensator.

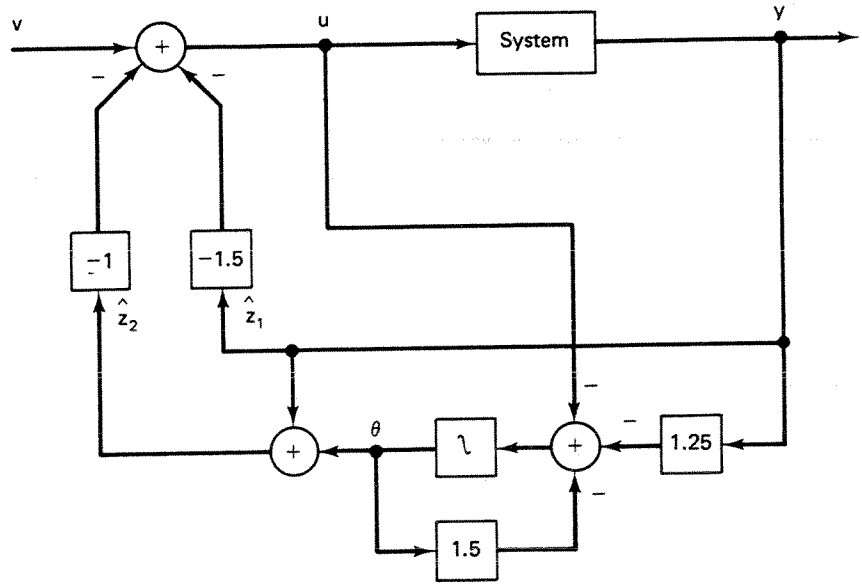



**Figure 4.3-3a.** Compensator structure.

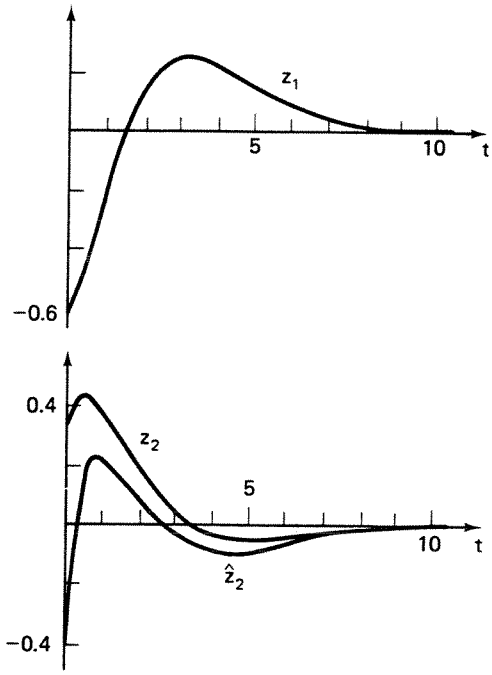



**Figure 4.3-3b.** Response using compensator of (a).

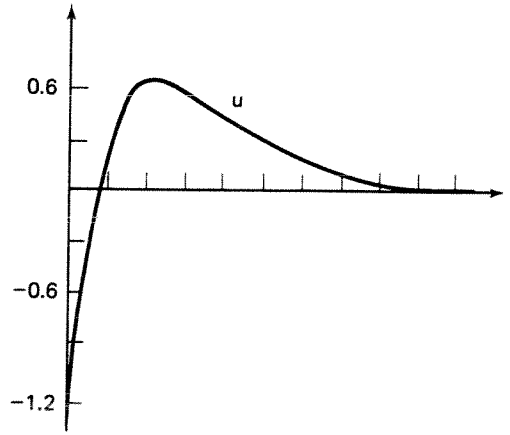



**Figure 4.3-3c.** Control generated by compensator of (a).

**Example 4.3-2. Observers and Some Classical Designs**

This example is solved in a different way (using several canonical forms) in Chen [9, p. 296–298]. The problem is to design a combined observer-controller that will relocate the poles of a system with transfer function $H(s) = 1/s(s + 1)$ at the new locations $\{s = -1 \pm j\}$, and to do this with a reduced-order observer that has a pole at $s = -2$. Because of the separation property, we can consider the controller and observer problems separately.

**Design of Controller.** We shall work with a realization of $H(s)$ in controller canonical form. This form can be obtained by inspection,

$$A_c = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}, \qquad b_c = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad c_c = [0 \quad 1]$$

Using state feedback $k_c x_c$, the row vector $k_c$ must be such that the characteristic polynomial of $[A_c - b_c k_c]$ will be

$$a_{\text{cont}}(s) = (s + 1 + j)(s + 1 - j) = s^2 + 2s + 2$$

It follows easily that

$$k_{c2} = 2$$

**Design of Reduced-Order Observer.** Without loss of generality we can assume that $u(\cdot) \equiv 0$. Then $x_{c2}(\cdot) = y(\cdot)$, and we only need to estimate $x_{c1}(\cdot)$. We have

$$\dot{x}_{c1}(t) = -x_{c1}(t), \qquad \dot{y}(t) = \dot{x}_{c2}(t) = x_{c1}(t)$$

and we define

$$y_r(t) = \dot{y}(t) = x_{c1}(t)$$

Therefore the reduced-order observer is

$$\dot{\hat{x}}_{c1}(t) = -\hat{x}_{c1}(t) + l[y_r(t) - \hat{x}_{c1}(t)]$$

where $l$ is to be such that $a_{\text{obs}}(s) = s + 2 = \det(sI - A_r + lc_r) = s + 1 + l$, which gives $l = 1$.

**Combined Observer-Controller.** We can combine the above results as shown in Fig. 4.3-4. Note that the input $u(\cdot) = v(\cdot) - k\hat{x}(\cdot)$ has to be fed to the observer. We have also moved $y(\cdot)$ across the integrator in the observer system so as to obviate the need for differentiation.

It will be of interest to consider various alternative configurations obtainable by transfer function manipulation. We begin by computing the transfer functions from $y(\cdot)$ and $v(\cdot)$ to $u(\cdot)$, the actual input to the original system. From the block diagram (Fig. 4.3-4) we can write the equations

$$\dot{\theta}(t) = -2\theta(t) - 2y(t) + u(t)$$
$$u(t) = v(t) - 2y(t) - [y(t) + \theta(t)]$$

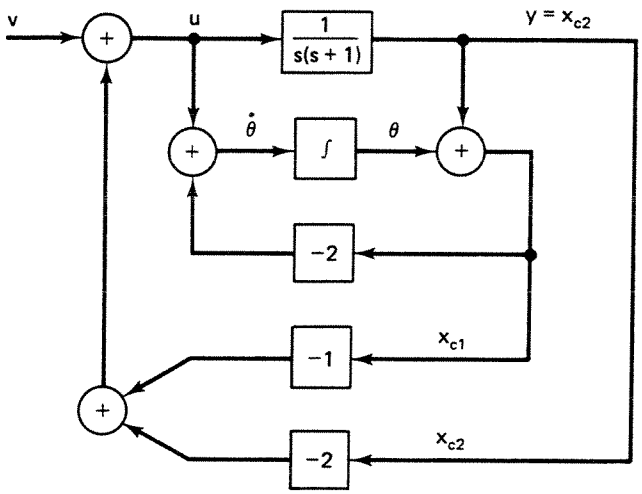



**Figure 4.3-4.** Combined observer-controller for Example 4.3-2.

Taking Laplace transforms and eliminating $\theta$, we obtain

$$U(s) = \left(-3 + \frac{2}{s+2}\right)Y(s) + V(s) - U(s)(s+2)^{-1} \tag{16}$$

Therefore

$$U(s) = -\frac{3s+4}{s+3}Y(s) + \frac{s+2}{s+3}V(s) \tag{17}$$

Relation (16) can be realized as shown in Fig. 4.3-5(a), which may be rearranged as in Fig. 4.3-5(b), a classical configuration. By the methods of Sec. 3.1.2 we can check that this realization has two hidden natural frequencies at $s = -2$, which are cancelled out of the transfer function from $v(\cdot)$ to $y(\cdot)$. The configuration is stable but uses one integrator more than necessary.

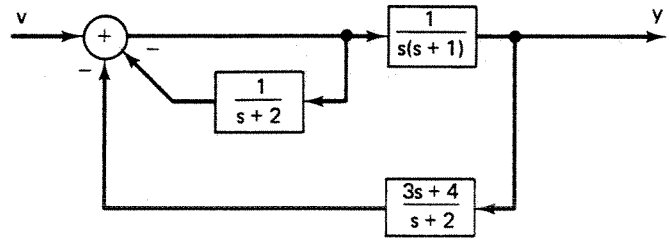
Relation (17) suggests the realization of Fig. 4.3-5(c), which can be shown to have hidden natural frequencies at $s = -2$ and $s = -3$. The frequency $s = -3$ is fixed by our original choice of transfer function poles and observer poles; in this case it *happens* to be a stable natural frequency, so the configuration is usable, though again it has one more integrator than necessary.

Another classical configuration, with *unity* feedback, is shown in Fig. 4.3-6, where $\mathcal{Y}(s)$ is determined by choosing it to make the overall transfer function equal to its desired value of $(s^2 + 2s + 2)^{-1}$. We have

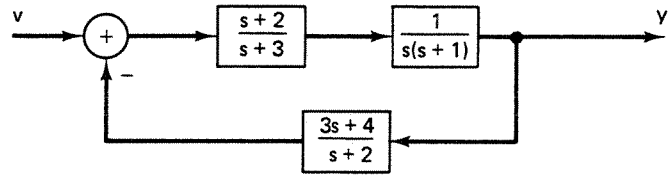
$$\frac{\mathcal{Y}(s)H(s)}{1 + \mathcal{Y}(s)H(s)} = \frac{1}{s^2 + 2s + 2}$$

which yields

$$\mathcal{Y}(s) = s(s+1)^{-1}$$




(a)




(b)

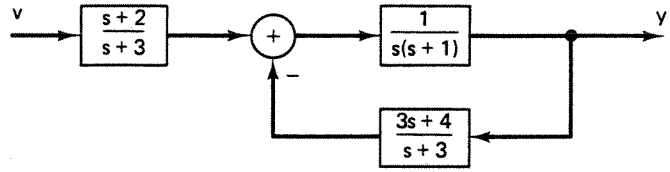



**Figure 4.3-5.** Various transfer function arrangements of the realization in Fig. 4.3-4. (a) Direct implementation of (16). (b) A classical configuration. (c) Direct implementation of (17).

This is a very simple solution which seems to avoid all the fuss about separate controller design, the design of reduced-order observers, and the combination of controller and observer. Why don't we just use it directly? The reason is (cf. the discussion in Sec. 3.1) that this particular procedure cannot give us full control over the hidden modes; in fact, we can see (by writing down a state-space realization or, more simply, by using the transfer function method of Exer. 2.4-7) that the configuration of Fig. 4.3-6 will have an unstable hidden mode at $s = 0$. However by a

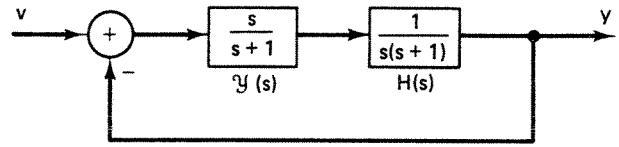



**Figure 4.3-6.** Classical unity-feedback configuration.

slightly more complicated design procedure, we can in fact achieve fairly satisfactory results with this configuration, as we shall show in Sec. 4.5.2.

***Exercises***

***4.3-1.*** *Eigenvalues of Submatrices* [3]

Suppose we have a realization $\{c, A\}$ in the (column-transposed observer) form with $c = [0 \cdots 0\ 1]$ and $A$ a *right-companion* matrix with $-[a_n \cdots a_1]$ in the last column. Suppose also that we wish to find a nonsingular matrix $T$ so that $\bar{A}_r$, defined as the $(n-1) \times (n-1)$ left-hand submatrix of $T^{-1}AT$, has characteristic polynomial $\det(sI - \bar{A}_r) = \alpha(s) = s^{n-1} + \alpha_1 s^{n-2} + \cdots + \alpha_{n-1}$. Show that choosing $T$ to have last column $[\alpha_{n-1} \cdots \alpha_1\ 1]'$ with 1s for the remaining diagonal elements and 0s elsewhere will yield $\bar{c} = c$ and $\bar{A}_r$ as a right-companion matrix with last column $-[\alpha_{n-1} \cdots \alpha_1]'$.

***4.3-2.*** *Another Approach to Reduced-Order Observers*

Show how to use the result of Exercise 4.3-1 to determine a reduced-order observer for an arbitrary but observable realization $(A, b, c)$. Do Examples 4.3-1 and 4.3-2 by this method.

***4.3-3.***

Show that the subsystem $\{c_r, A_r\}$ defined by Eqs. (4.3-3) will be observable if $\{c, A\}$ is observable. Do this

a. By using the result of Exercise 4.3-1,
b. By direct algebraic manipulation of $\Theta(c_r, A_r)$, and
c. By using the PBH tests of Sec. 2.4.3.

***4.3-4.*** *Updating an Inertial Navigator with a Velocity Measurement.*

An error model of the east-velocity channel of an inertial navigator is (in normalized variables)

$$\begin{bmatrix} \dot{v} \\ \dot{\varphi} \\ \dot{\epsilon} \end{bmatrix} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} v \\ \varphi \\ \epsilon \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ w \end{bmatrix}$$

where $v$ = east-velocity error, $\varphi$ = platform tilt about north axis, $\epsilon$ = north-gyro drift, and $w$ = gyro drift rate of change.

a. Show that the open-loop eigenvalues are $\lambda_1 = 0$, $\lambda_{2,3} = \pm j$.

b. Construct an observer using $z = v$ as the observation [made by a Doppler radar (airplane) or an EM log (ship)], placing the estimate-error poles at $\{-10, -0.1, -0.1\}$.

c. Construct a *second-order* observer using $z = v$ as the observation, placing the estimate-error poles at $\{-0.1, -0.1\}$.

***4.3-5.***

We are given a system with transfer function $H(s) = 1/s(s-2)$ and wish to use the unity-feedback configuration of Fig. a in order to design a compensator $\mathcal{Y}(s)$ that will shift both poles of $H(s)$ to $-1$. No requirement is put on the zeros. You can choose any polynomial $\beta(s)$ (of degree not greater than 2) that you wish.

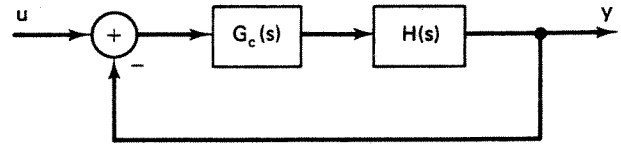



**Figure a.**

a. Find such a compensator. Is $G_c$ stable?

b. Do you think that the compensator procedure is satisfactory? Give reasons.

c. If we set up some minimal realizations for $G_c$ and $H$ and connect them as in Fig. a, then the overall system will be a realization of $\beta(s)/(s+1)^2$. Is it a minimal realization? A controllable realization? An observable realization?

d. Suppose that in some other design calculation we have computed a block $G_f(s) = (13s+1)/(s+1)$ and that we now want to find $G_c(s)$ as in Fig. b so that the overall transfer function is $\beta(s)/(s+1)^2$ and the overall system is stable. Find such a $G_c(s)$.

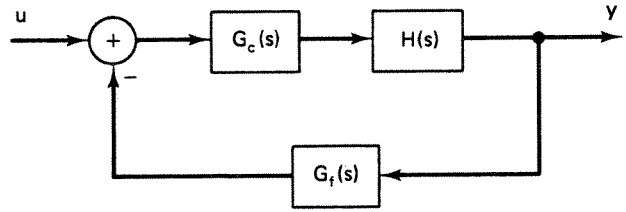



**Figure b.**

Compare the results of parts b and d.

***4.3-6.***

A system is described by the transfer function $H(s) = (s+1)/s^2 = Y(s)/U(s)$.

a. Find a first order dynamic feedback compensator so that the transfer function from reference input $V$ to output $Y$ is $Y(s)/V(s) = [(s^2+2s+4)(s+2)]^{-1}(s+8)(s+1)$.

b. Draw a block diagram of the resulting design.

c. Find a first order dynamic feedback compensator so that the transfer function from reference input to output is $Y(s)/V(s) = (s^2+2s+4)^{-1}(s+1)$.

d. Draw a block diagram of the design of part c.

e. Is the system of part a controllable from $V$? Observable from $Y$?

***4.3-7.***

Carry out the steps required to obtain reduced-order observers for discrete-time systems.

## 4.4 AN OPTIMALITY CRITERION FOR CHOOSING OBSERVER POLES

In Sec. 3.4.1, we discussed the relation between the modal controller and the so-called *optimal quadratic regulator*. Since the asymptotic observer is *dual* to the modal controller, it is natural to ask for its relationship to the dual of the quadratic regulator, and we shall briefly discuss this here.

We begin by noting that we could make the observer error decay as rapidly as we wished by putting the observer poles sufficiently far into the left half plane. In fact, one can show that in the limit the asymptotic observer reduces exactly to the ideal differentiating observer of Sec. 2.3.1. Of course, this fact brings up one reason we should not try to get too high a speed of reconstruction for the state estimates: The resulting observer will have (a high bandwidth and) a high susceptibility to the almost-inevitable measurement noise.

The conflict between speed of reconstruction and protection against measurement noise can be introduced into a quadratic optimality criterion as follows. Suppose that our actual state equations are

$$\dot{x}(t) = Ax(t) + bu(t) + gv_1(t) \tag{1a}$$

$$y(t) = cx(t) + v_2(t) \tag{1b}$$

where $v_1(\cdot)$ and $v_2(\cdot)$ are uncorrelated zero-mean wide-band (white) noise processes with spectral intensities unity and $r$, respectively; i.e., the noise power in a frequency band $(-B/2, B/2)$ is $B$ or $rB$, respectively. We assume that the unknown initial condition $x(0)$ is also random but that it is not correlated with the noises $v_1(\cdot)$ and $v_2(\cdot)$. We have no direct knowledge of $v_1(\cdot)$ and $v_2(\cdot)$ and therefore can only set up an observer as before,

$$\dot{\hat{x}}(t) = A\hat{x}(t) + bu(t) + l[y(t) - c\hat{x}(t)], \qquad \hat{x}(0) = 0 \tag{2}$$

But now the error will depend not only on $x(0) - \hat{x}(0)$ but also on the noises $v_1(\cdot)$ and $v_2(\cdot)$,

$$\dot{\tilde{x}}(t) = (A - lc)\tilde{x}(t) + gv_1(t) - lv_2(t), \qquad \tilde{x}(0) = x(0) \tag{3}$$

Again, by choosing $l$ sufficiently large we can make the expected or mean value of $\tilde{x}$ tend toward zero, $E[\tilde{x}] \to 0$, as fast as desired. But we still have to worry about the fluctuations of $\tilde{x}(\cdot)$—choosing large values for $l$ will accentuate the driving noise term $lv_2(\cdot)$. One criterion is to choose $l$ so as to minimize the mean-square error, i.e., to minimize

$$J = E\|\tilde{x}(t)\|^2 \tag{4}$$

This problem can be attacked in many different ways and in fact has an extensive literature (see, e.g., [1] and [10]–[12] among many others).

We shall not pursue the calculations here—they need some knowledge of statistics and of the solution of stochastic differential equations—but shall just quote the most relevant result. Suppose that

$$\{A, g\} \text{ is controllable}, \qquad \{c, A\} \text{ is observable} \tag{5}$$

Then the roots of the characteristic polynomial $\det(sI - A + \tilde{l}c)$ of the optimal observer can be found as the left-half-plane roots of the $2n$-degree polynomial

$$g(s)g(-s) + ra(s)a(-s) = 0 \tag{6}$$

where

$$a(s) = \det(sI - A), \qquad g(s) = c\,Adj\,(sI - A)g$$

This should be compared with the rule (3.4-7)–(3.4-8) for the optimal regulator of Sec. 3.4.1—the two solutions will be identical if we make the dual interchanges†

$$br^{-1}b' \longleftrightarrow gg', \qquad c'c \longleftrightarrow c'r^{-1}c \tag{7}$$

The implications of (6) for the noisy observer can therefore be studied for various values of the noise parameter $r$ by the same arguments as in Sec. 3.4.1. For example, if $r \longrightarrow \infty$ (high observation noise), the optimal observer poles are obtained as the stable roots of the original characteristic polynomial $a(s)$ along with the unstable roots reflected across the $j\omega$ axis. On the other hand, when $r \longrightarrow 0$, we can have $n - m$ poles going off to infinity in a Butterworth configuration, with the other $m$ poles going to the zeros of $g(s) = c\,Adj\,(sI - A)g$. For intermediate values of $r$, the symmetric root-locus plot can be useful, as we shall now illustrate.

**Example 4.4-1. Harmonic Oscillator**

For the undamped harmonic oscillator driven by a disturbance input $w$ of unit intensity,

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} w$$

determine the locus of optimum modes for an observer that uses a noisy measurement $y = x_2 + v$, where $v$ is observation noise of intensity $r$.

†It should be noted that more than one (variational) dual problem can be associated with the quadratic regulator problem and that they can be useful in obtaining approximate solutions and bounding solutions—see, for example, References [13] and [26].

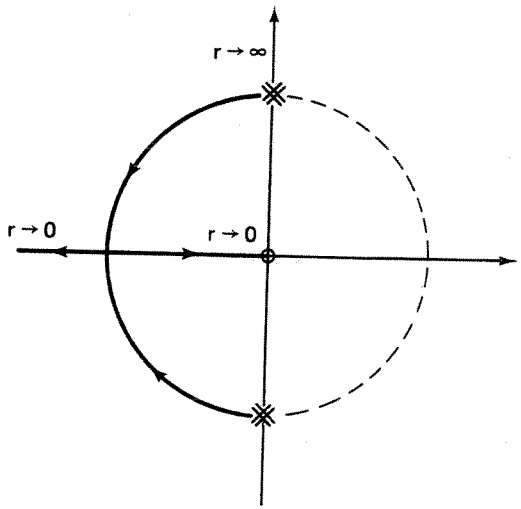
**Solution**

$$\frac{y(s)}{w(s)} = [0 \quad 1]\begin{bmatrix} s & -1 \\ \omega_0^2 & s \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ 1 \end{bmatrix} = \frac{s}{s^2 + \omega_0^2}$$

so the modes of the optimum observer are the stable solutions of

$$-s^2/(s^2 + \omega_0^2)^{-2} + r = 0$$

The corresponding root loci are shown in the figure.




Note that as the observation noise tends toward zero one observer mode approaches the origin and hence is shielded from the driving noise $w(\cdot)$ by the system zero at the origin; the other mode tends toward $-\infty$ and results effectively in differentiation of the system output to determine the state.

**Other Factors in Selecting Observer Poles.** It should be noted that nonstochastic factors are also important in the selection of observer dynamics. References [14]–[15] discuss how one can select them to compensate for the effect of deterministic modeling errors, component tolerances, and amplifier biases (see Exercise 4.1-6). Here we may remark that the reduced-order methods of Sec. 4.3 cannot be used when the observed process contains white (broadband) noise, because differentiation would greatly accentuate the noise. Observer designs for colored additive noise have also been studied (see, e.g., [16] and [17]).

**A Stochastic Separation Property.** As in the purely deterministic case studied in Section 4.2, the problem of optimal quadratic regulation when the states and the measurements are corrupted by noise as in (1) also reveals

a so-called *separation property*. That is, the expected cost

$$EJ = E\int_0^\infty \{|c\,x(t)|^2 + r|u(t)|^2\}\,dt$$

is minimized by a feedback law

$$u(t) = -\bar{k}\hat{x}(t)$$

where $\bar{k}$ is determined as in the case of complete state availability, and $\hat{x}(\cdot)$ is the optimal least-squares estimate of $x(\cdot)$ as determined by (6). We shall not enter further into this result and the special conditions under which it holds, except to refer to the derivations in, for example, [1, pp. 256–292]. We may remark also that, as mentioned in Sec. 3.4.1, the use of state-estimates may cause a deterioration in some of the special robustness (e.g., guaranteed gain and phase margins) properties of the optimal quadratic regulator with completely accessible states (see [18]).

***Exercises***

***4.4-1.*** *Variational Approach to the Estimator*

It can be shown that the solution of the following optimization problem also leads to the optimal noisy observer. Find $x(\cdot)$ and $v_1(\cdot)$ to minimize $J = \int_0^\infty \{[y(t) - cx(t)]'r^{-1}[y(t) - cx(t)] + v_1'(t)v_1(t)\}\,dt$ subject to the equations $\dot{x}(t) = Ax(t) + gv_1(t)$, $y(t) = cx(t) + v_2(t)$. It might be argued that this is a reasonable criterion for the observer problem, whether or not its minimization is equivalent to minimizing the mean-square error $E[x(t) - \hat{x}(t)]'[x(t) - \hat{x}(t)]$; this equivalence is hard to show a priori, but it can be justified *ex post facto* by verifying that its solution yields the same result as the minimum mean-square-error problem. Therefore, solve this problem by using the (Lagrange multiplier, calculus of variations) method of Sec. 3.4.1 and obtain the solution rule described in Sec. 4.4. *Note:* This method yields the linear two-point equations [$x(t_0)$ and $\lambda(t_f)$ are given]

$$\begin{bmatrix}\dot{x}(t)\\ \dot{\lambda}(t)\end{bmatrix} = \begin{bmatrix}A & -gg'\\ -c'r^{-1}c & -A'\end{bmatrix}\begin{bmatrix}x(t)\\ \lambda(t)\end{bmatrix} + \begin{bmatrix}0\\ c'r^{-1}y(t)\end{bmatrix}$$

which can be very effectively studied by scattering-theory methods. (See Reference 27 of Chapter 3.)

***4.4-2.*** *Optimum Root Loci*

Consider the station-keeping satellite of Example 3.3-2, with input disturbances $w_x$ and $w_y$, and $\ddot{x} = 9x + 2\dot{y} + w_x$, $\ddot{y} = -2\dot{x} - 4y + w_y$. Determine the loci of optimum observer modes when we have measurements of $y$ corrupted by white noise of intensity $r$ and

a. $w_x$ = unit intensity white noise and $w_y = 0$ and

b. $w_x = 0$ and $w_y$ = unit intensity white noise.

c. Extend (6) to the case of systems with multiple uncorrelated white noise inputs. Now find the loci when both $w_x$ and $w_y$ are (uncorrelated) unit intensity white noise disturbances. *Hint:* If $H_x(s)$ and $H_y(s)$ are the transfer functions from $w_x$ and $w_y$, respectively, to the output $y$, show that the observer modes are the left-half-plane roots of $H_x(s)H_x(-s) + H_y(s)H_y(-s) + r = 0$.

## 4.5 DIRECT TRANSFER FUNCTION DESIGN PROCEDURES

The combined observer-controller gives a complete solution, at least in theory, to the problem we raised in Sec. 3.1 of using feedback to modify the dynamic behavior of a given system. Let us briefly review the situation. We pointed out in Sec. 3.1 that simple output feedback or even feedback of the output and some derivatives did not give us enough information to carry out our desired objectives in all situations. Moreover, it was not very clear exactly what was to be done and exactly what could be achieved. The fact that the states gave a complete description of the system then led us to examine the possibility of feeding back not the output and its derivatives but the states of a realization of the given system. We showed in Sec. 3.2 that for a *controllable* realization such feedback could, in fact, allow us to relocate the poles (or rather, the natural frequencies) arbitrarily. But, of course, in general the states may not be directly available, and this led us in Sec. 4.1 to the use of asymptotic observers, which can provide asymptotically good estimates of the states of any *observable* realization. The difference between using output feedback (by which term we shall henceforth also imply possible feedback of some derivatives of the output) and using an observer is that in the latter both input and output are fed back (see Fig. 4.2-1, redrawn more schematically as Fig. 4.5-1). However, by some simple block diagram manipulations, Fig. 4.5-1 can be redrawn in the form of several classical output-feedback configurations, which is perhaps why it was never explicitly considered. Of course, the fact that we are starting from the special observer-controller configuration imposes certain constraints on the derived classical configurations. Moreover, we now know that the point is that such block diagram manipulations may impair stability by introducing additional *hidden modes*, which may or may not be stable. [We should stress that block diagram manipulations are not in general equivalent to similarity transformations between different realizations—in particular, block diagram manipulations can change not only the natural frequencies but also the number of states.]

However, we should not abandon the possibilities of transfer function analysis so easily. We know that by watching for cancellations in the *nominal* transfer function we can keep track of hidden modes (cf. Exercise 2.4-7) and therefore have full knowledge of the internal behavior of the system.

This thought suggests that we carefully examine the special observer-controller configuration from a transfer function point of view to see whether it may be possible to achieve a direct transfer function design without using internal state-space descriptions. This can indeed be done, as we shall show in Sec. 4.5.1, and with definite advantages in simplicity and directness. Moreover, the analysis in Sec. 4.5.1 also suggests certain flexibilities in the design procedure and objectives that are by no means as evident in the state-space approach (Sec. 4.5.2). A more purely transfer function analysis, in which we work exclusively with polynomial equations, is developed in Sec. 4.5.3.

### *4.5.1 A Transfer Function Reformulation of the Observer-Controller Design*

As we might expect from the block diagram configurations shown in Figs. 4.2-1 and 4.3-3, the transfer function approach will be essentially the same whether we use an observer or a reduced-order observer. For simplicity of explanation, however, we shall start our analysis with the full-order observer and then show how easily the reduced-order observer design can be incorporated into the transfer function approach.

The observer-controller configuration of Fig. 4.2-1 can be redrawn more schematically as shown in Fig. 4.5-1. The transfer functions of the blocks denoted $H_u(s)$ and $H_y(s)$ can be calculated from Eq. (4.2-2) for $\hat{x}(\cdot)$,

$$\dot{\hat{x}}(t) = A\hat{x}(t) + l[y(t) - c\hat{x}(t)] + bu(t)$$

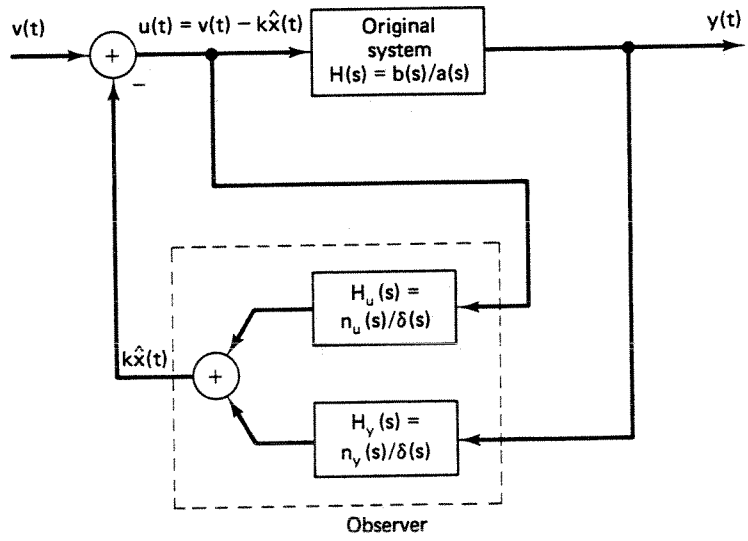



**Figure 4.5-1.** Block diagram of the combined observer-controller of Fig. 4.2-1.

so that

$$k\hat{X}(s) = k(sI - A + lc)^{-1}[lY(s) + bU(s)]$$
$$= H_y(s)Y(s) + H_u(s)U(s) \tag{1}$$

where

$$H_y(s) = k(sI - A + lc)^{-1}l, \qquad H_u(s) = k(sI - A + lc)^{-1}b \tag{2}$$

As we expect (see the more detailed diagram, Fig. 4.2-1), the natural frequencies of $H_u(s)$ and $H_y(s)$ are the same, viz., the roots of the characteristic polynomial of the observer, det $(sI - A + lc)$. We know also that the overall transfer function must be

$$\frac{Y(s)}{V(s)} = \frac{b(s)}{\det(sI - A + bk)} = \frac{b(s)}{\alpha(s)} \tag{3}$$

as compared to the transfer function of the original system:

$$H(s) = \frac{b_1 s^{n-1} + \cdots + b_n}{s^n + a_1 s^{n-1} + \cdots + a_n} = \frac{b(s)}{a(s)} \tag{4a}$$

Now once the basic structure of Fig. 4.5-1 has been given (or has been "guessed"), it seems reasonable to try to directly calculate transfer functions $\{H_y(s), H_u(s)\}$ that can yield the desired overall transfer function (3).

Therefore let us assume that these transfer functions have the form

$$H_y(s) = \frac{\beta_1 s^{n-1} + \cdots + \beta_n}{s^n + \delta_1 s^{n-1} + \cdots + \delta_n} = \frac{n_y(s)}{\delta(s)} \tag{4b}$$

and

$$H_u(s) = \frac{\gamma_1 s^{n-1} + \cdots + \gamma_n}{s^n + \delta_1 s^{n-1} + \cdots + \delta_n} = \frac{n_u(s)}{\delta(s)} \tag{4c}$$

Now $n_y(s)$, $n_u(s)$, and $\delta(s)$ are to be chosen so that the transfer function of the overall configuration of Fig. 4.5-1 has a desired form (3). The overall transfer function can be computed as follows: Note first that the loop through $H_u(s)$ can be replaced by a transfer function $\delta(s)/[n_u(s) + \delta(s)]$ in series with $H(s) = b(s)/a(s)$. Then we can write

$$\frac{Y(s)}{V(s)} = \frac{\delta(s)b(s)/[n_u(s) + \delta(s)]a(s)}{1 + ([\delta(s)b(s)n_y(s)]/\{[n_u(s) + \delta(s)]a(s)\delta(s)\})}$$
$$= \frac{\delta^2(s)b(s)}{\delta(s)\{[n_u(s) + \delta(s)]a(s) + b(s)n_y(s)\}}$$

Therefore we have a characteristic polynomial of degree $3n$. However, it is

obvious from Fig. 4.5-1 that we can use the same $n$ integrators to implement $H_u(s)$ and $H_y(s)$. If we do this, we can "cancel" $\delta(s)$ to have a realization with the transfer function

$$\frac{Y(s)}{V(s)} = \frac{b(s)\delta(s)}{a(s)\delta(s) + a(s)n_u(s) + b(s)n_y(s)} \tag{5}$$

and a characteristic polynomial of degree $2n$, i.e., with only $2n$ modes. Now by proper choice of $[n_y(s), n_u(s), \delta(s)]$ we can hope to obtain an arbitrary $2n$-th-degree characteristic polynomial, say, $p(s)$, where

$$p(s) = s^{2n} + \cdots + p_{2n}$$

While this degree of generality may be useful, the results of Sec. 4.2 suggest that one choice is to make

$$p(s) = \alpha(s)\delta(s) \tag{6}$$

where $\alpha(s)$ = the desired $n$th-degree characteristic polynomial and $\delta(s)$ = an arbitrary $n$th-degree polynomial, for then the overall transfer function will be

$$\frac{Y(s)}{V(s)} = \frac{b(s)\delta(s)}{\alpha(s)\delta(s)} = \frac{b(s)}{\alpha(s)} \tag{7}$$

as obtained with the observer-controller configuration of Sec. 4.2. Of course $\delta(s)$ should not be left completely free but should be chosen to have its roots sufficiently stable that the *hidden modes* we shall now have due to the cancelled $\delta(s)$ in (7) will not seriously perturb the overall system.

However, we have not yet said, in the present framework, what is needed to be able to achieve the choice $p(s) = \alpha(s)\delta(s)$. We shall prove that this choice of $p(s)$, and in fact *any other choice of* $p(s)$ *as an arbitrary 2nth-degree polynomial*, can be achieved if and only if

$$a(s) \text{ and } b(s) \text{ are relatively prime} \tag{8}$$

that is, whenever there are no cancellations in the original transfer function, $H(s) = b(s)/a(s)$. Of course, this result is consistent with the result of Sec. 4.2 because (cf. Sec. 2.4.1) in such a case any $n$th-order controllable realization (and such a realization can *always* be obtained) of $H(s)$ will also be observable, or (vice versa) any $n$th-order observable realization will also be controllable. Nevertheless, the present approach shows that no explicit knowledge of controllability or observability, or even of the concept of state, is necessary to solve the problem of modifying the dynamics of a given system by feedback. Of course this neglects the vital fact that it is the state-variable concept—and the analysis of modal controllability and asymptotic

observability—that led to consideration of the structure of Fig. 4.5-1 in the first place. The reader will recall that in Sec. 3.1.1, when we first introduced this problem, we did not really have much clue as to this particular structure. We tried some simple structures that did not work, and perhaps by trial and error we may have been led to the structure of Fig. 4.5-1. But as it happened, the development along state-variable lines took place first, and apparently it was only in 1968 that a transfer function interpretation was sought (cf. Chen [8] and [9]); related results had been obtained earlier by Mortensen [19] and Shipley [20] (see Exercises 3.1-2 and 3.1-3), but at that time there was not the full awareness of the importance of explicit attention to hidden modes.

We drop these speculations and return to the analysis of the basic relation

$$p(s) = a(s)[\delta(s) + n_u(s)] + b(s)n_y(s) \tag{9}$$

where $a(s)$, $b(s)$, and $\delta(s)$ are specified $n$th-degree polynomials, and we have to try to find the $n$th-degree polynomials $n_y(s)$ and $n_u(s)$ such that the right-hand side of (9) is some specified [e.g., by (6)] $2n$th-degree polynomial. We can explore this question in several ways. Perhaps the simplest way to begin is to see if we can obtain from (9) enough equations to uniquely specify the coefficients of $n_y(s)$ and $n_u(s)$. Therefore, we expand both sides of (9) and equate the coefficients of corresponding powers of $s$ to get (cf. also Exercise A-6)

$$S(a, b)z = w \tag{10}$$

where

$$z' = [\delta_1 + \gamma_1 \cdots \delta_n + \gamma_n \quad \beta_1 \cdots \beta_n]$$
$$w' = [p_1 - a_1 \cdots p_n - a_n \quad p_{n+1} \cdots p_{2n}]$$

and $S(a, b)$ is the $2n \times 2n$ matrix (shown for $n = 3$)

$$S(a, b) = \left[\begin{array}{ccc:ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ a_1 & 1 & 0 & b_1 & 0 & 0 \\ a_2 & a_1 & 1 & b_2 & b_1 & 0 \\ a_3 & a_2 & a_1 & b_3 & b_2 & b_1 \\ 0 & a_3 & a_2 & 0 & b_3 & b_2 \\ 0 & 0 & a_3 & 0 & 0 & b_3 \end{array}\right] \tag{11}$$

We thus have $2n$ equations for the $2n$ unknowns $\{\beta_1, \ldots, \beta_n, \gamma_1, \ldots, \gamma_n\}$, and these equations will have a solution for an *arbitrary* right-hand side $w$ if and only if $\det S(a, b) \neq 0$. But $S(a, b)$ is the well-known Sylvester matrix,

and it is known (see Sec. 2.4.4) that

$$\det S(a, b) \neq 0 \Longleftrightarrow \{a(s), b(s)\} \text{ are coprime} \tag{12}$$

which is the condition (8) that we quoted earlier. However, to keep this section self-contained, let us give a direct proof of result (12). Thus note that $S(a, b)$ will be nonsingular if and only if the only solution of the equation

$$S(a, b)[p_0 \quad p_1 \quad \cdots \quad p_{n-1} \quad q_0 \quad \cdots \quad q_{n-1}]' = [0 \quad \cdots \quad 0]' \tag{13}$$

is the trivial solution

$$p_0 = p_1 = \cdots = q_0 = \cdots = q_{n-1} = 0$$

Now (13) can be written in polynomial form as (cf. Exercise A-6)

$$a(s)p(s) + b(s)q(s) = 0 \tag{14}$$

where $p(s) = p_0 s^{n-1} + \cdots + p_{n-1}$ and similarly for $q(s)$, while $a(s)$ has the higher degree $n$. But if (14) is true for $p(s) \not\equiv 0$ and $q(s) \not\equiv 0$, we shall have

$$b(s)/a(s) = -p(s)/q(s)$$

which is impossible if and only if $b(s)$ and $a(s)$ are coprime. Therefore $p(s) \equiv 0 \equiv q(s)$, i.e., $S(a, b)$ will be nonsingular, if and only if $b(s)$ and $a(s)$ are coprime.

This completes the proof of (12) and justifies our new design procedure: Choose $\delta(s)$ arbitrarily, and solve for the coefficients of $n_y(s)$ and $n_u(s)$ from Eq. (10); now set up the configuration of Fig. 4.5-1 using the same $n$ integrators to implement the common denominator $\delta(s)$ of $H_y(s)$ and $H_u(s)$.

**Reduced-Order Compensators.** Actually, a little thought shows that only $n - 1$ integrators are really necessary, corresponding to the use of reduced-order observers in the state-space method.

Thus in (4) we used an $n$th-degree denominator polynomial $\delta(s)$ for $H_y(s)$ and $H_u(s)$, but clearly an $(n - 1)$th-degree polynomial will suffice [less than $n - 1$ would make $H_y(s)$ and $H_u(s)$ improper, since Eq. (10) shows that the degrees of $n_u(s)$ and $n_y(s)$ can be as high as $n - 1$]. Therefore, let us replace $\delta(s)$ by, say,

$$\delta_r(s) = s^{n-1} + \cdots + \delta_{n-1} \tag{15}$$

and notice that this will make the overall characteristic polynomial have degree $2n - 1$ rather than $2n$, say,

$$p_r(s) = p_1 s^{2n-1} + \cdots + p_{2n} \tag{16}$$

The new design equation is [cf. (9)]

$$p_r(s) = [\delta_r(s) + n_u(s)]a(s) + n_y(s)b(s) \tag{17}$$

which is equivalent to a somewhat simpler matrix equation than (10), viz.,

$$S(a, b)z_r = w_r \tag{18}$$

where

$$z_r' = [1 + \gamma_1 \cdots \delta_{n-1} + \gamma_n \quad \beta_1 \cdots \beta_n]$$

and

$$w_r' = [p_1 \cdots p_n \quad p_{n+1} \cdots p_{2n}]$$

Of course these changes do not affect the solvability of the equation, which will still yield a solution for arbitrary $w_r$ if and only if $\{a(s), b(s)\}$ are coprime.

**Example 4.5-1**

To illustrate the transfer function method, let us rework Example 4.3-2 by the present method. The problem was, given $H(s) = 1/s(s + 1)$, to shift the poles to $\{-1 \pm j\}$ using a reduced-order observer with pole at $-2$.

In the present method, this means that we choose

$$\delta_r(s) = s + 2$$

and

$$p_r(s) = (s + 1 + j)(s + 1 - j)(s + 2)$$

Then we have to find $n_y(s)$ and $n_u(s)$ such that

$$a(s)[n_u(s) + \delta_r(s)] + b(s)n_y(s) = p_r(s)$$

In such simple low-degree cases, it is best to assume the expected forms for $n_y(s)$ and $n_u(s)$ and compare coefficients. This leads to

$$(s^2 + s)(\gamma_1 s + \gamma_2 + s + 2) + (\beta_1 s + \beta_2) = s^3 + 4s^2 + 6s + 4$$

and

$$\gamma_1 = 0, \qquad \gamma_2 = 1, \qquad \beta_1 = 3, \qquad \beta_2 = 4$$

Therefore the compensator is defined by connecting

$$H_y(s) = \frac{3s + 4}{s + 2}, \qquad H_u(s) = \frac{1}{s + 2}$$

as in Fig. 4.5-1. We can verify that this coincides with the results obtained with considerably more effort in Example 4.3-2.

The transfer function method will always be simpler when the original system is specified by its transfer function $b(s)/a(s)$, and, depending on the circumstances, the transfer function method might be simpler even otherwise. Note that the highly structured (Toeplitz) form of $S(a, b)$ will lend itself to efficient solution methods (see, e.g., F. Gustavsson and D. Yun, *IEEE Transactions on Circuits and Systems*, **CAS-26**, Sept. 1979).

### 4.5.2 Some Variants of the Observer-Controller Design

The transfer function analysis also suggests some interesting variants that are not quite so obvious in a state-space analysis.

We return to the block diagram of Fig. 4.5-1 and recall that [cf. (17)]

$$\frac{Y(s)}{V(s)} = \frac{b(s)\delta_r(s)}{p_r(s)} \qquad p_r(s) = [\delta_r(s) + n_u(s)]a(s) + n_y(s)b(s) \tag{19}$$

where $\delta_r(s)$ and $p_r(s)$ are arbitrary (of degrees $n-1$ and $2n-1$, respectively) and $n_u(s)$ and $n_y(s)$ (of degrees not greater than $n-1$) are to be determined. We showed that if $\{a(s), b(s)\}$ were relatively prime we could choose $p_r(s)$ as an arbitrary polynomial of degree $2n-1$. In particular, by choosing it to have the special form

$$p_r(s) = \alpha(s)\delta_r(s) \tag{20a}$$

we obtained the observer-controller design,

$$\frac{Y(s)}{V(s)} = \frac{b(s)}{\alpha(s)} \tag{20b}$$

where $\alpha(s)$ is an arbitrary $n$th-degree polynomial and the cancelled (or hidden) modes corresponding to $\delta_r(s)$ are also completely under our control.

However, it should be emphasized that unlike what may appear from the state-space analysis, we have choices other than always choosing $p_r(s)$ to cancel out $\delta_r(s)$ from the overall transfer function; in particular, we could leave in all or some of the factors of $\delta_r(s)$. We still always have $2n-1$ natural frequencies, but the number of poles could vary from 1 to $2n-1$ and we could add up to $n-1$ additional zeros [beyond those left uncancelled in $b(s)$]. It is not quite clear exactly how this additional freedom should be exploited, but in Exercise 4.5-2 we shall indicate one possibility of interest in classical control system theory.

If we are not at all interested in the zero locations but only wish to have $2n-1$ arbitrary natural frequencies, then the configuration of Fig. 4.5-1 can be simplified somewhat to the one shown in Fig. 4.5-2. Namely, from

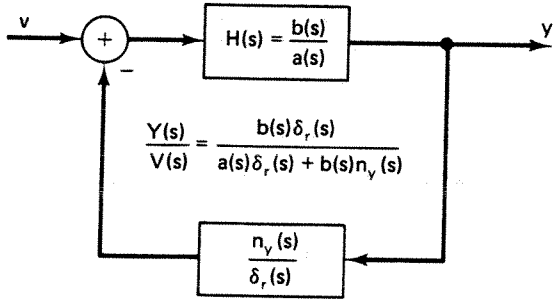



**Figure 4.5-2.** Configuration that can provide an arbitrary set of natural frequencies but with a fixed added set of zeros.

(17) we see that we can set

$$n_u(s) \equiv 0 \tag{21}$$

and still find $\{\delta_r(s), n_y(s)\}$ to set up an arbitrary $(2n - 1)$th-degree characteristic polynomial $p_r(s)$. Therefore, feedback of the input $u(\cdot)$ is not essential in setting up an internally stable realization. What is lost by making assumption (21) is that now $\delta_r(s)$ is fixed by our choice of $p_r(s)$, and therefore we cannot in general obtain an overall transfer function with denominator polynomial of degree $n$ [the degree of the denominator $a(s)$ of the original system $H(s)$]. Moreover, the design procedure introduces some fixed zeros [the roots of $\delta_r(s)$] into the overall transfer function.

We leave it to the reader to show that results similar to these can be obtained for the classical unity-feedback compensation scheme of Fig. 4.5-3. The results associated with Figs. 4.5-2 and 4.5-3 were first obtained by Pearson ([21] and [22]), who used state-space proofs.

We note that the role of the explicit feedback of $u(\cdot)$ is to give us some extra degrees of freedom—namely, if $n_u(s) \not\equiv 0$, then we can choose $\delta_r(s)$ as an arbitrary $(n - 1)$th-order polynomial and still solve Eq. (17) for any $p_r(s)$, e.g., in the observer-controller form (20) or otherwise.

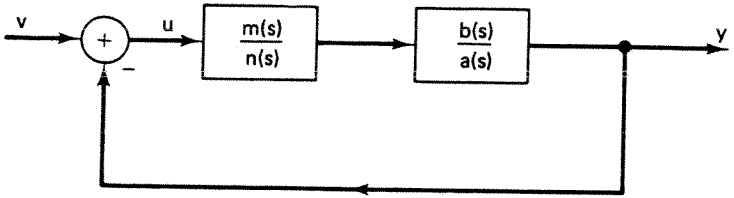



**Figure 4.5-3.** The classical unity-freedback scheme can provide $2n - 1$ arbitrary modes, but there is no freedom in choosing the zeros.

However, we should note here that in the schemes presented so far the so-called *relative order* of the (nominal or actual) transfer function is the same; i.e., the difference between the number of poles and zeros is constant.

### 4.5.3 Design Via Polynomial Equations

In Sec. 4.5.1, we saw that the basic equations to be solved were of the form [cf. (9) and (17)]

$$p(s)a(s) + q(s)b(s) = c(s) \tag{22}$$

where $\{a(s), b(s), c(s)\}$ were given polynomials and $\{p(s), q(s)\}$ are polynomials to be determined under certain degree constraints. In Sec. 4.5.1 we studied these equations by converting them to equivalent matrix form [cf. (10) or (18)]. However, we can also give a more direct solution of these so-called Diophantine† equations. This solution will suggest another approach to the compensation problem, which will lead us more directly (without appealing to state-space insights) to the basic design equations (9) and (17); this approach will also extend nicely to the multivariable case (cf. Sec. 7.5.1).

**Direct Analysis of the Diophantine Equation (22).** Let us start with Eq. (22). We see first of all that any common factor of $\{a(s), b(s)\}$ must also be present in $c(s)$, and therefore if $c(s)$ is to be arbitrary, a *necessary* condition for the success of our design method is that $\{a(s), b(s)\}$ are coprime.

To show the sufficiency of this condition is not hard either if we recall that coprimeness is equivalent to the fact (cf. Sec. 2.4.4) that there exist polynomials $\{z(s), w(s)\}$ such that

$$z(s)a(s) + w(s)b(s) = 1 \tag{23}$$

But then clearly

$$f(s) \triangleq c(s)z(s), \qquad g(s) \triangleq c(s)w(s) \tag{24}$$

will satisfy the original equation (22). However, in our special problem there have to be degree constraints on any solution pair $\{\bar{f}(s), \bar{g}(s)\}$ because they will be used as numerators and denominators of certain transfer functions, which must be proper if they are to be physically realizable. Such constraints are not hard to achieve.

We shall show that when $c(s)$ is of degree $2n - 1$, $a(s)$ of degree $n$, and $b(s)$ of degree $n - 1$, then we can replace $\{\bar{f}(s), \bar{g}(s)\}$ by polynomials $\{f_0(s), g_0(s)\}$ of degree not greater than $n - 1$. This will be a satisfactory solution

†Because of the restriction that $\{p(s), q(s)\}$ be polynomials. Diophantus of Alexandria wrote a book in the third century A.D. about the (mathematically identical) problem of finding integer solutions to $pa + qb = c$ when $\{a, b, c\}$ are given integers. More recent discussions can be found in books on number theory or continued fractions (see, e.g., the very elementary books [23, Chaps. 1, 2] and [24, pp. 1–8] and also [25]).

for the observer and reduced-order observer problems. We do this by using the Euclidean division algorithm (in a way already illustrated in Sec. 2.4.4). Thus we start by writing

$$\begin{bmatrix} \bar{f}(s) & \bar{g}(s) \\ b(s) & -a(s) \end{bmatrix} \begin{bmatrix} a(s) \\ b(s) \end{bmatrix} = \begin{bmatrix} c(s) \\ 0 \end{bmatrix} \tag{25}$$

By polynomial division, we can write

$$\bar{g}(s) = l(s)a(s) + g_0(s), \qquad \deg g_0 < n \tag{26a}$$

Then the (elementary row) operation of adding $l(s)$ times the second equation in (25) to the first equation will give

$$\begin{bmatrix} f_0(s) & g_0(s) \\ b(s) & -a(s) \end{bmatrix} \begin{bmatrix} a(s) \\ b(s) \end{bmatrix} = \begin{bmatrix} c(s) \\ 0 \end{bmatrix} \tag{26b}$$

where

$$f_0(s) = \bar{f}(s) + l(s)b(s) \tag{26c}$$

As for the degree of $f_0(s)$, we note that [cf. (26b)]

$$\deg f_0(s) + \deg a(s) = \deg [c(s) - g_0(s)b(s)] \leq 2n - 1$$

so that

$$\deg f_0(s) \leq n - 1$$

Therefore $\{f_0(s), g_0(s)\}$ have the claimed properties.

**Physical Interpretations and an Alternative Approach.** Reviewing the above, we see that the polynomial criterion (23) for the coprimeness of $\{a(s), b(s)\}$ was at the heart of our solution.

To better understand the physical significance of this condition, we note that $\{b(s), a(s)\}$ are not arbitrary polynomials but are related to the system input and output by

$$y(s) = \frac{b(s)}{a(s)} u(s) \tag{27}$$†

Now we can rewrite (27) as

$$y(s) = b(s)\xi(s), \qquad a(s)\xi(s) = u(s) \tag{28}$$

where we should recall that the *partial state* $\xi(\cdot)$ and its derivatives completely

†For notational convenience, we do not use capital letters for the transforms of $\{y(\cdot), x(\cdot), u(\cdot)\}$.

determine the state of any realization.† The interesting thing about (23) and (28) is that they show immediately how to recover the partial state if it is not directly accessible, namely

$$z(s)u(s) + w(s)y(s) = z(s)a(s)\xi(s) + w(s)b(s)\xi(s) = \xi(s) \tag{29}$$

This suggests the configuration of Fig. 4.5-4, where we have shown the feedback through an additional operation—multiplication by a possibly rational function, $m(s)$. The point is that the resulting transfer function is

$$\frac{y(s)}{v(s)} = \frac{b(s)}{a(s) + m(s)} \tag{30}$$

and therefore can be made to have an arbitrary denominator by proper choice of $m(s)$; e.g., $m(s) = \alpha(s) - a(s)$ (cf. Sec. 3.2.2).

Of course, the fact that $\xi(s)$ can be reconstructed is not unexpected, since we are effectively using the *ideal observer*, i.e., differentiation, to obtain the (partial) state. However, it is not hard to obtain a physically realizable (asymptotic observer) version of the above scheme. Thus, if $m(s)\xi(s)$ is the quantity desired for feedback to the input, let us introduce a *denominator polynomial* $\delta(s)$ as shown in Fig. 4.5-5(a). Now $w(s)m(s)\delta(s)$ and $z(s)m(s)\delta(s)$ can be *reduced* as in (26) to have degree less than or equal to that of $\delta(s)$, yielding physically realizable implementation as shown in Fig. 4.5-5(b), which is identical to Fig. 4.5-1. In more detail, note from (23) that

$$[z(s)m(s)\delta(s)]a(s) + [w(s)m(s)\delta(s)]b(s) = m(s)\delta(s) \tag{31a}$$

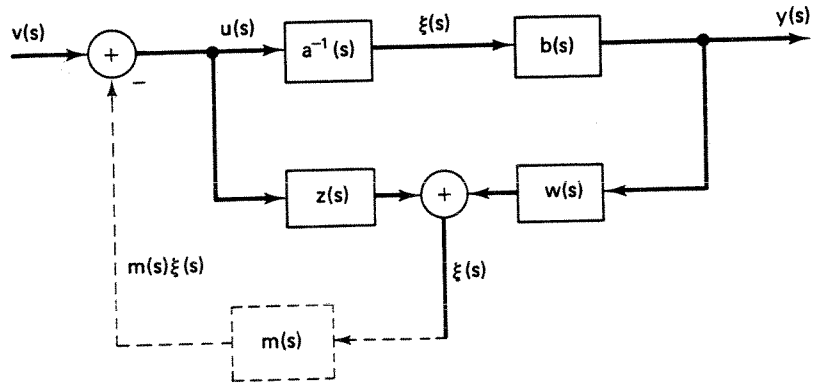



**Figure 4.5-4.** Reconstruction and feedback of the state to get a transfer function $b(s)[a(s) + m(s)]^{-1}$.

†See Exercise 3.2-5. Note first that $\xi(s)$ determines the state vector $x_c(\cdot)$ of a controller-form realization; then the states of any other minimal realization can be obtained by a nonsingular (similarity) transformation.

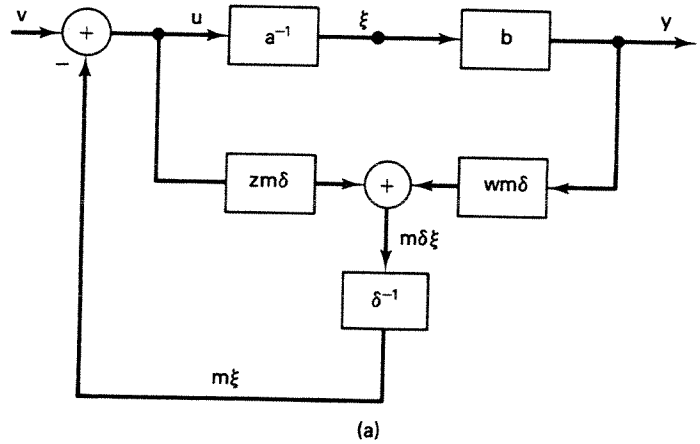



**Figure 4.5-5.** Realizable compensator. (a) Toward a realizable compensator. (b) Reducing the order in (a)

which we can reduce as in (26) to

$$n_u(s)a(s) + n_y(s)b(s) = m(s)\delta(s) \tag{31b}$$

Let us assume that the *feedback* function $m(s)$ is a polynomial, and let

$$\deg m(s) \leq n - 1 \tag{32a}$$

to correspond to the fact that no more than $n - 1$ derivatives of the partial state $\xi$ are needed to yield the states of the realization.

Now by construction, $\deg n_y(s) < n = \deg a(s)$, and from (31)–(32) we

can see that

$$\deg n_u(s) < \deg \delta(s) \tag{32b}$$

Therefore, choosing

$$\deg \delta(s) \geq n - 1 \tag{33}$$

will suffice to make $n_u(s)/\delta(s)$ and $n_y(s)/\delta(s)$ realizable, i.e., proper rational functions. Now (31b) directly yields Fig. 4.5-5(b). To complete the identification with the results of Sec. 4.5.1, we just have to choose

$$m(s) = \alpha(s) - a(s) \tag{34}$$

With this substitution in (31b) we see that

$$a(s)[n_u(s) + \delta(s)] + b(s)n_y(s) = \alpha(s)\delta(s) \tag{35}$$

which is precisely Eq. (9) of Sec. 4.5.1, thus yielding the promised rederivation of the results of Sec. 4.5.1.

The insight gained in this calculation as to the significance of (23) can be applied to obtain the results of Sec. 4.5.2 and in fact some useful extensions thereof, especially to changing the relative order. However we shall not pursue these extensions here.

***Exercises***

***4.5-1.***

Given $H(s) = (s - 1)/s^2$, find a compensator that will yield a new third-order realization with characteristic polynomial $(s + 5)(s^2 + s + 1)$. Is there a unique solution of this problem? If not, give some criteria for choosing among the different possible solutions.

***4.5-2.*** (*Franklin*)

Consider a system with the transfer function $1/s(s + 1)$, which we wish to modify using feedback so as to have a new characteristic polynomial $s^2 + 4s + 8$. Show that the steady-state error in the response to a unit ramp is $1/K_v$, $K_v = 2$. It can be shown that for a system with closed-loop poles at $\{-p_i\}$ and closed-loop zeros at $\{-z_i\}$,

$$\frac{1}{K_v} = \sum \frac{1}{p_i} - \sum \frac{1}{z_i}$$

Show that by choosing $p_3 = 0.1$ and $-z_3 = -1/10.4$, we can achieve $K_v = 10.$. This illustrates one way in which flexibility in zero assignment might be used.

***4.5-3.***

Given a system with the transfer function $b(s)/a(s)$ and a desired stable transfer function $b(s)\delta(s)/p(s)$, where $\deg p(s) - \deg \delta(s) \geq n$, show that we

can obtain an internally stable realization as follows: Use state feedback (or an observer-controller compensator) to change $b(s)/a(s)$ to $b(s)/\alpha(s)$, where $\alpha(s)$ contains $n$ roots of $p(s)$; then cascade this with a realization of $\delta(s)/\epsilon(s)$, where $\alpha(s)\epsilon(s) = p(s)$.

*4.5-4. Idealized Compensators*

In Eqs. (4.5-2) and (4.5-3), suppose we assume that the denominator polynomial $\delta(s)$ is absent; i.e., assume $\delta(s) = 1$. Show that we can achieve a desired transfer function $H(s) = b(s)/\alpha(s)$, where $\alpha(s)$ is a specified $n$th-order polynomial, by choosing the coefficients of $\beta(s)$ and $\gamma(s)$ so that $[\beta' \ \gamma']S' = [0 \ \alpha' - a']$, where $S$ is the Sylvester matrix of $\{a(s), b(s)\}$, $\beta$ is the column vector of the coefficients of $\beta(s)$, and similarly for $a$, $\alpha$, and $\gamma$. Compare with the results of Sec. 4.5.3.

*4.5-5. Reconciliation to State-Space Designs*

Suppose the original transfer function is realized in controller form. Calculate the feedback gain $k_c$ required to give a new characteristic polynomial $\alpha(s)$, and assume that the states $x(t)$ are provided by an ideal observer. Calculate $H_u(s)$ and $H_y(s)$ for this compensator (cf. Fig. 4.3-1), and show that they coincide with $\beta(s)$ and $\gamma(s)$ as found in Exercise 4.5-4.

*4.5-6.*

Consider the compensator scheme depicted in Fig. 4.3-6. Show that we can find a stable transfer function $\mathcal{Y}(s)$ that will make all the modes of the closed-loop system stable if and only if $H(s)$ has the following *parity interlacing property*: for every RHP zero of $H(s)$, there must be an even number (counted according to multiplicity) of poles of $H(s)$ to the right of it. *Reference:* D. C. Youla, J. J. Bongiorno, and C. N. Lu, *Automatica*, **10**, pp. 159–174, 1974; see also *Ibid.*, **12**, pp. 387–388, 1976.

## REFERENCES


1. K. J. ÅSTROM, *Introduction to Stochastic Control Theory*, Academic Press, New York, 1970.
2. R. E. KALMAN, P. FALB, and M. A. ARBIB, *Topics in Mathematical System Theory*, McGraw-Hill, New York, 1969.
3. D. G. LUENBERGER, "Observing the State of a Linear System," *IEEE Trans. Mil. Electron.*, **MIL-8**, pp. 74–80, 1964. (Also see Ph.D. thesis, Stanford University, Stanford, Calif., 1963.)
4. B. GOPINATH, "On the Control of Linear Multiple Input-Output Systems," *Bell Syst. Tech. J.*, **50**, pp. 1063–1081, March 1971. (Also see Ph.D. thesis, Stanford University, Stanford, Calif., 1968.)
5. D. G. CUMMING, "Design of Observers of Reduced Dynamics," *Electron. Lett.*, **5**, no. 10, pp. 213–214, May 15, 1969.
6. D. G. LUENBERGER, "An Introduction to Observers," *IEEE Trans. Autom. Control*, **AC-16**, pp. 596–603, Dec. 1971.

7. A. E. BRYSON and D. G. LUENBERGER, "The Synthesis of Regulator Logic Using State-Variable Concepts," *Proc. IEEE*, **58**, pp. 1803–1811, Nov. 1970.
8. C. T. CHEN, "A New Look at Transfer Function Design," *Proc. IEEE*, **59**, pp. 1580–1585, Nov. 1971. See also Proc. Natl. Electronics Conf., **25**, pp. 46–51, 1969.
9. C. T. CHEN, *Introduction to Linear System Theory*, Holt, Rinehart and Winston, New York, 1970.
10. R. E. KALMAN and R. S. BUCY, "New Results in Linear Filtering and Prediction Theory," *Trans. ASME Ser. D. J. Basic Eng.*, **83**, pp. 95–107, Dec. 1961.
11. H. KWAKERNAAK and R. SIVAN, *Linear Optimal Control Systems*, Wiley, New York, 1972.
12. T. KAILATH, *Lectures on Linear Least-Squares Estimation*, CISM Courses and Lectures No. 140, Springer-Verlag, New York, 1978.
13. W. L. CHAN, "Variational Dualities in the Linear Regulator and Estimation Problems," *J. Inst. Math. Appl.*, **18**, pp. 237–248, Oct. 1976.
14. J. D. POWELL, "Mass Center Estimation in Spinning Drag-Free Satellites," *J. Spacecr. Rockets*, **9**, pp. 399–405, 1972. (Also see Ph.D. thesis, Dept. of Aeronautics and Astronautics, Stanford University, Stanford, Calif., May 1970.)
15. F. E. THAU and A. KESTENBAUM, "The Effect of Modeling Errors on Linear State Reconstructors and Regulators," *J. Dyn. Syst. Meas. Control*, pp. 454–459, Dec. 1974.
16. T. KAILATH and R. GEESEY, "An Innovations Approach to Least-Squares Estimation, Part V: Innovations Representations and Recursive Estimation in Colored Noise," *IEEE Trans. Autom. Control*, **AC-18**, no. 5, pp. 435–453, Oct. 1973.
17. L. M. NOVAK, "Discrete-Time Optimal Stochastic Observers," in *Control and Dynamic Systems*, Vol. 12 (C. T. Leondes, ed.), Academic Press, New York, 1976, pp. 259–311.

**18a.** J. C. DOYLE, "Guaranteed Margins for LQG Regulators," *IEEE Trans. Autom. Control*, **AC-23**, pp. 756–757, Aug. 1978.

**18b.** J. C. DOYLE and G. STEIN, "Robustness with Observers," *IEEE Trans. Automat. Contr.*, **AC-24**, Oct. 1979.

19. R. E. MORTENSEN, "The Determination of Compensation Functions for Linear Feedback Systems To Produce Specified Closed-Loop Poles," *Internal Technical Rept. TR-59-0000-00781*, Space Tech. Laboratories, Los Angeles, Aug. 1959. (See also *IEEE Trans. Autom. Control*, **AC-8**, p. 386, Oct. 1963.)
20. P. P. SHIPLEY, "A Unified Approach to Synthesis of Linear Systems," *IEEE Trans. Autom. Control*, **AC-8**, pp. 114–120, April 1963.
21. J. B. PEARSON, "Compensator Design for Dynamic Optimization," *Int. J. Control*, **9**, pp. 413–482, 1969.
22. F. M. BRASCH and J. B. PEARSON, "Pole Placement Using Dynamic Compensators," *IEEE Trans. Autom. Control*, **AC-15**, pp. 34–43, 1970.
23. C. D. OLDS, *Continued Fractions*, Random House, New York, 1963.

24. A. YA. KHINCHIN, *Continued Fractions*, Chicago University Press, Chicago, 1964 (Russian ed., 1935).

25. V. KUČERA, *Discrete Linear Control: The Polynomial Equation Approach*, J. Wiley, London, 1979.

26. U. B. DESAI and H. L. WEINERT, "Generalized Control-Estimation Duality and Inverse Projections," *Proceedings 1979 Conference on Information Sciences and Systems*, Johns Hopkins University, Baltimore, Maryland, March 1979.