

Design of Digital Control Systems Using Transform Techniques

5.1 INTRODUCTION

The idea of controlling processes that evolve in time is ubiquitous. Systems from airplanes to the national rate of unemployment, from unmanned space vehicles to human blood pressure, are considered fair targets for control. Over a period of three decades from about 1930 until 1960, a body of control theory was developed based on electronic feedback amplifier design modified for servomechanism problems. This theory was coupled with electronic technology suitable for implementing the required dynamic compensators to give a set of approaches to solve control problems now often called *classical techniques*. The landmark references to this theory are Nyquist (1932), Bode (1945), and Evans (1950). For random inputs, the work of Wiener (1948) should be added. An excellent pedagogical presentation of these methods is given in Truxal (1955). The unifying theme of these methods is the use of Laplace or Fourier transform representations of the system dynamics and the control specifications; hence, we refer to them here as *transform techniques* after the central role of the frequency domain in the approach.

In this chapter, we discuss the use of transform techniques in the design of digital control systems. First, we describe the use of discrete equivalents to construct a digital controller indirectly from a continuous design. This method is referred to as *emulation* or *s-plane design*. It is attractive because the *design* is carried out exactly as if the system were continuous, the only change due to the digital implementation being the extra step of emulating the resulting compensation in a discrete form. Another method of design is to discretize the system model at the outset, then to perform the design entirely using the discrete representation. This method is referred to as *z-plane design*, *discrete design*, or *direct digital design*.

Modifications of the transform techniques are necessary to make them directly applicable to discrete design. We find that the root locus can be transferred unchanged to the z -plane, but the interpretation of the results is different than in the s -plane. Frequency-response methods can also be used in a way similar to the way they were used in continuous systems, but the Bode hand-plotting methods are no longer useful and the calculation of gain and phase margins requires the use of a computer.

Sometimes discrete frequency response design is carried out by using a bilinear (Tustin) transformation similar to the one used in Chapter 4. It is called the w -transform. This technique is often referred to as *w-plane design*. It allows the use of Bode's techniques for hand plotting the magnitude and phase and is, therefore, particularly useful when not using a computer.

In many cases, use of both emulation and discrete design methods gives the best result. The discrete controller is initially determined using the emulation method, and then z -plane analysis tools are used to verify or to modify the design.

As with any control design, a subsequent step consisting of numerical simulation of the system including all known delays and nonlinearities is an important additional effort. It often identifies deficiencies in the design that arose because of the approximations made to arrive at the linear model that is required for the transform techniques. A simulation offers the opportunity to modify the design based on more detailed models before committing to hardware. This is discussed further in Chapter 11.

5.2 CONTROL SYSTEM SPECIFICATIONS

Before describing how the transform techniques can be applied to digital control designs, we must first review the control specification ideas for continuous systems and discuss how these specifications are interpreted and modified for discrete systems.

Example 5.1: The specifications for the design of the azimuth control of an antenna to pick up signals from a low-altitude communications satellite (discussed in Appendix A) are:

1. Tracking error to a ramp input less than 0.01 rad.
2. Overshoot to a step input $\leq 16\%$.
3. Settling time to within 1% ≤ 10 sec.

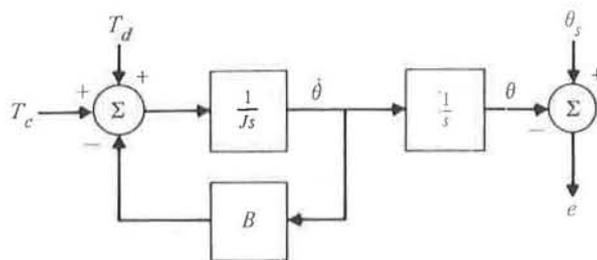


Figure 5.1 Model of an antenna tracking control.

The equations of motion for the antenna are

$$J\ddot{\theta} + B\dot{\theta} = T_c + T_d, \quad (5.1)$$

where θ is the antenna pointing angle, T_c is the drive-motor torque, and T_d is the wind torque. The system parameters are the moment of inertia of the moving parts, J , and the damping coefficient, B , consisting of the mechanical friction component and the back emf effect of the electric motor. We will assume that the time constant, J/B , is equal to 10 sec. A block diagram of the antenna system is shown in Fig. 5.1.

The transfer function of the system, or *plant*, can be written as

$$G(s) = \frac{\theta(s)}{U(s)} = \frac{1}{s(10s + 1)}, \quad (5.2)$$

where $u = T_c/B$.

The aim of the design is to measure the error between the angle of the satellite θ_s and the antenna and compute T_c so that the error, e , ($= \theta_s - \theta$), is always less than 0.01 rad during tracking. The satellite angle that must be followed can be adequately approximated by a fixed velocity,

$$\theta_s(t) = (0.01 \text{ rad/sec})t. \quad (5.3)$$

Fig. 5.2 shows a generic block diagram of a *unity feedback* closed-loop system that applies directly to this example and many other systems. In the figure, the wind disturbance enters in normalized form as $w = T_d/B$, the reference input, θ_s , is referred to as r , and the antenna angle, θ , is referred to as y . The block containing D is the controller that we will be designing.

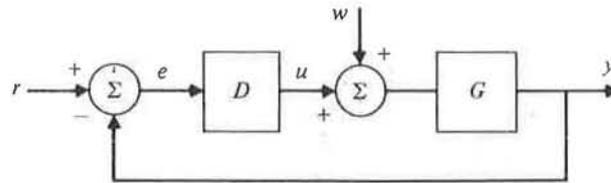


Figure 5.2 A unity feedback system.

The worst possible wind torque can be approximated as a gust that comes suddenly and holds constant for several seconds. We will approximate this by a step function and require that the transients must be settled out in less than 10 sec to leave a total steady-state error within the tracking-error specifications.

Performance characteristics that must be specified for this system, which are characteristic of many others, can now be listed. They are:

1. Steady-state tracking accuracy
2. Transient accuracy (dynamic response):
 - a) stability
 - b) rise time
 - c) overshoot
 - d) settling time
3. Disturbance rejection:
 - a) steady state
 - b) transient
4. Control effort required:
 - a) maximum magnitude of u
 - b) energy $K \int u^2 dt$
5. Sensitivity to parameter changes

Steady-state accuracy refers to the requirement that after all transients are negligible, the error $r - y$ or, for the antenna, $\theta_s - \theta$, must be acceptably small. The two causes of nonzero error are the reference r and the disturbance w . Consider first the reference. Some control systems have a finite nonzero steady-state error when the reference is a constant. Such systems are labeled "Type 0," because there is finite error with a zero-order polynomial input. The reason for the error can be seen from Fig. 5.2: The finite dc gain between e and y requires that e be nonzero in order for y to be nonzero. Similarly, a control system that has finite nonzero steady-state

error to a first-order polynomial input (a ramp) is called a "Type I" system. And then comes "Type II," *mutatis mutandis*.¹ In each case, the disturbance w is taken to be zero. However, in evaluating the system we must include the effects of w in the final calculations. In general, we add the errors due to reference and disturbance to find a total system error, which must be within acceptable limits (in this case, beam width of the tracker antenna).

To compute the steady-state error due to r or w , we assume the system is stable and use the final-value theorem. Suppose the unity feedback system shown in Fig. 5.2 has a reference input that is a step function and that the disturbance is zero. The error will have the transform,

$$\begin{aligned} E(s) &= \frac{R(s)}{1 + D(s)G(s)} \\ &= \frac{1}{s} \frac{1}{1 + D(s)G(s)}. \end{aligned} \quad (5.4)$$

The final value of $e(t)$, if the closed-loop system is stable, is

$$e(\infty) = \lim_{s \rightarrow 0} s \frac{1}{s} \frac{1}{1 + D(s)G(s)},$$

and, therefore, for the example with a step input, $e(\infty) = 0$ because

$$\lim_{s \rightarrow 0} G(s) = \infty.$$

For a unit ramp input,

$$e(\infty) = \lim_{s \rightarrow 0} s \frac{1}{s^2} \frac{1}{1 + D(s)G(s)},$$

and the error is finite if

$$\lim_{s \rightarrow 0} sD(s)G(s)$$

is finite. This will be true for the example if D has a finite dc gain (no integrator). In fact, in this case, because

$$\lim_{s \rightarrow 0} sG(s) = 1$$

¹With necessary changes, i.e., with "second" for "first" and "II" for "I."

the error to a unit ramp is

$$e(\infty) = \lim_{s \rightarrow 0} \frac{1}{D(s)}.$$

Thus, if the system were being controlled with a continuous controller, the required specification of maintaining the error less than 0.01 rad for a θ_s ramp input of 0.01 rad/sec would be met if

$$\lim_{s \rightarrow 0} \frac{1}{D(s)} \leq 1.$$

In other words, the dc gain of $D(s)$ must be greater than or equal to 1. Another way of stating these results is that the velocity error constant, K_v , must be greater than or equal to 1, where

$$K_v = \lim_{s \rightarrow 0} sD(s)G(s).$$

All these same ideas can be applied when D is implemented in a computer and represented by its discrete transfer function $D(z)$. Analysis of this case, however, requires that we find the discrete transfer function of the plant, that is, the portion referred to as G in Fig. 5.2.

In order to do this, we need to know how the computer's output is transferred to act upon the *continuous* portion of the system; that is, we need to know what kind of *hold* is used between the outputs of D at discrete instances and the continuous $G(s)$. Although different holds are possible, by far the most common is the *zero-order hold* (ZOH), discussed in Chapters 1, 3, and 4. It consists of simply holding the computer output at a constant value throughout the sample period until a new sample is obtained, at which time a new output is determined and held, and so forth. This results in a discontinuous signal with steps at each sample instant. Integrated circuits to perform this function, called *sample-and-hold amplifiers*, are readily available.

It is sometimes useful to pass the output of a ZOH through a low-pass filter to remove the discontinuities. This can be desirable for reducing vibration and extending the lifetime of certain actuators, especially hydraulic ones. However, nothing comes free! The smoother signals have more lag, with the associated detrimental effect on stability.

The portion of the system for which we desire a discrete representation, $G(z)$, is shown in Fig. 5.3. Its inputs are the sampled signals, $u(n)$, from D (the computer) and its outputs are the samples from the plant, $y(n)$. There

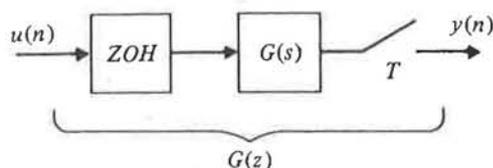


Figure 5.3 The discrete model of the continuous part of the system.

is an exact discrete representation of this, which has already been studied in Chapter 2. Applying (2.39) we find that

$$G(z) = (1 - z^{-1}) \mathcal{Z} \left\{ \frac{G(s)}{s} \right\}. \quad (5.5)$$

The closed-loop system can now be represented in a purely discrete manner. The discrete transfer functions of the controller, $D(z)$, and the plant, $G(z)$, are combined as before according to Fig. 5.2, where it is now understood that the reference r and the disturbance w are sampled versions of their continuous counterparts.

Proceeding as we did for the continuous system, suppose the input r is a step, $r(n) = 1(n)$, and the disturbance w is zero. The transform of the error is computed using the same block-diagram reduction tools that apply for continuous systems represented by their Laplace transforms, except that now we use $D(z)$ and $G(z)$. Doing this yields the transform of the error

$$\begin{aligned} E(z) &= \frac{R(z)}{1 + D(z)G(z)} \\ &= \frac{z}{z-1} \frac{1}{1 + D(z)G(z)}. \end{aligned}$$

The final value of $e(k)$, if the closed loop system is stable with all roots of $1 + DG = 0$ inside the unit circle, is, by (2.100),

$$\begin{aligned} e(\infty) &= \lim_{z \rightarrow 1} (z-1) \frac{z}{z-1} \frac{1}{1 + D(z)G(z)} \\ &= \frac{1}{1 + D(1)G(1)} \\ &\triangleq \frac{1}{1 + K_p}. \end{aligned} \quad (5.6)$$

Thus, $D(1)G(1)$ is the position constant, K_p , of the Type 0 system. If DG has a pole at $z = 1$, then the error given by (5.6) is zero. Suppose there is a single pole at $z = 1$. Then we have a Type I system and we can compute the error to a unit ramp input, that is, $r = t1(t)$. Then

$$E(z) = \frac{Tz}{(z-1)^2} \frac{1}{1 + D(z)G(z)}$$

Now the steady-state error is

$$\begin{aligned} e(\infty) &= \lim_{z \rightarrow 1} (z-1) \frac{Tz}{(z-1)^2} \frac{1}{1 + DG} \\ &= \lim_{z \rightarrow 1} \frac{Tz}{(z-1)(1 + D(z)G(z))} \\ &\triangleq \frac{1}{K_v} \end{aligned} \quad (5.7)$$

Thus the velocity constant of a Type I discrete system with unity feedback (as shown in Fig. 5.2) is

$$K_v = \lim_{z \rightarrow 1} \frac{(z-1)(1 + D(z)G(z))}{Tz},$$

which simplifies to

$$K_v = \lim_{z \rightarrow 1} \frac{(z-1)D(z)G(z)}{Tz}. \quad (5.8)$$

Although it appears from (5.8) that K_v is inversely proportional to the sample period, in fact, this is not the case if comparing for the same $G(s)$. The reason is that the loop gain of $G(z)$ is essentially proportional to the sample period. This proportionality is exact for the very simple case where $G(s) = 1/s$, as can be seen by using (5.5) and inspecting entry 4 in Appendix B.2. For systems with a finite K_v and fast sample rates, this proportionality will be approximately correct. The result of this proportionality is that K_v of a continuous plant alone preceded by a ZOH is essentially the same whether evaluated with the discrete representation or the continuous one (see Problem 5.16).

Because systems of Type I occur frequently, it is useful to observe that the value of K_v is fixed by the *closed-loop* poles and zeros by a relation given, for the continuous case, by Truxal (1955). Suppose the overall transfer

function Y/R is $H(z)$ and that $H(z)$ has poles p_i and zeros z_i . Then we can write

$$H(z) = K \frac{(z - z_1)(z - z_2) \cdots (z - z_n)}{(z - p_1)(z - p_2) \cdots (z - p_n)}. \quad (5.9)$$

Now suppose that $H(z)$ is the transfer function of a Type I system, which implies that the steady-state error of this system to a step is zero and requires that

$$H(1) = 1. \quad (5.10)$$

Furthermore, by definition we can express the error to a ramp as

$$\begin{aligned} E(z) &= R(1 - H(z)) \\ &= \frac{Tz}{(z - 1)^2} (1 - H(z)), \end{aligned}$$

and the final value of this error is given by

$$e(\infty) = \lim_{z \rightarrow 1} (z - 1) \frac{Tz}{(z - 1)^2} (1 - H(z)) = \frac{1}{K_v};$$

therefore (omitting a factor of z in the numerator, which makes no difference in the result)

$$\frac{1}{TK_v} = \lim_{z \rightarrow 1} \frac{1 - H(z)}{z - 1}. \quad (5.11)$$

Because of (5.10), the limit in (5.11) is indeterminate, and so we can use L'Hôpital's rule

$$\begin{aligned} \frac{1}{TK_v} &= \lim_{z \rightarrow 1} \frac{(d/dz)(1 - H(z))}{(d/dz)(z - 1)} \\ &= \lim_{z \rightarrow 1} -\frac{dH(z)}{dz}. \end{aligned}$$

However, note that by (5.10) again, at $z = 1$, we have

$$\frac{d}{dz} \ln H(z) = \frac{1}{H} \frac{d}{dz} H(z) = \frac{d}{dz} H(z),$$

so that

$$\begin{aligned}
 \frac{1}{TK_v} &= \lim_{z \rightarrow 1} -\frac{d}{dz} \ln H(z) \\
 &= \lim_{z \rightarrow 1} -\frac{d}{dz} \left\{ \ln K \frac{\prod(z - z_i)}{\prod(z - p_i)} \right\} \\
 &= \lim_{z \rightarrow 1} -\frac{d}{dz} \left\{ \sum \ln(z - z_i) - \sum \ln(z - p_i) + \ln K \right\} \\
 \frac{1}{TK_v} &= \lim_{z \rightarrow 1} \left\{ \sum \frac{1}{z - p_i} - \sum \frac{1}{z - z_i} \right\} \\
 &= \sum_{i=1}^n \frac{1}{1 - p_i} - \sum_{i=1}^n \frac{1}{1 - z_i}.
 \end{aligned}$$

We note especially that the farther the poles of the closed-loop system are from $z = 1$, the larger the velocity constant and the smaller the errors. Similarly, K_v can be increased and the errors decreased by zeros *close* to $z = 1$. From the results of Chapter 2 on dynamic response, we recall that a zero close to $z = 1$ usually signals large overshoot and poor dynamic response. Thus is expressed one of the classic trade-off situations: we must balance small steady-state errors against good transient response.

Example 5.2: For our antenna problem, we have specified that a unit ramp produce a steady-state error no more than 0.01 rad. From (5.7) we see that a $K_v \geq 1$ will satisfy the specification for this problem, an identical result to the continuous case. In fact, ideas of steady-state error from continuous systems generally carry over to discrete systems with very little change.

Transient accuracy, or dynamic response, refers to the ability of the system to keep the error small as $r(t)$ changes. Specifications of transient performance can be made in the time domain and then translated to the frequency domain either in terms of characteristic pole locations in s or z , or in terms of frequency-response features such as bandwidth and phase margin. We will aim to consider specifications in terms of characteristic root locations in the z -plane by transforming the desired s -plane specifications to equivalent locations in the z -plane. This is accomplished by using the relation $z = e^{sT}$ to map the poles in the s -plane to the z -plane.

First we need to transfer the transient specifications from a time description to an s -plane pole-location requirement. In Fig. 5.4(a) are plotted the step responses of a second-order system with unity dc gain, no finite zeros, and various damping ratios. We see immediately that the major influence on percent overshoot is the damping ratio, ζ . In Fig. 5.4(b) is plotted the value of this feature against ζ . A specification on percent overshoot can, for the second-order system, be translated into a specification of ζ . Note also that we should refer to Chapter 2 and Figs. 2.30 and 2.31, where we plotted our finding that an extra zero can also greatly influence overshoot. We must consider both ζ and the zero locations to help meet specifications for transient response on percent overshoot. For the case without an extra zero, Fig. 5.4(b) shows that, very roughly,

$$\% \text{ overshoot} \cong (1 - \zeta/0.6)100$$

for a second-order system. Thus, given a requirement on percent overshoot, we require

$$\zeta \geq (0.6) \left(1 - \frac{\% \text{ overshoot}}{100} \right).$$

Another feature of interest is the rise time of the response toward its final value. By inspection of Fig. 5.4(a) we see that the time scale is in terms of ω_n , the distance of the poles from the origin of the s -plane. Thus the rise time will certainly be shorter as ω_n is increased. Although there is some dependence of rise time on ζ , we can take the curve for $\zeta = 0.5$ to be about the center of the distribution and thus approximate the rise time by

$$t_r \cong 1.8/\omega_n,$$

where we take t_r to be the time necessary for the response to rise from 0.1 to 0.9. A requirement on t_r thus becomes a requirement that ω_n satisfy

$$\omega_n \geq 1.8/t_r. \quad (5.12)$$

The final time-domain feature of importance to us is the settling time. This is the time required for the response to settle to within some small fraction of its steady-state value and stay there. For the prototype second-order system, we can return to the mathematics of the solution to conclude that the transient is of the form

$$y(t) = 1 - e^{-\zeta\omega_n t} \cos(\omega_d t + \phi),$$

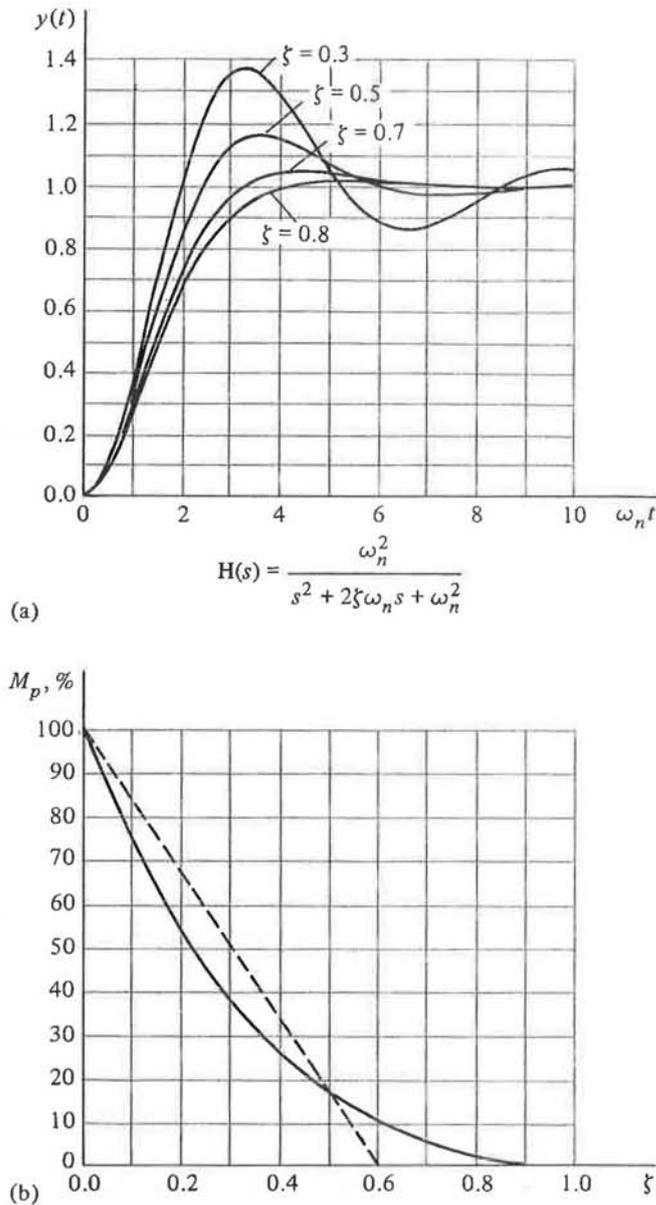


Figure 5.4 (a) Typical time responses for a second-order system. (b) Dependence of overshoot on damping ratio for second-order system. (From Franklin, G. F., Powell, J. D., and Emami-Naeini, A., *Feedback Control of Dynamic Systems*, Addison-Wesley, 1986.)

where $\omega_d = \omega_n\sqrt{1-\zeta^2}$. The point is that the transient portion of this signal is contained in an envelope of $e^{-\zeta\omega_n t}$, where $-\zeta\omega_n$ is the real part of the root location. Thus, we can require that $\zeta\omega_n$ be large enough that the

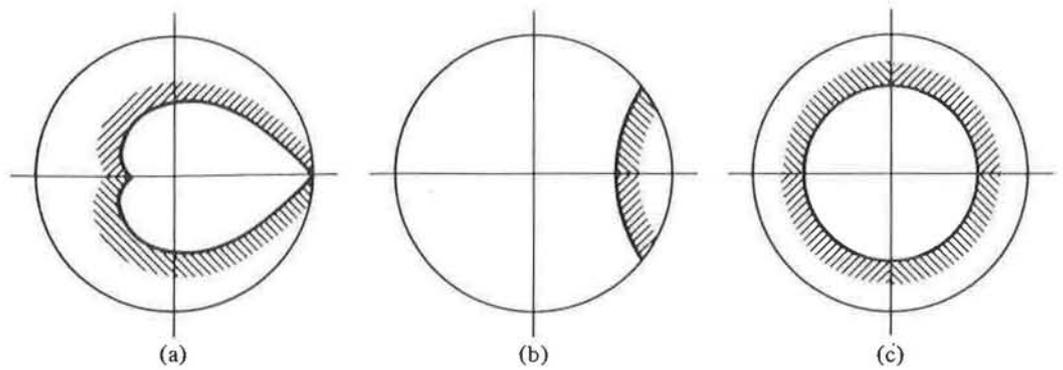


Figure 5.5 The mapping of s -plane specifications to the z -plane. (a) Damping; (b) frequency; (c) settling time.

transient will be squeezed into whatever error-tolerance band we choose. A typical value of the error-tolerance is 1%, for which we compute the envelope function to be

$$e^{-\zeta\omega_n t_s} = 0.01,$$

and thus

$$\zeta\omega_n t_s = 4.6,$$

$$\boxed{t_s = 4.6/\zeta\omega_n} \quad (5.13)$$

or, in order to settle in t_s sec or less, we require that

$$\operatorname{Re}\{s_i\} = \zeta\omega_n \geq 4.6/t_s.$$

We now need to convert these specifications into guidelines on the placement of poles in the z -plane in order to guide the design of digital controls. We do so by mapping via $z = e^{sT}$. Thus the restriction on percent overshoot has been expressed as a restriction on damping ratio, ζ . In the z -plane, curves of pole locations for constant ζ are logarithmic spirals, as sketched in Fig. 5.5(a) for $\zeta = 0.5$. The forbidden region is indicated by the partial hatching. The restriction on rise time is the requirement that the natural frequency be greater than a certain value. In the z -plane the curves of constant ω_n are lines drawn at right angles to the constant ζ spirals. A given value is sketched in Fig. 5.5(b), again with the hatching on the undesirable side of the line. The final time-domain specification was in terms of settling time. In this case, the real parts of the roots, $-\zeta\omega_n$, were restricted. Because the s - z

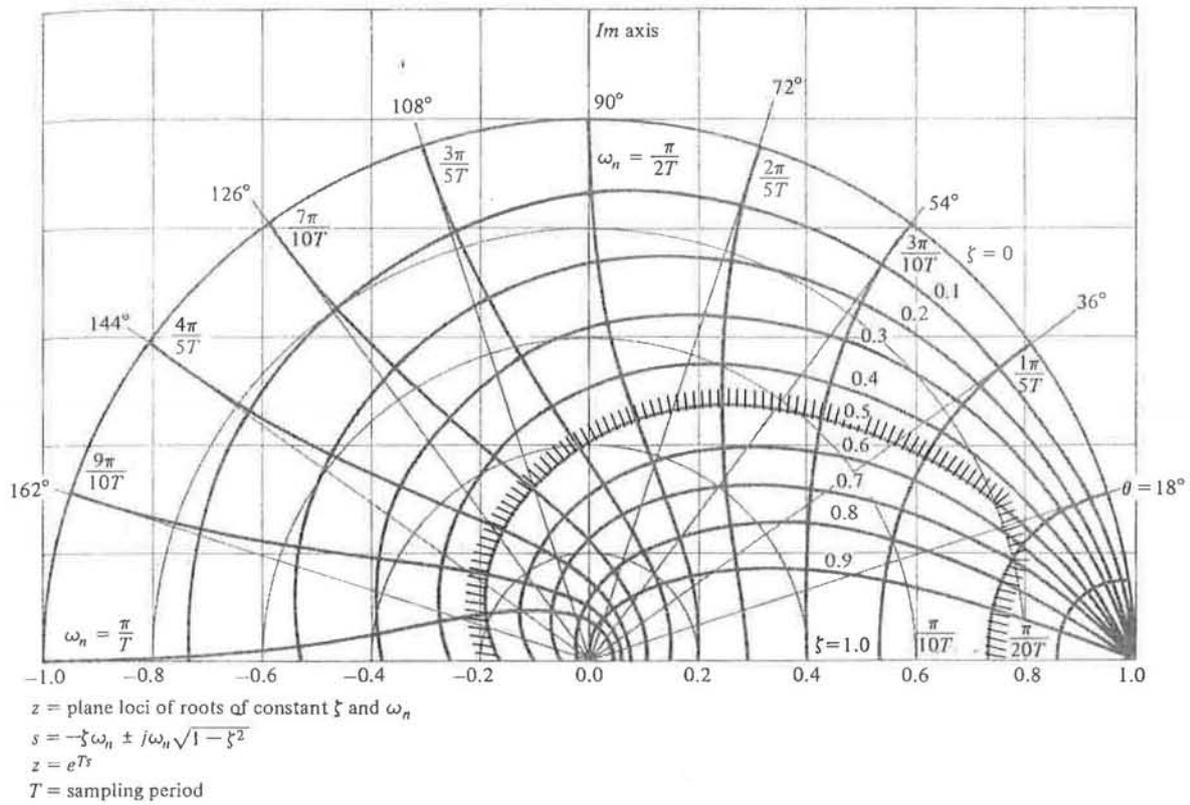


Figure 5.6 Plot of acceptable region for poles of a second-order system to satisfy dynamic response specification.

mapping has the z -plane root radius at $r = e^{-\zeta\omega_n T}$, we see at once that a settling-time restriction maps into a restriction that the z -plane poles should be inside a circle given by

$$r_0 = e^{-4.6T/t_s},$$

which, when sketched, looks like Fig. 5.5(c).

The final effect is drawn in Fig. 5.6, where we have arbitrarily picked the desired specifications as overshoot $\leq 16\%$, $t_r \leq 6$ sec, and $t_s \leq 20$ sec. Assuming $T = 1$ sec, we find that

$$\% \text{ overshoot} \leq 16\% \Rightarrow \zeta \geq 0.5,$$

$$t_r \leq 6 \Rightarrow \omega_n \geq \frac{1.8}{6},$$

$$t_s \leq 20 \Rightarrow r \leq 0.8.$$

The region forbidden by these specifications is indicated by the partial hatching along the line for $\omega_n = \pi/10T$ to the circle of radius 0.8 to the spiral of $\zeta = 0.5$ and then on around the spiral. These curves are approximate as befits a design process that is essentially a set of guidelines to the pole locations of a closed-loop system. The final design must be checked by simulation and/or experiment, and modifications must be made as indicated by the manner in which the first design fails to meet the specifications. For example, if a trial design has a settling time that is too long, then the radius of the poles should be reduced. Likewise, if corrections to overshoot and rise time are needed, Fig 5.6 indicates how the specified z -plane pole locations should be modified to accomplish the desired result.

The effectiveness of the system in *disturbance rejection* is readily studied with the topology of Fig. 5.2. From the figure, if we take $r = 0$, we find

$$E(z) = \frac{WG}{1 + DG}. \quad (5.14)$$

If the loop gain, $|DG|$, is large compared to 1, then (5.14) reduces to

$$E(z) \cong W/D(z).$$

Thus the extent of disturbance reduction is given by the amount of *gain* that precedes the disturbance in the loop. In particular, if $w(k)$ is a constant, then an integrator in D (pole at $z = 1$) will cause the steady-state error due to the disturbance to be zero. From the point of view of frequency response, a disturbance at w in Fig. 5.2 will be rejected over the frequency range where $|DG| \gg 1$ and $|D| \gg 1$, and also over the range where $|G| \ll 1$ and $|DG| \leq 1$.

The *control effort* required to perform a control task is important on several counts. Because all physical variables are bounded, the device that provides the control, such as the motor that drives the tracking antenna, can put out only a certain maximum torque even when turned on fully. It is pointless to try to get 100 N-m out of a 10 mN-m motor! Conversely, after completion of a design meeting dynamic-response specifications, we can simulate a worst-case transient; and from the size of the control signal required, we can determine the size of the motor necessary to meet these specifications. In addition to peak control, $|u|$, we are sometimes interested in the total heat generated by the drive motor. This, too, will influence

the size and design (and expense) of the motor. Usually this number is proportional to $\int u^2 dt$, over a typical transient period. Another measure of control effort arises in gas jets used for attitude control of satellites, where the total fuel used is a proper measure of control effort, and where the fuel expenditure is proportional to $\int_0^\infty |u| dt$. The theory and applications of optimal control are an effort to include these objectives directly in the design. In this chapter, we restrict ourselves to analysis of control effort after the fact, and we suggest that a simulation of the final design will determine whether design is satisfactory from the point of view of control effort. If given a choice, we will prefer that design which gives the smallest value of control effort while meeting the error (dynamic and steady-state) specifications. Many of these ideas are discussed further in Chapters 9 and 11.

Finally, *sensitivity to parameter changes* needs to be studied separately for changes in plant parameters and for changes in controller parameters. As to changes in the parameters of the plant, the situation is very much like the disturbance-signal rejection, and both features are contained in the concept of robustness as discussed briefly in Chapter 2. The larger the gain of the feedback loop around the offending parameter, the lower the sensitivity of the transfer function to changes in that parameter. Because in the most common cases we have very slowly varying parameters, we are led to design for high gain in the vicinity of $z = 1$, which corresponds to very slow or constant signals. If this high gain is in front of the disturbance, then we will also achieve good disturbance rejection. The second aspect concerns the effects of changes in the controller, $D(z)$. Here, we have control over the topology, and a design choice can be made to minimize the effects of parameter changes in $D(z)$. Furthermore, in a digital control, the effects of round-off errors and truncation in realization of parameters in $D(z)$ are important; in Chapter 7 we discuss selection of canonical realizations to minimize these effects.

The designer's job is to meet the specifications. This can typically be done in many different ways. There are different kinds of actuators that can be used, different sensors and sensed quantities are possible for selection, choices in the plant design are a design variable, and the control law can be selected. The designer must pick the most cost-effective combination of these options to meet the desired system performance. This chapter and Chapters 6 and 9 will concentrate on the options available in designing the control law; however, the reader is encouraged to examine the design example in Section 12.7 in order to have a more accurate picture of the kind of options that a designer typically has available.

5.3 DESIGN USING EMULATION

The first part of this design procedure should already be familiar to the reader because it consists of the design of a continuous-control system. Many textbooks have been written on the subject, for example, Franklin, Powell, and Emami (1986). The design is done in the s -plane, using root-locus or frequency-response techniques to derive a satisfactory $D(s)$ as the controller. This step totally ignores the fact that a sampler and digital computer will eventually be used. Having $D(s)$, we then convert the design to a digital control by considering $D(s)$ to be a filter transfer function and applying one of the techniques from Chapter 4 to obtain an equivalent $D(z)$. The example designs will show that the method produces a good controller for the case where the sampling rate is 30 times faster than the system bandwidth but produces a controller needing further refinement for the case where the sampling rate is 6 times the bandwidth.

Carrying out the initial design using continuous methods is a good idea independent of whether it will be used in a subsequent emulation step or merely as a guide for a direct discrete design. Knowing how the system could perform if implemented with continuous hardware provides a target for how well the digital system should perform and aids in selecting the sample rate.

Another method for emulating a continuous design is to use an optimal control formulation. This is discussed in Section 9.4.5.

Example 5.3: We will now apply the design technique to the task of determining a discrete controller for the antenna tracker discussed in examples 5.1 and 5.2.

The specification that the overshoot be less than 16% requires, according to Fig. 5.4(a), that ζ must be ≥ 0.5 for a second order system. The specification of settling time, $t_s \leq 10$ sec, with (5.13),

$$t_s = \frac{4.6}{\zeta\omega_n}$$

shows that a closed loop $\omega_n \cong 1$ rad/sec will satisfy the requirements.

The process starts by applying the s -plane techniques for control-system design. The root locus in Fig. 5.7 shows that canceling the one pole of $G(s)$ at $s = -0.1$ with a zero in a lead compensator, placing the compensator pole at $s = -1$, and using a dc gain of 1,

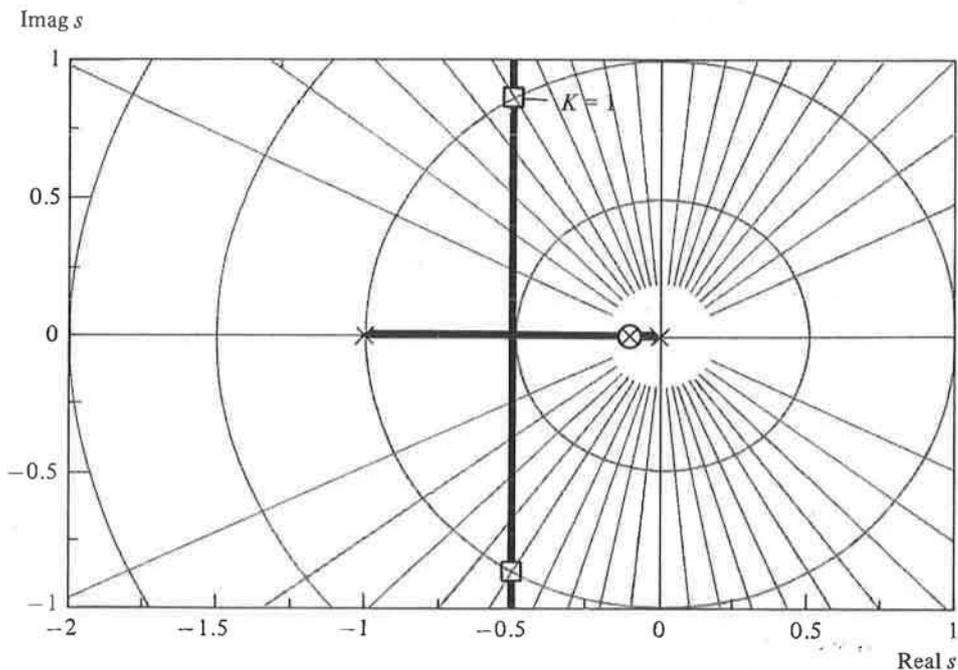


Figure 5.7 Root locus in the s -plane for the antenna tracking control.

that is,

$$D(s) = \frac{10s + 1}{s + 1} \quad (5.15)$$

provides closed-loop roots that satisfy the requirements for ζ and ω_n . Because this is a second-order system, this ensures that the overshoot limit of 16% will also be met. We have already seen that a $D(s)$ with a dc gain of 1 produces $K_v = 1$, which will satisfy the steady-state error requirements. This controller results in a closed-loop block diagram as shown in Fig. 5.8.

What should $D(z)$ be? We could use any of the discrete equivalent methods of Chapter 4. Due to its simplicity of use, we choose

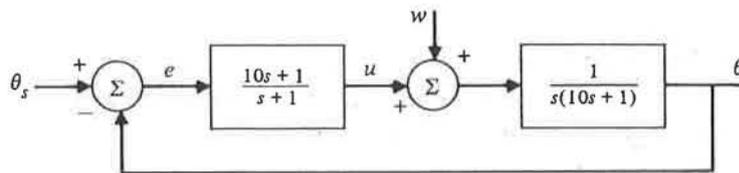


Figure 5.8 Closed-loop block diagram of the antenna control.

to illustrate the zero-pole mapping technique. First, we must select the sampling rate ω_s . We have designed the system to have $\omega_n = 1$ rad/sec or 0.16 Hz. A safe sample rate, if plenty of computer power is available, is a factor of 20 or more times the closed-loop bandwidth of the system.² For this second-order case, the bandwidth is approximately equal to the natural frequency, and therefore a conservative sample rate would be 3 Hz or higher. We initially choose a very safe $\omega_s = 5$ Hz, that is, $T = 0.2$ sec.

The compensation, $D(s)$, has two first-order factors (5.15); the zero is at $s = -0.1$, and the pole is at $s = -1$. The pole-zero mapping technique requires that each singularity is mapped according to $z = e^{sT}$; therefore, where $D(z)$ is of the form

$$D(z) = K \frac{z - z_1}{z - z_2},$$

there will be a zero at

$$z_1 = e^{(-0.1)(0.2)} = 0.9802,$$

a pole at

$$z_2 = e^{(-1)(0.2)} = 0.8187,$$

and because the dc gain of $D(z)$ and $D(s)$ must be identical, we have that

$$\begin{aligned} \text{dc gain} &= \lim_{z \rightarrow 1} D(z) = \lim_{s \rightarrow 0} D(s) = 1 \\ &= K \frac{1 - 0.9802}{1 - 0.8187}. \end{aligned} \quad (5.16)$$

Therefore we have

$$K = 9.15,$$

and the design of the discrete compensation is

$$D(z) = 9.15 \frac{z - 0.9802}{z - 0.8187}. \quad (5.17)$$

²Selection of sampling rates is discussed with greater detail in Chapter 10.

This is converted into a difference equation for implementation into a computer using the ideas developed in Chapter 2. Specifically, we first multiply top and bottom by z^{-1} to obtain

$$D(z) = \frac{u(z)}{e(z)} = 9.15 \frac{1 - 0.9802z^{-1}}{1 - 0.8187z^{-1}},$$

which can be restated as

$$u(z)(1 - 0.8187z^{-1}) = 9.15 e(z)(1 - 0.9802z^{-1}).$$

The z -transform expression above is converted to the difference equation form by noting that z^{-1} represents a 1-cycle delay. Thus

$$u(k) = 0.8187u(k-1) + 9.15(e(k) - 0.9802e(k-1)).$$

This equation can be directly evaluated by a computer provided that one past value of the control output and the error input have been saved. The actual code to implement the equation in a control computer might look something like that described in Table 5.1. Note that the calculation of u' in the table is coded so as to minimize the time between sample and output. Generally, one performs all the calculations that do not depend on *that* cycle's sample before the actual sample so as to minimize the delay. More details on implementation are contained in Chapter 12.

A description of a digital controller that appears to satisfy the specifications for the antenna is now complete. The designer has three options at this point. First, the controller could be implemented in a control computer, connected to the antenna system, turned on, and its performance observed to verify whether it really meets the desired specifications. A second option would be to perform a z -plane analysis of the entire system in order to theoretically determine the effect of the approximation brought about by the discretization of the continuous $D(s)$. A third option would be to simulate the entire system (controller plus antenna) in a computer and to observe the computed response of the system.

Let's proceed with the second and third options. To analyze the behavior of this compensation, we must determine the z -transform of the continuous plant (Fig. 5.1) preceded by a zero-order hold (ZOH)

Table 5.1 Real-time controller implementation.

-
-
1. Initialize.
 2. $u_{\text{old}} = 0$
 3. $u' = 0$
 4. $e_{\text{old}} = 0$
 5. Start control loop.
 6. Sample A/D converter to obtain y ($= y(k)$).
 7. Sample A/D converter to obtain r ($= r(k)$).
 8. $e = y - r$
 9. $u = u' + 9.15 e$
 10. Send u to D/A converter.
 11. $u_{\text{old}} = u$
 12. $e_{\text{old}} = e$
 13. $u' = 0.8187 u_{\text{old}} - (8.969) e_{\text{old}}$
 14. Wait for end of sample period, T sec.
 15. Go to 5.
-
-

by using (5.5). Applying (5.5) to the $G(s)$ in (5.2), we obtain

$$G(z) = \frac{z-1}{z} \mathcal{Z} \left\{ \frac{a}{s^2(s+a)} \right\}, \quad (5.18)$$

which is

$$G(z) = \frac{z-1}{z} \mathcal{Z} \left\{ \frac{1}{s^2} - \frac{1}{as} + \frac{1}{a} \frac{1}{s+a} \right\}.$$

Using the tables in Appendix B, we find

$$\begin{aligned} G(z) &= \frac{z-1}{z} \left\{ \frac{Tz}{(z-1)^2} - \frac{z}{a(z-1)} + \frac{1}{a} \frac{z}{z-e^{-aT}} \right\} \\ &= \frac{Az+B}{a(z-1)(z-e^{-aT})}. \end{aligned}$$

$$A = e^{-aT} + aT - 1, \quad B = 1 - e^{-aT} - aTe^{-aT}.$$

For this example, $T = 0.2$ and $a = 0.1$ (see X-C2D in Table E.1), and this evaluates to

$$G(z) = 0.00199 \frac{z + 0.9934}{(z - 1)(z - 0.9802)}. \quad (5.19)$$

The z -plane roots of a system with a digital controller are found by solving the closed-loop characteristic equation,

$$1 + D(z)G(z) = 0.$$

For the antenna system, this becomes

$$1 + 9.15 \frac{(z - 0.9802)}{(z - 0.8187)} \frac{(0.00199)(z + 0.9934)}{(z - 1)(z - 0.9802)} = 0,$$

which has roots at (see RLOCUS in Table E.1)

$$z = 0.900 \pm j0.162.$$

These roots can be evaluated in terms of the corresponding ζ and ω_n by using Fig. 5.6, or they could be converted back to the s -plane using the inverse of $z = e^{sT}$,

$$s = \frac{1}{T} \ln(z).$$

This calculation shows that the system has equivalent s -plane roots³ at

$$s = -0.446 \pm j0.891,$$

which indicates that the specified values have been modified by the approximations in the discrete equivalent to

$$\zeta = 0.447 \text{ (from } \zeta = 0.5),$$

$$t_s = 10.3 \text{ sec (from 10 sec),}$$

$$\text{overshoot} = 20.8\% \text{ (from 16\%),}$$

³The equivalence is exact only for root locations; the time response and frequency response will not match exactly.

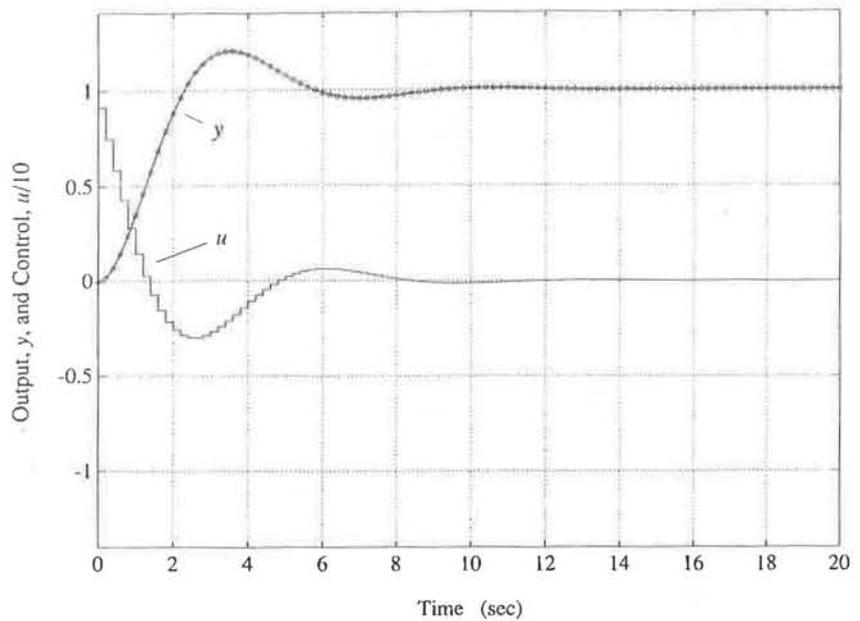


Figure 5.9 Step response of the 5-Hz controller (5.17).

which are only slightly degraded. Furthermore, the simulated output of the system represented by $D(z)$ and $G(z)$ is shown in Fig. 5.9, and it confirms that the controller will perform as indicated by the calculations above. This simulation was carried out using the linear, discrete model of the system; however, simulations are often embellished with the important nonlinearities and computation delays in the system in order to assess their effect in addition to the effect of the discretization approximations.

Sampling at a rate that is over 20 times faster than the bandwidth is a good, safe rule of thumb; however, it is sometimes necessary to sample more slowly to minimize the task's demand on the computer. If sampling is done more slowly, the quality of the dynamic response goes down somewhat because of the increased delay of $T/2$ added by the ZOH, and random disturbances have a greater influence on the system. Furthermore, when using the emulation design method, the approximations from the discrete equivalent can produce substantial deviations from the desired control performance and could even destabilize an otherwise stable system. We will see in Section 5.4 how to fix the performance degradation from the emulation design method when used at sample rates less than 20 times the bandwidth.

As an illustration, let's repeat the antenna design with a sample rate of 1 Hz ($T = 1$ sec), that is, where the sample rate is approximately 6 times the bandwidth.

Repeating the calculations in (5.16) with $T = 1$ sec, we obtain

$$D(z) = 6.64 \frac{z - 0.9048}{z - 0.3679}. \quad (5.20)$$

Furthermore, repeating the calculations in (5.18) with $T = 1$ sec results in

$$G(z) = 0.0484 \frac{z + 0.9672}{(z - 1)(z - 0.9048)}. \quad (5.21)$$

Combining (5.20) and (5.21), we find that the closed-loop roots are at

$$\begin{aligned} z &= 0.523 \pm j0.636, \\ s &= -0.194 \pm j0.883, \end{aligned}$$

which indicates that the system should have

$$\begin{aligned} \zeta &= 0.21 \quad \text{instead of the desired } \zeta = 0.5, \\ t_s &= 23.7 \text{ sec} \quad \text{instead of 10 sec,} \\ \text{overshoot} &= 52\% \quad \text{instead of 16\%}. \end{aligned} \quad (5.22)$$

A plot of the step response of the resulting system, (5.20) and (5.21), is shown in Fig. 5.10 and generally verifies the degradation predicted by the discrete analysis in (5.22).

The response in Fig. 5.10 and the analysis both show a degradation of the overshoot from 16% to about 50%, which corresponds to a damping ratio decrease from $\zeta = 0.5$ to $\zeta = 0.2$. Clearly the accuracy of the approximation is not adequate in this case. And in general, the emulation design method will not yield accurate results for sample rates slower than 20 times the bandwidth. However, as we shall see in the next section, the emulation method can be used with slower sample rates to determine an initial compensation, which is then "patched up" to obtain good response using z -plane design methods.

The explanation of the performance degradation shown in the example lies in the fact that even if $D(z)$ generates nearly the same sample values

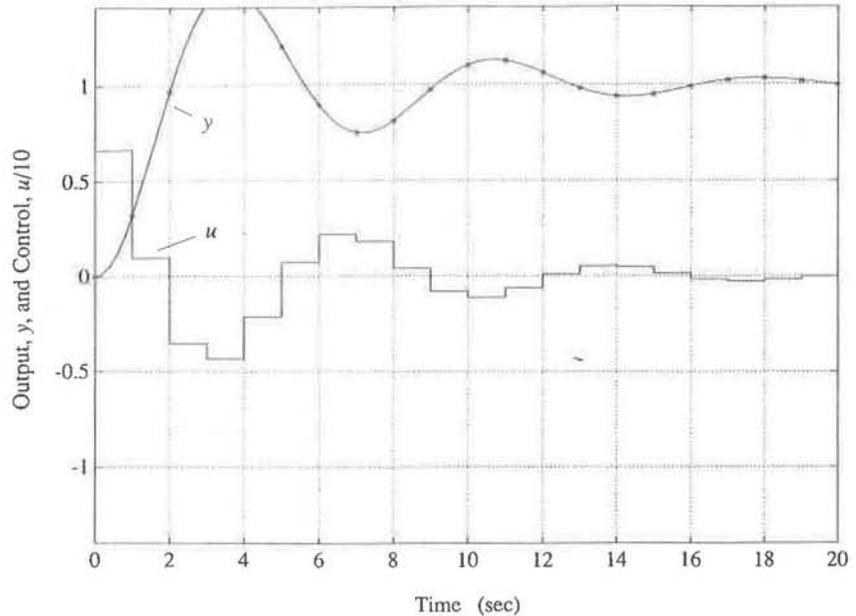


Figure 5.10 Step response of the 1-Hz controller (5.20).

as $D(s)$, the zero-order hold reconstruction of u is only an approximation to the continuous u assumed in the design of $D(s)$. In fact, Fig. 1.3 shows that the u from the ZOH is at best an approximation to a delayed version of $u(t)$, delayed about $T/2$ sec. A time delay is well known to produce phase lag and generally leads to a less stable design if not taken properly into account.⁴ The effect of the ZOH lag is sometimes anticipated when doing the original s -plane design, thus causing extra lead to be placed in the $D(s)$; therefore, the discretization simply degrades the results back to the point that the designer originally intended.

Although the simple second-order system described in this section can be designed by hand calculations, design of more complex systems requires the availability of computer-aided design tools. In fact, when sampling at 20 times bandwidth or more, we saw from the example in (5.16) that the calculations need to be carried out to four-place accuracy so as to retain two-place accuracy in the final digital controller. The reason for this sensitivity is that derivative information is being obtained by differencing two sequential

⁴At $\omega_n = 1$ and $T = 1$ sec, a delay of $T/2$ introduces a phase lag of $1/2$ rad, or about 30° . Because $\zeta = 0.5$ corresponds to a phase margin (see Section 5.5) of about 50° , the delay reduces that to a phase margin of 20° corresponding to $\zeta = 0.2$, which generally agrees with (5.22) and the step response of Fig. 5.10.

values that become almost identical for fast sampling. Therefore, even the second-order case with a fairly fast sample rate is very tedious without computer-based design tools. The standard routines that exist in most CAD packages that would aid in the calculations for this section and/or would generally be useful in design are listed below with generic names referring to entries in Table E.1 that give the specific names in several CAD packages.

1. RLOCUS: A program to compute the root locus of a closed-loop system given the open-loop transfer function. The routine is equally applicable to s -plane and z -plane system descriptions; however, when working in the z -plane, it is useful to provide a square aspect ratio and to include the unit circle, features that have been incorporated in the routines under ZLOCUS in Table E.1.
2. X-C2D: A program to convert a continuous system preceded by a ZOH to a discrete system.
3. EQUIVNT: A program to determine the various discrete equivalents, (bilinear, zero-pole mapping, etc.).
4. SIMULATE: A program to compute the response of a discrete system to impulses, steps, or initial conditions.

5.4 z -PLANE DESIGN USING ROOT LOCUS

The second method of design is to determine the controller directly in the z -plane. At the outset, the plant model is transformed to a discrete system using (5.5), and the design iterations to achieve the desired system specifications are carried out using z -plane analysis tools. By carrying out the digitization on the plant model instead of on the $D(s)$ as was done for the emulation design method, the approximate nature of the process can be eliminated. This is so because the actual plant must be preceded by a hold (usually a ZOH) and, therefore, has an exact discrete equivalent that includes the lagging effect of the hold. There is no exact equivalent for a $D(s)$ because its continuous response is dependent on the input signal between sample instances as well as at sample instances. Each discrete equivalent approximation technique is essentially an assumption on what the input signal is doing between the samples, and no one assumption can ever anticipate all types of input signals. The impact of these ideas is that a $D(z)$ found using discrete design methods will yield performance when implemented which is very close to the desired specifications for fast or slow sample rates.

To demonstrate the z -plane design method, we will initially employ the root-locus design tool. In subsequent sections and chapters, we will demonstrate z -plane design using frequency-response, direct-design, state-space, and optimal-control methods.

The root locus is the locus of points where roots of a characteristic equation can be found as some real parameter varies from zero to large values.⁵ From Fig. 5.2 and block-diagram analysis, the characteristic equation of the single-loop system is

$$1 + D(z)G(z) = 0. \quad (5.23)$$

The significant thing about (5.23) is that this is exactly the same equation as that found for the s -plane root locus. The implication is that the mechanics of drawing the root loci are exactly the same in the z -plane as in the s -plane; that is, the rules for the real axis, asymptote construction, and arrival/departure angles are all unchanged from those developed for the s -plane. The difference lies in the interpretation of the results because the pole locations mean different things when we come to interpret the system stability and dynamic response.

Example 5.4: Suppose we design the antenna system for the slow sampling case with $T = 1$ sec. The exact discrete model of the plant plus hold is given by the $G(z)$ in (5.21). If the controller consisted simply of a proportional gain [$u = K(\theta_s - \theta)$], the locus of roots versus K can be found by solving the characteristic equation

$$1 + 0.0484 K \frac{z + 0.9672}{(z - 1)(z - 0.9048)} = 0$$

for many values of K . The result is shown in Fig. 5.11 as the dashed arc marked (a). From study of the root locus we should remember that this locus, with two poles and one zero, is a circle centered at the zero ($z = -0.9672$) and breaking away from the real axis between the two real poles at the point where K versus z is a maximum for real z (see Problem 5.3). Here it is almost halfway, a bit closer to 0.9048 than 1.0.

⁵Sometimes we are interested in negative parameter values and look at root loci for the parameter in the entire range $-\infty \leq K < \infty$. The loci for positive gain are the most common.

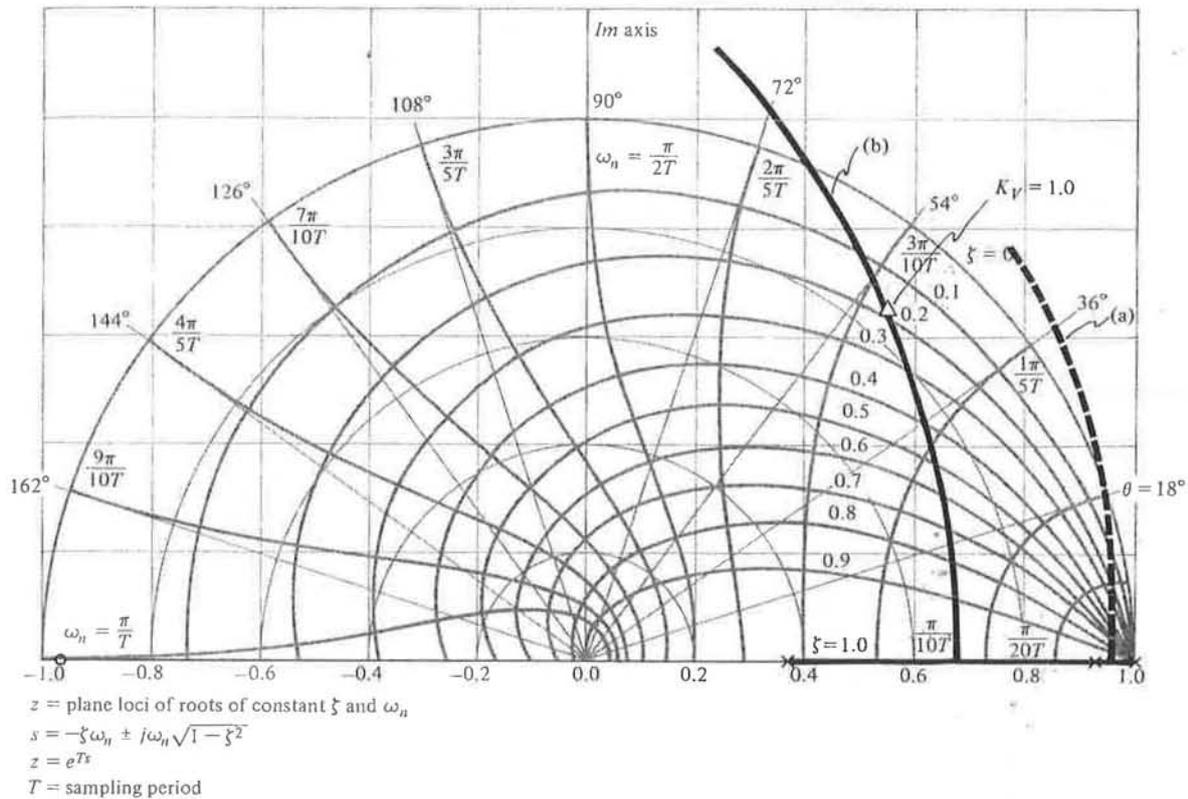


Figure 5.11 Root loci for antenna design: (a) uncompensated system; (b) locus for $D(z)$ with poles and zeros of (5.20).

From the root locus of the uncompensated system (Fig. 5.11a) it is clear that some dynamic compensation is required if we are to get satisfactory response from this system. The radius of the roots never gets less than 0.95, preventing the t_s specification from being met; and the system goes unstable at $K \cong 19$ [where $K_v = 0.92$, as can be verified by using (5.8)], which means that there is no stable value of gain that meets the steady-state error specification.

If we cancel the plant pole at 0.9048 with a zero and add a pole at 0.3679, we are using the compensation of (5.20). The root locus for this versus the gain K [K was equal to 6.64 in (5.20)] is also sketched in Fig. 5.11 as the solid curve (b). The point where $K = 6.64$ is marked by a triangle, and we can see that a damping ratio of about 0.2 is to be expected, as we have previously seen from (5.22) and the step response of Fig. 5.10. This point, however, does have the specified value of $K_v = 1$ because this criterion was used

in arriving at (5.20). The locus shows that increasing the gain, K , would lower the damping ratio still further. Better damping could be achieved by decreasing the gain, but then the criterion of steady-state error would be violated. It is therefore clear that this choice of compensation pole and zero cannot meet the specifications.

A better choice of compensation can be expected if we transform the specifications into the z -plane and select the compensation so that the closed loop roots meet those values. The original specifications were $K_v \geq 1$, $t_s \leq 10$ sec, and overshoot less than 16%, which translated into $\zeta \geq 0.5$. In the s -plane, these specifications would be met if $K_v \geq 1$, the roots were to the left of the line $\text{Re}(s) = -0.5$ rad/sec and to the left of the $\zeta = 0.5$ lines (30° from vertical). These specifications were met with $K_v = 1$ and root locations at

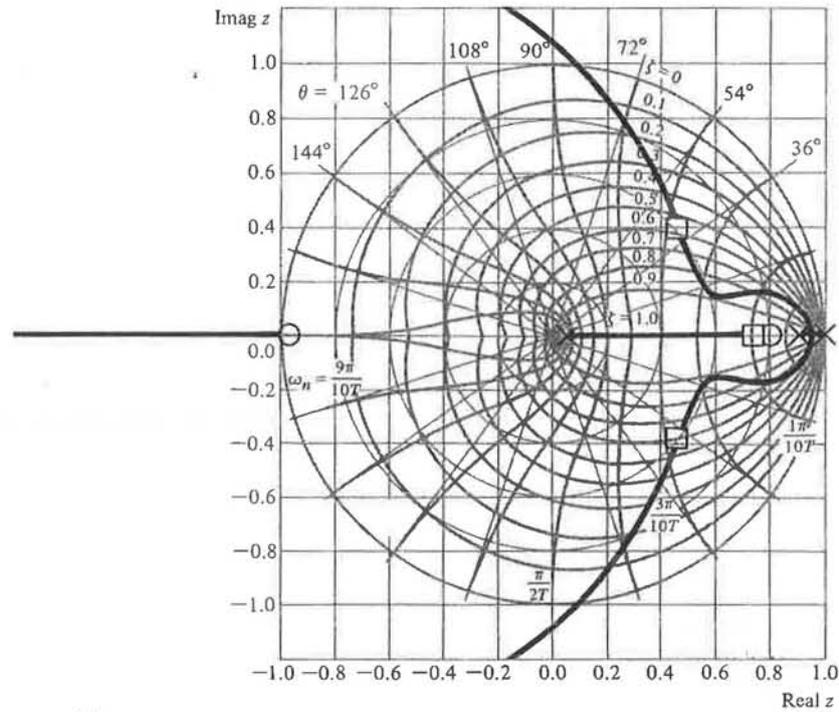
$$s = -0.5 \pm j0.867.$$

A guide for arriving at acceptable compensation in the z -plane design is to transform these root locations to the z -plane using $z = e^{sT}$. This yields

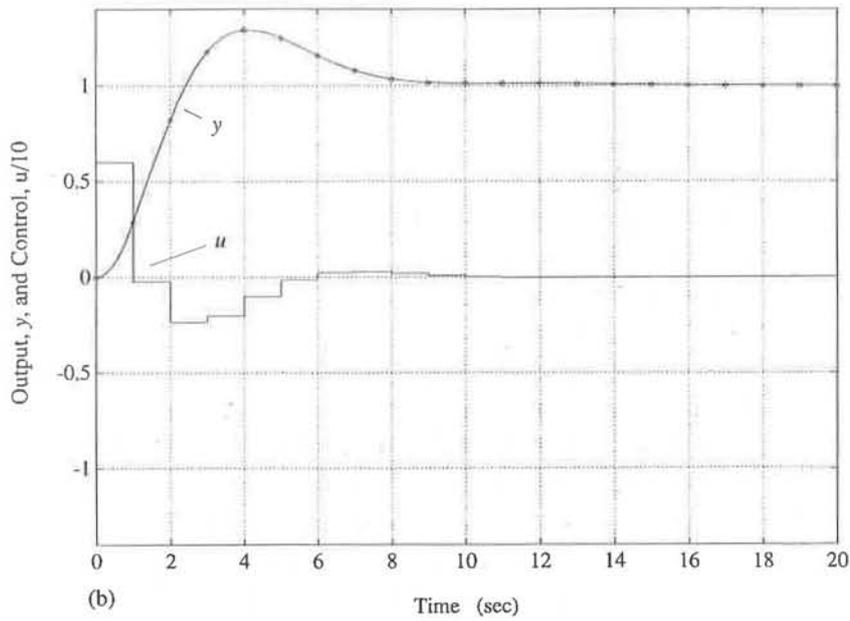
$$z = 0.392 \pm j0.462.$$

One could also transform the specifications directly to the z -plane as was done in Fig. 5.6. This approach leads us to conclude that the t_s specification requires that the roots be inside the radius $r = e^{-0.5} = 0.61$ and the overshoot requires that the roots are inside the $\zeta = 0.5$ spiral. The requirement that $K_v \geq 1$ applies in either plane but is computed by (5.8) for the z -plane.

It is typically advantageous to use the design obtained using emulation and to modify it using discrete design methods so that it is acceptable. The problem with the emulation-based design is that the damping is too low at the mandated gain, a situation that is typically remedied by adding more "lead" in the compensation. More lead is obtained in the s -plane by increasing the separation between the compensation's pole and zero; and the same holds true in the z -plane. Therefore, for a first try, let's keep the zero where it is (canceling the plant pole) and move the compensation pole to the left until the roots and K_v are acceptable. After a few trials, we find that there is no pole location that satisfies all the requirements! Although moving the pole to the left of $z \cong +0.05$ will produce acceptable z -plane pole locations, the gain K_v is not sufficiently high to meet the



(a)



(b)

Figure 5.12 Antenna design $D(z)$ given by (5.24): (a) root locus, (b) step response.

criterion for steady-state error. The only way to raise K_v , and to meet the requirements for damping and settling time is to move the zero to the left also.

After some trial and error, we see that

$$D(z) = 6 \frac{z - 0.80}{z - 0.05} \quad (5.24)$$

meets the required z -plane constraints for the complex roots and has a $K_v = 1.26$. The root locus for (5.24) is shown in Fig. 5.12(a), and the roots corresponding to $K = 6$ are marked by squares. The fact that all requirements seem to be met is encouraging, but there is an additional real root at $z = 0.74$ and a zero at $z = 0.8$, which may degrade the actual response from that expected if it were a second-order system. The actual time history is shown in Fig. 5.12(b). It shows that the overshoot is 29% and the settling time is 15 sec. Therefore, further iteration is required to improve the damping and to prevent the real root from slowing down the response.

A compensation that achieves the desired result is

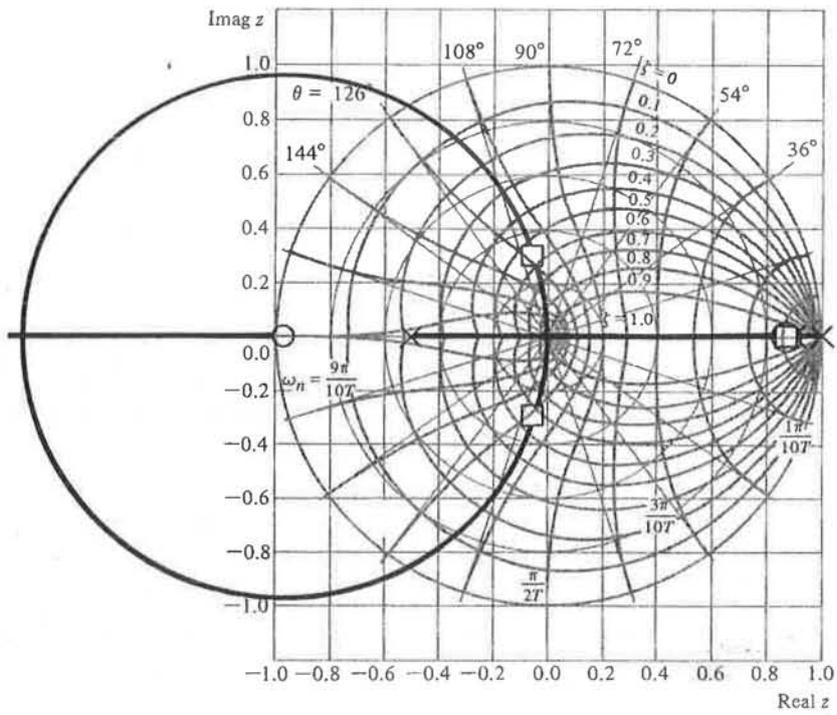
$$D(z) = 13 \frac{z - 0.88}{z + 0.5} \quad (5.25)$$

The damping and radius of the complex roots substantially exceed the specified limits, and $K_v = 1.04$. Although the real root is slower than the previous design, it is very close to a zero that attenuates its contribution to the response. The root locus for all K 's is shown in Fig. 5.13(a) and the time response for $K = 13$ in Fig. 5.13(b).

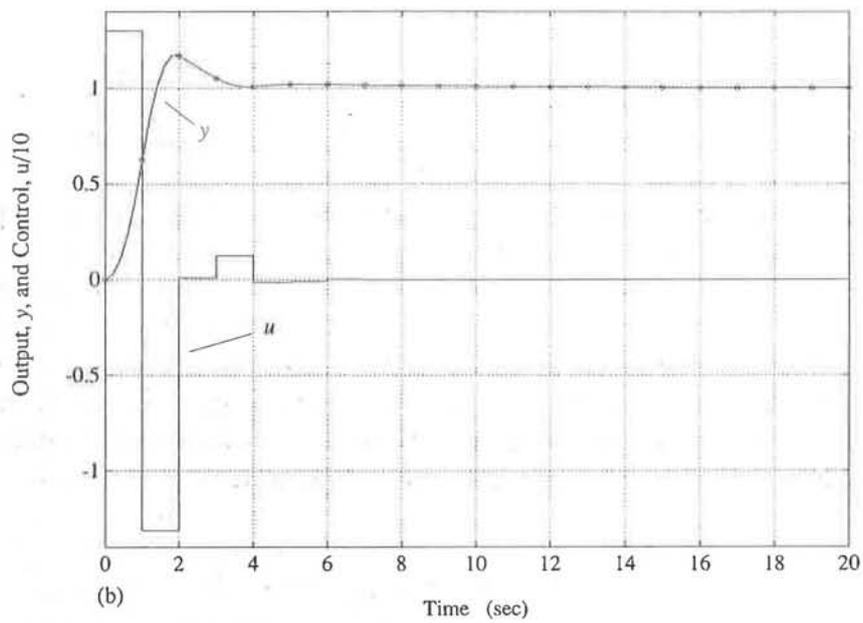
Note that the pole of (5.25) is on the negative real z -plane axis. In general, placement of poles on the negative real axis should be done with some caution. In this case, however, no adverse effects resulted because all roots were in well damped locations. As an example of what could happen, consider the compensation

$$D(z) = 9 \frac{(z - 0.8)}{(z + 0.8)} \quad (5.26)$$

Its root locus versus K and step response are shown in Fig. 5.14. All roots are real with one root at $z = -0.59$. But this negative real axis root has $\zeta = 0.2$ and represents a damped sinusoid with



(a)



(b)

Figure 5.13 Antenna design $D(z)$ given by (5.25): (a) root locus, (b) step response.

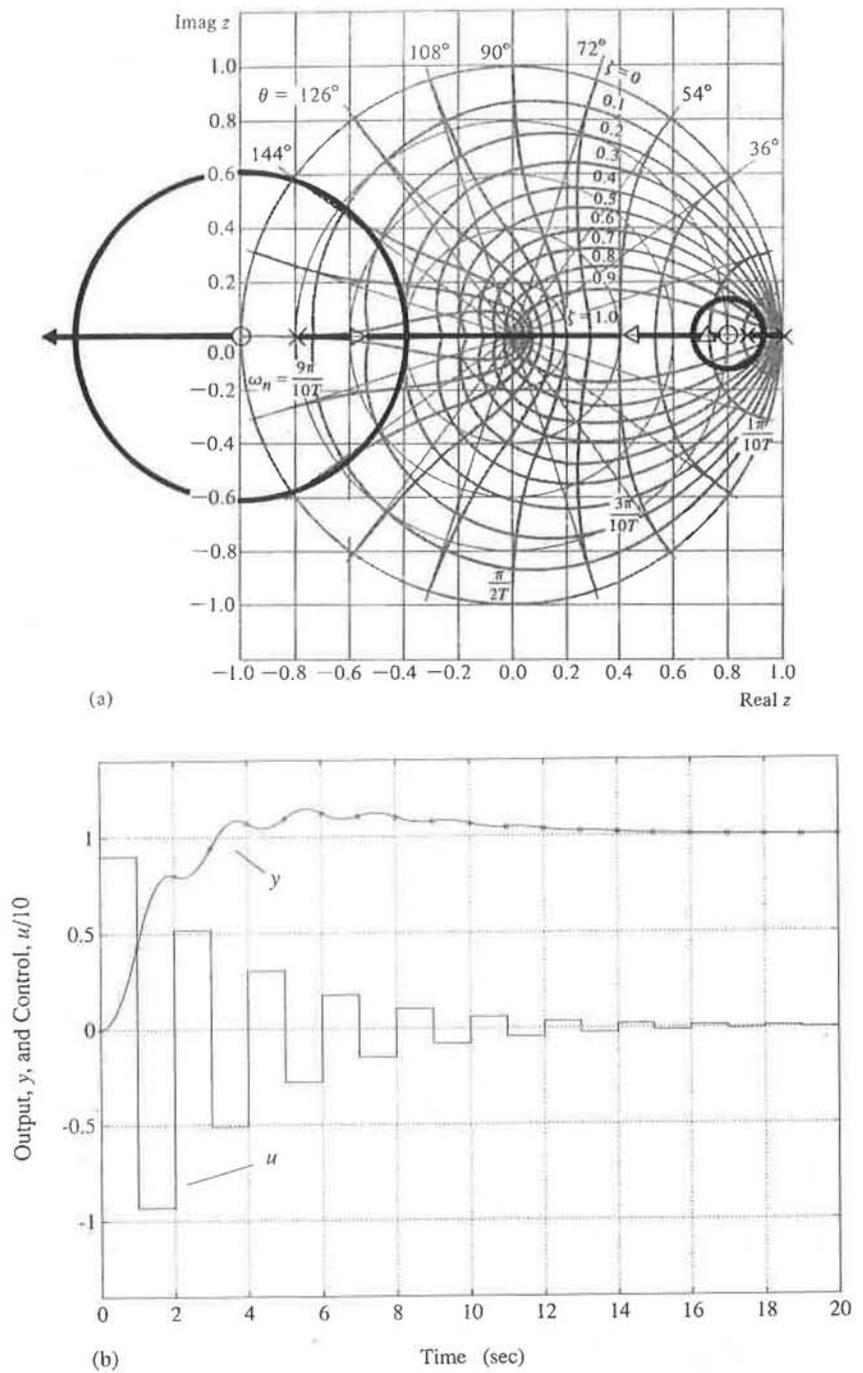


Figure 5.14 Antenna design $D(z)$ given by (5.26): (a) root locus, (b) step response.

frequency of $\omega_s/2$. The output has very low overshoot, it comes very close to meeting the settling time specification, and its $K_v = 1$; however, the control, u , has large oscillations with a damping and frequency consistent with the negative real root. This indicates that there are "hidden oscillations" or "intersample ripple" in the output that are only apparent by computing the continuous plant output between sample points as is done in Fig. 5.14. The computation of the intersample behavior was carried out by computing it at a much higher sample rate than the digital controller, taking care that the control value was constant throughout the controller sample period. (see OUTPUT in Table E.1). Note that, if only the output at the sample points had been determined, the system would appear to have very good response. Furthermore, this design uses much more control effort than that of Fig. 5.13 although, on the surface, its sampled output appears to meet the specifications. So we see that a compensation pole in a lightly damped location on the negative real axis could lead to a poorly damped system root and undesirable performance.

In the design examples to this point, the computed output time histories have assumed that the control, $u(k)$, was available from the computer at the sample instant. However, in a real system this is not precisely true. In the control implementation example in Table 5.1, we see that some time must pass between the sample of $y(k)$ and the output of $u(k)$ for the computer to calculate the value of $u(k)$. This time delay is called *latency* and usually can be kept to a small fraction of the sample period with good programming and computer design. Its effect on performance can be evaluated precisely using the transform analysis of Section 2.4.2, the state-space analysis of Section 2.4.4, or frequency response. However, it is easier and quicker to evaluate the effect of a full-cycle delay using root locus; assuming a potential delay of one full cycle bounds the possible effects of a partial cycle of latency. Furthermore, in computers not specifically designed for real-time control, one full cycle of delay between the sampling and the output is sometimes the case.

Because a one-cycle delay has a z -transform of z^{-1} , the effect of a full-cycle delay can be analyzed by adding z^{-1} to the numerator of the controller representation. This will result in an additional pole at the origin of the z -plane. If there is a delay of two cycles, two poles will be added to the z -plane origin, and so on.

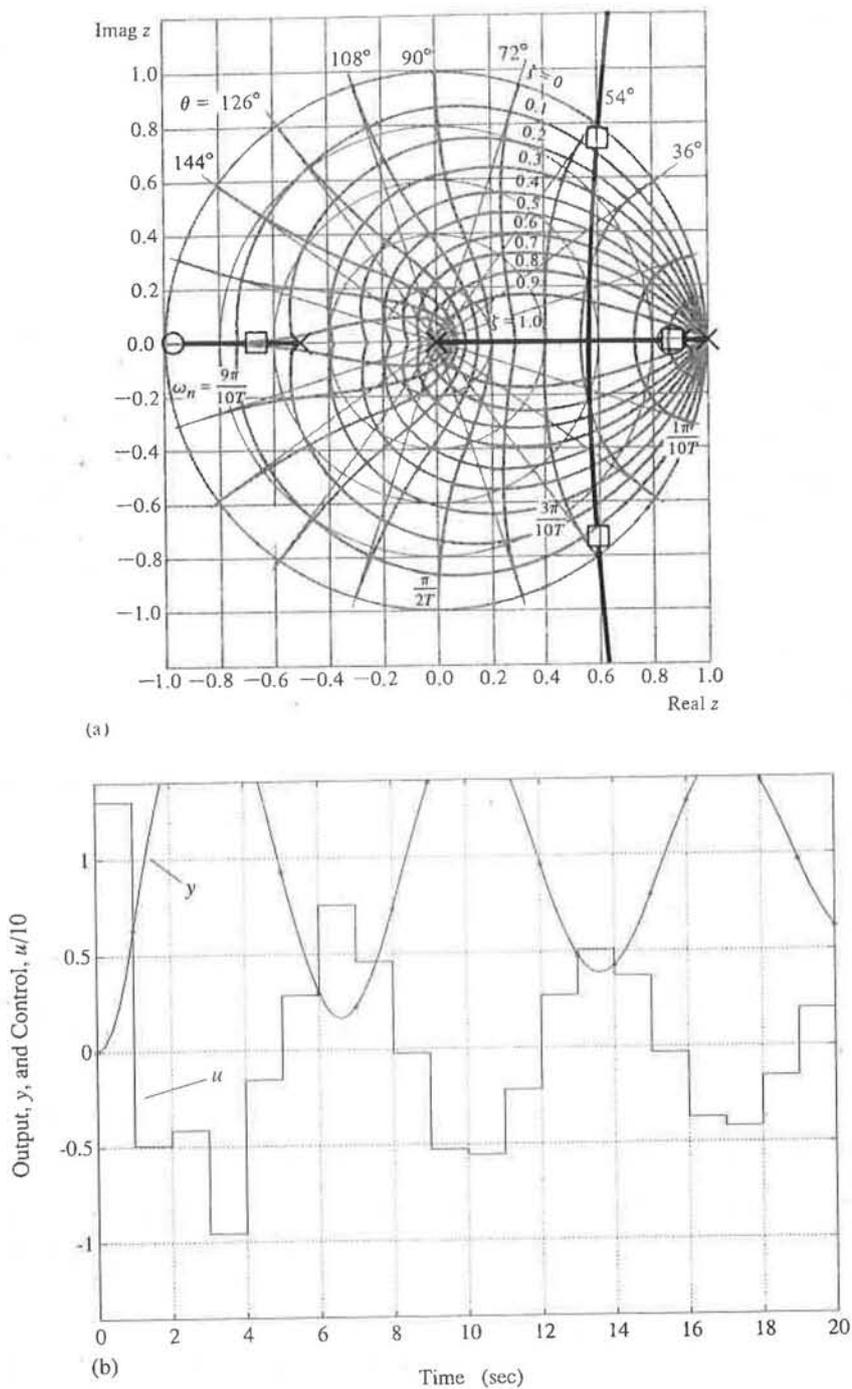


Figure 5.15 One-cycle-delay antenna design $D(z)$ given by (5.27): (a) root locus, (b) step response.

Example 5.5: To get some idea of the effect of delays, let's add one cycle delay to the compensation of (5.25). The new controller representation is

$$D(z) = 13 \frac{z - 0.88}{z(z + 0.5)}. \quad (5.27)$$

It results in the root locus and time response in Fig. 5.15, which are both substantially changed from the same controller without the delay in Fig. 5.13. The only difference is the new pole at $z = 0$. The severity of the one-cycle delay is due to the fact that this controller is operating at a very slow sample rate (six times the closed loop bandwidth). In fact this sensitivity to delays is one of many reasons why one would prefer to avoid sampling at this slow a rate because a one-cycle delay is sometimes a feature of a control computer.

5.5 FREQUENCY-RESPONSE METHODS WITH THE z -TRANSFORM

The frequency-response methods for control-system design were developed from the original work of Bode (1945) on feedback-amplifier techniques.

Their attractiveness for design depends on several ideas.

1. The gain and phase curves can be easily plotted *by hand*.
2. Nyquist's stability criterion can be applied, and dynamic response specifications can be readily interpreted in terms of gain and phase margins, which are easily seen on the plot of log gain and phase versus log frequency.
3. The system error constants, mainly K_p or K_v , can be read directly from the low-frequency asymptote of the gain plot.
4. The corrections to the gain and phase curves (and thus the gain and phase margins) introduced by a trial pole or zero of a compensator can be quickly and easily computed, *using the gain curve alone*.
5. The effect of pole, zero, or gain changes of a compensator on the speed of response (i.e., the crossover frequency) can be quickly and easily determined using the gain curve alone.

We will briefly review these points here as they apply to continuous systems. However, the books by James, Nichols, and Phillips (1947), Clark (1962), Ogata (1970), and Franklin, Powell, and Emami-Naeini (1986) give

pedagogic treatments of the ideas, and they should be referred to for a more complete review. We will concentrate the discussion on illustrating the changes required for applying frequency-response design methods to discrete systems. Section 2.6 introduces the idea of the response of a discrete system to a sinusoidal input. The basic concepts are the same as for continuous systems, but the evaluation of the magnitude and phase of a transfer function, $H(z)$, is accomplished by letting z take on values around the unit circle, $z = e^{j\omega T}$, that is,

$$\begin{aligned} \text{magnitude} &= |H(z)|_{e^{j\omega T}} \\ \text{phase} &= \angle H(z)|_{e^{j\omega T}}. \end{aligned} \quad (5.28)$$

These relationships make useless the hand-plotting procedures developed by Bode and his proof relating the phase to the magnitude curve on a log-log plot. The inability to use these ideas detracts from the ease with which a designer can predict the effect of pole and zero changes on the frequency response. Therefore, points 1, 4, and 5 above are less true for discrete frequency-response design using the z -transform than they are for continuous systems. With some care in the interpretations, points 2 and 3 are essentially unchanged. All these points will be discussed further in this section as they pertain to design using the z -transform.

The following section, Section 5.6, will discuss the use of frequency-response methods using the w -transform. The w -transform approach was developed so that the points above are almost as easy to use for discrete systems as they are for continuous ones. Therefore, in practice, frequency-response design of systems using a discrete model is often carried out using the w -transform; however, the need to replace the z -plane with the w -plane is less obvious in today's environment, where good software tools are universally available to perform the plotting for the designer.

5.5.1 Gain and Phase Plotting

For continuous systems, one can use Bode's hand-plotting techniques to generate plots of amplitude and phase versus frequency for the open-loop system. For discrete systems represented in the z -plane, the hand-plotting rules do not apply because z takes on values around the unit circle instead of the imaginary axis evaluation of s that is the basis for the hand-plotting

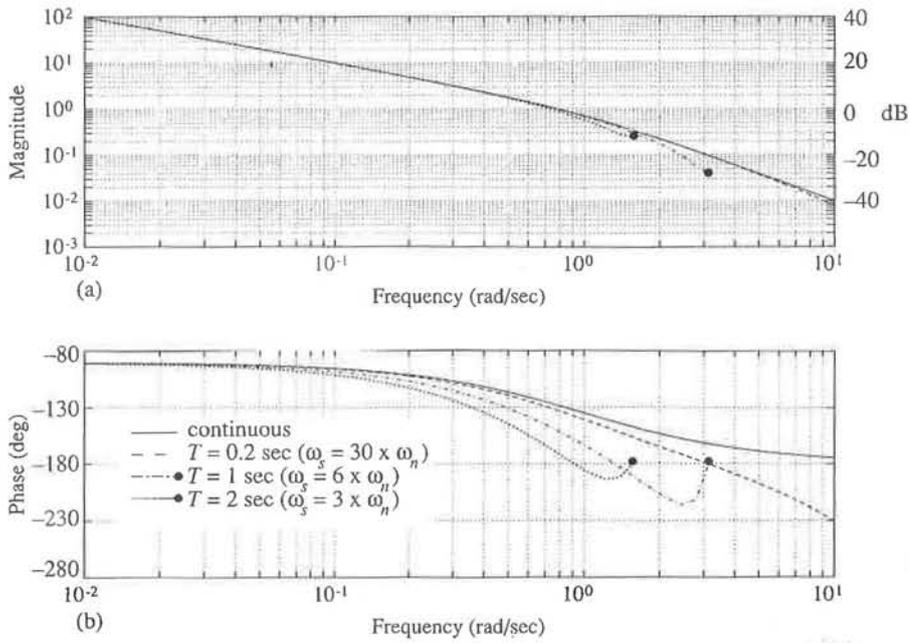


Figure 5.16 Frequency response of (5.29): (a) magnitude and (b) phase.

rules. The lack of hand-plotting capability has little impact if computer-based tools are available to perform the task. However, it is important for the designer to retain the ability to perform hand plotting of continuous-system frequency response because the general nature of the curves is similar to the discrete case, and the intuition gained can be used as a check. In fact, for fast sampling, the curves are virtually identical. As the sample rate slows to four times the frequency of interest, the phase curve departs from that of an equivalent continuous system.

Example 5.6: To illustrate the ideas above, Fig. 5.16 shows the magnitude and phase of

$$G(s) = \frac{1}{s(s + 1)} \tag{5.29a}$$

for s taking on values from $0 \leq j\omega < \infty$. Also shown are the magnitude and phase of its ZOH discrete equivalent at three sample rates,

that is,

$$\begin{aligned} G(z) &= 0.0187 \frac{(z + 0.9355)}{(z - 1)(z - 0.8187)} && \text{for } T = 0.2 \text{ sec,} \\ G(z) &= 0.368 \frac{(z + 0.718)}{(z - 1)(z - 0.368)} && \text{for } T = 1 \text{ sec,} \\ G(z) &= 1.135 \frac{(z + 0.523)}{(z - 1)(z - 0.135)} && \text{for } T = 2 \text{ sec,} \end{aligned} \quad (5.29b)$$

with z taking on values at $z = e^{j\omega T}$ for $0 \leq \omega T \leq \pi$.

The $G(s)$ magnitude curve could be approximated by two straight-line asymptotes intersecting at the breakpoint of $\omega = 1$ rad/sec. The phase hand-plotting rules would show quickly that the curve starts at -90° with a transition to -180° at $\omega = 1$ rad/sec. No analogous methods exist for plotting the $G(z)$ curves; they need to be done by computer (see FREQRESP in Table E.1). As should be expected, the discrete and continuous curves are very similar in nature because they all represent the response of the same physical process but at different sample rates. The fastest sampling case is extremely close to the continuous one, whereas the slower sampling cases progressively degrade from it.

Note from Example 5.6 that the primary effect of sampling is to cause an additional phase lag, whereas the amplitude response is affected very little. Fig. 5.17 shows the phase difference, $\Delta\phi$, between the continuous case and the discrete cases. The approximation to the discrete phase lag of

$$\Delta\phi = \frac{\omega T}{2} \quad (5.30)$$

is also shown and demonstrates the accuracy of the notion that the primary effect of the sampling in a digital controller is to delay the input by one half the sample period. The assumption is excellent for sample rates up to $\omega T = \pi/2$ which represents frequencies up to $1/4$ the sample rate. Crossover frequencies (magnitude = 1) will almost always be slower than $1/4$ the sample rate; therefore, one could obtain a good estimate of the phase margin of a sampled continuous system by simply subtracting the $\omega T/2$ factor from the phase of a continuous analysis.

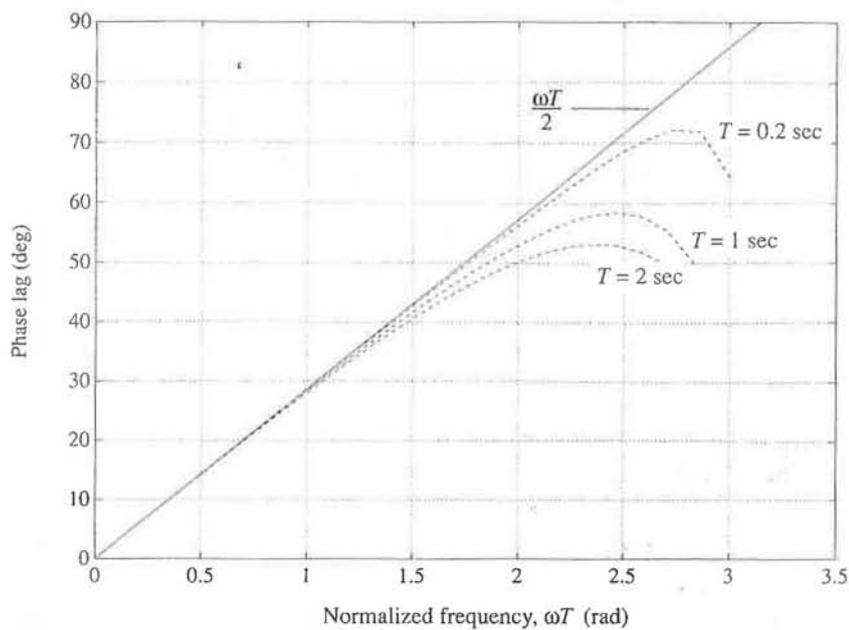


Figure 5.17 Phase lag due to sampling.

5.5.2 Nyquist Stability

For continuous systems, the Nyquist criterion establishes stability by determining whether there are any singularities within a contour that encloses the entire right-hand side (unstable region) of the s -plane. This leads to the determination of whether there are any zeros of

$$1 + D(s)G(s) = 0 \quad (5.31)$$

in the right hand plane, that is, whether the closed-loop system has any unstable roots. The entire contour evaluation is fixed by examining $D(s)G(s)$ over $s = j\omega$ for $0 \leq \omega < \infty$, which is also the frequency-response evaluation of the open-loop system. Figure 5.18(a) shows the full contour and the portion of the contour for $0 \leq j\omega < \infty$. The specific statement of the Nyquist stability criterion for continuous systems is

$$Z = N + P$$

where

Z = the number of unstable roots,

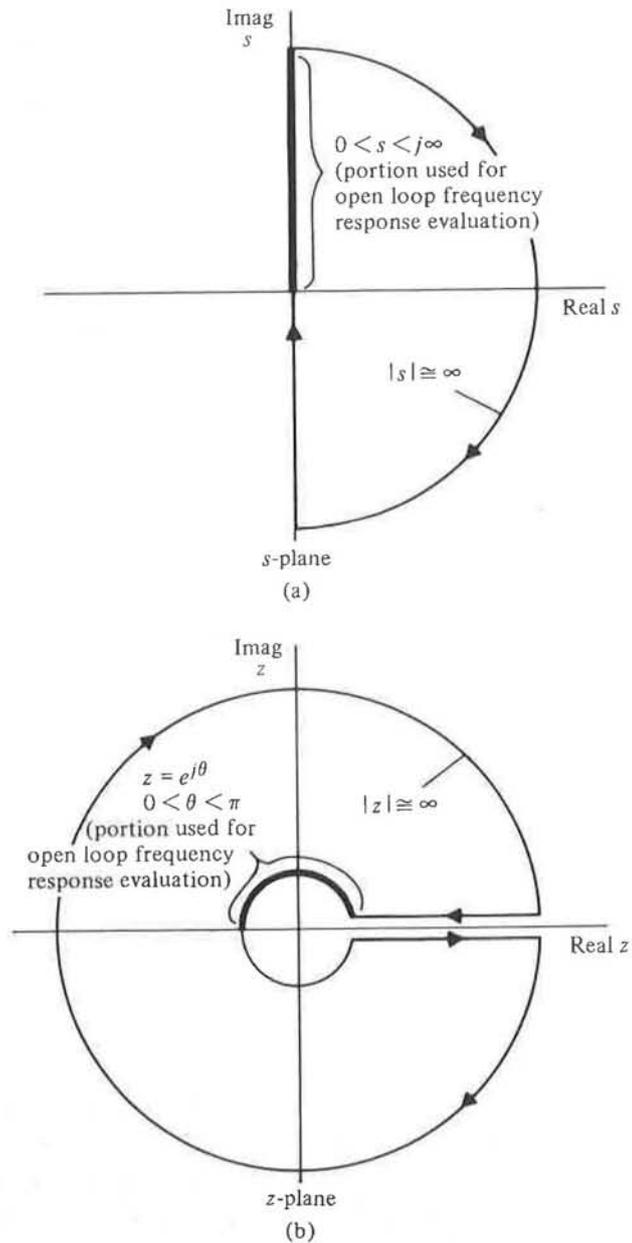


Figure 5.18 Nyquist criterion evaluation contour: (a) continuous systems, (b) discrete z -plane systems.

N = the number of clockwise encirclements of the -1 point for a contour evaluation of $D(s)G(s)$ with s taking on values as shown in Fig 5.18(a), and

P = the number of unstable poles of $D(s)G(s)$.

Therefore, for the predominant case of a stable open-loop system ($P = 0$) the closed-loop system is stable if the contour evaluation of $D(s)G(s)$ does not encircle the -1 point. For unstable open-loop systems, the closed-loop system is stable if the contour evaluation encircles the -1 point once for each unstable open-loop pole. The proof of this criterion relies on Cauchy's principle of the argument and is given in most textbooks on continuous control systems, including all those referred to above. For the discrete case, the ideas are identical. The only difference is that the contour enclosing the unstable region of the z -plane is the space outside the unit circle as shown in Fig. 5.18(b). Therefore the statement of the Nyquist stability criterion for discrete systems represented in the z -plane is

$$Z = N + P$$

where

Z = the number of unstable roots,

N = the number of clockwise encirclements of the -1 point for a contour evaluation of $D(z)G(z)$ with z taking on values as shown in Fig 5.18(b), and

P = the number of unstable poles of $D(z)G(z)$.

Example 5.7: To illustrate the criterion for the discrete case, let's evaluate the stability of the open-loop system

$$G(s) = \frac{1}{s(s+1)}$$

with proportional discrete feedback [$D(z) = K$] at a very slow sample rate of $1/2$ Hz. The discrete representation (5.29b) results in the magnitude and phase shown in Fig. 5.16 for $0 \leq \omega T \leq \pi$. The complete z -plane contour used in the evaluation of $D(z)G(z)$ is labeled in Fig. 5.19(a) to facilitate comparison with the results of the evaluation (called the *Nyquist plot*) in Fig. 5.19(b). Note that the portion from $a \rightarrow b$ is directly from Fig. 5.16, whereas the section from $b \rightarrow c$ is the same information with the phase reflected about 180° . All other portions were inferred from Fig. 5.19(a) based on the pole-zero configurations. Had the plot been generated by the evaluation of $KG(s)$ with s taking on values of the contour of Fig. 5.18(a), the differences would be that point b would have had zero magnitude, and the segment $defgh$ located along the positive real axis

of the discrete Nyquist plot would be absent. Because there are no -1 point encirclements, the system is stable for the $K = 1$ case as plotted. Note that the necessary information to determine stability is contained in the portion of the plot corresponding to the discrete frequency response of the open-loop system, that is, the portion from $a \rightarrow b$ in Fig. 5.19(b).

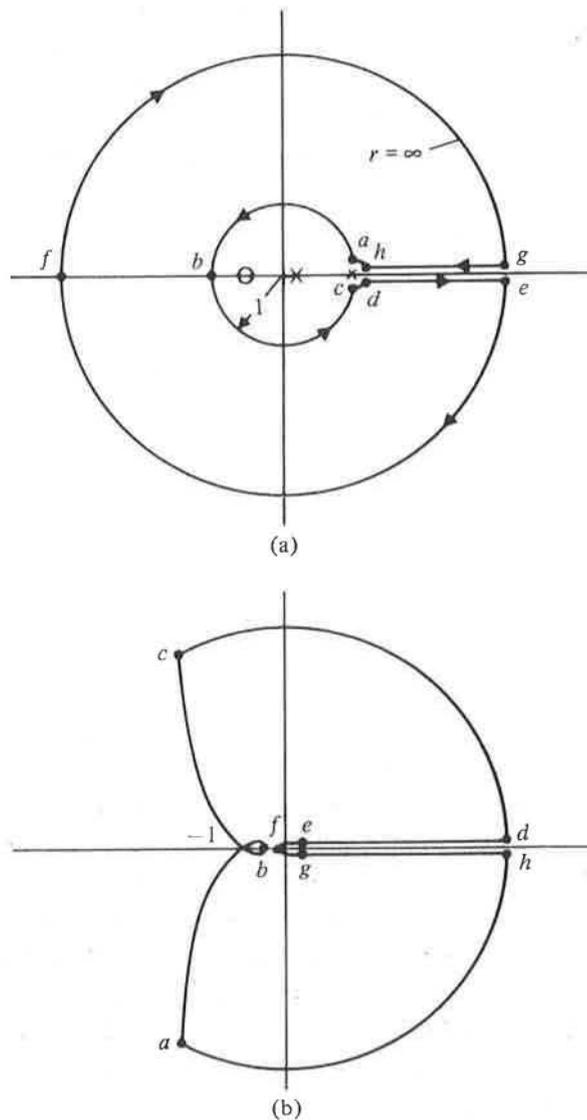


Figure 5.19 (a) Contour in z-plane for Example 5.7; (b) Nyquist plot of Example 5.7 using contour from (a).

The Nyquist plot shows the number of encirclements and thus the stability of the closed-loop system. The encirclements can be determined by the portion of the plot representing the frequency-response evaluation, specifically, whether it passes to the right or left of the -1 point. Gain and phase margins are defined so as to provide a two-dimensional measure of how close the Nyquist plot is to encircling the -1 point, and they are identical to the definitions developed for continuous systems. The Gain Margin (GM) is the inverse of the amplitude of $D(z)G(z)$ when its phase is 180° and is a measure of how much the gain of the system can increase before instability results. The Phase Margin (PM) is the difference between 180° and the phase of $D(z)G(z)$ when its amplitude is 1. It is a measure of how much additional phase lag or time delay can be tolerated before instability results because the phase of a system is highly related to these characteristics.

Example 5.8: To illustrate gain and phase margins, let's consider the open-loop system

$$G(s) = \frac{1}{s(s+1)^2}$$

with proportional discrete feedback ($D(z) = K$) at a more typical sample rate of 5 Hz. Use of (5.5) results in

$$G(z) = 0.0012 \frac{(z+3.38)(z+0.242)}{(z-1)(z-0.8187)^2}$$

The portion of the Nyquist plot representing the frequency response in the vicinity of -1 is plotted in Fig. 5.20. Unlike the previous example with a very slow sample rate, the higher sample rate causes the magnitude to be essentially zero at $\omega T = \pi$, and hence the Nyquist plot goes to the origin. The plot is very similar to what would result for a continuous controller.⁶ Furthermore, just as in the continuous case, there are no -1 point encirclements if $K = 1$ as plotted. This would also be the case for any value of $K \leq 1.8$, that is, low enough so that the magnitude of the Nyquist plot is less than 1 as it crosses the negative real axis. The system is then stable. For values of $K \geq 1.8$ the -1 point lies within the loop on the negative real axis, thus creating two encirclements ($N = 2$) and two unstable roots ($Z = 2$).

⁶See Example II in Section 5.4 of Franklin, Powell, and Emami-Naeini (1986).

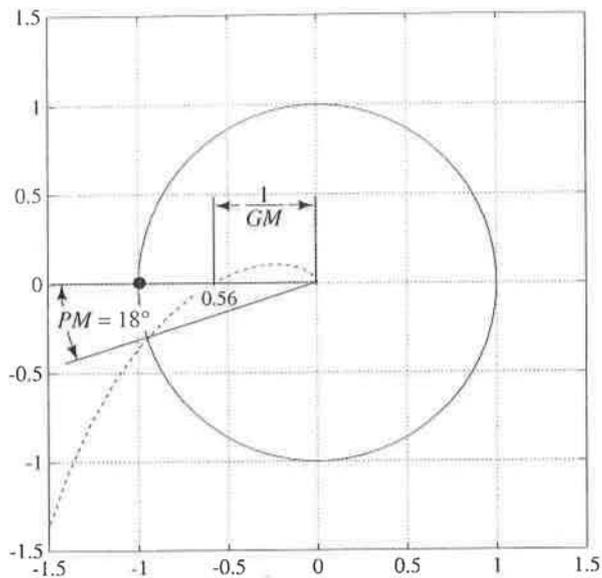


Figure 5.20 Gain and phase margins for Example 5.8.

The GM and PM are indicated in the figure and show that the system is stable because the $GM = 1.8$ and $PM = 18^\circ$.

For continuous systems, it is often pointed out that the phase margin is related to the damping ratio, ζ , for a second-order system by the approximate relation, $\zeta \cong PM/100$. This relationship is examined in Fig. 5.21 for the continuous case and for discrete systems with two values of the sample rate.

Figure 5.21 was generated by evaluating the damping ratio of the closed-loop system that resulted when discrete proportional feedback was used with the open-loop system

$$G(s) = \frac{1}{s(s+1)}.$$

A z -transform analysis of this system resulted in z -plane roots that were then transformed back to the s -plane via the inverse of $z = e^{sT}$. The ζ of the resulting s -plane roots are plotted in the figure. As the feedback gain was varied, the damping ratio and phase margin were related as shown in Fig. 5.21. The actual sample rates used in the figure are 1 Hz and 5 Hz, which represent 6 and 30 times the open-loop system root at 1 rad/sec.

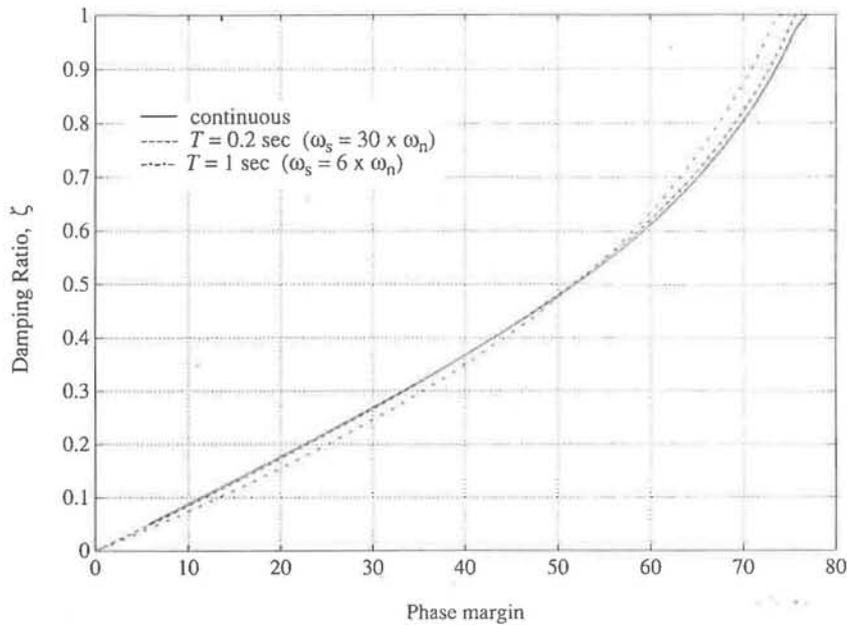


Figure 5.21 Damping ratio of a second-order system versus phase margin (PM).

The conclusion to be drawn from Fig. 5.21 is that the PM from a discrete z -plane frequency response analysis carries the same implications about the damping ratio of the closed-loop system as it does for continuous systems. For second-order systems without zeros, the relationship between ζ and PM in the figure shows that the approximation of $\zeta \cong PM/100$ is equally valid for continuous and discrete systems. For higher order systems, the damping of the individual modes needs to be determined using other methods.

5.5.3 Low-Frequency Gains and Error Coefficients

For continuous systems, we define the error constants as

$$K_p = \lim_{s \rightarrow 0} D(s)G(s)$$

and

$$K_v = \lim_{s \rightarrow 0} s D(s)G(s),$$

which relate directly to the system errors for step inputs and ramp inputs, respectively. From a magnitude frequency-response plot, one can find the

value of K_p for Type 0 systems by simply determining the gain of the low-frequency asymptote that has zero slope. For Type I systems, the gain versus frequency approaches a slope of -1 on the low-frequency asymptote. K_v is most easily determined by evaluating the gain on the low-frequency asymptote at $\omega = 1$. If there are singularities around $\omega = 1$ or at lower frequencies, the magnitude plot won't be on the asymptote at $\omega = 1$, and it is necessary to extrapolate the asymptote to $\omega = 1$ and use that value as K_v .

The error constants for discrete systems were established in Section 5.2 and are

$$K_p = \lim_{z \rightarrow 1} D(z)G(z)$$

and

$$K_v = \lim_{z \rightarrow 1} \frac{(z-1)D(z)G(z)}{Tz}. \quad [5.8]$$

For a Type 0 system, the procedure is identical to the continuous case. Since $z = e^{j\omega T}$, $z \rightarrow 1$ implies that $\omega T \rightarrow 0$, and the magnitude frequency-response plot will show a constant value on the low-frequency asymptote which is equal to K_p .

For a Type I system, the procedure is again identical to the continuous case in that the magnitude of $D(z)G(z)$ at $\omega = 1$ on the low-frequency asymptote is equal to K_v . This can be seen from (5.8) if we note that for $\omega T \rightarrow 0$, $e^{j\omega T} \cong 1 + j\omega T$. Therefore

$$\lim_{z \rightarrow 1} \frac{(z-1)}{Tz} = \lim_{j\omega \rightarrow 0} \omega,$$

thus establishing the fact that evaluation of the low-frequency asymptote of $D(z)G(z)$ at $\omega = 1$ yields K_v . This fact is most easily used if the frequency-response magnitude is plotted versus ω in units of rad/sec so that $\omega = 1$ rad/sec is readily found. If the magnitude is plotted versus ω in units of Hz or versus ωT , one would need to perform a calculation to find the $\omega = 1$ rad/sec point. However, the error constants could be calculated directly with good software tools; therefore the issues in their calculation are of passing interest only. But no matter how the constants are found, the fact remains for discrete and continuous frequency response alike, the higher the magnitude curve at low frequency, the lower the steady-state errors.

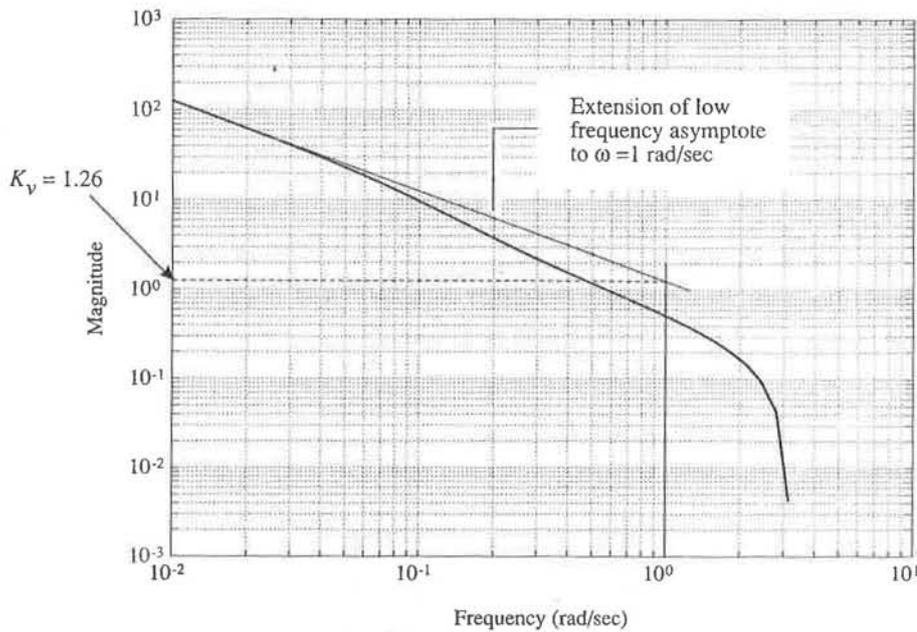


Figure 5.22 Determination of K_v from frequency response.

Example 5.9: Let us apply this to the determination of K_v for the antenna system with the compensation given by (5.24). The open-loop discrete transfer function is

$$G(z)D(z) = (0.0484) \frac{z + 0.9672}{(z - 1)(z - 0.9048)} (6) \frac{z - 0.80}{z - 0.05},$$

which yields the magnitude versus frequency in Fig. 5.22. Also note in the figure that the extension of the low-frequency asymptote at $\omega = 1$ has a magnitude of 1.26, thus indicating that $K_v = 1.26$.

5.5.4 Compensator Design

The amplitude and phase curves can be used to determine the stability margins based on the Nyquist stability criterion for either continuous or discrete systems. In the continuous case with minimum-phase transfer functions, Bode showed that the phase is uniquely determined by an integral of the slope of the magnitude curve on a log-log plot. If the function is rational, these slopes are readily and adequately approximated by constants! Thus we have the result that the amplitude curve must cross unity gain (zero log) at

a slope (in approximation) of -1 for a reasonable phase margin. The ability to predict stability from the amplitude curve alone is an important contributor to the ease at which designers can evaluate changes in compensator parameters.

For discrete systems, Bode's relationship between the amplitude and phase curve is lost because z takes on values around the unit circle instead of s traversing the imaginary axis as in continuous systems. Fig. 5.16 illustrates the degree to which the relationship is lost and indicates that the error would be small for frequencies slower than $1/20$ th of the sample rate. However, it is typically necessary to determine both magnitude and phase for discrete z -plane systems for an accurate assessment of the stability.

In carrying out designs, the z -plane poles and zeros on the real axis are located by their fractional location between zero and ± 1 . The equivalent idea in the z -plane for the "breakpoint" in Bode's hand-plotting rules is that the magnitude will change slope at a frequency when ωT , the *angular* position on the unit circle in radians, has the same value as the fractional distance of the singularity on the real axis to $z = +1$. For example, a pole at $z = 0.9$ will produce a slope change at $\omega T = 0.1$ rad. This equivalence is very accurate for low angular values ($\omega T \leq 0.1$ rad, i.e., sampling at more than 60 times the frequency) and is a reasonable approximation for angular values less than 0.8 rad (i.e., sampling at more than 8 times the frequency). In order for a designer to arrive at trial compensations with potential for better PM , GM , steady-state errors, or crossover frequency, it is useful to understand how a pole or zero placement will affect the magnitude and phase curves. Because of the equivalence of the break-point concept between the continuous and discrete cases, this is accomplished for discrete systems using the ideas from the continuous Bode hand-plotting techniques, keeping in mind that their fidelity degrades for slow sampling. It is easiest to select compensator break points if the frequency-response magnitude and phase is plotted versus ωT so that the correspondence between those curves and the location of the compensation parameters is retained.

Example 5.10: Let us design the discrete controller for the antenna system one more time. Using the slow sample rate of 1 Hz to illustrate the discrete aspects more clearly, we have the system transfer function

$$G(z) = 0.0484 \frac{z + 0.9672}{(z - 1)(z - 0.9048)} \quad [5.21]$$

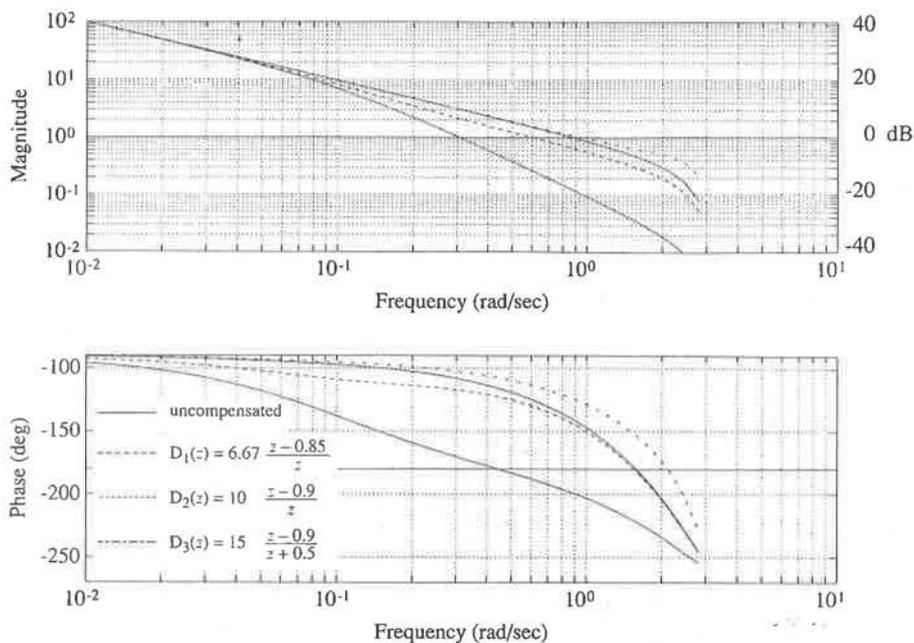


Figure 5.23 Frequency response with D_1 , D_2 , and D_3 for Example 5.10.

The magnitude and phase of the uncompensated system $[G(z)]$ shown in Fig. 5.23 indicate that the system has a PM of 8° and crossover frequency (ω_{co}) of 0.3 rad/sec. The specifications for the design are that the overshoot should be less than 16%, the settling time should be less than 10 sec, and $K_v \geq 1$. The 16% overshoot translates into $\zeta \geq 0.5$ from Fig. 5.4, which translates into the requirement that the PM be $\geq 50^\circ$ from Fig. 5.21. The specification for settling time translates via (5.13) into the requirement that $\omega_n \geq 0.92$.

Because K_v of $G(z) = 1$, the compensated system will also have $K_v = 1$ provided the dc gain of $D(z) = 1$. In terms of the frequency response, this means that the extension of the low-frequency-magnitude asymptote passes through the value 1 at $\omega = 1$ for the uncompensated case (in Fig. 5.23), and the gain of this low-frequency asymptote should not decrease with any candidate compensation. To maintain an acceptable K_v , we will evaluate only $D(z)$'s with a dc gain of 1.

The uncompensated system's PM of 8° indicates poor damping, and the ω_{co} of 0.3 rad/sec indicates that it will be too slow. Just as for continuous systems, ω_{co} occurs approximately at the system bandwidth and dominant natural frequency; therefore, we should try

to change the design so that it has a ω_{co} of about 0.9 rad/sec in order to meet the $t_s \leq 10$ sec. Once we find a compensation that meets the guidelines of $PM = 50^\circ$ and $\omega_{co} = 0.9$ rad/sec, we will need to check whether the t_s and overshoot specifications are actually met, because the translations made are based on second-order systems with no zeros.

Fig. 5.23 shows several attempts. The breakpoint of the first attempt [$D_1(z)$ in Fig. 5.23] was at 0.15 rad/sec⁷ and did not increase the slope of the magnitude curve at a low enough frequency to bring about the desired ω_{co} . This was remedied in $D_2(z)$, where the breakpoint was lowered to 0.1 rad/sec (zero at $z = 0.9$) causing a ω_{co} of 0.9 rad/sec, but the resulting PM of 40° was still lower than desired. By moving the pole out to $z = -0.5$ in $D_3(z)$, we had very little effect on the ω_{co} but achieved an increase in the PM to 50° . Because both goals are met, $D_3(z)$ has a reasonable chance to meet the specifications; in fact, the calculation of a time history of the system response to a step input shows that the t_s is 7 sec, but, alas, the overshoot is 27%. The guidelines were not successful in meeting the specifications because the system is third order with a zero, whereas the rules were derived assuming a second-order system without a zero.

The necessary revisions to our design guidelines are clear; we want more than a 50° PM and do not require a 0.9 rad/sec ω_{co} . Fig. 5.24 shows the system frequency response using $D_3(z)$ along with two revisions of $D(z)$ that satisfy our revised goals. $D_4(z)$ has a 60° PM and a 0.6 rad/sec ω_{co} , and $D_5(z)$ has a 58° PM and a 0.8 rad/sec ω_{co} . The time history of the system response to a step using $D_5(z)$ in Fig. 5.25 shows that it exactly meets the requirements for 16% overshoot and $t_s = 10$ sec. Furthermore, the design of the system was so that $K_v = 1$; therefore, all the requirements are met and the design is complete.

To implement this compensation,

$$D_5(z) = 12.8 \frac{z - 0.883}{z + 0.5}, \quad (5.32)$$

⁷The zero at $z = 0.85$ translates into a 0.15 rad/sec breakpoint only because the sample period, T , is 1 sec. For $T = 0.1$ sec, a zero at $z = 0.85$ would translate into a 1.5 rad/sec breakpoint, etc.

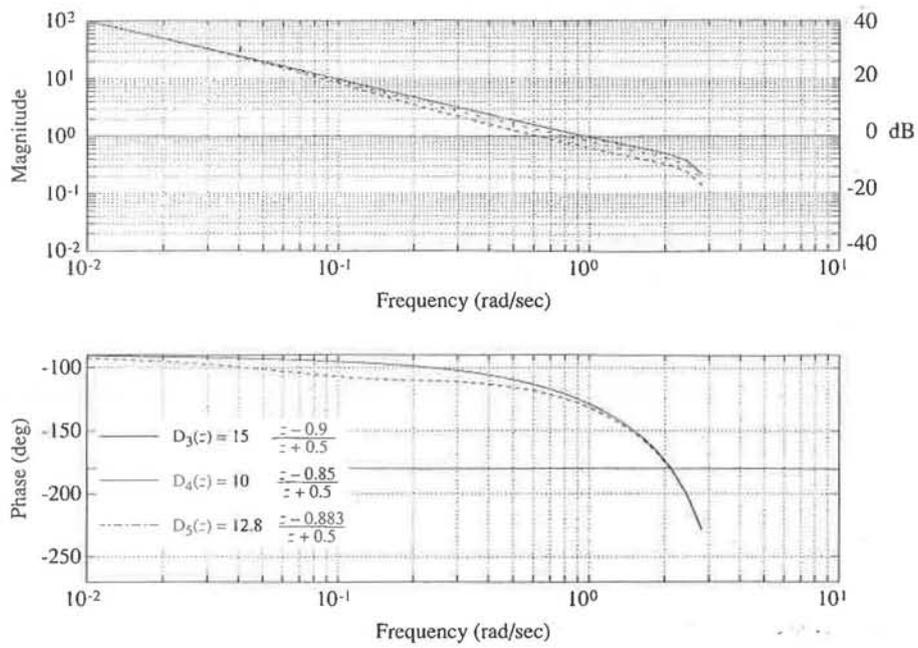


Figure 5.24 Frequency response with D_3 , D_4 , and D_5 for Example 5.10.

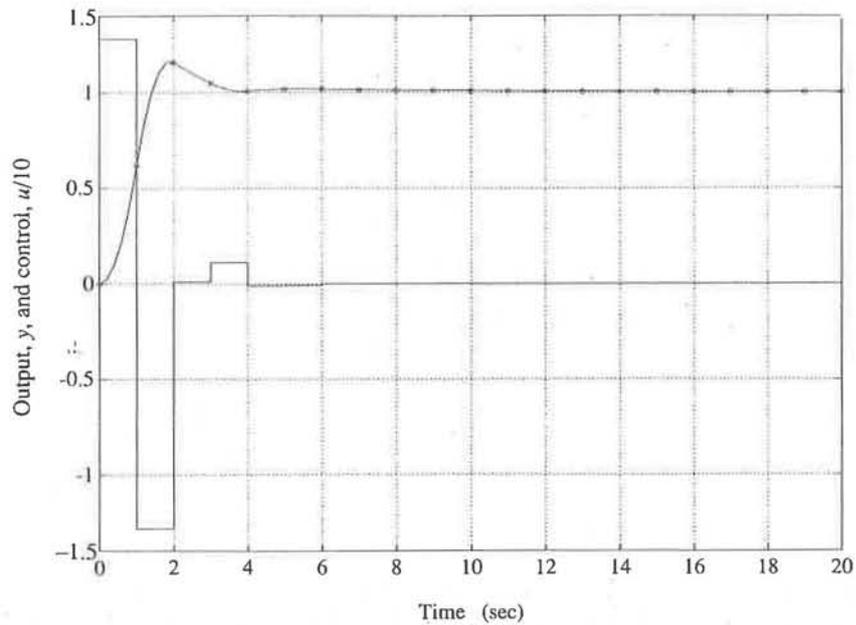


Figure 5.25 Step response of Example 5.10 with $D_5(z)$.

we can use the same technique when (5.17) was inverted, carrying out the process by inspection to arrive at

$$u(k) = -0.5 u(k-1) + 12.8 (e(k) - 0.883 e(k-1)), \quad (5.33)$$

which can be directly coded in the control computer.

5.6 FREQUENCY RESPONSE METHODS WITH THE w -TRANSFORM

The discrete system w -transform was developed in order to retain many of the design features from continuous systems. The essential idea of the method is to transform the discrete model of the plant, $G(z)$, by substituting a new variable, w , with the bilinear mapping

$$w \triangleq \frac{2}{T} \frac{z-1}{z+1}, \quad (5.34)$$

and to perform the compensator design in this " w -plane." The resulting compensation is then converted back to the z -plane, and hence the required difference equation is easily obtained. Much of the flavor of continuous system design is retained because the stability boundary in the w -plane is the imaginary axis, just as in the s -plane. It is interesting to note that the transformation (5.34) is identical to that used for Tustin's filter approximation in Chapter 3,

$$s = \frac{2}{T} \frac{z-1}{z+1}, \quad [3.8]$$

except that in that case, it was used to convert rational functions of s into approximate but realizable functions of z .

The transformation of the stability boundary between the z -plane and the w -plane can be seen by writing (5.34) as

$$w = \frac{2}{T} \frac{e^{sT} - 1}{e^{sT} + 1} = \frac{2}{T} \tanh \frac{sT}{2}.$$

If s is pure imaginary ($=j\omega$) and therefore $z = e^{j\omega T}$ (on the unit circle), then

$$w \triangleq j\nu = j \frac{2}{T} \tan \frac{\omega T}{2} \quad (5.35)$$

and we see that while z goes around the unit circle, the w -plane frequency, ν , stays *real* and goes from 0 to ∞ . We chose the scale factor⁸ of $2/T$ in (5.34) to make sure that the error constant would come out correctly and so that w -plane transfer functions would approach those in the s -plane as T went to zero.

The design process begins with the conversion of the continuous model, $G(s)$, to a z -plane discrete model, $G(z)$, using (5.5). We then transform to the w -plane using the inverse of (5.34), or

$$z = \frac{1 + wT/2}{1 - wT/2} \quad (5.36)$$

to obtain $G(w)$.⁹ The design is then carried out with the $G(w)$ as if the system were continuous. The magnitude and phase of $G(j\nu)$ are the magnitude and phase of $G(z)$ as z takes on values around the unit circle; and because $G(j\nu)$ is a rational function of ν , we can apply all the standard straight-line approximations to the log magnitude and phase curves. Nyquist's stability criterion applies to $G(j\nu)$ in the w -plane just as it does to $G(j\omega)$ in the s -plane because in both cases we are determining the number of zeros of $(1 + G)$ in the right-hand plane (or unstable roots). Therefore, the gain and phase margins of classical Bode designs apply directly to $G(j\nu)$. The result of the design will be a $D(w)$ that achieves the desired PM , and so on; this $D(w)$ is then converted to a $D(z)$ using 5.34 to complete the process.

Example 5.11: Consider again the antenna design used in the previous sections. Its discrete model, with $T = 1$ sec, is, from (5.21),

$$G(z) = 0.0484 \frac{z + 0.9672}{(z - 1)(z - 0.9048)} \quad [5.21]$$

⁸This transformation is sometimes done without the $2/T$ scale factor, in which case the transformed frequency ($\triangleq \nu'$) differs from the real frequency, ω , by $\nu' = \tan \omega T/2$ and is essentially a frequency that is scaled to the sample rate; that is, $\nu' = 1$ when $\omega = \omega_s/4$. See Whitbeck and Hofmann (1978).

⁹We use the same symbol, G , for three distinct functions, $G(s)$, $G(z)$, and $G(w)$. The arguments s , z , and w identify the function as well as the variable.

This is transformed to the w -plane using (5.36) to yield

$$G(w) = -\frac{(w/120 + 1)(w/2 - 1)}{w(w/0.0999 + 1)}. \quad (5.37)$$

The algebra in computing $G(w)$ from $G(z)$ can be somewhat tedious; Appendix B contains a table of a few common transfer functions and a general formula for making the conversion. Design software packages also typically contain routines to perform this conversion. (See X-Z2W in Table E.1.)

Note that the gain of $G(w)$ is precisely the same as that of $G(s)$; it is unity in both cases. This will always be true for a $G(w)$ computed using the definition of w given in (5.34). The gain of 1 in (5.37) is the K_v of the uncompensated discrete system, as the reader can verify using (5.8); and it also applies to the continuous system, as can be verified from (5.2). We also note that in (5.37) the denominator looks very similar to that of $G(s)$ and that the denominators will be the same as T approaches zero. This would also have been true for any zeros of $G(w)$ that corresponded to zeros of $G(s)$, but our example did not have any. The example also shows the creation of a right-hand-plane zero of $G(w)$ at $2/T$ and the creation of a fast left-hand-plane zero when compared to the original $G(s)$. The two "created" zeros can be attributed to the sampling and hold operations and thus depend on the sample rate. They both become faster and thus less important to the design problem as the sample rate is increased. In general, one or both of these additional zeros usually occur.

In our example, the most important feature added to $G(s)$ in transforming to $G(w)$ is the right-hand-plane zero at $w = 2$ rad/sec ($= 2/T$), which will introduce serious distortion as ν approaches 2 rad/sec. From (5.35) we see that this point corresponds to $\omega T/2 = \tan^{-1}(1) = 45^\circ$ or $\omega T = 90^\circ$. Because the limit of frequency occurs at $\omega = \omega_s/2$ ($\omega T = 180^\circ$), frequencies of interest can approach $\omega = \omega_s/4$ and will be affected by this zero. More significantly, the zero at 2 is in the right half of the plane so that Bode's gain-phase integral does not apply to this one term. We will need to be especially cautious as we get close to $\nu = 2$.

We want a compensation that will keep $K_v = 1$ and that will have a damping of about $\zeta = 0.5$. Using the rule of thumb that the phase margin (PM) $\cong 100\zeta$ (see Fig. 5.21) suggests that the PM should be 50° . A plot of the magnitude of (5.37) is shown in Fig. 5.26

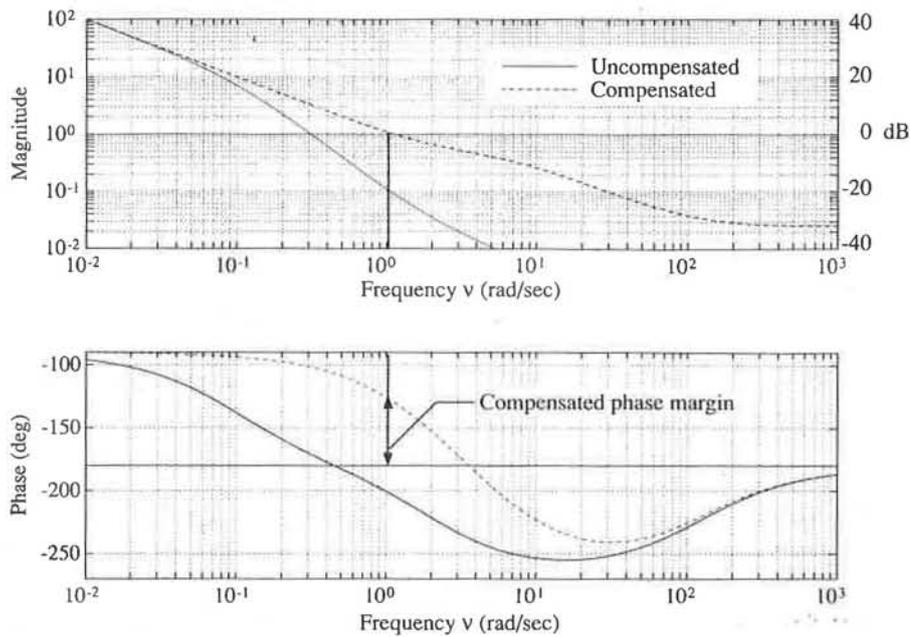


Figure 5.26 Magnitude and phase of $G(w)$ and $D(w)G(w)$.

by the solid line and is set to give $K_v = 1$. The phase margin for the system without compensation is on the order of 10° , just as it was for the z -plane frequency-response design in Section 5.5. Note that the zero at $\nu = 2$ contributes phase lag because it is in the right-hand plane. Let us try compensating by placing a zero on top of the pole at $\nu = 0.0999$. How about the compensation pole. . . or do we need a pole? In designing continuous systems, we always include a pole to avoid noise amplification and to make them easier to build. A w -plane compensation with no pole yields a z -plane pole at $z = -1$, which leads to a marginally unstable compensation and an unstable closed-loop system in spite of a positive phase margin;¹⁰ therefore, to stay out of trouble, always use a zero and a pole together.

Where should we place the pole? The crossover frequency will be at approximately $\nu = 1$, where the phase of the uncompensated system was -21° and the compensation zero added 84° ; therefore the pole can cause a phase lag of up to 13° in meeting the desired 50° phase margin. Placing the pole at $\nu = 6$ meets this requirement,

¹⁰The resulting system has a higher-order numerator than denominator, and one must revert to Nyquist's -1 encirclement criterion to determine stability.

so our compensation is

$$(w) = \frac{1 + w/0.0999}{1 + w/6}, \quad (5.38)$$

and the compensated loop gain and phase are the dashed curves in Fig. 5.26. Note that the phase in the vicinity of $\nu = 10$ corresponds to a magnitude slope of -3 rather than -1 ; thus compensating non-minimum phase systems (as most w -plane designs are) based on magnitude alone is treacherous.

Reversing the w -transformation, we now use (5.34) to convert the $D(w)$ of (5.38) to

$$D(z) = \frac{15.8(z - 0.9048)}{z + 0.5}. \quad (5.39)$$

A root locus of the system

$$\begin{aligned} DG &= (15.8) \frac{(1 - 0.9048z^{-1})(0.0484)(z^{-1})(1 + 0.9672z^{-1})}{(1 + 0.5z^{-1})(1 - z^{-1})(1 - 0.9048z^{-1})} \\ &= 0.758 \frac{z + 0.9672}{(z - 1)(z + 0.5)} \end{aligned} \quad (5.40)$$

is the circle centered at -0.97 in Fig. 5.27(a). The closed-loop pole corresponding to the root locus gain of 0.758 is marked Δ . The resulting roots do not have the desired damping of $\zeta = 0.5$. In fact, the damping is $\zeta = 0.37$. The step response is shown in Fig. 5.27(b). This breakdown in the phase-margin/damping rule of thumb can sometimes occur in discrete frequency-response design when the sampling is extremely slow or when there is an added pole or zero. In this particular design example, the sample rate is 3.4 times faster than the closed-loop root on the positive real z -plane axis and is slower than one would typically select. If one does sample very slowly, however, the rule of thumb that $PM \cong 100\zeta$ is no longer valid.

To further illustrate this point, Fig. 5.28 compares the damping of the example (5.40) with varying loop gains and thus phase margins. The numbers shown adjacent to the discrete curve indicate the ratio of the sample rate to the closed-loop root frequency. Although the curve is specialized for this example, it illustrates the danger in relying on the rule of thumb and indicates a necessity to actually check whether the specifications are met.

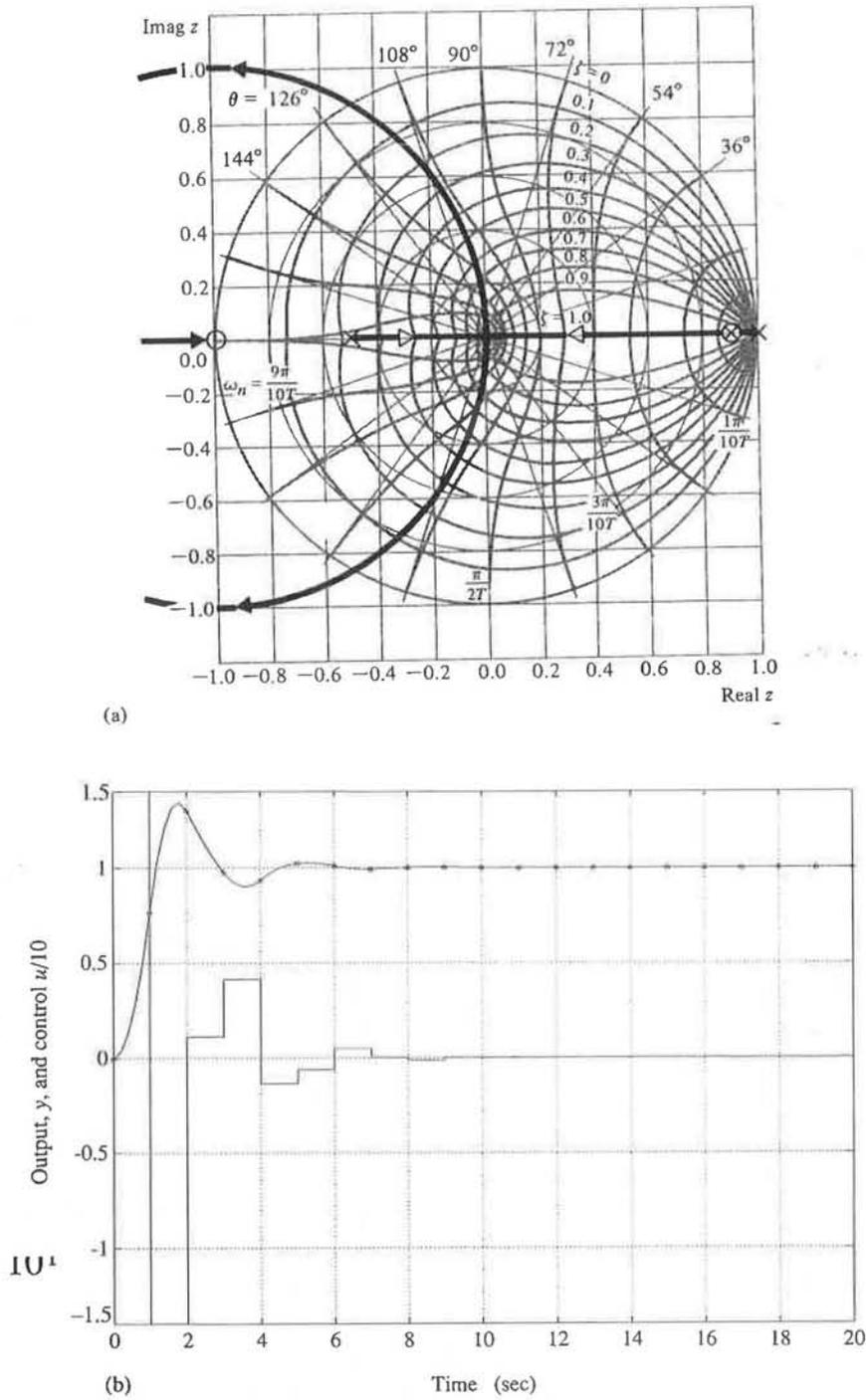


Figure 5.27 (a) Root locus of system designed in the w -plane. (b) Step response of closed-loop antenna control designed in the w -plane.

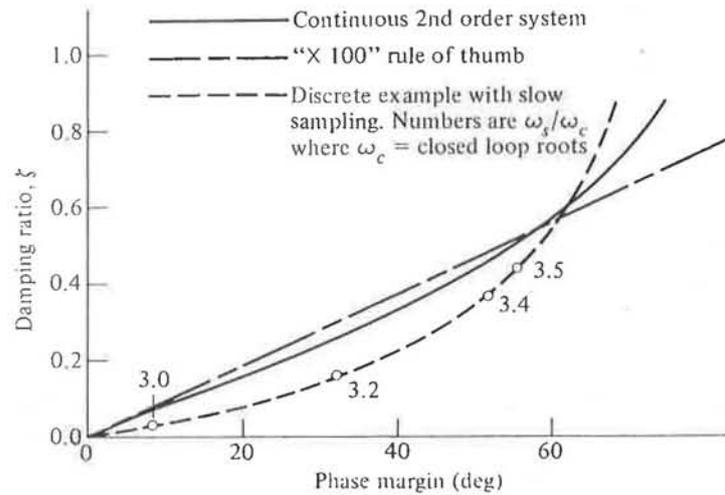


Figure 5.28 System damping versus phase margin.

5.7 DIRECT DESIGN METHOD OF RAGAZZINI

[Ragazzini and Franklin (1958)]

Much of the style of the transform design techniques we have been discussing in this chapter grew out of the limitations of technology that was available for realization of the compensators with pneumatic components or electric networks and amplifiers. In particular, many constraints were imposed in order to assure the realization of electric compensator networks $D(s)$ as networks consisting only of resistors and capacitors.¹¹ In the digital computer, such limitations on realization are, of course, not relevant, and one can ignore these particular constraints. One design method that eliminates these constraints begins from the very direct point of view that we are given a plant (plus hold) discrete transfer function $G(z)$, that we want to construct a desired transfer function, $H(z)$, between R and Y , and that we have the computer transfer function, $D(z)$, to do the job. The overall transfer function is given by the formula

$$H(z) = \frac{DG}{1 + DG},$$

¹¹In the book by Truxal (1955), where much of this theory is collected at about the height of its first stage of development, a chapter is devoted to RC network synthesis.

from which we get the design formula

$$D(z) = \frac{1}{G(z)} \frac{H(z)}{1 - H(z)}. \quad (5.41)$$

From (5.41) we can see that this design calls for a $D(z)$ that will cancel the plant effects and that will add whatever is necessary to give the desired result. The problem is to discover and implement constraints on $H(z)$ so that we do not ask for the impossible.

First, let us consider the constraint of causality. From z -transform theory we know that if $D(z)$ is causal, then as $z \rightarrow \infty$, its transfer function is well behaved; it does not have a pole at infinity. Looking at (5.41), we see that if $G(z)$ were to have a zero at infinity, then $D(z)$ would have a pole there unless we request an $H(z)$ that is such as to cancel it. Thus we have the constraint that for $D(z)$ to be causal

$$H(z) \text{ must have a zero at infinity of the same order} \\ \text{as the zero of } G(z) \text{ at infinity.} \quad (5.42)$$

This requirement has an elementary interpretation in the time domain: $G(z)$ has a zero at infinity because the pulse response of the plant has a delay of at least one sample time. If there is a transportation lag in the plant, then the delay can be several samples and $G(z)$ can start with $z^{-\ell}$. The causality requirement on $H(z)$ is that the closed-loop system must have at least as long a delay as the plant has.

Considerations of stability add a second constraint. The roots of the characteristic equation of the closed-loop system are the roots of the equation

$$1 + D(z)G(z) = 0. \quad (5.43)$$

We can express (5.43) as a polynomial if we identify $D = c(z)/d(z)$ and $G = b(z)/a(z)$ where a , b , c , and d are polynomials. Then the characteristic polynomial is

$$ad + bc = 0. \quad (5.44)$$

Now suppose there is a common factor in DG , as would result if $D(z)$ were called upon to cancel a pole or zero of $G(z)$. Let this factor be $z - \alpha$ and suppose it is a pole of $G(z)$, so we can write $a(z) = (z - \alpha)\bar{a}(z)$, and to

cancel it we have $c(z) = (z - \alpha)\bar{c}(z)$. Then (5.44) becomes

$$\begin{aligned}(z - \alpha)\bar{a}(z)d(z) + b(z)(z - \alpha)\bar{c}(z) &= 0, \\ (z - \alpha)[\bar{a}d + b\bar{c}] &= 0.\end{aligned}\tag{5.45}$$

In other words—perhaps it was obvious from the start—a common factor *remains a factor of the characteristic polynomial*. If this factor is outside the unit circle, the system is unstable! How do we avoid such cancellation? Considering again (5.41), we see that if $D(z)$ is not to cancel a pole of $G(z)$, then that factor of $a(z)$ must also be a factor of $1 - H(z)$. Likewise, if $D(z)$ is not to cancel a zero of $G(z)$, such zeros must be factors of $H(z)$. Thus we write the constraints:¹²

$1 - H(z)$ must contain as zeros all the poles of $G(z)$ that are outside the unit circle. $H(z)$ must contain as zeros all the zeros of $G(z)$ that are outside the unit circle.	(5.46)
--	--------

Consider finally the constraint of steady-state accuracy. Because $H(z)$ is the overall transfer function, the error transform is given by

$$E(z) = R(z)(1 - H(z)).\tag{5.47}$$

Thus if the system is to be Type I with velocity constant K_v , we must have zero steady-state error to a step and $1/K_v$ error to a unit ramp. The first requirement is

$$e(\infty) = \lim_{z \rightarrow 1} (z - 1) \frac{1}{z - 1} [1 - H(z)] = 0,\tag{5.48}$$

which implies

$$\boxed{H(1) = 1}.\tag{5.49}$$

The velocity constant requirement is that

$$e(\infty) = \lim_{z \rightarrow 1} (z - 1) \frac{Tz}{(z - 1)^2} [1 - H(z)] = \frac{1}{K_v}.\tag{5.50}$$

¹²Roots on the unit circle are also unstable by some definitions, and good practice indicates that we should not cancel singularities outside the radius of desired settling time. See Fig. 5.5 and the discussion associated with it.

From (5.47) we know that $1 - H(z)$ is zero at $z = 1$, so that to evaluate the limit in (5.50), it is necessary to use L'Hôpital's rule with the result [see (5.11) and following]

$$\boxed{-T \frac{dH}{dz} \Big|_{z=1} = \frac{1}{K_v}} \quad (5.51)$$

An example will best illustrate the application of these constraints. Consider again the plant described by the transfer function (5.21) and suppose we ask for the same design that led to (5.15) as a continuous controller. The continuous closed-loop system has a characteristic equation

$$s^2 + s + 1 = 0.$$

With a sampling period $T = 1$ sec, this maps to the discrete characteristic equation

$$z^2 - 0.7859z + 0.36788 = 0. \quad (5.52)$$

Let us therefore ask for a design that is stable, has $K_v = 1$, and has poles at the roots of (5.52) plus, if necessary, additional poles at $z = 0$, where the transient is as short as possible. The form of $H(z)$ is thus

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + \dots}{1 - 0.7859z^{-1} + 0.3679z^{-2}}. \quad (5.53)$$

The causality design constraint, using (5.42) and (5.18), requires that

$$H(z) \Big|_{z=\infty} = 0$$

or

$$b_0 = 0. \quad (5.54)$$

Equations (5.46) add no constraints because $G(z)$ has all poles and zeros inside the unit circle except for the single zero at ∞ , which is taken care of by (5.54). The steady-state error requirement leads to

$$\begin{aligned} H(1) &= 1 \\ &= \frac{b_1 + b_2 + b_3 + \dots}{1 - 0.7859 + 0.3679} = 1. \end{aligned} \quad (5.55)$$

Therefore

$$b_1 + b_2 + b_3 + \cdots = 0.5820$$

and

$$-T \frac{dH}{dz} \Big|_{z=1} = \frac{1}{K_v}.$$

Because in this case both T and K_v are 1, we use (5.55) and the derivative with respect to z^{-1} to obtain

$$\begin{aligned} 1 &= \frac{1}{K_v} = \frac{dH}{dz^{-1}} \Big|_{z=1} \\ &= \frac{(0.5820)[b_1 + 2b_2 + 3b_3 + \cdots] - [0.5820][-0.7859 + 0.3679(2)]}{(0.5820)(0.5820)} \end{aligned}$$

or

$$\frac{b_1 + 2b_2 + 3b_3 + \cdots - [-0.05014]}{0.5820} = 1. \quad (5.56)$$

Because we have only two equations to satisfy, we need only two unknowns and we can truncate $H(z)$ at b_2 . The resulting equations are

$$b_1 + b_2 = 0.5820, \quad b_1 + 2b_2 = 0.5318,$$

which have the solution

$$b_1 = 0.6321, \quad b_2 = -0.05014. \quad (5.57)$$

Thus the final design gives an overall transfer function

$$H(z) = \frac{0.6321z - 0.05014}{z^2 - 0.7859z + 0.3679}. \quad (5.58)$$

We shall also need

$$1 - H(z) = \frac{(z-1)(z-0.4180)}{z^2 - 0.7859z + 0.3679}. \quad (5.59)$$

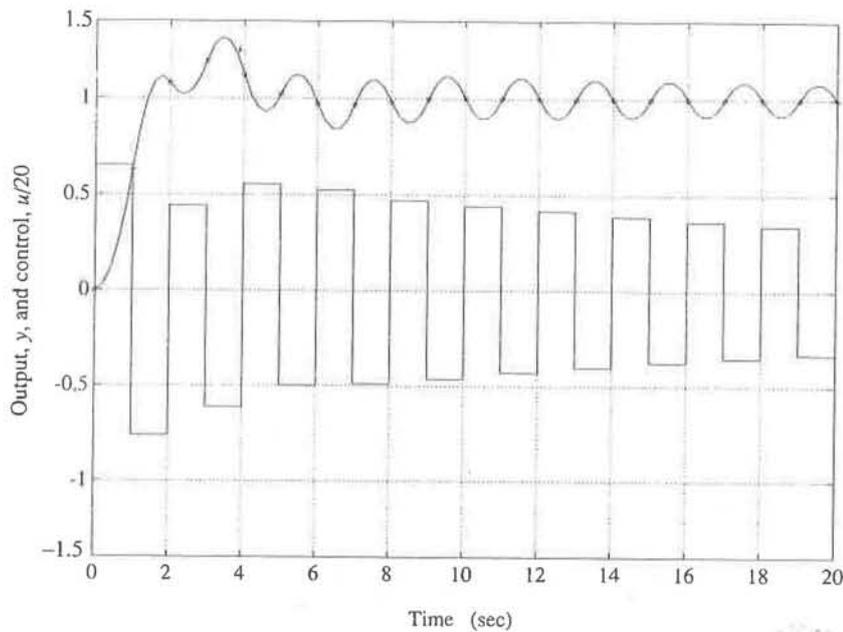


Figure 5.29 Step response of antenna system from direct design.

We know that $H(1) = 1$ so that $1 - H(z)$ must have a zero at $z = 1$. Now, turning to the basic design formula, (5.41), we compute

$$\begin{aligned} D(z) &= \frac{(z-1)(z-0.9048)(0.6321)}{(0.04837)(z+0.9672)} \frac{(z-0.07932)}{(z-1)(z-0.4180)} \\ &= 13.07 \frac{(z-0.9048)}{(z+0.9672)} \frac{(z-0.07932)}{(z-0.4180)}. \end{aligned}$$

A plot of the step response of the resulting design is provided in Fig. 5.29 and verifies that the response samples behave as specified by $H(z)$. However, as can be seen also from the figure, large oscillations occur in the control that causes the system response to oscillate considerably between samples. You may question how this can be for a system response transfer function,

$$\frac{Y(z)}{R(z)} = H(z) = \frac{DG}{1 + DG},$$

that is shown by (5.58) to have only two well damped roots. The answer lies

in the fact that the control response is determined from

$$\frac{U(z)}{R(z)} = \frac{D}{1 + DG} = \frac{H(z)}{G(z)},$$

which for this example is

$$\frac{U(z)}{R(z)} = 13.06 \frac{z - 0.0793}{z^2 - 0.7859z + 0.3679} \frac{(z - 1)(z - 0.9048)}{z + 0.9672}.$$

There is a root at $z = -0.9672!$. This is the source of the oscillation in the control response, but it did not show up in the output response because it was exactly canceled by a zero. The control oscillation causes the “intersample ripple” in the output response, and the designer should be alert to this if poorly behaved roots arise in the control response. An actual prediction of the output intersample ripple based on linear analysis was not possible with the z -transform method described so far; rather, one would need to apply the “modified z -transform,” which is beyond the scope of this text. Alternatively, one can use a CAD simulation to find such oscillations quite easily, as was done here. To avoid this oscillation, we could introduce another term in $H(z)$, b_3z^{-3} , and require that $H(z)$ be zero at $z = -0.9672$, so this zero of $G(z)$ is not canceled by $D(z)$. The result will be a simpler $D(z)$ with a slightly more complicated $H(z)$. However, rather than pursue this method further, we will wait until the more powerful method of pole assignment by state-variable analysis is developed in the next chapter, where computer algorithms are more readily provided.

5.8 PID CONTROL

Just as in continuous systems, there are three basic types of control: Proportional, Integral, and Derivative, hence the name, PID. In the design examples so far, we have been using the discrete equivalent of lead compensation, which is essentially a combination of proportional and derivative control. Let us now review these three controls as they pertain to a discrete implementation. The term PID is widely used because there are commercially available modules that have knobs for the user to turn that set the values of each of the three control types.

5.8.1 Proportional Control

A discrete implementation of proportional control is identical to continuous; that is, where the continuous is

$$u(t) = K_p e(t) \Rightarrow D(s) = K_p,$$

the discrete is

$$u(k) = K_p e(k) \Rightarrow \boxed{D(z) = K_p}$$

where $e(t)$ is the error signal as shown in Fig 5.2.

5.8.2 Derivative Control

For continuous systems, derivative or rate control has the form

$$u(t) = K_p T_D \dot{e}(t) \Rightarrow D(s) = K_p T_D s$$

where T_D is called the *derivative time*. Differentiation can be approximated in the discrete domain as the first difference, that is,

$$u(k) = K_p T_D \frac{(e(k) - e(k-1))}{T} \Rightarrow \boxed{D(z) = K_p T_D \frac{1 - z^{-1}}{T} = K_p T_D \frac{z - 1}{Tz}}$$

In many designs, the compensation is a sum of proportional and derivative control (or PD control). In this case, we have

$$D(z) = K_p \left(1 + \frac{T_D(z-1)}{Tz} \right).$$

or, equivalently,

$$\boxed{D(z) = K \frac{z - \alpha}{z}}$$

which is similar to the lead compensations that have been used in the designs in the previous sections. The difference is that the pole is at $z = 0$, whereas the pole has been placed at various locations along the z -plane real axis for the previous designs. In the continuous case, pure derivative control represents the ideal situation in that there is no destabilizing phase lag from the differentiation, or, equivalently, the pole is at $s = -\infty$. This s -plane pole maps into $z = 0$ for discrete rate control; however, the $z = 0$ pole does

add some phase lag because of the necessity to wait for one cycle in order to compute the first difference. Any other stable pole location, whether on the positive or negative real axis, would also have some delay or phase lag associated with it for the same reason.

5.8.3 Integral Control

For continuous systems, we integrate the error to arrive at the control,

$$u(t) = \frac{K_p}{T_I} \int_{t_0}^t e(t) dt \Rightarrow D(s) = \frac{K_p}{T_I s},$$

where T_I is called the *integral*, or *reset time*. The discrete equivalent is to sum all previous errors, yielding

$$u(k) = u(k-1) + \frac{K_p T}{T_I} e(k) \Rightarrow \boxed{D(z) = \frac{K_p T}{T_I (1 - z^{-1})} = \frac{K_p T z}{T_I (z - 1)}} \quad (5.60)$$

Just as for continuous systems, the primary reason for integral control is to reduce or eliminate steady-state errors, but this typically occurs at the cost of reduced stability.

5.8.4 PID Control

Combining all the above yields the PID controller

$$\boxed{D(z) = K_p \left(1 + \frac{Tz}{T_I(z-1)} + \frac{T_D(z-1)}{Tz} \right)} \quad (5.61)$$

This form of control law is able satisfactorily to meet the specifications for a large portion of control problems and is therefore packaged commercially and sold for general use. The user simply has to determine the best values of K_p , T_D , and T_I .

5.8.5 Ziegler-Nichols PID Tuning

The parameters in the PID controller could be selected by any of the design methods previously discussed. However, these methods require a dynamic model of the process which is not always readily available. Ziegler-Nichols tuning is a method for picking the parameters based on fairly simple experiments on the process and thus bypasses the need to determine a complete dynamic model.

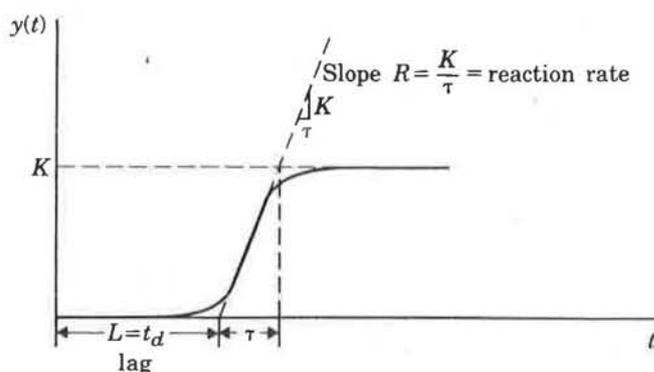


Figure 5.30 Process open-loop step response.

There are two methods. The first, called the *transient-response method*, requires that a step response of the open-loop system is obtained which looks something like that in Fig. 5.30. The response is reduced to two parameters, the time delay, L , and the steepest slope, R , which are defined in the figure. In order to achieve a damping of about $\zeta = 0.2$, the parameters are selected according to those in Table 5.2.

The second method is called the *stability-limit method*. The system is first controlled using proportional control only. The gain, K_p , is slowly increased until continuous oscillations result, at which point the gain and oscillation period are recorded and called K_u and P_u . The PID gains are then determined from Table 5.3.

The rules are based on continuous systems and will apply to the discrete case for very fast sampling (more than 20 times the bandwidth) provided the designer uses the value of T in (5.61) that reflects the actual sample period being used by the controller. For slower sampling, a response degradation similar to that in Example 5.3 should be expected, and additional rate control (higher T_D) would likely be required to make up for the sampling lag.

Table 5.2 Ziegler-Nichols tuning parameters using transient response.

	K_p	T_I	T_D
P	$1/RL$		
PI	$0.9/RL$	$3L$	
PID	$1.2/RL$	$2L$	$0.5L$

Table 5.3 Ziegler-Nichols tuning parameters using stability limit.

	K_p	T_I	T_D
P	$0.5K_u$		
PI	$0.45K_u$	$P_u/1.2$	
PID	$0.6K_u$	$P_u/2$	$P_u/8$

Example 5.12: Let us apply integral control to the system controlling temperature through mixing described in Appendix A. The transfer function is

$$G_3(s) = \frac{e^{-s\tau_d}}{s/a + 1}$$

For the digital implementation, we assume a zero-order hold, sampling period $T = 1$ sec, system time constant $a = 1$ sec, and a $1\frac{1}{2}$ period delay ($\tau_d = 1.5$ sec). The transfer function for this example was determined in (2.42) as

$$G_3(z) = 0.3935 \frac{z + 0.6065}{z^2(z - 0.3679)} \quad (5.62)$$

As it stands, this transfer function has unity gain to a constant control and will have a steady-state error to a constant command or disturbance. If we assume that such behavior in the steady state is unacceptable, we can correct the problem by including integral control by using (5.60) to arrive at the effective system transfer function of

$$DG_3 = 0.3935 K \frac{z + 0.6065}{z(z - 1)(z - 0.3679)}, \quad (5.63)$$

where $K = (K_p T)/T_I$ following (5.60).

The root locus of this system versus K is sketched in Fig. 5.31(a) with the roots corresponding to $K = 1$ marked with a square and those corresponding to $K = 0.3$ marked with a triangle. The $K = 0.3$ roots have $\zeta = 0.5$ and, therefore, should have about a 15% overshoot. Fig. 5.31(b) shows the step response, which verifies the overshoot and indicates that the system has a settling time of

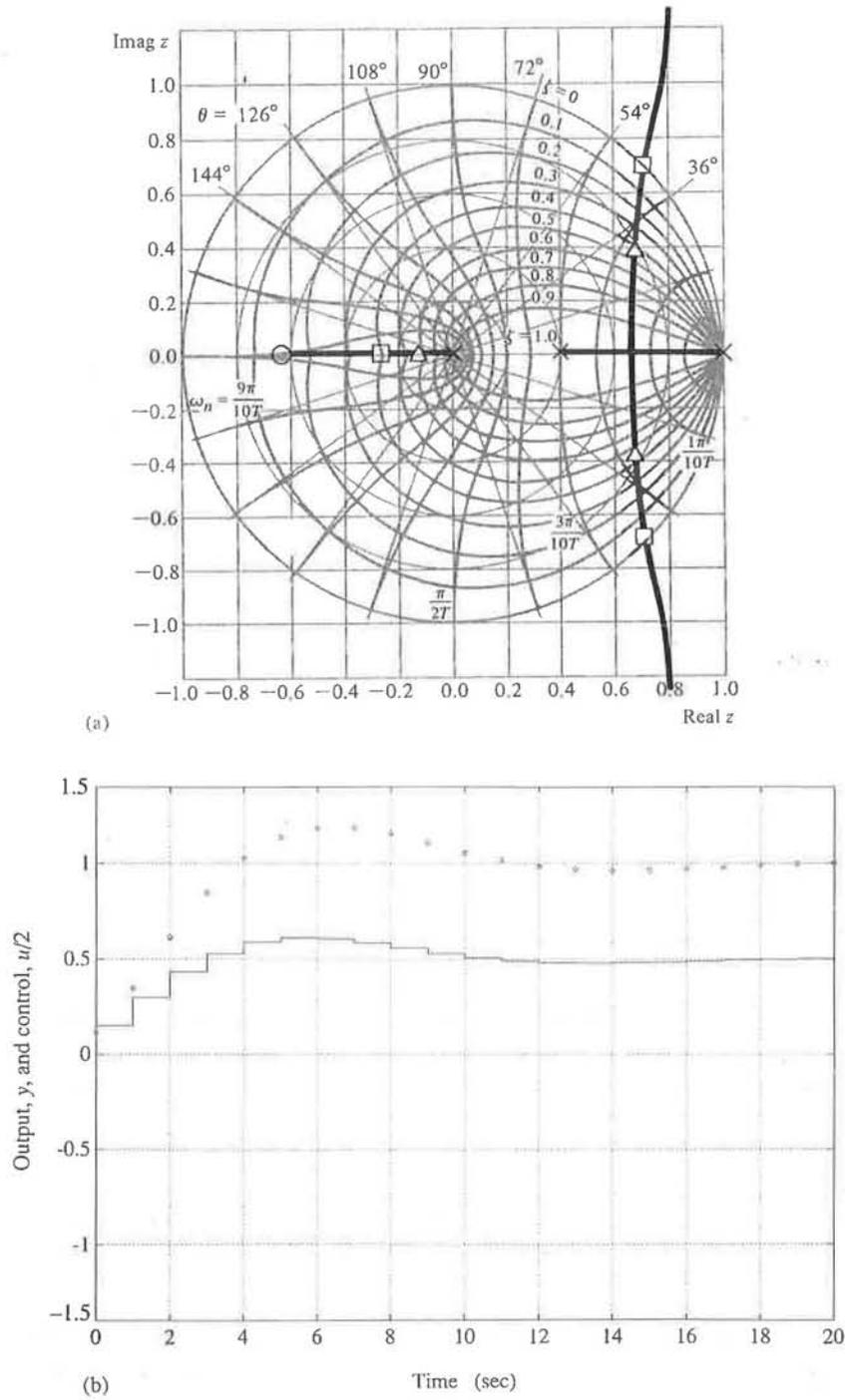


Figure 5.31 Mixing-flow plant with pure discrete integral control: (a) root locus, (b) step response.

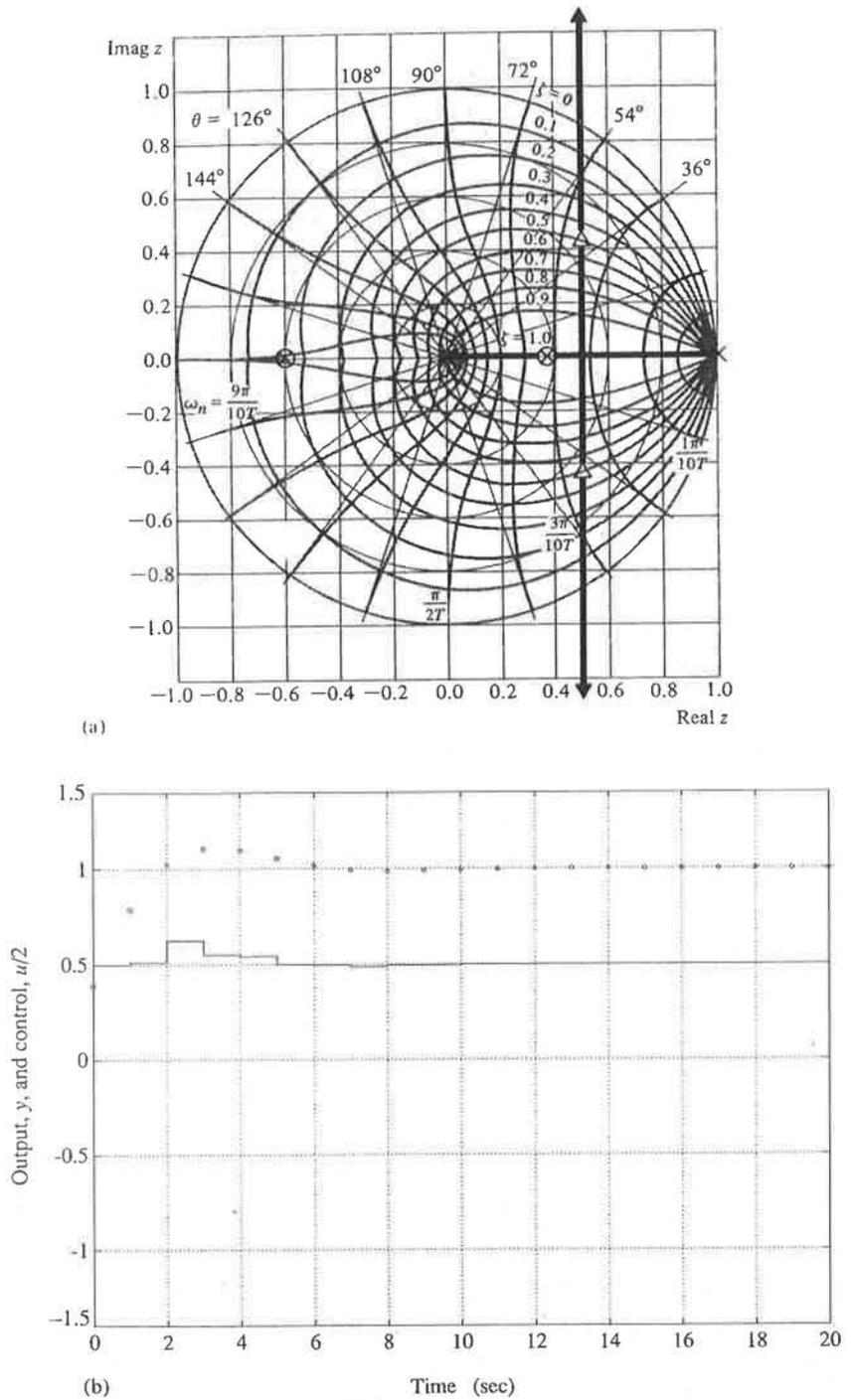


Figure 5.32 Mixing flow plant with lead compensation and discrete integral control: (a) root locus, (b) step response.

$t_s = 18$ sec. A system with an open-loop time constant of 1 sec is capable of a much faster response than this; however, the delay will cause stability problems if we ask for too much. Let's simply add a lead compensation in order to investigate how to speed up the system. One that cancels the plant pole at $z = 0.3679$ and the plant zero at $z = -0.6065$ is

$$D_L = K \frac{z - 0.3679}{z + 0.6065}.$$

With this addition, the complete compensation becomes

$$D(z) = K \frac{z(z - 0.3679)}{(z - 1)(z + 0.6065)};$$

and the system open-loop transfer function reduces to

$$DG_3(z) = 0.3935K \frac{1}{z(z - 1)},$$

whose root locus versus K is sketched in Fig. 5.32(a). The triangle marks the location of $\zeta = 0.5$, which occurs for $K = 1$ and yields a step response as shown in Fig. 5.32(b). Note that the overshoot has slightly improved and the settling time has been cut in half to 9 sec.

In this example, the transfer function was available for design purposes; therefore, it was possible to determine the integral control using the design methods discussed in Section 5.4. Had we not had a model, the Zeigler-Nichols method could have been applied to help determine the gains.

5.9 SUMMARY

In this chapter we have reviewed the philosophy and specifications of the design of control systems by transform techniques and discussed four such methods. First we developed the relations between the time-domain specifications of overshoot, rise time, and settling time and poles in the z -plane. Using the theory and techniques of discrete equivalents, we then showed how a continuous design can be converted into a discrete design. This design process was called *emulation*. With a sample rate of six times the bandwidth, we found that the approximation was quite coarse and would require substantial adjustment to meet the design specifications. As a second design

approach, we discussed the root locus in the z -plane. We saw that the root locus is the same as for s -plane designs, but the relations to time-domain response must refer to the z -plane. Our third design method was based on frequency-response techniques. The process was carried out in the z -plane, where extensive reliance on a computer is required, and in the w -plane, where much of the experience from continuous design can be used more readily. Our final method was a direct transfer-function calculation wherein we found causality and stability constraints on an overall transfer function so that an acceptable compensator can be derived. Here we found that canceling poles near the unit circle can have undesirable effects. In the final section we presented a design by root locus methods for a plant which required the introduction of discrete integral control.

PROBLEMS AND EXERCISES

5.1 Use the $z = e^{sT}$ mapping function and prove that the curve of constant ζ in s is a logarithmic spiral in z .

5.2 Sketch the acceptable region in the s -plane for the specification on the antenna given before (5.15) and sketch the s -plane root locus corresponding to the controller of (5.15).

5.3 *Root locus review.* The following root loci illustrate important features of the root locus technique. All are capable of being done by hand, and it is recommended that they be done that way in order to develop skills in verifying a computer's output. Once sketched roughly by hand, it is useful to fill in the details with a computer.

a) The locus for

$$1 + K \frac{s + 1}{s^2(s + P_1)}$$

is typical of the behavior near $s = 0$ of a double integrator with lead compensation or a single integration with a lag network and one additional real pole. Sketch the locus for values of P_1 of 5, 9, and 20. Pay close attention to the real axis break-in and break-away points.

b) The locus for

$$1 + K \frac{1}{s(s + 1)((s + a)^2 + 4)}$$

illustrates the possibility of complex multiple roots and shows the value of departure angles. Plot the locus for $a = 0, -1$, and -2 . Be sure to note the departure angles from the complex poles in each case.

- c) The locus for

$$1 + K \frac{(s+1)^2 + \omega^2}{s(s^2 + 4)} = 0$$

illustrates the use of complex zeros to compensate for the presence of complex poles due to vibration modes. Be sure to compute (estimate) the angles of departure and arrival. Sketch the loci for $\omega = 1$ and $\omega = 3$. Which case is unconditionally stable (stable for all positive K less than the design value)?

- d) For

$$1 + K \frac{s}{(s - P_1)(s - P_2)} = 0$$

show that the locus is a circle of radius $\sqrt{P_1 P_2}$ centered at the origin (location of the zero). Can this result be translated to the case of two poles and a zero on the negative real axis?

- 5.4 Appendix A gives the transfer function of a satellite attitude control as

$$G_1(z) = K \frac{z + 1}{(z - 1)^2}$$

- a) Sketch the root locus of this system as a function of K with unity feedback. What is the type of the uncompensated system?
- b) Add a lead network so that the dominant poles are at $\zeta = 0.5$ and $\theta = 45^\circ$. Plot the closed-loop step response.
- 5.5 Use Ragazzini's direct-design method to find a compensation for the satellite transfer function of Appendix A with $T = 1.0$ sec such that all the closed loop poles are at $z = 0$. Sketch the root locus and the step response of the resulting design.
- 5.6 Repeat the design of the antenna control system described in Examples 5.1 through 5.4 but use a sample period of $T = 0.1$ sec.
- a) Use emulation with pole-zero mapping of (5.15).
- b) Use the z -plane root locus.
- 5.7 Design the antenna control system with a sample period of $T = 0.5$ sec using emulation and check the resulting compensation with a z -plane analysis.
- a) Use the method as described in Section 5.3
- b) Augment the plant model with an approximation of the sampler consisting of

$$H_s(s) = \frac{2/T}{s + 2/T}$$

then design $D(s)$ and find the discrete equivalent.

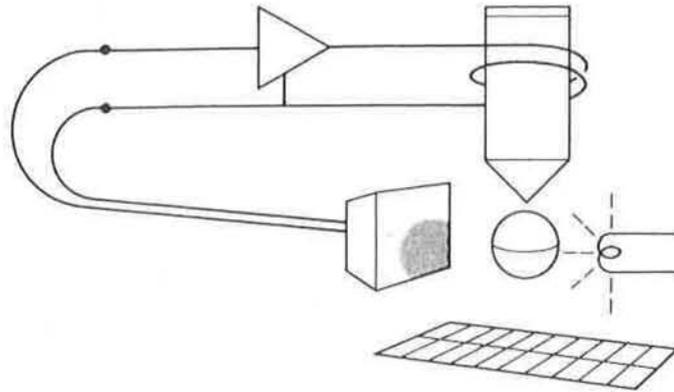


Figure 5.33 A steel ball suspended by means of an electromagnet.

- c) Compare the degradation of the damping ratio, ζ , due to sampling for both design methods.

5.8 For

$$G(s) = \frac{1}{(s + 0.1)(s + 3)}$$

being controlled with a digital controller using a sample period of $T = 0.1$ sec, design compensation using z -plane root locus that will respond to a step with a rise time of ≤ 1 sec and an overshoot $\leq 5\%$. What can be done to reduce the steady state error?

5.9 It is possible to suspend a mass of magnetic material by means of an electromagnet whose current is controlled by the position of the mass [Woodson and Melcher (1968)]. A schematic of a possible setup is shown in Fig. 5.33. The equations of motion are

$$m\ddot{x} = -mg + f(x, \mathbf{I}),$$

where the force on the ball due to the electromagnet is given by $f(x, \mathbf{I})$. At equilibrium, the magnet force balances the gravity force; suppose we call the current there \mathbf{I}_0 . If we write $\mathbf{I} = \mathbf{I}_0 + i$ and expand f about $x = 0$ and $\mathbf{I} = \mathbf{I}_0$, and if we neglect higher-order terms, we obtain

$$m\ddot{x} = k_1x + k_2i.$$

Reasonable values are $m = 0.02$ kg, $k_1 = 20$ N/m, $k_2 = 0.4$ N/A.

- Compute the transfer function from i to x and draw the (continuous) root locus for simple feedback $i = -Kx$.
- Let the sample period be 0.02 sec and compute the plant discrete transfer function when used with a zero-order hold.

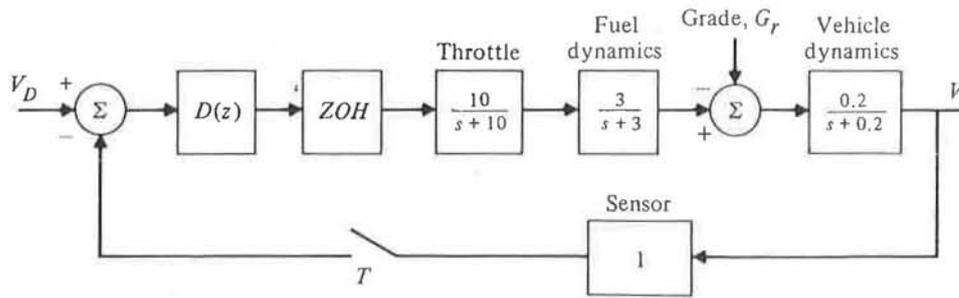


Figure 5.34 An automotive cruise-control system.

- c) Design a digital control for the magnetic levitation to meet the specifications $t_r \leq 0.1$ sec, $t_s \leq 0.4$ sec, and overshoot $\leq 20\%$.
- d) Plot a root locus of your design versus k_1 and discuss the possibility of balancing balls of various masses.
- e) Plot a step response of your design to an initial disturbance displacement on the ball and show both x and the control current i . If the sensor can measure x over a range of only $\pm \frac{1}{4}$ cm, and if the amplifier can provide a current of only 1 A, what is the *maximum* displacement possible for control, neglecting the nonlinear terms in $f(x, I)$?

5.10 Design compensation using w -plane frequency-response methods for a $1/s^2$ plant (Appendix A.1) that yields a bandwidth of approximately 10 rad/sec. Pick two candidate sample rates, one where $\nu T/2$ at crossover is approximately 1.3, and one fairly fast sample rate yielding a $\nu T/2$ of approximately 0.2. Design each case to have the same phase margin (approximately 30°), then compare the damping of the equivalent s -plane roots resulting from the two designs.

5.11 The transfer function for pure derivative control is

$$D(z) = K_p T_D \frac{z-1}{Tz},$$

where the pole at $z = 0$ adds some destabilizing phase lag. It therefore seems that it would be advantageous to remove it, that is, to use derivative control of the form

$$D(z) = K_p T_D \frac{(z-1)}{T}.$$

Can this be done? Support your answer with the difference equation that would be required and discuss the requirements to implement it.

5.12 For the automotive cruise-control system shown in Fig. 5.34,

- a) design a PD controller to achieve a t_r of 5 sec with no overshoot,
- b) determine the speed error on a 3% grade (i.e., $G_r = 3$ in Fig. 5.34), and
- c) design a PID controller to meet the same specifications as part (a) and that has no error on grades.

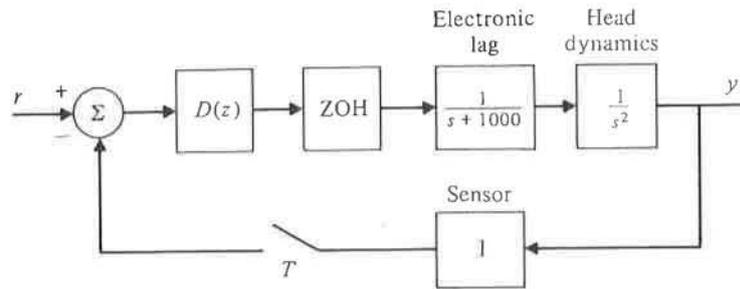


Figure 5.35 A disk drive tracker head.

5.13 For the disk drive tracker head described in Fig. 5.35, design compensation for a 20 msec t_s to a step input with overshoot $\leq 20\%$ and $T = 1$ msec. Use

- s -plane frequency-response design and emulation,
- z -plane frequency-response design, and
- w -plane frequency-response design.

5.14 The tethered satellite system shown in Fig. 5.36 has a moveable tether attachment point so that torques can be produced for attitude control. The block diagram of the system is shown in Fig. 5.37. Note that the integrator in the actuator block indicates that a constant-voltage command to the servo motor will produce a constant velocity of the attachment point.

- Is it possible to stabilize this system with θ feedback to a PID controller? Support your answer.

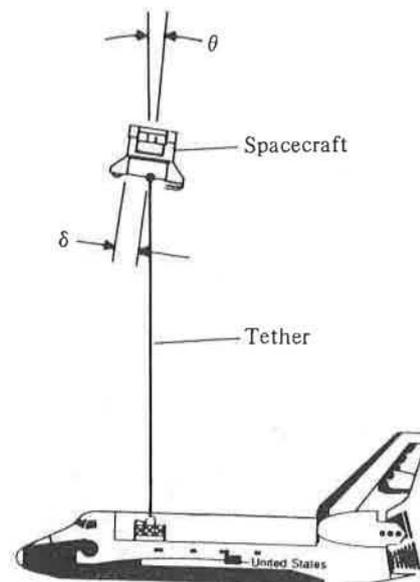


Figure 5.36 A tethered satellite system.

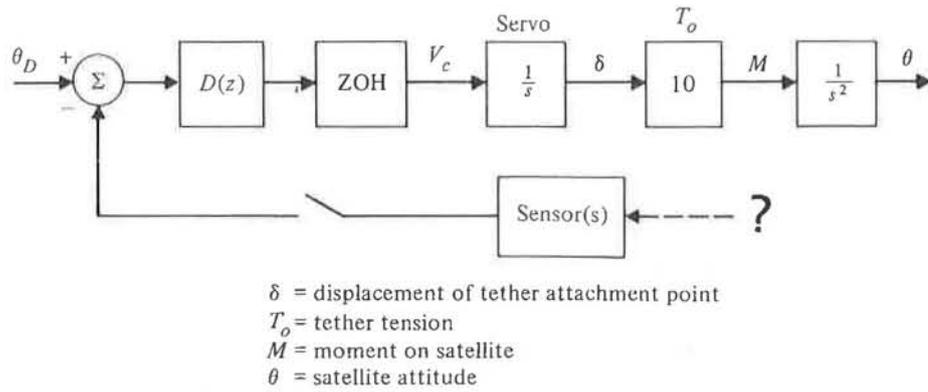


Figure 5.37 Block diagram for the tethered satellite system.

- b) Select what sensors would be most useful.
- c) Design compensation for the system using the sensor(s) that you selected in part (b) so that it has a 2-sec rise time and $50^\circ PM$.

5.15 The excavator shown in Fig. 5.38 has a sensor measuring the angle of the stick as part of a control system to control automatically the motion of the bucket through the earth. The sensed stick angle is to be used to determine the control signal to the hydraulic actuator moving the stick. The schematic diagram for this control system is shown in Fig 5.39, where $G(s)$ is the system transfer function



Figure 5.38 An excavator with an automatic control system.

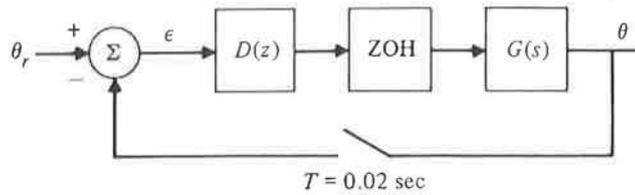


Figure 5.39 Schematic diagram for the control system of the excavator.

given by:

$$G(s) = \frac{1000}{s(s+10)(s^2+1.2s+144)},$$

and where the compensation is implemented in a control computer sampling at $f_s = 50$ Hz and is of the form

$$D(z) = K(1 + K_v(1 - z^{-1}) + K_a(1 - 2z^{-1} + z^{-2})).$$

The oscillatory roots in $G(s)$ arise from the compressibility of the hydraulic fluid (with some entrained air) and is often referred to as the *oil-mass resonance*.

a) Show that the steady-state error, $\epsilon (= \theta_r - \theta)$, is

$$\epsilon(\infty) = \frac{1.44}{K}$$

when θ_r is a unit ramp.

- b) Determine the highest K possible (i.e., at the stability boundary) for proportional control ($K_a = K_v = 0$).
- c) Determine the highest K possible (i.e., at the stability boundary) for PD ($K_a = 0$) control.
- c) Determine the highest K possible (i.e., at the stability boundary) for PD plus acceleration ($K_a \neq 0$) control.¹³

5.16 For a system given by

$$G(s) = \frac{a}{s(s+a)}$$

determine the conditions under which the K_v of the continuous system is approximately equal to the K_v of the system preceded by a ZOH and represented by its discrete transfer function.

¹³For further reading on damping the oil-mass resonance of hydraulic systems, see Viersma (1980).

5.17 Design a digital controller for

$$G(s) = \frac{1}{s(s + 0.4)}$$

preceded by a ZOH so that the response has a rise time of approximately 0.5 sec, overshoot $< 25\%$, and zero steady-state error to a step command. [*Hint:* Cancel the plant pole at $s = 0.4$ with a compensator zero; a second-order closed loop system will result, making the transient response comparison between experiment and theory much easier.]

- a) Determine a $D(z)$ using emulation with the pole-zero mapping technique. Do two designs, one for $T = 100$ msec and one for $T = 250$.
- b) Repeat part (a) using the z -plane root locus method.
- c) Simulate the closed-loop system response to a unit step with the $D(z)$'s obtained in parts (a) and (b). Use the discrete equivalent of the plant in your calculations. Compare the four digitally controlled responses with the original specifications. Explain any differences that you find.