4.2 z-PLANE SPECIFICATIONS OF CONTROL SYSTEM DESIGN

Different properties that describe the performance of a feedback control system (Fig. 4.2) are listed below.

- Stability
- Steady-state accuracy
- Transient accuracy
- Disturbance rejection
- Insensitivity and robustness

We will discuss each of these in turn. Our attention will be focused on the unity feedback systems[1] of the form shown in Fig. 4.3.

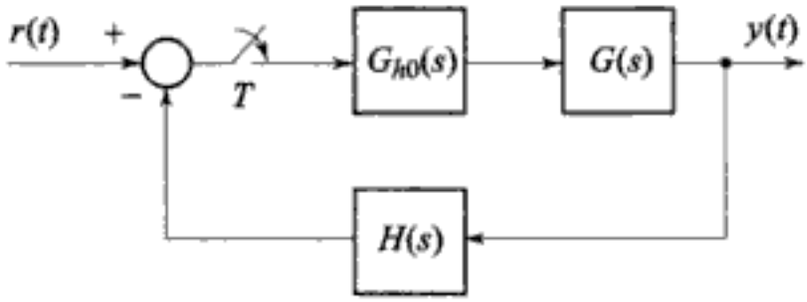

Fig. 4.2 A nonunity feedback discrete-time system

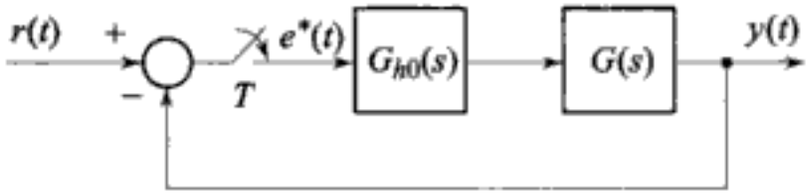

Fig. 4.3 A unity feedback discrete-time system

Stability

Stability is a very important property of closed-loop systems. Almost every working system is designed to be stable. We seek to improve the system performance within the constraints imposed by stability considerations.

Analytical study of stability of discrete-time systems can be carried out using the Jury stability criterion (Section 2.7) and/or the Routh stability criterion (Section 2.13).

1. It is assumed that the reader is familiar with the design of unity and nonunity feedback continuous-time systems. With this background, the results presented in this chapter for unity feedback discrete-time systems can easily be extended for the nonunity feedback case.

Steady-State Accuracy

Steady-state accuracy refers to the requirement that after all transients become negligible, the error between the reference input r and the controlled output y must be acceptably small. The specification on steady-state accuracy is often based on polynomial inputs of degree k: $r(t) = \frac{t^k}{k!}\mu(t)$. If $k = 0$, the input is a step of unit amplitude; if $k = 1$, the input is a ramp with unit slope; and if $k = 2$, the input is a parabola with unit second derivative. From the common problems of mechanical motion control, these inputs are called, respectively, *position, velocity,* and *acceleration* inputs.

For quantitative analysis, we consider the unity-feedback discrete-time system shown in Fig. 4.3. The steady-state error is the difference between the reference input $r(k)$ and the controlled output $y(k)$ when steady-state is reached, i.e., steady-state error

$$e_{ss}^* = \lim_{k\to\infty} e(k) = \lim_{k\to\infty} [r(k) - y(k)] \tag{4.1a}$$

Using the final value theorem (Eqn. (A.45) in Appendix A),

$$e_{ss}^* = \lim_{z\to 1} [(z-1)E(z)] \tag{4.1b}$$

provided that $(z-1)\,E(z)$ has no poles on the boundary and outside of the unit circle in the z-plane.

For the system shown in Fig. 4.3, define

$$G_{h0}G(z) = (1 - z^{-1})\,\mathcal{Z}\left[\frac{G(s)}{s}\right]$$

Then we have $$\frac{Y(z)}{R(z)} = \frac{G_{h0}G(z)}{1 + G_{h0}G(z)}$$

and $$E(z) = R(z) - Y(z) = \frac{R(z)}{1 + G_{h0}G(z)} \tag{4.2}$$

By substituting Eqn. (4.2) into Eqn. (4.1b), we obtain

$$e_{ss}^* = \lim_{z\to 1} [(z-1)E(z)] \tag{4.3a}$$

$$= \lim_{z\to 1} \left[(z-1)\frac{R(z)}{1 + G_{h0}G(z)}\right] \tag{4.3b}$$

Thus, the steady-state error of a discrete-time system with unity feedback depends on the reference input signal $R(z)$, and the forward-path transfer function $G_{h0}G(z)$. By the nature of the limit in Eqns (4.3), we see that the result of the limit can be zero or can be a constant different from zero. Also the limit may not exist, in which case the final-value theorem does not apply. However,

it is easy to see from basic definition (4.1a) that $e_{ss}^{*} = \infty$ in this case anyway because $E(z)$ will have a pole at $z = 1$ that is of order higher than one. Discrete-time systems having a finite nonzero steady-state error when the reference input is a zero-order polynomial input (a constant) are labelled 'Type-0'. Similarly, a system that has finite nonzero steady-state error to a first-order polynomial input (a ramp) is called a 'Type-1' system, and a system with finite nonzero steady-state error to a second-order polynomial input (a parabola) is called a 'Type-2' system.

Let the reference input to the system of Fig. 4.3 be a step function of magnitude unity. The z-transform of $r(t) = \mu(t)$ is

$$R(z) = \frac{z}{z-1} \tag{4.4a}$$

Substituting $R(z)$ into Eqn. (4.3b), we have

$$e_{ss}^{*} = \lim_{z \to 1} \frac{1}{1 + G_{h0}G(z)} = \frac{1}{1 + \lim_{z \to 1} G_{h0}G(z)}$$

In terms of the *position error constant* K_p, defined as

$$K_p = \lim_{z \to 1} G_{h0}G(z) \tag{4.4b}$$

the steady-state error to unit-step input becomes

$$e_{ss}^{*} = \frac{1}{1 + K_p} \tag{4.4c}$$

For a ramp input $r(t) = t\mu(t)$, the z-transform

$$R(z) = \frac{Tz}{(z-1)^2} \tag{4.5a}$$

Substituting into Eqn. (4.3b), we get

$$e_{ss}^{*} = \lim_{z \to 1} \frac{T}{(z-1)[1 + G_{h0}G(z)]} = \frac{1}{\lim_{z \to 1}\left[\frac{z-1}{T} G_{h0}G(z)\right]}$$

In terms of *velocity error constant* K_v, defined as

$$K_v = \frac{1}{T} \lim_{z \to 1} [(z-1)G_{h0}G(z)] \tag{4.5b}$$

the steady-state error to unit-ramp input becomes

$$e_{ss}^{*} = \frac{1}{K_v} \tag{4.5c}$$

For a parabolic input $r(t) = (t^2/2)\,\mu(t)$, the z-transform

$$R(z) = \frac{T^2 z(z+1)}{2(z-1)^3} \tag{4.6a}$$

Substituting into Eqn. (4.3b), we get

$$e_{ss}^{*} = \lim_{z\to 1} \frac{T^2}{(z-1)^2[1+G_{h0}G(z)]} = \frac{1}{\lim_{z\to 1}\left[\left(\frac{z-1}{T}\right)^2 G_{h0}G(z)\right]}$$

In terms of *acceleration error constant* K_a, defined as

$$K_a = \frac{1}{T^2}\lim_{z\to 1}[(z-1)^2\, G_{h0}G(z)] \tag{4.6b}$$

the steady-state error to unit-parabolic input becomes

$$e_{ss}^{*} = \frac{1}{K_a} \tag{4.6c}$$

▲▲

As said earlier, discrete-time systems can be classified on the basis of their steady-state response to polynomial inputs. We can always express the forward-path transfer function $G_{h0}G(z)$ as

$$G_{h0}G(z) = \frac{K\prod_i (z-z_i)}{(z-1)^N \prod_j (z-p_j)};\ p_j \neq 1,\ z_i \neq 1 \tag{4.7}$$

$G_{h0}G(z)$ in Eqn. (4.7) involves the term $(z-1)^N$ in the denominator. As $z \to 1$, this term dominates in determining the steady-state error. Digital control systems are therefore classified in accordance with the number of poles at $z = 1$ in the forward-path transfer function as described below.

Type-0 system

If $N = 0$, the steady-state errors to various standard inputs obtained from Eqns (4.1)–(4.7) are

$$e_{ss}^{*} = \begin{cases} \dfrac{1}{1+K_p} & \text{in response to unit-step input;}\quad K_p = \left.\dfrac{K\prod_i (z-z_i)}{\prod_j (z-p_j)}\right|_{z=1} \\ \infty & \text{in response to unit-ramp input} \\ \infty & \text{in response to unit-parabolic input} \end{cases} \tag{4.8a}$$

Thus a system with $N = 0$, or no pole at $z = 1$ in $G_{h0}G(z)$, has a finite non-zero position error, infinite velocity and acceleration errors at steady-state.

Type-1 system

If $N = 1$, the steady-state errors to various standard inputs are

$$e_{ss}^{*} = \begin{cases} 0 & \text{in response to unit-step input} \\ \dfrac{1}{K_v} & \text{in response to unit-ramp input; } K_v = \left.\dfrac{\dfrac{K}{T}\prod_i (z - z_i)}{\prod_j (z - p_j)}\right|_{z=1} \\ \infty & \text{in response to unit-parabolic input} \end{cases} \tag{4.8b}$$

Thus a system with $N = 1$, or one pole at $z = 1$ in $G_{h0}G(z)$, has zero position error, a finite non-zero velocity error and infinite acceleration error at steady-state.

Type-2 system

If $N = 2$, the steady-state errors to various standard inputs are

$$e_{ss}^{*} = \begin{cases} 0 & \text{in response to unit-step input} \\ 0 & \text{in response to unit-ramp input} \\ \dfrac{1}{K_a} & \text{in response to unit-parabolic input; } K_a = \left.\dfrac{\dfrac{K}{T^2}\prod_i (z - z_i)}{\prod_j (z - p_j)}\right|_{z=1} \end{cases} \tag{4.8c}$$

Thus a system with $N = 2$, or two poles at $z = 1$ in $G_{h0}G(z)$, has zero position and velocity errors and a finite non-zero acceleration error at steady-state.

Steady-state errors for various inputs and systems are summarized in Table 4.1.

Table 4.1 Steady-state errors for various inputs and systems

Type of input	*Steady-state error*		
	Type-0 system	*Type-1 system*	*Type-2 system*
Unit-step	$\dfrac{1}{1+K_p}$	0	0
Unit-ramp	∞	$\dfrac{1}{K_v}$	0
Unit-parabolic	∞	∞	$\dfrac{1}{K_a}$

$$K_p = \lim_{z\to 1} G_{h0}G(z); \; K_v = \frac{1}{T}\lim_{z\to 1}[(z-1)\,G_{h0}G(z)];$$

$$K_a = \frac{1}{T^2}\lim_{z\to 1}[(z-1)^2\,G_{h0}G(z)]$$

▲▲

The development above indicates that, in general, increased system gain K and/or addition of poles at $z = 1$ to the open-loop transfer function $G_{h0}G(z)$ tend to decrease steady-state errors. However, as will be seen later in this chapter, both large system gain and the poles at $z = 1$ in the loop transfer function have destabilizing effects on the system. Thus a control system design is usually a trade off between steady-state accuracy and acceptable relative stability.

Example 4.1

In the previous chapter we have shown that sampling usually has a deterimental effect on the transient response and the relative stability of a control system. It is natural to ask what the effect of sampling on the steady-state error of a closed-loop system will be? In other words, if we start out with a continuous-time system and then add S/H to form a digital control system, how would the steady-state errors of the two systems compare, when subject to the same type of input?

Let us first consider the system of Fig. 4.3 without S/H. Assume that the process $G(s)$ is represented by Type-1 transfer function:

$$G(s) = \frac{K(1+\tau_a s)(1+\tau_b s)\cdots(1+\tau_m s)}{s(1+\tau_1 s)(1+\tau_2 s)\cdots(1+\tau_n s)}$$

having more poles than zeros.

The velocity error constant

$$K_v = \lim_{s\to 0} sG(s) = K$$

The steady-state error of the system to unit-step input is zero, to unit-ramp input is $1/K$, and to unit-parabolic input is ∞.

We now consider the system of Fig. 4.3 with S/H:

$$G_{h0}G(z) = (1 - z^{-1})\,\mathcal{Z}\left[\frac{K(1+\tau_a s)(1+\tau_b s)\cdots(1+\tau_m s)}{s^2(1+\tau_1 s)(1+\tau_2 s)\cdots(1+\tau_n s)}\right]$$

$$= (1 - z^{-1})\,\mathcal{Z}\left[\frac{K}{s^2} + \frac{K_1}{s} + \text{terms due to the non-zero poles}\right]$$

$$= (1 - z^{-1})\left[\frac{KTz}{(z-1)^2} + \frac{K_1 z}{z-1} + \text{terms due to the non-zero poles}\right]$$

It is important to note that the terms due to the non-zero poles do not contain the term $(z - 1)$ in the denominator. Thus, the velocity error constant is

$$K_v = \frac{1}{T}\lim_{z\to 1}\,[(z-1)G_{h0}G(z)] = K$$

The steady-state error of the discrete-time system to unit-step input is zero, to unit-ramp input is $1/K$, and to unit-parabolic input is ∞. Thus for a Type-1 system, the system with S/H has exactly the same steady-state error as the continuous-time system with the same process transfer function (this, in fact, is true for Type-0 and Type-2 systems also).

Equations (4.5b) and (4.6b) may purport to show that the velocity error constant and the acceleration error constant of a digital control system depend on the sampling period T. However, in the process of evaluation, T gets cancelled and the error depends only on the parameters of the process and the type of inputs.

Transient Accuracy

Figure 4.4 shows a typical unit-step response of a digital control system. Specifications of transient performance may be made in the time domain in terms of rise time, peak time, peak overshoot, settling time, etc. The use of root locus plots for the design of digital control systems necessitates the translation of time-domain performance specifications into desired locations of closed-loop poles in the z-plane. Whereas the use of frequency response plots necessitates the translation of time-domain specifications in terms of frequency response features such as bandwidth, phase margin, gain margin, resonance peak, resonance frequency, etc.

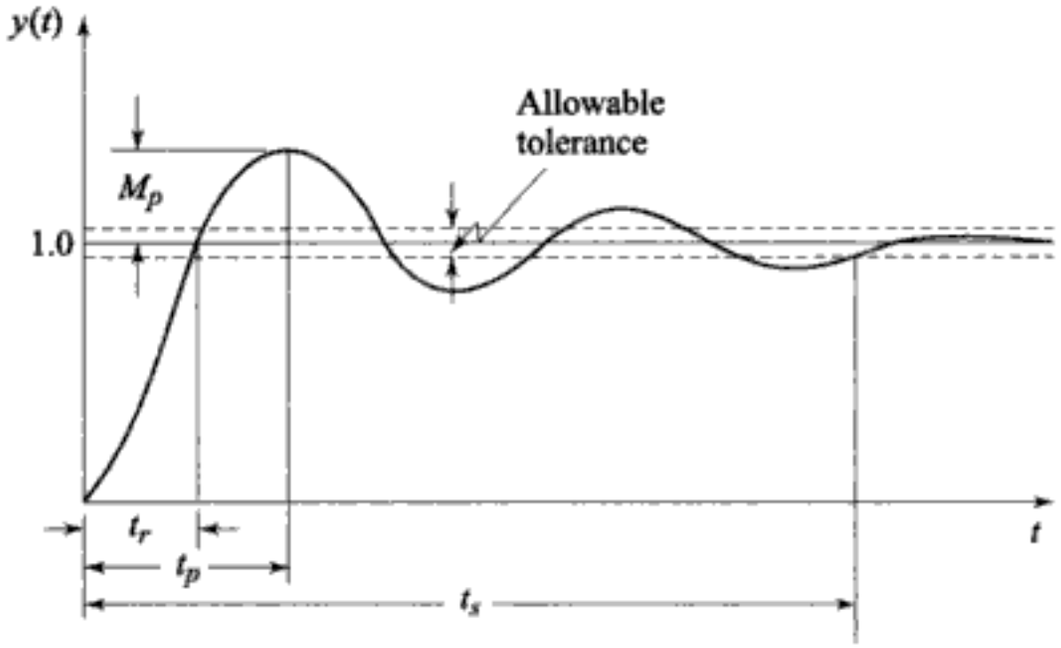

Fig. 4.4 Typical unit-step response of a digital control system

Specifications in terms of root locations in z-plane

Our approach is to first obtain the transient response specifications in terms of characteristic roots in the s-plane, and then use the relation

$$z = e^{sT} \tag{4.9}$$

to map the s-plane characteristic roots to the z-plane.

The transient response of Fig. 4.4 resembles the unit-step response of an underdamped second-order system

$$\frac{Y(s)}{R(s)} = \frac{\omega_n^2}{s^2 + 2\zeta\omega_n s + \omega_n^2} \tag{4.10}$$

where

ζ = damping ratio, and

ω_n = undamped natural frequency.

The transient response specifications in terms of rise time t_r, peak time t_p, peak overshoot M_p, and settling time t_s can be approximated to the parameters ζ and ω_n of the second-order system defined by Eqn. (4.10) using the following correlations[2]:

$$t_r(0\% \text{ to } 100\%) = \frac{\pi - \cos^{-1}\zeta}{\omega_n\sqrt{1-\zeta^2}} \tag{4.11}$$

$$t_p = \frac{\pi}{\omega_n\sqrt{1-\zeta^2}} \tag{4.12}$$

$$M_p = \exp(-\pi\zeta / \sqrt{1-\zeta^2}) \tag{4.13}$$

$$t_s(2\% \text{ tolerance band}) = \frac{4}{\zeta\omega_n} \tag{4.14}$$

Peak overshoot is used mainly for relative stability. Values in excess of about 40% may indicate that the system is dangerously close to absolute instability. Many systems are designed for 5 to 25% overshoot. No overshoot at all is sometimes desirable. However, this usually penalizes the speed of response needlessly.

The specification on speed of response in terms of t_r, t_p and/or t_s should be consistent as all these depend on ζ and ω_n. The greater the magnitude of ω_n when ζ is constant, the more rapid the response approaches the desired steady-state value. The value of ω_n is limited by measurement noise considerations—a system with large ω_n has large bandwidth and will therefore allow the high frequency noise signals to affect its performance.

We need to now convert the specifications on ζ and ω_n into guidelines on the placement of poles and zeros in the z-plane in order to guide the design of digital controls. We do so through the mapping (4.9).

Figure 4.5 illustrates the translation of specifications on ζ and ω_n to the characteristic root locations in the z-plane (referring to Section 2.12 will be helpful). The s-plane poles

$$s_{1,2} = -\zeta\omega_n \pm j\omega_n\sqrt{1-\zeta^2} = -\zeta\omega_n \pm j\omega_d \tag{4.15a}$$

for constant ζ, lie along a radial line in the s-plane (Fig. 4.5a). In the z-plane,

$$z_{1,2} = e^{-\zeta\omega_n T} e^{\pm j\omega_n T\sqrt{1-\zeta^2}} = r\, e^{\pm j\theta} \tag{4.15b}$$

2. Chapter 6 of reference [180].

The magnitude of z (i.e., the distance to the origin) is $r = e^{-\zeta\omega_n T}$ and the angles with the positive real axis of the z-plane, measured positive in the counterclockwise direction, are $\theta = \omega_n T\sqrt{1-\zeta^2}$. It should be observed that the z-plane pole locations depend on the s-plane positions as well as the sampling interval T.

As ω_n increases for a constant-ζ, the magnitude of z decreases and the phase angle increases; constant-ζ locus is a logarithmic spiral in the z-plane (Fig. 4.5b). Increasing ω_n negatively, gives the mirror image.

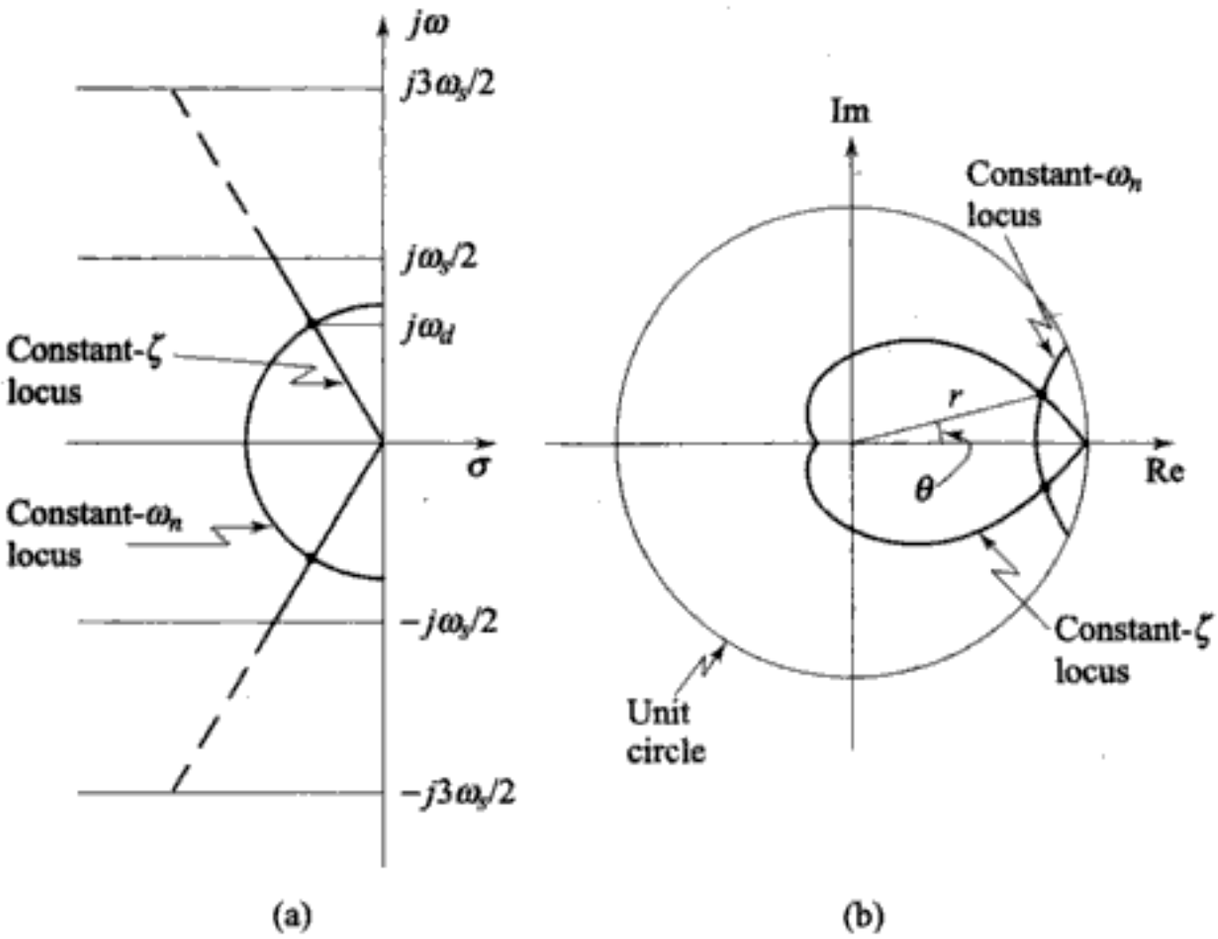

Fig. 4.5 Mapping of s-plane patterns on the z-plane

In Fig. 4.5a, the s-plane has been divided into strips of width ω_s, where $\omega_s = 2\pi/T$ is the sampling frequency. The primary strip extends from $\omega = -\omega_s/2$ to $+\ \omega_s/2$, and the complimentary strips extend from $-\ \omega_s/2$ to $-3\omega_s/2$, ... , for negative frequencies, and from $\omega_s/2$ to $3\omega_s/2$,..., for positive frequencies. We will assume that the low-pass analog filtering characteristics of the continuous-time plant and the ZOH device attenuate the responses due to the poles in the complimentary strips; only the poles in the primary strip, generally, need be considered.

Figure 4.5 illustrates the mapping of constant-ζ locus in the primary strip of the s-plane to the z-plane. As the imaginary parts $\pm j\omega_d = \pm j\omega_n\sqrt{1-\zeta^2}$ of the s-plane poles move closer to the limit $\pm\ j\omega_s/2$ of the primary strip, the angles

$\theta = \pm\, \omega_d T = \pm\, \omega_n T\sqrt{1-\zeta^2}$ of the z-plane poles approach the direction of the negative real axis. The negative real axis in the z-plane thus corresponds to the boundaries of the primary strip in the s-plane. Figure 4.5 also shows the mapping of a constant-ω_n locus in the primary strip of the s-plane to the z-plane.

In the z-plane, the closed-loop poles must lie on the constant-ζ spiral to satisfy peak overshoot requirement, also the poles must lie on constant-ω_n curve to satisfy speed of response requirement. The intersection of the two curves (Fig. 4.5b) provides the preferred pole locations, and the design aim is to make the root locus pass through these locations.

If one chooses the following boundaries for the system response:

$$T = 1 \text{ sec}$$

$$\text{Peak overshoot} \le 15\% \Rightarrow \zeta \ge 0.5$$

$$\text{Settling time} \le 25 \text{ sec} \Rightarrow \omega_n \ge \frac{8}{25},$$

the acceptable boundaries for the closed-loop pole locations in z-plane are shown in Fig. 4.6.

In the chart of Fig. 4.6, the patterns are traced for various natural frequencies and various damping ratios. Such a chart is a useful aid in root-locus design technique. We will be using this chart in our design examples.

Dominant poles

Most control systems found in practice are of high order. The preferred locations of closed-loop poles given by Fig. 4.5b realize the specified transient performance only if the other closed-loop poles and zeros of the system have negligible effect on the dynamics of the system, i.e., only if the closed-loop poles corresponding to specified ζ and ω_n are dominant.

In the following, we examine the relationship between the pole-zero patterns and the corresponding step-responses of discrete-time systems. Our attention will be restricted to the step responses of the discrete-time system with transfer function

$$\frac{Y(z)}{R(z)} = \frac{K(z-z_1)(z-z_2)}{(z-p)(z-re^{j\theta})(z-re^{-j\theta})} = \frac{K(z-z_1)(z-z_2)}{(z-p)(z^2-2r\cos\theta z+r^2)} \qquad (4.16)$$

for a selected set of values of the parameters K, z_1, z_2, p, r and θ.

We assume that the roots of the equation

$$z^2 - 2r\cos\theta z + r^2 = 0$$

are the preferred closed-loop poles corresponding to the specified values of ζ and ω_n. Complex-conjugate pole pairs corresponding to $\zeta = 0.5$ with $\theta = 18°$, 45° and 72° will be considered in our study. The pole pair with $\theta = 18°$ is shown in Fig. 4.7.

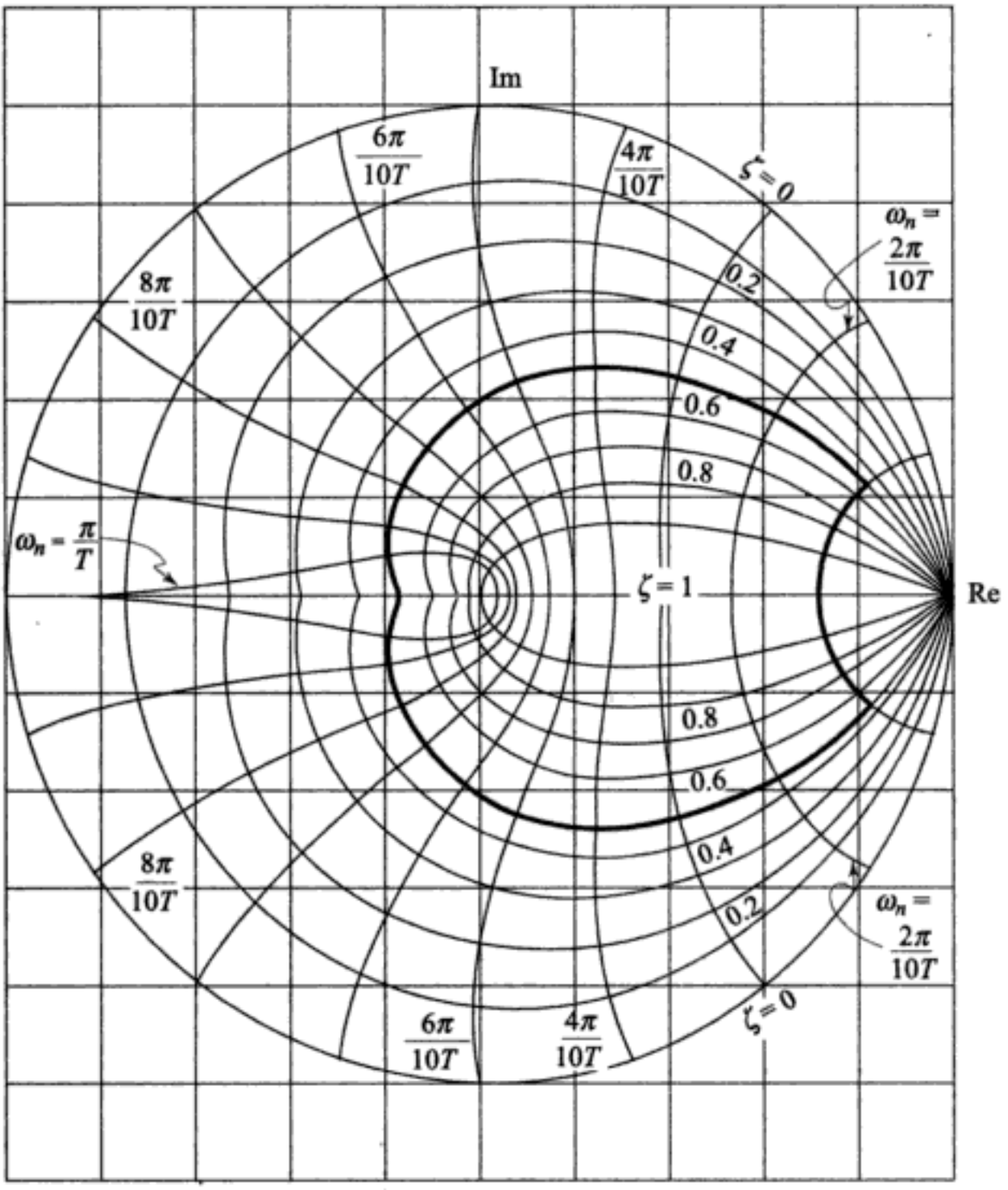

Fig. 4.6 Plot of an acceptable region for poles of a second-order system to satisfy dynamic response specifications

To study the effect of zero location, we let $z_2 = p$ and explore the effect of the (remaining) zero location z_1 on the transient performance. We take the gain K to be such that the steady-state output value equals the step size. For a unit-step input,

$$Y(z) = \left[\frac{K(z - z_1)}{z^2 - 2r\cos\theta\, z + r^2}\right]\left(\frac{z}{z-1}\right) \tag{4.17}$$

with

$$K = \frac{1 - 2r\cos\theta + r^2}{(1 - z_1)}$$

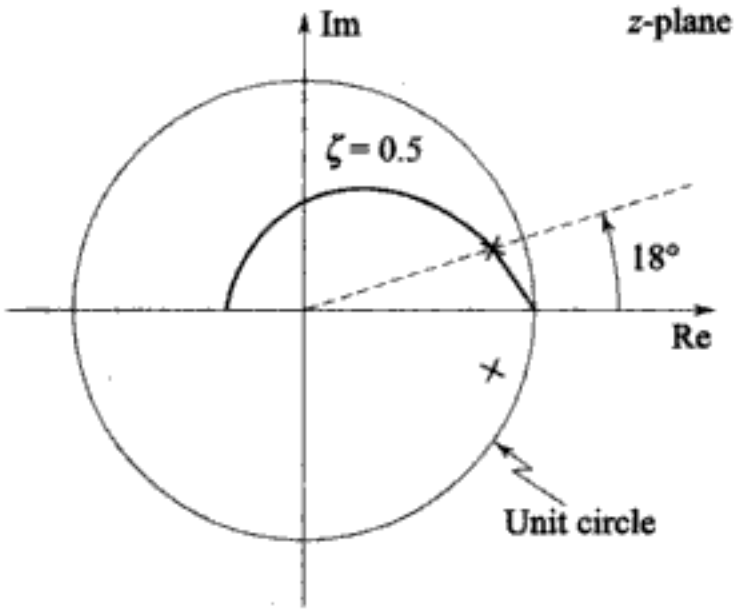

Fig. 4.7 A complex-conjugate pole pair of the system (4.16)

The major effect of the zero z_1 on the step response $y(k)$ is to change the peak overshoot, as may be seen from the step responses plotted in Fig. 4.8a. Figure 4.8b shows plots of peak overshoot *versus* zero location for three different cases of complex-conjugate pole pairs. The major feature of these plots is that the zero has very little influence when on the negative real axis, but its influence is dramatic when it comes near +1.

To study the influence of a third pole on a basically second-order response, we again consider the system (4.16), but this time we fix $z_1 = z_2 = -1$ and let p vary from near -1 to near $+1$. In this case, the major influence of the moving singularity is on the rise time of the step response. Figure 4.8c shows plots of rise time *versus* extra pole location for three different cases of complex-conjugate pole pairs. We see here that the extra pole causes the rise time to get very much longer as the location of p moves toward $z = +1$ and comes to dominate the response.

Our conclusions from these plots are that the addition of a pole or a zero to a given system has only a small effect if the added singularities are in the range 0 to -1. However, a zero moving toward $z = +1$ greatly increases the system overshoot. A pole placed toward $z = +1$ causes the response to slow down and thus primarily affects the rise time which is being progressively increased. The pole pair corresponding to specified ζ and ω_n is a dominant pole pair of the closed-loop system only if the influence of additional poles and zeros is negligibly small on the dynamic response of the system.

Specifications in terms of frequency response features

The translation of time-domain specifications into desired locations of pair of dominant closed-loop poles in z-plane is useful if the design is to be carried out by using root locus plots. The use of frequency response plots necessitates the translation of time-domain performance specifications in terms of frequency response features.

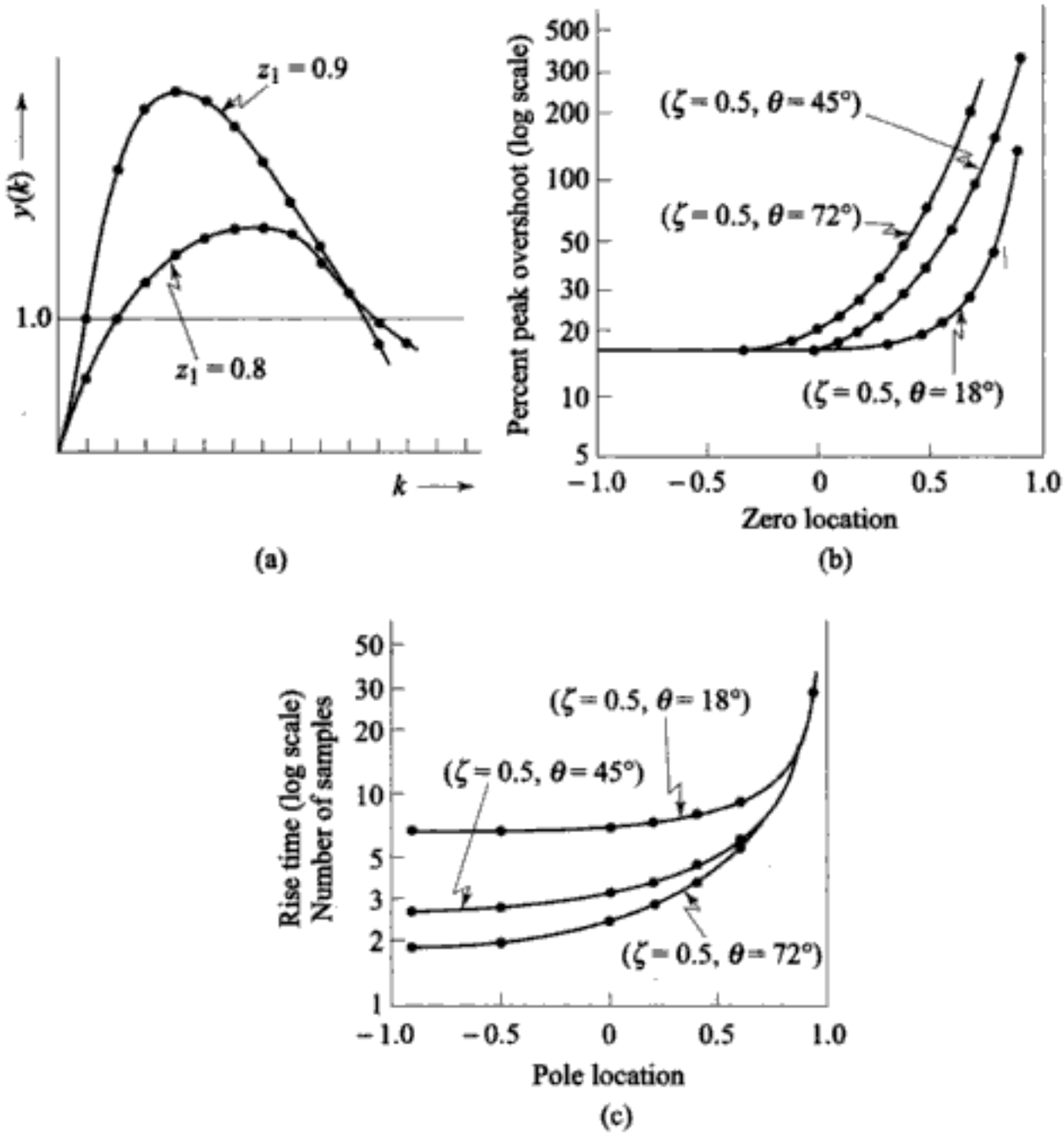

Fig. 4.8 (a) Effect of an extra zero on a discrete-time second-order system, $\zeta = 0.5, \theta = 18°$; (b) effect of an extra zero on a discrete-time second-order system; and (c) effect of an extra pole on a discrete-time second-order system

All the frequency-domain methods of continuous-time systems can be extended for the analysis and design of digital control systems. Consider the system shown in Fig. 4.3. The closed-loop transfer function of the sampled-data system is

$$\frac{Y(z)}{R(z)} = \frac{G_{h0}G(z)}{1 + G_{h0}G(z)} \tag{4.18}$$

Just as in the case of continuous-time systems, the absolute and relative stability conditions of the closed-loop discrete-time system can be investigated by making the frequency response plots of $G_{h0}G(z)$. The frequency response

plots of $G_{h0}G(z)$ are obtained by setting $z = e^{j\omega T}$ and then letting ω vary from $-\omega_s/2$ to $\omega_s/2$. This is equivalent to mapping the unit circle in the z-plane onto the $G_{h0}G(e^{j\omega T})$-plane. Since the unit circle in the z-plane is symmetrical about the real axis, the frequency response plot of $G_{h0}G(e^{j\omega T})$ will also be symmetrical about the real axis, so that only the portion that corresponds to $\omega = 0$ to $\omega = \omega_s/2$ needs to be plotted.

A typical curve of (refer Eqn. (4.18))

$$\frac{Y}{R}(e^{j\omega T}) = \frac{G_{h0}G(e^{j\omega T})}{1 + G_{h0}G(e^{j\omega T})}, \tag{4.19}$$

the closed-loop frequency response, is shown in Fig. 4.9. The amplitude ratio and phase angle will approximate the ideal $1.0 \angle 0°$ for some range of 'low' frequencies but will deviate for high frequencies. The height M_r (resonance peak) of the peak is a relative stability criterion; the higher the peak, the poorer the relative stability. Many systems are designed to exhibit a resonance peak

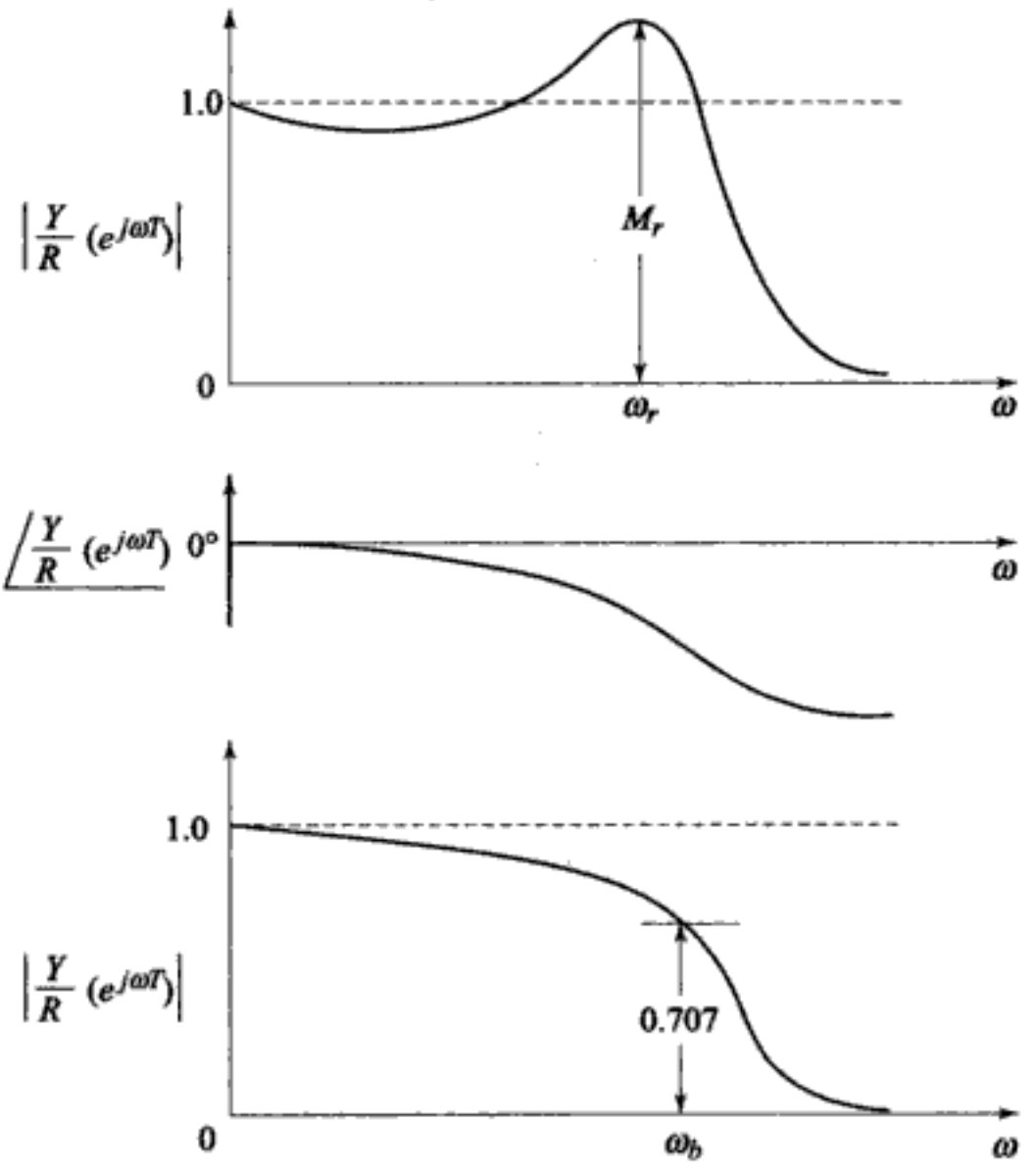

Fig. 4.9 Closed-loop frequency response criteria

in the range 1.2 to 1.4. The frequency ω_r (resonance frequency) at which this peak occurs is a speed of response criterion; the higher the ω_r, the faster the system. For systems that exhibit no peak (sometimes the case), the *bandwidth* ω_b is used for speed of response specifications. Bandwidth is the frequency at which amplitude ratio has dropped to $1/\sqrt{2}$ times its zero-frequency value. It can, of course, be specified even if there is a peak.

Two open-loop performance criteria are in common use to specify relative stability. These are gain margin GM and phase margin ΦM. A typical curve of $G_{h0}G(e^{j\omega T})$, the open-loop frequency response, is shown on the polar plane in Fig. 4.10. Gain margin is the multiplying factor by which the steady-state gain of $G_{h0}G(e^{j\omega T})$ could be increased so as to put the system on the edge of instability. Phase margin is the number of degrees of additional phase lag required to drive the system to the edge of instability.

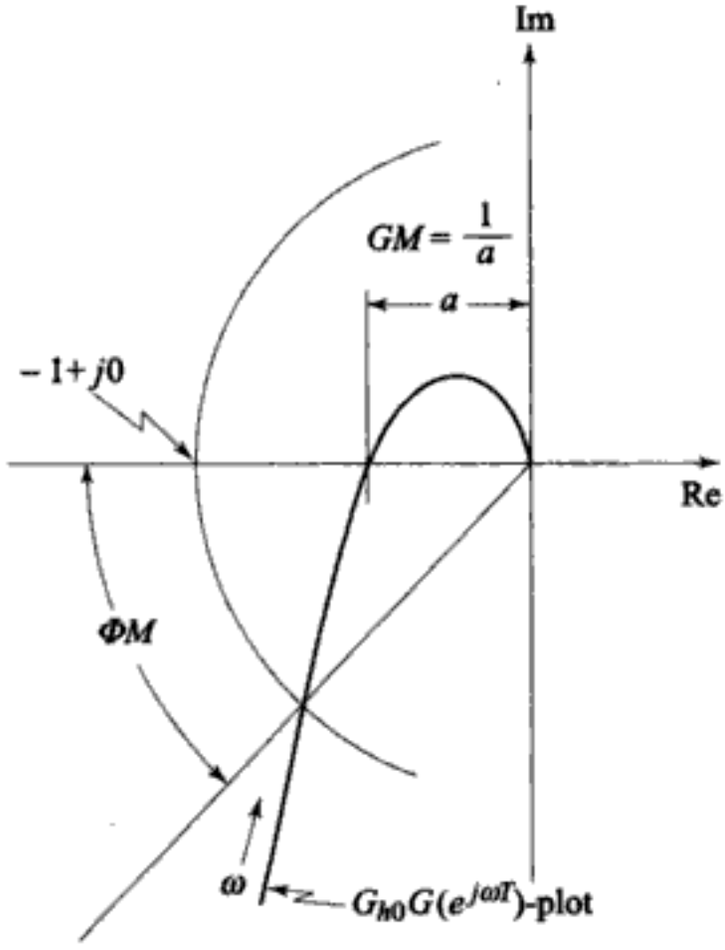

Fig. 4.10 Gain margin and phase margin

Both a good gain margin and a good phase margin are needed; neither is sufficient by itself. General numerical design goals for gain and phase margins cannot be given since systems that satisfy other specific performance criteria may exhibit a wide range of these margins. It is possible, however, to give useful lower bounds—the gain margin should usually exceed 2.5 and the phase margin should exceed 30°.

The translation of time-domain specifications in terms of frequency response features is carried out by using the explicit correlations for second-order system (4.10). The following correlations are valid approximations for higher-order systems dominated by a pair of complex conjugate poles[3].

$$M_r = \frac{1}{2\zeta\sqrt{1-\zeta^2}}; \ \zeta \leq 0.707 \tag{4.20}$$

$$\omega_r = \omega_n\sqrt{1-2\zeta^2} \tag{4.21}$$

$$\omega_b = \omega_n\left[1-2\zeta^2+\sqrt{(2-4\zeta^2+4\zeta^4)}\right]^{\frac{1}{2}} \tag{4.22}$$

$$\Phi M = \tan^{-1}\left\{2\zeta\Big/\left[\sqrt{1+4\zeta^4}-2\zeta^2\right]^{\frac{1}{2}}\right\}$$

$$\cong 100\zeta \tag{4.23}$$

Disturbance Rejection

The effectiveness of a system in disturbance signal rejection is readily studied with the topology of Fig. 4.11a. The response $Y(z)$ to disturbance $W(z)$ can be found from the closed-loop transfer function

$$\frac{Y(z)}{W(z)} = \frac{1}{1+D(z)G_{h0}G(z)} \tag{4.24a}$$

We now introduce the function

$$S(z) = \frac{1}{1+D(z)G_{h0}G(z)} \tag{4.24b}$$

which we call the *sensitivity function* of the control system for reasons to be explained later. To reduce the effects of disturbances, it turns out that $S(e^{j\omega T})$ must be made small over the frequency band of disturbances. If constant disturbances are to be suppressed, $S(1)$ should be made small. If $D(z)G_{h0}G(z)$ includes an integrator (which means that $D(z)$ or $G_{h0}G(z)$ has a pole at $z = 1$), then the steady-state error due to constant disturbance is zero. This may be seen as follows. Since for a constant disturbance of amplitude A, we have

$$W(z) = \frac{Az}{z-1},$$

the steady-state value of the output is given by

$$y_{ss} = \lim_{z\to 1}(z-1)Y(z) = \lim_{z\to 1}(z-1)S(z)W(z) = \lim_{z\to 1}AS(z)$$

which is equal to zero if $D(z)G_{h0}G(z)$ has a pole at $z = 1$.

3. Chapter 9 of reference [180].

Note that the point where the disturbance enters the system is very important in adjusting the gain of $D(z)G_{h0}G(z)$. For example, consider the system shown in Fig. 4.11b. The closed-loop transfer function for the disturbance is

$$\frac{Y(z)}{W(z)} = \frac{G_{h0}G(z)}{1 + D(z)G_{h0}G(z)}$$

In this case, the steady-state error due to constant disturbance $W(z)$ is not equal to zero when $G_{h0}G(z)$ has a pole at $z = 1$. This may be seen as follows.

Let $$G_{h0}G(z) = Q(z)/(z - 1)$$

where $Q(z)$ is a rational polynomial of z such that $Q(1) \neq 0$ and $Q(1) \neq \infty$; and $D(z)$ is a controller which does not have pole at $z = 1$. Then

$$y_{ss} = \lim_{z \to 1} \frac{(z-1)\, G_{h0}G(z)}{1 + D(z)\, G_{h0}G(z)}\, W(z) = \lim_{z \to 1} \frac{AzQ(z)}{z - 1 + D(z)Q(z)} = \frac{A}{D(1)}$$

Thus the steady-state error is non-zero; the magnitude of the error can be reduced by increasing the controller gain.

Figure 4.11c gives a block diagram of the situation where measurement noise $W_n(z)$ enters the system through feedback link. The closed-loop transfer function for this disturbance is

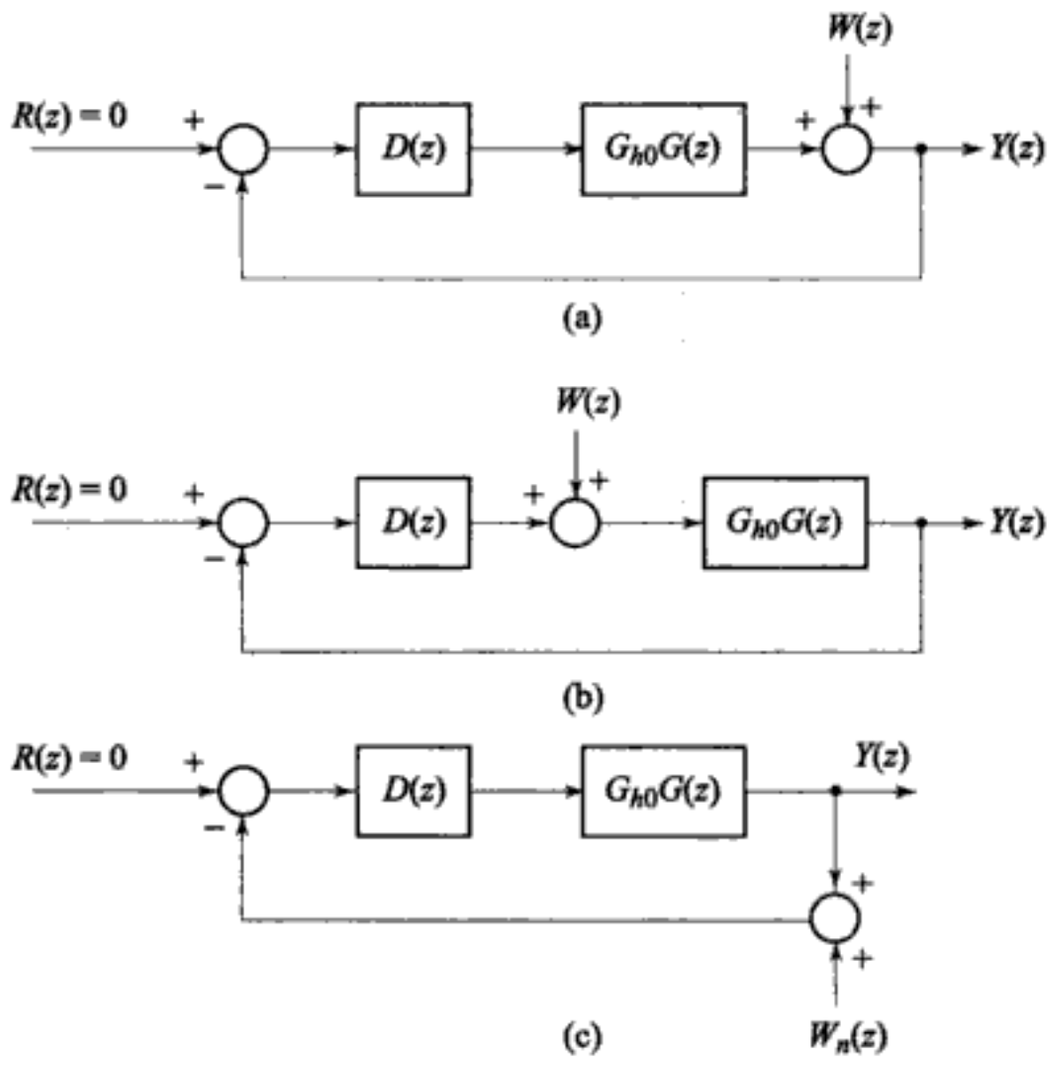

Fig. 4.11 Disturbance rejection

$$\frac{Y(z)}{W_n(z)} = \frac{D(z)G_{h0}G(z)}{1+D(z)G_{h0}G(z)} \tag{4.25}$$

Thus the measurement noise is transferred to the output whenever $|D(z)G_{h0}G(z)| > 1$. Hence large gains of $D(z)G_{h0}G(z)$ will lead to large output errors due to measurement noise. This is in conflict with the disturbance-rejection property with respect to configurations of Figs 4.11a and 4.11b. To solve this problem, we can generally examine the measuring instrument and modify the filtering so that it satisfies the requirements of a particular control problem.

Insensitivity and Robustness

Finally, in our design we must take into account both the small and often the large differences between the derived process model and the real process behaviour. The differences may appear due to modelling approximations and the process behaviour changes with time during operation. If, for simplicity, it is assumed that the structure and order of the process model are chosen exactly and they do not change with time, then these differences are manifested as parameter errors.

Parameter changes with respect to nominal parameter vector $\boldsymbol{\theta}_n$ are assumed. The closed-loop behaviour for parameter vector

$$\boldsymbol{\theta} = \boldsymbol{\theta}_n + \Delta\boldsymbol{\theta}$$

is of interest. If the parameter changes are small, then sensitivity methods can be used. For controller design, both good control performance (steady-state accuracy, transient accuracy, and disturbance rejection) and small parameter sensitivity are required. The resulting controllers are then referred to as *insensitive controllers*. However, for large parameter changes, the sensitivity design is unsuitable. Instead, one has to assume several process models with different parameter vectors $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, ..., $\boldsymbol{\theta}_M$, and try to design a *robust controller* which for all process models will maintain stability and certain control performance range.

As to design of insensitive controllers, the situation is very much like the disturbance-signal rejection. The larger the gain of the feedback loop around the offending parameter, the lower the sensitivity of the closed-loop transfer function to changes in that parameter.

Consider the digital control system of Fig. 4.11. The closed-loop input-output behaviour corresponding to the nominal parameter vector is described by

$$M(\boldsymbol{\theta}_n, z) = \frac{Y(z)}{R(z)} = \frac{D(z)G_{h0}G(\boldsymbol{\theta}_n, z)}{1+D(z)\,G_{h0}G(\boldsymbol{\theta}_n, z)} \tag{4.26}$$

The process parameter vector now changes by an infinitesimal value $d\boldsymbol{\theta}$. For the control loop, it follows that

$$\left.\frac{\partial M(\boldsymbol{\theta}, z)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} = \frac{D(z)}{[1+D(z)\,G_{h0}G(\boldsymbol{\theta}_n, z)]^2} \left.\frac{\partial G_{h0}G(\boldsymbol{\theta}, z)}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_n} \tag{4.27}$$

From Eqns. (4.26)–(4.27), it follows that

$$\frac{dM(\boldsymbol{\theta}_n,z)}{M(\boldsymbol{\theta}_n,z)} = S(\boldsymbol{\theta}_n, z)\frac{dG_{h0}G(\boldsymbol{\theta}_n,z)}{G_{h0}G(\boldsymbol{\theta}_n,z)} \tag{4.28a}$$

with the sensitivity function $S(\boldsymbol{\theta}_n, z)$ of the feedback control given as

$$S^{M}_{G_{h0}G} = S(\boldsymbol{\theta}_n, z) = \frac{1}{1+D(z)G_{h0}G(\boldsymbol{\theta}_n,z)} \tag{4.28b}$$

This sensitivity function shows how relative changes of input/output behaviour of a closed loop depend on changes of the process transfer function. Small parameter-sensitivity of the closed-loop behaviour can be obtained by making $S(\boldsymbol{\theta}_n, e^{j\omega T})$ small in the significant frequency range.

The Case for High-Gain Feedback

Control system design with high-gain feedback results in:

(i) good steady-state tracking accuracy,
(ii) good disturbance-signal rejection, and
(iii) low sensitivity to process-parameter variations.

There are, however, factors limiting the gain:

(i) High gain may result in instability problems.
(ii) Input amplitudes limit the gain; excessively large magnitudes of control signals will drive the process to saturation region of its operation, and the control system design based on linear model of the plant will no longer give satisfactory performance.
(iii) Measurement noise limits the gain; with high gain feedback, measurement noise appears unattenuated in the controlled output.

Therefore in design we are faced with trade-offs.

4.3 DIGITAL COMPENSATOR DESIGN USING FREQUENCY RESPONSE PLOTS

All the frequency response methods of continuous-time systems[4] are directly applicable for the analysis and design of digital control systems. For a system with closed-loop transfer function

$$\frac{Y(z)}{R(z)} = \frac{G_{h0}G(z)}{1+G_{h0}G(z)} \tag{4.29}$$

the absolute and relative stability conditions can be investigated by making the frequency response plots of $G_{h0}G(z)$. The frequency response plots of $G_{h0}G(z)$ can be obtained by setting

$$z = e^{j\omega T};\ T = \text{sampling interval} \tag{4.30}$$

and then letting the frequency ω vary from $-\omega_s/2$ to $\omega_s/2$; $\omega_s = 2\pi/T$. Computer assistance is normally required to make the frequency response plots.

4. Chapters 8–10 of reference [180].